Discussion Paper No. 1312

ISSN (Print) 0473-453X
ISSN (Online) 2435-0982

AN EMPIRICAL ASSESSMENT OF DATA SHARING AND COMPUTATIONAL REPRODUCIBILITY IN SOCIOLOGY

Rio Ikeuchi
Taisuke Imai

May 2026

The Institute of Social and Economic Research
The University of Osaka
6-1 Mihogaoka, Ibaraki, Osaka 567-0047, Japan

An Empirical Assessment of Data Sharing and Computational Reproducibility in Sociology *

Rio Ikeuchi[1] Taisuke Imai[2,†]

[1]Osaka School of International Public Policy, The University of Osaka
[2]Institute of Social and Economic Research, The University of Osaka
[†]Corresponding author: taisuke.imai@iser.osaka-u.ac.jp

May 5, 2026

Abstract

Public sharing of data and analysis code enables independent verification of published findings, yet systematic evidence on such sharing in sociology remains scarce, even as audits have been conducted in psychology, economics, and political science. We examine 730 empirical articles published between January 2019 and June 2024 in three leading general-interest journals: *American Sociological Review*, *American Journal of Sociology*, and *Social Forces*. For each article, we record whether a publicly accessible replication package is provided, and for those that share materials, we attempt to computationally reproduce the main-text tables and figures. Across the sample, 9.9% of articles provide replication packages, with substantial variation across journals (5.2% to 22.9%) and research types (0% for qualitative studies, 30% for experiments). These rates fall well below those reported for political science and economics. Among 72 packages examined in detail, more than half cannot be verified due to missing or incomplete materials. At the same time, no article is wholly non-reproducible, and 22% are fully or largely reproducible. Sociology is not unusually prone to errors when materials are shared and runnable—it simply shares them far less often. We discuss implications for journal policies and for transparency standards that reflect the field's methodological diversity.

Keywords: open science, transparency, data sharing, computational reproducibility

*Financial support from the Japan Society for the Promotion of Science (KAKENHI No. JP22K21358 to T.I.) is gratefully acknowledged.

1 Introduction

The credibility of empirical research depends on the ability of independent researchers to verify reported findings. Over the past decade, large-scale replication efforts in psychology and economics have shown that many published results do not replicate when tested with new data from independent samples (Open Science Collaboration, 2015; Camerer et al., 2016, 2018; Holzmeister et al., 2025), and a related concern has emerged about *computational reproducibility*: whether the results reported in a published article can be regenerated using the original data and code (Christensen and Miguel, 2018; Dreber and Johannesson, 2025; Nosek et al., 2026). Computational reproducibility is a prerequisite for replication. If a published result cannot be reproduced even with the original materials, there is no meaningful basis for assessing whether it would replicate in new settings. For this reason, public sharing of data and code has become a core element of the open science movement. A growing body of meta-scientific research has examined how often researchers share materials and how often those materials actually allow the published results to be reproduced (Nosek et al., 2015; Hardwicke et al., 2020; Miguel, 2021; Breznau et al., 2025).

This evidence base has expanded rapidly across the social sciences. In economics and management, leading journals now require authors to deposit replication packages, employ data editors to verify reproducibility prior to publication, and have been the focus of several large-scale reproduction exercises (Vilhuber, 2020; Brodeur et al., 2023, 2026; Fišar et al., 2024). In political science, recent audits find that rates of public data and code sharing range from about 21% across the discipline as a whole to 74% in top-tier journals (Stockemer, Koehler and Lentz, 2018; Scoggins and Robertson, 2024; Rainey et al., 2025). Similar audits have been carried out in demography (Basellini, 2024), criminology (Greenspan, Baggett and Boutwell, 2024), finance (Pérignon et al., 2024), management (Prosser et al., 2024), and psychology (Hardwicke et al., 2020, 2021). Across this literature, two patterns consistently emerge. First, posting rates increase substantially where journal policies require it. Second, even when materials are publicly available, a non-trivial share of published results cannot be reproduced from them.

