

# How (Not) to Reproduce: Practical Considerations to Improve Research Transparency in Political Science

R. Michael Alvarez, *California Institute of Technology, USA*

Simon Heuberger, *American University, USA*

## ABSTRACT

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

Research transparency has become a central concern in political science (Clemens 2017; Colaresi 2016; Coughlin 2017; Freese 2007; Freese and Peterson 2017; Gertler, Galiani, and Romero 2018; Miguel et al. 2014; Open Science Collaboration 2012). Transparency greatly strengthens the quality of research, heightens accountability, and increases trust in the discipline. Data transparency concerns two related but significantly different goals: (1) using the data and code from a published paper to obtain the same results reported in the paper; and (2) taking the protocol for a study to obtain the same results with a new or different dataset. The first goal refers to reproducibility, which verifies published results and code; the second goal refers to replication, which tests the validity of published findings (Plesser 2018; Shepherd et al. 2017). This article is concerned with reproducibility.

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, thereby further advancing scientific knowledge. (For a

general discussion, see Gherghina and Katsanidou 2013; Gleditsch, Metelits, and Strand 2003; Ishiyama 2014; and Nosek et al. 2015. Specific examples are in Lall 2016 and in the extended debate in Harden, Sokhey, and Wilson 2018; Heuberger 2018; Muchlinski et al. 2018; Neunhoeffler and Sternberg 2018; and Wang 2018.) Archival code and data also are used pedagogically, particularly in graduate methodology courses (Janz 2016). As Gary King (1995, 444) 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.”

Indeed, the issue of research transparency has become so important that *PS: Political Science & Politics* published seven papers in a collection titled “Opening Political Science.” These papers all advance important arguments about how political science can improve research transparency (Brezna 2021; Engzell and Rohrer 2020; Janz and Freese 2020; Kapiszewski and Karcher 2020; Lupia 2020; Rinke and Wuttke 2021; Rohlfing et al. 2020). However, 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. Also missing is guidance for journals to conceptualize replication guidelines that aid successful reproduction.

Today, many journals request or require authors to submit reproduction materials to data archives (discussed further in this article’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

R. Michael Alvarez  is professor of political and computational social science at the California Institute of Technology and past co-editor of *Political Analysis*. He can be reached at [rma@caltech.edu](mailto:rma@caltech.edu).

Simon Heuberger received his PhD in political science from American University and supervised the paper-reproduction process currently in place at *Political Analysis*. He can be reached at [heuberger.simon@alumni.american.edu](mailto:heuberger.simon@alumni.american.edu).

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 (e.g., 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, of the dozens of reproduction datasets submitted to *Political Analysis* during a recent 18-month period, all except

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

These requirements are relatively simple and not controversial.<sup>2</sup> However, many if not most authors fail to meet these minimal requirements when they provide initial reproduction

*Our experience has shown that disorganized and virtually unusable reproduction material is still the norm rather than the exception in political science.*

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. These examples showcase what authors can do better and also provide recommendations that can inform journals' replication guidelines.

#### SHORTCOMINGS IN JOURNAL REPRODUCTION MATERIALS

This section discusses the most common problems that occur during the reproduction review process at *Political Analysis*—problems that we suspect are typical for political science journals with research-transparency requirements. We use the example of

materials for their manuscript. By not initially providing the appropriate materials, authors often cause the failure of the journal editorial team's efforts to use those materials to produce the results reported in the manuscript. This necessitates additional communication with the authors and the subsequent revision(s) of their reproduction materials, with the process reiterating until a point at which editors can release well-documented and well-organized reproduction materials on Dataverse. Authors often face delays in their manuscript's production process because *Political Analysis* will not send a paper into production until the reproducibility requirement has been met. It is concerning that authors often fail to produce usable materials when they are

*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 that a journal's research-transparency team can use successfully? We argue that both should bear some costs.*

*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.<sup>1</sup>

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 data access of the American Political Science Association and the Society for Political Methodology. To meet this goal, *Political Analysis* requests that authors provide the following:

1. A README file that describes the materials that the author 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. Software packages and other materials that are necessary to reproduce the results reported in the manuscript.

required to publish their paper, which implies that the principle of research transparency likely has not been incorporated 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 that 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 (i.e., the replicators) must have a generalist's understanding of the primary software languages in use in their field (i.e., for political science, primarily Stata, R, and Python); they must have current 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. However, we do not believe that replicators 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 workflow the type of practices that will produce good code and documentation; as more authors engage with this, we believe that these issues will largely

disappear. More important, however, is that writing code for many studies is a significant component of the research endeavor. Because the research conducted is the author's responsibility, we do not believe that it is inappropriate to expect authors to pay the same close attention to their code as they do to collecting accurate data, to provide appropriate citation of previous work, and to follow best practices that generally govern high-quality social science research.

