

Research Replication: Practical Considerations

R. Michael Alvarez, *California Institute of Technology*

Ellen M. Key, *Appalachian State University*

Lucas Núñez, *California Institute of Technology*

ABSTRACT

With the discipline's push toward data access and research transparency (DA-RT), journal replication archives are becoming increasingly common. As researchers work to ensure that replication materials are provided, they also should pay attention to the content—rather than simply the provision—of journal archives. Based on our experience in analyzing and handling journal replication materials, we present a series of recommendations that can make them easier to understand and use. The provision of clear, functional, and well-documented replication materials is key for achieving the goals of transparent and replicable research. Furthermore, good replication materials enhance the development of extensions and related research by making state-of-the-art methodologies and analyses more accessible.

More than two decades ago, Gary King (1995) published an important and then-provocative paper titled “Replication, Replication.” In his paper, King presented a simple claim: “[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, 444, emphasis in original). At the time, most social scientists tried to provide important details about their empirical analysis in footnotes and appendices. However, there often was insufficient space in journals and books for these important details. Instead, King argued for a different approach: that is, for authors to make available the actual data used to make an empirical claim as well as all of the materials necessary to manipulate and analyze that data to reproduce the results underpinning the empirical claim.¹

When King wrote his paper, the general principle of research replication was widely discussed, and many researchers began to think about ways to make their research materials available to other scholars. Two decades ago, smartphones and the “cloud” did not exist, electronic storage and sharing of large datasets or other materials were not straightforward, and there were few mechanisms that scholars could use to make their replication materials available. There certainly were some places where authors could

store replication materials, in particular the Inter-University Consortium for Political and Social Research (ICPSR) “Publication-Related Archive,”² which launched about the time that King published his paper. However, few scholars at that time used facilities like the ICPSR to store their replication materials, and few journals or funding agencies required researchers to provide easily accessible replication materials.

In recent years, however, the situation has changed dramatically. First, cloud-based computing has made data- and code-sharing simple—some would say trivial. Researchers can easily share their materials from their own file-sharing archives (e.g., Dropbox, Box, or Google Drive); from code-sharing archives (e.g., GitHub or Bitbucket); or by using shared archives either at their institution or those provided for researchers (e.g., Dataverse). Second, researchers in many graduate programs are now trained to build replication into their workflows, and many have recognized that providing data, code, and other research materials boosts their visibility and increases citations. Third, many journals and funding agencies now *require* that research materials be made available on publication. Fourth, due to a number of highly publicized issues regarding research transparency in the social sciences, professional organizations, colleges and universities, and other advocates are encouraging researchers to be more open regarding details of their research.

Thus, research replication and transparency have become central concerns in the social sciences, which we argue strengthens research done in political science and other disciplines. That another scholar can easily and quickly confirm the published results in a paper helps that scholar gain confidence in the integrity of the findings. When that scholar can use replication

R. Michael Alvarez is professor of political science at California Institute of Technology and co-editor of *Political Analysis*. He can be reached at rma@hss.caltech.edu.

Ellen M. Key is assistant professor of political science at Appalachian State University. She can be reached at keyem@appstate.edu.

Lucas Núñez is a graduate student at California Institute of Technology and graduate editorial assistant at *Political Analysis*. He can be reached at lnunez@caltech.edu.

materials to test the robustness of those published results in various ways and possibly improve on the methodology or analysis previously published, it builds the type of cumulative knowledge that makes for a better social science.

JOURNALS AND REPLICATION POLICIES

Political Analysis—the journal of the Society for Political Methodology (SPM)—was among the first journals in social science to

That another scholar can easily and quickly confirm the published results in a paper helps that scholar gain confidence in the integrity of the findings. When that scholar can use replication materials to test the robustness of those published results in various ways and possibly improve on the methodology or analysis previously published, it builds the type of cumulative knowledge that makes for a better social science.

develop a replication policy for papers published in the journal, beginning after the publication of King's 1995 paper. For example, in 1996, some papers published in *Political Analysis* stored their replication materials in the ICPSR archive (Box-Steffensmeier and Lin 1996). However, because the policy was largely voluntary, other articles published in the journal at that time make no mention of replication materials. By 2000, the journal began a more systematic collection of replication materials, storing them on the SPM website.³