Sociology has remained largely outside this growing audit literature. The discipline has long called for greater transparency (Winship, 2007; Freese, 2007; Firebaugh, 2007), and a growing body of commentary argues that sociology lags neighboring fields in adopting open science practices (Freese and Peterson, 2017; Moody, Keister and Ramos, 2022; Weeden, 2023; Miske et al., 2026). Survey evidence supports this view: Ferguson et al. (2023) find that sociologists report lower rates of data sharing, code sharing, and preregistration than psychologists, economists, or political scientists, despite expressing broadly favorable attitudes toward these practices. The institutional context also differs in important ways. The American Sociological Association's (2018) Code of Ethics states that sociologists "share data and pertinent documentation as an integral part of a research plan," but leaves implementation largely to individual researchers rather than journals or repositories. Among 11 sociology journals reviewed by

Weeden (2023), only one mandates data and code sharing; eight encourage it, and two are silent. Despite these signs of an availability gap, systematic, article-level evidence on what is actually being shared, whether sharing varies by research method, and whether shared materials enable the reproduction of published results, remains limited (Auspurg and Brüderl, 2022; Basellini, 2024).

This paper addresses that gap. We examine 730 empirical articles published between January 2019 and June 2024 in three leading general-interest sociology journals: *American Sociological Review*, *American Journal of Sociology*, and *Social Forces*. For each article, we record whether a publicly accessible replication package is provided and document its contents. For articles that provide such materials, we attempt to computationally reproduce the tables and figures in the main text using the supplied data and code, classifying each table or figure according to the framework developed by Fišar et al. (2024).

Three findings stand out. First, only 9.9% of articles in our sample provide a publicly accessible replication package, with adoption concentrated almost entirely in *American Sociological Review* (22.9%). *American Journal of Sociology* and *Social Forces* each post packages for only about 5% of articles. These rates fall well below those reported in recent audits of political science and economics. Second, among the 72 packages we examined in detail, more than half are effectively unverifiable due to missing data, missing code, or insufficient documentation. Third, among the packages we could run, 22% of articles are fully or largely reproducible at the article level, and reproducibility outcomes vary by research type, with experimental studies substantially more reproducible than observational ones.

These findings provide the first systematic, article- and package-level evidence on transparency in sociology and situate the discipline within a broader cross-disciplinary context. They also speak to an important question in ongoing journal policy debates: what would change if sociology journals adopted the practices that have improved availability and reproducibility in neighboring fields? Our results point to substantial room for improvement, but also suggest that effective standards will need to accommodate the methodological pluralism that characterizes sociology, including the sizable share of work for which a conventional data-and-code deposit is not a natural unit of transparency. We return to this issue in the discussion.

The remainder of the paper is organized as follows. Section 2 positions our study within the broader audit literature and the institutional context of sociology. Section 3 describes the sample and our coding procedure. Section 4 presents findings on data and code availability, as well as computational reproducibility. Section 5 considers the implications for journal policy, the limits of what package audits can reveal, and potential directions for transparency practices tailored to sociology's research traditions. Section 6 concludes.

2 Background

2.1 Cross-Disciplinary Audits of Data Sharing and Reproducibility

Empirical audits of open science practices tend to follow a common template. Researchers draw a sample of articles from a defined set of journals and a time period, code each paper for whether it shares data, code, or both, and then attempt to reproduce at least a subset of the reported results. Implementation details vary—coding may be automated or manual, reproduction attempts may be strictly "push-button" or allow limited adjustments, and output matching may be more or less stringent. Despite these differences, the literature has begun to converge on a set of stylized facts.

In economics, journal policy has shifted markedly over the past decade. The American Economic Association now requires authors to deposit replication packages prior to publication and verifies reproducibility through dedicated data editors (Vilhuber, 2020; American Economic Association, 2024). Reflecting these policies, recent samples from leading economics journals report posting rates approaching 100% (Matthews and Rantanen, 2025). Reproducibility, however, remains imperfect even when materials are available. Brodeur et al. (2026), reproducing 110 papers in leading economics and political science journals from 2022 onward, find that more than 85% are fully computationally reproducible but identify coding errors in approximately 25% of cases. Similarly, Fišar et al. (2024) report for *Management Science* that, when data access and software pose no obstacles, 95% of articles can be fully or largely reproduced. Data constraints, stemming from proprietary, restricted-use, or subscription-based sources, reduce the overall reproduction rate to 68%.

In political science, audit results vary with the sample and methodology. Stockemer, Koehler and Lentz (2018) examine quantitative articles published in three journals in 2015 and find that, although 56.6% of authors share data upon private request, only 16% deposit data in a publicly accessible form. Using machine learning to classify roughly 25,000 articles across 160 political science and international relations journals between 2010 and 2021, Scoggins and Robertson (2024) estimate that 21% can be linked to publicly available data, rising from 11% in 2010 to 26% by 2021. More recent hand-coded studies show a similar overall pattern but with sharp variation by journal tier. Rainey et al. (2025) report that 31% of quantitative articles published in 2022 share supporting data and code, with the rate reaching 74% in top-quartile journals.