Although it is beyond the scope of this article to describe in detail how authors can build these workflows, we note that funding agencies require detailed data-management plans, which often is an important part of a research workflow. Moreover, it is increasingly common for social scientists to use version-control and code-collaboration tools such as Git, Bitbucket, and GitHub. Finally, introductory graduate methods courses are beginning to include research transparency and other ethical issues in the curriculum, thereby training the next generation of scholars in these best practices. Also, professional societies can provide short courses and other educational materials for those who want to learn how to build research transparency into their workflows.

#### DOCUMENTATION: THE IMPORTANCE OF THE README FILE

To be usable by other researchers, reproduction materials require documentation, which means that all reproduction materials must provide a simple README file. The README is the first file that a user opens after downloading the data material; therefore, it must contain all of the information that a user needs to run the code. It is obvious that this depends to some degree on the material in question; however, any README file must contain five basic sections: (1) a list of all folders, subfolders, and data files contained in the material; (2) 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 the output used in the paper; and (5) the approximate running time of each code file based on hardware specifications. Figure 1 is an anonymized example of an insufficient README submitted to *Political Analysis*. It omits all five sections and provides no information about what to expect or how to proceed with the data analysis.

#### Reproduction Data

It often is assumed for the purposes of article reproduction that authors must provide the complete original dataset—this is untrue.

An author often is working with secondary data—for example, the American National Election Survey and the Cooperative Congressional Election Survey. In these cases, the author must provide documentation and code that extracts from these public data sources only the rows and columns used in the published study. That code also should provide details about all processing and manipulation that transforms the original data into the data used in the paper.

A second common problem is proprietary data—that is, authors who do not have permission to share their data because of copyright restrictions or other legal restrictions concerning public dissemination of the data. In this case, authors may provide a percentage sample of the data for reproduction purposes. This typically allows reproduction of the main findings while still protecting and respecting data ownership and privacy.

A third regularly occurring problem with reproduction data is the inclusion of identifying information in the reproduction materials. A good example is reproduction information that may contain names and addresses from voter-registration datasets or names and contact information provided as metadata in datasets from manual text-processing studies. There are many 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.

#### Code and Output

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

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

Figure 1

#### Example of Insufficient README File

```
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)
```

This is an anonymized example of an insufficient README file submitted to *Political Analysis*.

Figure 2

### Working Directories Should Not Be Set Locally

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

Local working directories with setwd().

Figure 3

### Confusingly Named Files and R Objects

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

This is an example in which two .csv source files and three R objects are all based on the word "data."

Figure 4

### Code Output (Left), Manuscript Table (Right)

x	warstds	predictors.1	predictors.2
1	1	0	0.036354228
2	2	0	0.009281192
3	3	0	0.017708231
4	4	0	0.034197142
5	5	1	0.078753771
6	6	0	0.025539793
7	7	1	0.013429648
8	8	0	0.027755230
9	9	1	0.012395045
10	10	0	0.006440910
11	11	1	0.007812825
12	12	0	0.002278077
13	13	0	0.011487555

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

The saved output bears no resemblance to the corresponding manuscript table.

The code must save output files for every figure and table in the main text and the manuscript appendix. 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

resampling fail to set random-number seeds. Failure to use the same random-number seed when trying to reproduce manuscript results is problematic because the reproduction will not generate the exact results reported. Thus, authors should always set the

*Therefore, we urge all political science journals to shift their focus from the mere implementation of research-transparency requirements to a rigorous evaluation of the quality, executability, and user-friendliness of the research materials.*

manuscript (e.g., Figure1.pdf, Table3.tex) to render the output clearly and easily identifiable. Crucially, the saved output must show identical content as the resulting respective figure or table. Figure 4 is an example of saved R output in .csv form that bears no resemblance to the corresponding manuscript table. Whereas the information presented in the manuscript table may be part of the produced .csv file, it is not possible to interpret this information in the data's current form.