Beginning in 2012, *Political Analysis* developed a new replication policy: all papers reporting data analyses (including simulation modeling) would be *required* to store the materials necessary to replicate the results reported in each published paper in the journal's Dataverse, prior to publication.⁴ This made *Political Analysis* one of only a few journals before 2015 requiring the provision of replication materials prior to publication.⁵ Currently, replication materials are requested before final acceptance of a manuscript for publication, which are reviewed by both the editor overseeing the manuscript and one of the journal's graduate editorial assistants. Only after the replication materials have been reviewed are they released to the public on the journal's Dataverse and the paper sent to the publisher for production. Although many issues arise during our review of replication materials, we have not had authors who refused to meet the journal's current replication requirement.

MEETING REPLICATION REQUIREMENTS

In April 2016, one of us published a study in *PS: Political Science & Politics* titled "How Are We Doing? Data Access and Replication in Political Science" (Key 2016a). This study examined the replication policies of six major journals, including *Political Analysis*, and determined how many recently published papers in those top journals had replication materials. The research found that during the study period, there was a significant bifurcation in the percentages of papers published with available replication materials. Three journals had a high percentage of studies that did not have them: 67.6% of the papers published in *American Political Science Review*, followed by *Journal of Politics* (51.1%) and *British Journal of Political Science* (50.0%). At the other end of the distribution were *International Organization* (10.2%), *American Journal of Political Science* (9.3%), and *Political Analysis* (1.9%).

Ordinarily, the fact that 98.1% of papers published in this study period in *Political Analysis* had available replication materials might be cause for celebration, especially given that the journal had the highest compliance rate among the major political science journals. However, from the journal's perspective, *Political Analysis* has universal compliance with the replication requirement.⁶ This discrepancy highlights the importance of journal replication archives and raises the issue of accessibility.

Because Key (2016b) published a replication archive, it was possible to identify the *Political Analysis* article by Bowers, Fredrickson, and Panagopoulos (2013) that had been coded as unavailable. Although Bowers et al. had provided a replication archive on the journal's Dataverse, additional software was needed to extract the file. In other words, the issue was not with the availability of the replication material; rather, it was that the single file containing the materials was in a format not easily accessible using standard software available on Macs and PCs.⁷ The more steps involved in retrieving replication materials, the less useful these materials become.

Like a ramp that is too steep, journal replication archives that require users to download not only the data they want to access but also specialty software to open the replication file make the archives inaccessible to a wider audience. Although many researchers at R1 institutions have access to a variety of statistical packages and powerful machines, graduate students and those at smaller institutions may not be as fortunate. This resource discrepancy highlights the tension between the admirable goals of DA-RT and the realities faced by authors, editors, and replication-archive users. This led us to understand that there are important, pressing problems regarding replication materials—in particular, ensuring that the materials are provided in usable and accessible formats.

COMMON ISSUES THAT ARISE WITH JOURNAL REPLICATION MATERIALS

By now, we have much experience with replication policies and materials. Recently, *Political Analysis* began to devote more time and resources to the process of reviewing and using the replication materials before final acceptance of a paper for publication. It is during the verification stage that we often find many common issues that arise with replication materials. This section discusses many of those issues, followed by a presentation of several recommendations for authors and other journals to improve the practice of replication and minimize accessibility issues that arose with the Bowers et al. replication materials. The most common issues found can be grouped broadly in three categories: organization, clarity, and usability.

Organization

Many replication packages include several files with data, codes, and codebooks. In some cases, these files are simply bundled together

in a folder or a compressed file without any indication of what each file contains and how a user of the replication materials should proceed. The problem usually is compounded by the use of obscure file names that convey little information about the content.

Clarity

Clarity refers to problems in understanding codes and scripts. It is relatively common for authors to provide codes with little or no annotation. Some codes are straightforward because the data manipulations and the estimation techniques used are fairly simple; in other cases, neither are standard. In these cases, the

Third, with the proliferation of Markov Chain Monte Carlo (MCMC) and other related estimation techniques, some replication codes require multiple hours (or days) to produce the estimates. In many cases, this is not indicated by the authors, which can lead users to assume that the code is not functioning or that there is another problem. Additionally, authors are increasingly using parallel computing, the operationalization of which can differ across operating systems and typically depends on the capabilities of the computer being used. For users unfamiliar with parallel computing, this can create a significant barrier for replicating and using the materials provided by the author(s).