In psychology, Hardwicke et al. (2024) analyze 400 articles published in 2022, half drawn from the field at large and half from the top 50 journals by impact factor, and distinguish between claimed availability and *functional* availability, meaning that materials can be accessed without contacting the authors. Functionally available data appear in 14% of field-wide articles and 16% of top-journal articles, while analysis code is available in 8.5% and 14%, respectively. These figures mark clear improvements over the 2014–2017 baseline reported by Hardwicke

et al. (2020), but most psychological research still does not provide the materials needed for independent verification.

Comparable audits in other fields reveal similar patterns. In demography, Basellini (2024) examines four leading journals between 2010 and 2023 and finds a sharp within-discipline contrast. In three journals without dedicated reproducibility infrastructure, about 12% of articles provide open code, whereas *Demographic Research*, which introduced a "Replicable" article category, reaches roughly 31%. In criminology, Greenspan, Baggett and Boutwell (2024) report that 7.7% of articles in five leading journals make data openly available and just 1.3% share analysis code. This is notable because Chin et al. (2023) find that 43% of criminologists report having shared data at some point in their careers. This gap between self-reported behavior and observed practice recurs across disciplines (Young and Horvath, 2015; Ferguson et al., 2023).

The recent cross-field assessment comes from Miske et al. (2026), who examine 600 papers published between 2009 and 2018 across 62 journals in the social and behavioral sciences. Their findings confirm the field-specific patterns documented above. Data availability was highest in political science (54.1%) and economics (37.3%) and lowest in education (2.9%). Among papers for which a reproduction could be attempted, 84.0% of political science papers and 77.2% of economics papers achieved approximate or precise reproducibility, substantially outpacing other disciplines.

Two patterns emerge from this literature. First, posting rates respond strongly to journal policy. Where mandates are in place and enforced, sharing is nearly universal. Where they are absent, it remains the exception, and the contrast between *Demographic Research* and its peer journals shows that this pattern can operate even within a single discipline. Second, posting alone does not guarantee reproducibility. Even in fields with high posting rates, a meaningful share of articles cannot be reproduced from the supplied materials, and data unavailability, rather than coding errors, is the most common barrier. Both patterns motivate the design of the present study, which separates availability from reproducibility and examines them in turn.

2.2 The Institutional Context of Sociology

Sociology's institutional context for data sharing differs from that of neighboring disciplines in several respects. The American Sociological Association's (2018) Code of Ethics includes a section on data sharing (§12.5) stating that sociologists "share data and pertinent documentation as an integral part of a research plan," while allowing exceptions for proprietary agreements and for cases in which confidentiality cannot be protected. This reflects a longstanding professional norm and explicitly recognizes that some forms of research, such as ethnography and interview-based work, may not be readily shareable. At the same time, the Code does not specify when or where data should be shared, and responsibility for implementation rests

with individual researchers rather than journals or repositories.

Journal policies in sociology vary widely. Weeden (2023) reviews 11 empirical sociology journals and finds that only one requires data and code sharing as a condition of publication; eight encourage it but allow authors to opt out, and two do not address the issue. Online Appendix A reviews the policies of 16 general and specialist sociology journals, including the three analyzed here. The submission guidelines for *American Sociological Review* reproduce the ASA Code's data-sharing language but do not mandate the deposit of replication materials. *American Journal of Sociology* encourages sharing and notes that editors and reviewers may take data availability into account when evaluating submissions, but does not require it. *Social Forces* introduced a data availability statement requirement in 2023, asking authors to declare their stance on sharing, but statements may indicate that data are "available upon request" rather than publicly deposited. This formulation has been widely criticized as difficult to enforce. Young and Horvath (2015) report that more than three-quarters of authors of sociological articles who had promised to provide data on request declined to do so when asked, often for reasons unrelated to legal or confidentiality constraints. Freese (2007) further notes that such promises can expire when authors leave the field, leaving readers without remedy.

A distinct feature of sociology is the breadth of its methodological approaches. Unlike economics, where empirical work is predominantly quantitative, sociology's leading journals regularly publish qualitative, ethnographic, historical, and mixed-methods research alongside quantitative analyses (Lucas, Morrell and Posard, 2013; Weeden, 2023). This pluralism complicates the uniform application of data-sharing requirements. In ethnographic research, for example, the materials that constitute "data," such as field notes, recordings, and interview transcripts, are often highly identifying and difficult to anonymize. The analytic process is also iterative and interpretive in ways that do not map cleanly onto computational reproduction (Murphy, Jerolmack and Smith, 2021; Khan, Hirsch and Zeltzer-Zubida, 2024). The ASA Code's allowance for cases in which confidentiality cannot be preserved reflects these challenges. In this study, we therefore distinguish between transparency standards for quantitative research, where computational reproducibility is well defined and where our reproduction efforts are concentrated, and the broader question of transparency in qualitative work, for which our evidence is descriptive rather than evaluative.