Finally, we have seen numerous examples in which the authors of papers with simulations or methods that involve sampling or

random-number seed when conducting simulations or using sampling methods, and this should be well documented in their code.

#### 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, but much work remains. This article identifies two practices that social science should adopt to help resolve the crisis: (1) journals

must provide detailed replication guidance and run provided material; and (2) authors must begin their work with replication in mind.

Many journals have been building 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 much 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 that data and code be provided before a paper's publication. However, some of the journals do not provide or use a permanent and public archive for these materials, and few at this point actually confirm that the data and code reproduce the claims reported in a paper. Therefore, we urge all political science journals to shift their focus from the mere implementation of research-transparency requirements to 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 (see the supplementary materials). Journals such as *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 to assist reproduction efforts. The guidance can range from elaborated bullet points to templates that showcase what is required (demonstrated in the supplementary materials). Additionally, we urge journals to establish a research-transparency team in which all members are trained to efficiently run Stata, R, and Python code and to diagnose common problems. With these training structures in place, graduate students are sufficiently equipped to conduct cost-effective journal reproductions. Journals also must ensure that their reproducers have access to computational resources that will reliably run complex simulations, machine- and deep learning, and that can handle larger-scale datasets.

On the author's side, highly disorganized and virtually unusable reproduction material is still consistently the norm. Virtually none of the reproduction datasets submitted to *Political Analysis* during the past 18 months ran without producing errors. Scholars must pay closer attention to the documentation and usability of the research materials that they make available to journals and other scholars. They should begin a quantitative study with reproducibility in mind to avoid a frantic rush to collect and document their research late in the publication process. This will ensure that their material meets transparency requirements when it is submitted to journals.

## SUPPLEMENTARY MATERIALS

To view supplementary material for this article, please visit <http://dx.doi.org/10.1017/S1049096521001062>. ■

## NOTES

1. Alvarez was co-editor of *Political Analysis* from 2010 to 2018, when the journal began requiring that authors provide research-transparency materials and then began the process of validating those materials. Heuberger was a graduate editorial assistant at the journal under the current editor, Jeff Gill, and he was in charge of validation of research-transparency materials and replication of all code between 2017 and 2021.
2. Due to space limitations, the requirements are presented in rudimentary form. Actual guidance given to authors at *Political Analysis* explains each point in detail.