Journal replication materials should be stored in permanent archives that ensure their availability after long periods. Although there are many data-storage options available, some are more durable than others.

lack of guidance in the code can render the replication materials extremely difficult to understand and use. This can significantly hinder validation of the data manipulation and analysis, as well as the ability of users to improve on the analyses or techniques presented in the paper.

A second factor that affects the clarity of replication materials is the way the statistical software output relates to figures and tables presented in the paper. For example, tables containing the output from several regressions are sometimes not produced in the code; the code instead produces a separate output for each regression. The user must determine which column in a table corresponds to which line of code. This can lead the author to make mistakes in (manually) transferring multiple outputs to a common display. This is more challenging when they are not in the same order or do not always report the full outcome produced by the code.

Finally, simulation studies and many estimation techniques use randomly generated numbers or samples (e.g., some types of bootstrap). When replication materials fail to provide the random sequence (i.e., the seed) and the pseudorandom-number algorithm used to generate results in the paper, comparison of the output obtained by the user and those available in the paper is more difficult to assess.⁸

Usability

The final broad class of issues relates to the usability of replication materials. Many replication packages involve the use of multiple interdependent files. The dependencies among these files are not always clear. Moreover, in many cases, they require a folder structure that is not provided by the author. Often, there is no clear indication of what the user should modify in the codes for the dependencies to work.

A second usability problem concerns software and packages in general. Some statistical software and statistical packages work well in one operating system but not in others.⁹ In other situations, different versions of software and statistical packages can produce different results due to updates, bug fixes, and different optimization methods. Lack of information about the software version used by the author—as well as missing information about software and package dependencies needed to replicate the results in the paper—can be a barrier to replication.

RECOMMENDATIONS

Based on our experiences with journal replication archives, we offer the following recommendations to authors and other journal editors to improve organization, clarity, and usability. Whereas the widespread practice of providing replication materials is relatively recent, we hope that journals, professional societies, and publishers will coordinate in the near future to develop common standards for replication materials.¹⁰

1. Archive

- a) Journal replication materials should be stored in permanent archives that ensure their availability after long periods. Although there are many data-storage options available, some are more durable than others. Personal websites are susceptible to “link rot” caused by website reconfiguration, changing institutions, and other maintenance issues (Key 2016a). At present, the best archives are those with an institutional guarantee, such as Dataverse and other university- or consortium-backed archives.
- b) Authors should store their materials using file formats that are likely to be accessible in the future (e.g., comma-delimited or flat text files for data rather than specific proprietary software files).¹¹ In addition to the preservation advantage of using these types of files, they can be easily read on currently available statistical software, simplifying their use across different platforms. If it is important that files be compressed or stored in a particular format, authors should provide documentation for future users who may not be familiar with how to uncompress or access these files.

2. Readme Files. The creation of a clear and sufficiently (but not overly) detailed readme file is a key element of every replication package. Several items should be included in a readme file, as follows:

- a) A reference to the associated paper or publication.
- b) A short description of the files and file types included in the replication package: for example, raw data, processed data, scripts to manage the data, and scripts to produce estimates.