Attention to transparency is not new within sociology. A special section of *Sociological Methods & Research* on replication and data was published in 2007 (Winship, 2007; Freese, 2007; Firebaugh, 2007), and Moody, Keister and Ramos (2022) discuss reproducibility specifically in the context of computational social science. Zenk-Möltgen and Lepthien (2014) examine the availability of replication packages at ASR and AJS for articles published in 2012–2013, finding that 30.9% of empirical ASR articles and 9.1% of empirical AJS articles had data that could be verified as accessible. A cross-disciplinary benchmark comes from Miske et al. (2026), who include sociology among six fields in a broad audit of 600 papers across 62 social and behavioral science journals published between 2009 and 2018, finding that data and code sharing

Table 1: Articles published in three sociology journals (January 2019–June 2024), by research type.

	All		SF		ASR		AJS	
	N	Share	*N*	Share	*N*	Share	*N*	Share
Research method								
Observational	492	0.67	293	0.80	109	0.56	90	0.54
Qualitative	140	0.19	40	0.11	51	0.26	49	0.29
Mixed methods	49	0.07	13	0.04	18	0.09	18	0.11
Experimental	43	0.06	21	0.06	14	0.07	8	0.05
Other	6	0.01	1	0.00	2	0.01	3	0.02
Total	730	1.00	368	1.00	194	1.00	168	1.00

Notes: ASR = *American Sociological Review*, AJS = *American Journal of Sociology*, and SF = *Social Forces*.

in sociology runs at roughly 20%. By design, however, that study offers breadth rather than depth. Given the discipline's distinctive institutional context, these questions deserve closer attention than the cross-disciplinary literature has yet provided. The present study extends this line of work by combining a systematic accounting of package availability with article- and output-level reproduction attempts specifically within sociology, across its three leading general-interest outlets, with a focus on how practices vary across method types.

3 Data

Sample. We examine all empirical research articles published between January 2019 and June 2024 in three leading sociology journals: *American Sociological Review* (ASR, Vols. 84–89), *American Journal of Sociology* (AJS, Vols. 124–129), and *Social Forces* (SF, Vols. 97–102). These journals are widely regarded as the discipline's leading general-interest outlets and collectively reflect the field's methodological breadth. We exclude book reviews, editorials, and six articles that contain no empirical analysis. The final sample consists of 730 articles.

Each article is manually classified into one of five mutually exclusive categories based on its primary methodological approach. *Observational* studies rely on one or more non-experimental quantitative datasets. *Experimental* studies involve manipulation of conditions in the data-generating process, including laboratory, field, and survey experiments. *Qualitative* studies use methods such as ethnography, interviews, or content analysis. *Mixed-methods* studies combine quantitative and qualitative approaches within a single article. *Other* studies include review articles, meta-analyses, and simulation studies. Because these categories are defined differently across the sociological literature (Small, 2011), we provide detailed coding criteria in Online Appendix B.3. Table 1 reports the resulting distribution.

Identifying replication packages. As sociology journals do not follow a standardized format for declaring data and code availability, identifying replication materials requires manual review. We proceeded in two steps. First, we conducted a full-text keyword search of each article for terms associated with data and code availability, such as `data is available`, `replication package`, `Open Science Framework`, and `Dataverse`. The complete list of keywords is provided in Table B.1 in the Online Appendix. Second, we manually reviewed each article to verify and supplement the keyword search. For articles that mention availability without providing a direct link, we conducted a Google search using the article title and author names to locate the materials. We classified a package as *publicly accessible* if it can be reached without authentication and contains files relevant to the published analysis. Materials described as "available upon request," as well as supplementary files that include only documentation, broken links, or links to materials that are no longer available, are coded as not publicly accessible.

Assessing package contents. For each publicly accessible package, we downloaded all available files in November and December 2024 and reviewed them to assess the completeness of the materials. This step is important because replication packages often do not include everything needed to fully reproduce the published results. For each package, we recorded whether it includes (i) analysis code, (ii) the data required to run that code, (iii) a README file or equivalent documentation, and (iv) log files documenting code execution. For each component, we further coded whether it is complete, partial, or absent.