## REFERENCES

- Brezna, Nate. 2021. "I Saw You in the Crowd: Credibility, Reproducibility, and Meta-Utility." *PS: Political Science & Politics* 54 (2): 1–5. DOI:10.1017/S1049096520000980.
- Bryan, Jenny. 2018. "Ode to the Here Package." [https://github.com/jennyb/here\\_here](https://github.com/jennyb/here_here).
- Clemens, Michael A. 2017. "The Meaning of Failed Replications: A Review and Proposal." *Journal of Economic Surveys* 31 (1): 326–42.
- Colaresi, Michael. 2016. "Prepublication, Replication: A Proposal to Efficiently Upgrade Journal Replication Standards." *International Studies Perspectives* 17 (4): 367–78.
- Coughlin, Steven S. 2017. "Reproducing Epidemiologic Research and Ensuring Transparency." *American Journal of Epidemiology* 186 (4): 393–94.
- Dafoe, Allan. 2014. "Science Deserves Better: The Imperative to Share Complete Replication Files." *PS: Political Science & Politics* 47 (1): 60–66.
- Engzell, Per, and Julia M. Rohrer. 2020. "Improving Social Science: Lessons from the Open Science Movement." *PS: Political Science & Politics* 54 (2): 1–4.
- Eubank, Nicholas. 2016. "Lessons from a Decade of Replications at the *Quarterly Journal of Political Science*." *PS: Political Science & Politics* 49 (2): 273–76.
- Freese, Jeremy. 2007. "Replication Standards for Quantitative Social Science: Why Not Sociology?" *Sociological Methods & Research* 36 (2): 153–72.
- Freese, Jeremy, and David Peterson. 2017. "Replication in Social Science." *Annual Review of Sociology* 43 (1): 147–65.
- Gertler, Paul, Sebastian Galiani, and Mauricio Romero. 2018. "How to Make Replication the Norm." *Nature* 554 (7690): 417–19.
- Gherghina, Sergiu, and Alexia Katsanidou. 2013. "Data Availability in Political Science Journals." *European Political Science* 12:333–49.
- Gleditsch, Nils Petter, Claire Metelits, and Håvard Strand. 2003. "Posting Your Data: Will You Be Scooped or Will You Be Famous?" *International Studies Perspectives* 4 (1): 9–95.
- Harden, Jeffrey J., Anand E. Sokhey, and Hannah Wilson. 2018. "Replications in Context: A Framework for Evaluating New Methods in Quantitative Political Science." *Political Analysis* 27:119–25.
- Heuberger, Simon. 2018. "Insufficiencies in Data Material: A Replication Analysis of Muchlinski, Siroky, He, and Kocher (2016)." *Political Analysis* 27:114–18.
- Ishiyama, John. 2014. "Research Transparency, and Journal Publications: Individualism, Community Models, and the Future of Replication Studies." *PS: Political Science & Politics* 47 (1): 78–83.
- Janz, Nicole. 2016. "Bringing the Gold Standard into the Classroom: Replication in University Teaching." *International Studies Perspectives* 17:392–407.
- Janz, Nicole, and Jeremy Freese. 2020. "Replicate Others as You Would Like to Be Replicated Yourself." *PS: Political Science & Politics* 54 (2): 1–4.
- Kapiszewski, Diana, and Sebastian Karcher. 2020. "Transparency in Practice in Qualitative Research." *PS: Political Science & Politics* 54 (2): 1–7.
- King, Gary. 1995. "Replication, Replication." *PS: Political Science & Politics* 28 (3): 444–52.
- Lall, Ranjit. 2016. "How Multiple Imputation Makes a Difference." *Political Analysis* 24 (4): 414–33.
- Lupia, Arthur. 2020. "Practical and Ethical Reasons for Pursuing a More Open Science." *PS: Political Science & Politics* 54 (2): 1–4.
- Lupia, Arthur, and Colin Elman. 2014. "Openness in Political Science: Data Access and Research Transparency." *PS: Political Science & Politics* 47 (1): 19–42.
- Miguel, Edward, Colin Camerer, Katherine Casey, Jason Cohen, Kevin Esterling, Alexander Gerber, et al. 2014. "Promoting Transparency in Social Science Research." *Science* 343:30–31.
- Muchlinski, David A., David Siroky, Jingrui He, and Matthew A. Kocher. 2018. "Seeing the Forest Through the Trees." *Political Analysis* 27 (1): 111–13.
- Neunhoeffer, Marcel, and Sebastian Sternberg. 2018. "How Cross-Validation Can Go Wrong and What to Do About It." *Political Analysis* 27 (1): 101–106.
- Nosek, Brian A., George Alter, George C. Banks, Denny Borsboom, Sara D. Bowman, Steven J. Breckler, et al. 2015. "Promoting an Open Research Culture." *Science* 348 (6242): 1422–25. DOI: 10.1126/science.aab2374.
- Open Science Collaboration. 2012. "An Open, Large-Scale, Collaborative Effort to Estimate the Reproducibility of Psychological Science." *Perspectives on Psychological Science* 7 (6): 657–60.
- Plesser, Hans E. 2018. "Reproducibility vs. Replicability: A Brief History of a Confused Terminology." *Frontiers in Neuroinformatics* 11:76.
- Rinke, Eike M., and Alexander Wuttke. 2021. "Open Minds, Open Methods: Transparency and Inclusion In Pursuit of Better Scholarship." *PS: Political Science & Politics* 54 (2): 1–4.

Rohlfing, Ingo, Lea Königshofen, Susanne Krenzer, Jan Schwalbach, and Ayjeren Bekmuratovna R. 2020. "A Reproduction Analysis of 106 Articles Using Qualitative Comparative Analysis, 2016–2018." *PS: Political Science & Politics* 54 (2): 1–5.

Shepherd, Bryan E., Meredith B. Peratikos, Peter F. Rebeiro, Stephany N. Duda, and Catherine C. McGowan. 2017. "A Pragmatic Approach for

Reproducible Research with Sensitive Data." *American Journal of Epidemiology* 186 (4): 387–92.

Wang, Yu. 2018. "Comparing Random Forest with Logistic Regression for Predicting Class-Imbalanced Civil War Onset Data: A Comment." *Political Analysis* 27 (1): 107–10.