-
- c) An indication of the order in which the scripts are to be run, as well as noting where the different tables and figures found in the associated publication are generated and stored. Authors should indicate whether there are intermediate outputs generated by one script and used by another so that it is clear to users how to proceed (e.g., if a script handles the raw data to produce a processed dataset to be used in another script, this should be noted in the readme file).
 - d) A list of software and software packages (as well as the dependent packages on which they rely) and the operating system used to produce results in the paper. Technical information on the hardware used to produce the results also is helpful (e.g., the number of cores), especially if computationally demanding techniques were applied. Authors should indicate which versions were used because packages frequently are altered to fix bugs and resolve other issues. These version changes can produce different results when running the same code with a different version of the package or software. Some models also require additional software such as JAGS or a C++ compiler to be available in the computer. Indicating this in the readme file is useful for a less-familiar or inexperienced user.
 - e) If unusual file extensions are used, authors should clearly indicate—to the extent possible—how to proceed with these files.
3. *File Names.* File names should be easy to understand and provide information about the file content. This is particularly important for replication materials that include multiple scripts and data files. If multiple scripts are to be run in a certain order, then including the order in the file name is useful (e.g., “1_DataProcessing,” “2_Estimation”). Alternatively, authors can create a master file that calls on the different scripts.
 4. *User-Created Software and Packages.* Sometimes researchers create their own statistical packages to produce the results in the paper. These packages ideally should be archived in a stable location, such as CRAN for R. If possible, authors should include a copy of the package or software in the replication package.
 5. *Operating-System Compatibility.* Ideally, authors should ensure that their replication materials function under common operating systems. If this is not possible, they should indicate on which operating system(s) they do function.
 6. *Script Documentation*
 - a) The scripts should include a short description of what the code does (e.g., “data recoding”), what dependencies it has, which package(s) it requires, and what the outputs are.
 - b) Authors should avoid scripts that include several pages of commands with little or no indication of what each line (or group of lines) is doing. Code should be annotated where necessary to indicate the purpose of each line or group of lines. This facilitates identification of specific parts of the data-management or estimation process and also helps to identify potential issues or mistakes.
 - c) Authors should strive to organize their scripts well by avoiding interspersing data management and recoding with estimations, unless absolutely necessary for the analysis.
 7. *Outputs.* Replication scripts should have tables and figures as a clearly identified output (i.e., either a file or an object in the statistical software) exactly as they appear in the paper (to the greatest extent possible). There are many ways that authors can achieve this. For example, for tables with several models estimated in Stata, they can use a combination of *eststo* and *esttab*, among many other commands available. In R, authors can use *stargazer*. For tables that are more ad hoc, they can store the outputs of multiple commands in an object (typically a matrix) in the statistical software that can be printed to a file.
 8. *Excluded Outputs.* Many replication materials include scripts (or parts of scripts) and data for outputs that are not included in the final paper (which typically are excluded at some point in the review process) or are relegated to an online appendix. In this case, authors should clearly differentiate between what is and is not included in the paper. This is particularly important for excluded robustness checks that tend to appear similar to the main results included in the paper, thereby potentially causing confusion.
 9. *Intermediate Outputs.* In many cases, scripts generate intermediate outputs that are then further processed to produce the final outputs included in the paper. In certain circumstances, these intermediate outputs require a significant amount of time to be generated. Among many others, the outputs include the generation of simulated datasets in simulation studies and the Markov Chains when estimating via MCMC. In this case, it is useful to include these intermediate outputs in the replication materials.
 10. *Random Sequences.* For simulation studies or estimation strategies that use randomization, authors should always include the random sequence of numbers used to produce the results in the paper (i.e., the seed) and the pseudorandom-number algorithm. Although the exact seed used should not be critical in terms of results obtained, the availability of the seed simplifies replication of the exact numbers in the paper. The location of the seed should be clearly identified at the beginning of the code in case the user wants to change it.
 11. *Directories and Folders.* Directory paths should be easily identifiable in the scripts so that users can change them accordingly. Ideally, the directory setting should need to be specified only at the beginning of the code; replacing multiple directories in different parts of the code can become cumbersome. If the replication materials are organized in multiple folders, authors should indicate at the beginning of each script where to set the main directory. All directory changes within the scripts should be automatic.
 12. *Computing Time.* Authors should indicate (in both the code and the readme file) whether a particular script or part of a script requires a considerable amount of time to compute. This is important because unaware users experiencing a lengthy estimation might incorrectly assume that there is an issue with the materials.
 13. *Parallel Computing.* Authors are increasingly making use of parallel computing in their scripts because it can accelerate computations significantly. Unfortunately, not all parallelized scripts work on all computers and operating systems. Therefore, authors should take care in highlighting the part of the script that is being parallelized and suggest how to proceed in case of incompatibilities with the user’s computer or operating system.

14. *Warnings and Errors.* It is not unusual to find scripts that correctly reproduce the output in the paper but for some reason return warnings or errors (usually related to inappropriate use of a package or differences between operating systems). In this case, authors should note the reasons why the warnings and errors are not a concern.

All replication materials are different and creating a set of recommendations that applies to each one is a monumental task. A simple rule for authors to follow when creating their replication materials is to put themselves in the place of another person whose only knowledge about the materials is what the author provides. In the future, the construction of well-organized, clear, and usable replication materials will depend on the training of scholars so that research replication is effectively incorporated into their workflows. Researchers can find further advice, particularly relative to coding practices, in articles by Nagler (1995) and Bowers and Voors (2016).