We classified a package as *verifiable* if both code and data are fully or partially available. Packages missing one of these components were coded as *unverifiable*. We also recorded the apparent reason for the missing data, distinguishing among: (i) the data are restricted and cannot be obtained by the replicator; (ii) the data can be obtained for free from publicly available sources; (iii) the data can be obtained from a commercial provider; (iv) the data may be obtainable through alternative means (e.g., the authors may provide the data upon request, particularly when the data are original and there is no clear barrier to sharing); and (v) the data cannot be obtained, or the reason for unavailability is unknown.

Reproducibility assessment. For verifiable packages, we attempt to computationally reproduce each table and figure in the article's main text. We do not reproduce content in supplementary materials or appendices, and we do not re-execute data-cleaning pipelines when only cleaned analysis data are provided. The unit of reproduction is the analysis as documented by the package, rather than the full pipeline from raw data. We conducted all reproduction attempts using only the data, code, and materials included in each replication package, without downloading any external resources referenced in the README files. Within these bounds, we follow the conventions of recent reproduction exercises (Trisovic et al., 2022; Fišar et al., 2024; Brodeur et al., 2026). We run the supplied scripts, making only minimal corrections for

clear typographical or file-path errors, and then assess whether the resulting outputs match those reported in the published article. We excluded two packages from the analysis that exceeded our computational capacity. See Online Appendix B.2 for a description of the computational environment and the scope of the reproduction exercise.

Following Fišar et al. (2024), each table and figure is classified into one of seven categories: (i) fully reproducible; (ii) largely reproducible, with minor issues; (iii) largely not reproducible, with major issues; (iv) not reproducible; (v) not reproducible but consistent with the provided log file; (vi) unverifiable; and (vii) not based on data or analysis. The full set of classification criteria is provided in Online Appendix B.4. We then aggregate these classifications to the article level, counting an article as fully reproducible only if all of its main-text tables and figures are fully reproducible.

The reproduction sample includes 283 figures and 252 tables drawn from the set of verifiable packages. We also document any corrections made during the process and report the share of verifiable articles that required minor adjustments in order to run.

4 Results

The full sample contains 728 articles. Of these, 72 articles (9.9%) provide a publicly accessible package. Of those 72, 32 packages (44%) are verifiable in the sense that they include both code and data sufficient to attempt reproduction; the remaining 40 cannot be run as supplied. Reproduction outcomes are reported below at the article level (72 packages) and at the level of individual tables and figures (535 exhibits drawn from verifiable packages).

4.1 Availability of Replication Packages

Overall and by journal. Across the 728 articles for which replication package status could be determined, 9.9% provide a publicly accessible package containing data, code, or both. Adoption is heavily concentrated in ASR, where 22.9% of articles share materials. In contrast, AJS and SF provide packages for only 5.4% and 5.2% of articles, respectively (Figure 1A). These cross-journal differences are statistically significant ($\chi^2(2) = 49.66$, $p < 0.001$). These disparities cannot be explained solely by differences in methodological composition. ASR and AJS have similar distributions across research types (Fisher's exact test, $p = 0.790$), yet their posting rates differ by more than 17 percentage points (Fisher's exact test, $p < 0.001$). The gap is large enough that ASR accounts for roughly 61.1% of all replication packages in our sample.

Trends over time. Posting rates increase modestly over the sample period, rising from 7.6% in 2019 to 12.3% in 2023 and 12.7% among articles published through June 2024 (Figure 1B). This trend is consistent with the broader shift toward open science practices observed in other social science fields, although the overall level remains low. The 2023 policy change at

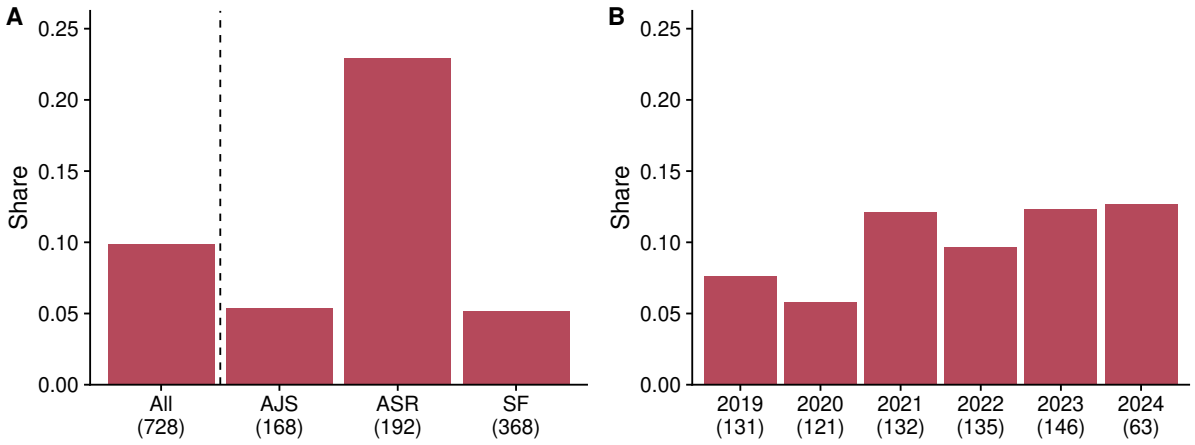

Figure 1: Share of empirical articles that provide a publicly accessible replication package. (A) By journal. (B) By year of publication. *Notes*: Numbers in parentheses indicate the number of articles in each category. The 2024 bar in panel B reflects articles published through June 2024.

Social Forces is informative in this regard. That year, the journal introduced a data availability statement requirement, asking authors to declare their position on sharing without requiring deposit. The most common statement indicates that data are "available upon request." Despite this relatively lenient design, the share of SF articles with a publicly accessible package increased from 10 of 302 articles (3.3%) before the policy to 9 of 66 (13.6%) after its introduction, representing a fourfold increase from a low baseline (Fisher's exact test, $p = 0.002$). Although the absolute level remains modest, this is one of the largest year-to-year changes in our data and suggests that even relatively soft transparency requirements can influence behavior. Whether this increase will persist and whether "available upon request" leads to meaningful access in practice remain open questions.

By research type. Posting rates differ markedly by research type. None of the 140 qualitative articles in our sample provides a publicly accessible replication package. Observational and mixed-methods articles, which make up most of the sample, post packages at rates of about 10%, close to the overall figure. Experimental articles have the highest rate, with 30% (of 43 articles) providing materials. Among the relatively small number of mixed-methods articles that do share materials, the posted content typically covers only the quantitative components.

The result for qualitative research requires careful interpretation. As discussed in Section 2.2, the ASA Code of Ethics explicitly acknowledges that field notes and other ethnographic materials cannot always be shared in the same way as structured datasets. Qualitative scholars have also argued that conventional standards of computational reproducibility are not well suited to evaluating interpretive research (Murphy, Jerolmack and Smith, 2021; Weeden, 2023; Khan, Hirsch and Zeltzer-Zubida, 2024). The absence of publicly accessible packages, therefore, reflects a practical constraint: qualitative articles in our sample do not provide materials in a form that supports the type of computational reproduction we study here. It should not be read as a judgment about the value or feasibility of other forms of

Table 2: Availability of replication materials in recent cross-disciplinary audits.

Discipline	Sample	Data	Code	Rate	Source
Sociology	ASR, AJS, SF, 2019–2024	✓	✓	10%	This study
Economics	Top 5 journals, 2018–2022	✓		99%	Matthews and Rantanen (2025)
Political science	160 journals, 2010–2021	✓		21%	Scoggins and Robertson (2024)
Political science	All journals, 2022 (overall)	✓	✓	31%	Rainey et al. (2025)
Political science	Top-quartile journals, 2022	✓	✓	74%	Rainey et al. (2025)
Political science	Top 3 journals, 2018–2022	✓		99%	Matthews and Rantanen (2025)
Law	Top 20 general law review journals, 2018–2022	✓		17%	Matthews and Rantanen (2025)
Demography	3 journals without reproducibility track, 2010–2023		✓	12%	Basellini (2024)
Demography	*Demographic Research*, 2010–2023		✓	31%	Basellini (2024)
Criminology	5 top journals, 2018–2022	✓		8%	Greenspan, Baggett and Boutwell (2024)
Psychology	Field-wide sample, 2022	✓		14%	Hardwicke et al. (2024)
Psychology	Top 50 journals, 2022	✓		16%	Hardwicke et al. (2024)

transparency, an issue we revisit in Section 5. For comparison with other disciplines, where audit samples typically include little or no qualitative work, we also report posting rates for quantitative studies alone. Excluding the 140 qualitative articles, 12.2% of the remaining 588 articles provide a replication package. This figure still remains below the corresponding rates reported in psychology, economics, and political science.