CONCLUSION

Replication materials are becoming increasingly common. As political scientists (and social scientists in general) work to ensure that journal replication archives are provided, researchers also must pay attention to the content—rather than merely the provision—of those archives. Based on our experience in analyzing and handling journal replication materials, we presented a series of recommendations that can make them easier to understand and manage. By improving organization, clarity, and usability, authors will bolster the accessibility of their replication materials. Although our list of recommendations is not intended to be exhaustive, we believe it can prevent many significant problems that arise with replication materials. The provision of clear, functional, and well-documented replication materials is key for achieving the goals of transparent and replicable research. Furthermore, good replication materials enhance the development of extensions and related research by making state-of-the-art methodologies and analysis more easily accessible. ■

NOTES

1. This article generally assumes that replication materials are from quantitative rather than qualitative research. We take that approach because the principles associated with research replication and transparency for quantitative materials are generally agreed on in the quantitative-research community. The statements in this article can (and should) generally apply to qualitative research.

2. Available at www.icpsr.umich.edu/icpsrweb/deposit/pr/index.jsp.
3. This has caused some confusion because SPM's website has much of this replication material, but the official website for the journal (run by the publisher) does not have it. Therefore, those interested in finding and using those materials must be diligent in their search. We also suspect that some authors have stored replication materials on their own personal or research websites, but we have not made a systematic search to determine how many of those websites still exist.
4. Available at <https://dataverse.harvard.edu/dataverse/pan>.
5. The *Quarterly Journal of Political Science* is another journal in the discipline that has a rigorous replication policy (Eubank 2014).
6. Some research articles or letters published in the journal are commentaries, reviews, and critiques; because they do not contain simulations or quantitative analyses, they have no materials subject to the journal's replication policy.
7. This was a TAR archive file, which had been further compressed using the GNU gz format. Although these formats are well known to Unix and Linux users, they may require the installation of special software by users of other platforms.
8. This issue should not present with major complications because any robust simulation study or estimation technique should not see its results affected by the random sequence used. The results obtained with any random sequence should be qualitatively the same and quantitatively extremely similar.
9. In some cases, the operating system used has an impact on how the statistical software interacts with other software. For example, using the package Rcpp, which allows for C++ operations in R, requires the user to follow different instructions in Windows and Mac for C++ and R to communicate correctly. In other situations, parallel computing requires different configurations for different operating systems.
10. Although providing a universally accessible archive is the gold standard, we hope researchers who have difficulty meeting all of the following criteria do not use that as an excuse to avoid providing replication materials. In other words, "something is always better than nothing."
11. Although codebooks may be developed internally (e.g., as part of a Stata .dta file), they will be lost when files are saved in a flat-file format. For this reason, we also recommend the inclusion of a separate codebook file.

REFERENCES

- Bowers, Jake, Mark M. Fredrickson, and Costas Panagopoulos. 2013. "Reasoning about Interference between Units: A General Framework." *Political Analysis* 21 (1): 97–124.
- Bowers, Jake, and Maarten Voors. 2016. "How to Improve Your Relationship with Your Future Self." *Revista de Ciencia Política* 36 (3): 829–48.
- Box-Steffensmeier, Janet M., and Tse-min Lin. 1996. "A Dynamic Model of Campaign Spending in Congressional Elections." *Political Analysis* 6 (1): 37–66.
- Eubank, Nicholas. 2014. "A Decade of Replications: Lessons from the Quarterly Journal of Political Science." *The Political Methodologist* 22 (1): 18–19.
- Key, Ellen M. 2016a. "How Are We Doing? Data Access and Replication in Political Science." *PS: Political Science & Politics* 49 (2): 268–72.
- . 2016b. "Replication Data for 'How Are We Doing? Data Access and Replication in Political Science'." Doi:10.7910/DVN/5LJAMC, Harvard Dataverse, V2.
- King, Gary. 1995. "Replication, Replication." *PS: Political Science & Politics* 28 (September): 444–52.
- Nagler, Jonathan. 1995. "Coding Style and Good Computing Practices." *PS: Political Science & Politics* 28 (3): 488–92.