Sociology in cross-disciplinary perspective. Table 2 compares the posting rate in our sample with recent audits from neighboring fields. Direct comparisons are complicated by differences in sampling frames, time periods, and coding approaches, so these results should be interpreted as rough orders of magnitude rather than precise benchmarks. Even so, three patterns are clear. First, sociology's 9.9% rate sits between the two most adjacent fields: the 7.7% data-sharing rate reported for criminology (Greenspan, Baggett and Boutwell, 2024) and the roughly 12% posting rate documented by Basellini (2024) for demography journals without dedicated reproducibility infrastructure. Given how closely these fields sit to sociology, this clustering is reassuring: our estimate falls where one would expect it to fall. Second, the comparison with demography also highlights the importance of journal-level policy within a single discipline. While the three demography journals studied by Basellini (2024) post at about 12%, *Demographic Research*, which introduced a "Replicable" article category, reaches 31%. Third, sociology lags well behind political science and economics. Audits of political science report overall sharing rates ranging from 21% (Scoggins and Robertson, 2024) to 31% (Rainey et al., 2025), with the latter reaching 74% in top-quartile journals, while pre-publication verification at leading economics journals has pushed posting rates close to universal levels (Matthews and Rantanen, 2025). Sociology also remains below the 14–16% functional data availability documented in psychology in 2022 (Hardwicke et al., 2024). These patterns are broadly consistent with the findings of Miske et al. (2026).

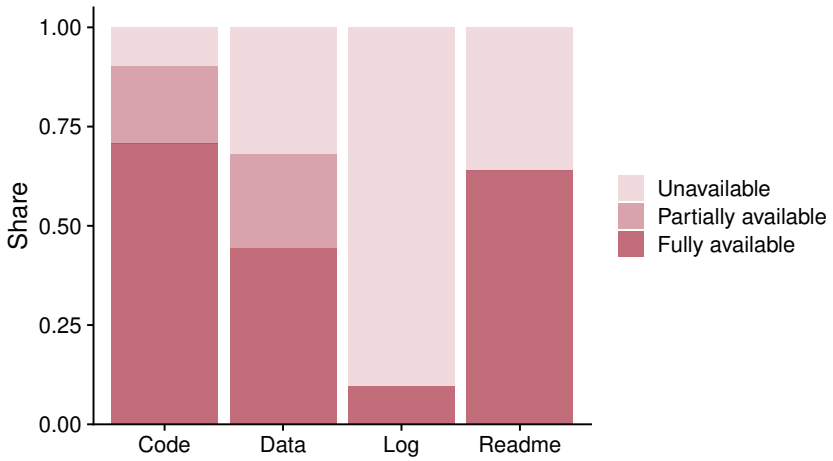

Figure 2: Availability of core replication components across 72 publicly accessible packages. *Notes*: Each bar reports the share of packages in which the indicated component is fully present, partially present, or absent.

4.2 Computational Reproducibility

We now turn to what the 72 publicly accessible packages in our sample actually contain and whether they support reproducing the published results.

What packages contain. Figure 2 summarizes the availability of the four core components of a replication package: code, data, README files, and log files. Approximately 90% of packages include analysis code, while the remaining 10% include only data or supplementary materials that do not support a runnable analysis. Among 65 packages that include code, 48% rely exclusively on Stata, 23% on R, 20% use a combination of Stata and R, and 6% use Python for at least part of the analysis. Data availability is lower than code availability. Roughly 30% of packages do not include analysis data, most often because the underlying datasets are restricted-use or proprietary. README files appear in 68% of packages, whereas only about 10% include log files documenting code execution and outputs.

Applying the verifiability criterion defined above, 40 of the 72 packages (56%) are classified as unverifiable because they lack at least one component needed to run the analysis, most often the data. This is a central finding of the reproduction exercise. Even among the relatively small share of articles that provide materials, more than half of the packages do not, on their own, support a reproduction attempt.

Article-level reproducibility. The 72 publicly accessible packages are partitioned into five article-level outcomes. Ten packages (14%) are fully reproducible, meaning that all main-text tables and figures match the published article, and a further six (8%) are largely reproducible with only minor issues. Together, these 16 packages constitute a combined "reproducible" share of 22%. Eleven packages (15%) are partially reproducible: the package supports reproduction of some parts of the analysis but not others, most often because publicly available

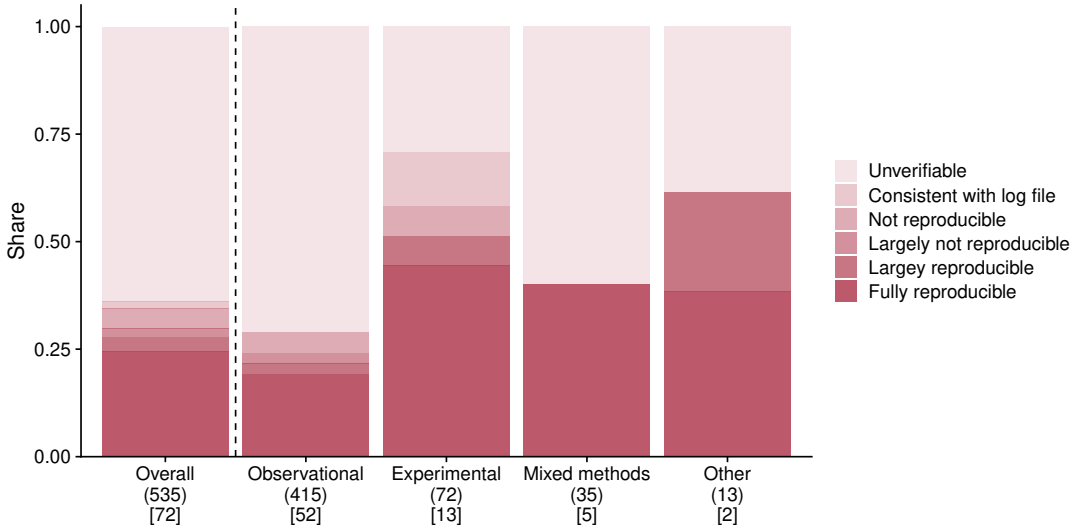

Figure 3: Reproducibility of individual tables and figures, by research type and by journal. *Notes*: Mixed-methods articles are grouped with observational studies because all available replication materials for mixed-methods articles cover the quantitative component. Numbers in parentheses denote the number of exhibits (tables and figures), while numbers in square brackets indicate the number of packages.

data are provided for certain analyses while sensitive or proprietary data are withheld for the rest. Five packages (7%) are runnable but contain at least one exhibit that fails to reproduce despite the materials appearing complete. The remaining 40 packages (56%) are unverifiable, as described above. Notably, no package is entirely non-reproducible at the article level: when reproduction fails, it does so for a subset of results rather than for the article as a whole. These outcomes reflect attempts to reproduce after minor corrections, such as typographical errors, broken file paths, and deprecated function calls, that we made in 38% of verifiable packages. A stricter "push-button" protocol that disallowed any modification would yield lower rates.

Reproducibility at the table and figure level. Figure 3 shifts the unit of analysis to individual tables and figures and disaggregates results by research type and journal. At the exhibit level, reproducibility is highest for experimental research, where 44% of tables and figures from experimental articles are fully reproducible, and lowest for observational research, where 23% of exhibits are fully or partially reproducible, and unverifiability is the most common outcome. This pattern is consistent with the greater reliance on restricted-use or proprietary data in observational sociology. Cross-journal differences are also evident at the output level. Reproducibility is highest in ASR, where 32% of articles in the verifiable subsample are fully reproducible, followed by SF at 28% and AJS at 8%.

Two interpretations of these results are worth distinguishing. An optimistic view is that, among the 32 verifiable packages in our sample, reproduction succeeds at rates broadly in line with those reported for psychology and economics (Hardwicke et al., 2021; Brodeur et al., 2026)—though we caution that direct cross-study comparison is complicated by differences in

sampling frames and reproduction protocols, and our runnable subsample is small and likely non-random. With these caveats, sociology does not appear unusually prone to error in the analyses we were able to run. A more pessimistic view focuses on how rarely such conditions are met. Only a small share of articles provide materials that allow a reproduction attempt, and a smaller share still can be fully reproduced. Taken together, the numbers are low: 14% of replication packages are fully reproducible, and only 9.9% of articles provide such packages. As a result, about 1.4% of the articles in the sample were fully reproducible using the supplied materials. Both interpretations describe the same underlying data. Which one is more relevant depends on the margin one aims to improve.

5 Discussion

5.1 Summary of Findings

This article provides the first systematic, package-level audit of data sharing and computational reproducibility in three leading general-interest sociology journals. Three findings stand out.

First, only 9.9% of empirical articles published between January 2019 and June 2024 provide publicly accessible replication packages. Adoption is concentrated in one journal, *American Sociological Review* (22.9%), and remains very low in the other two, *American Journal of Sociology* (5.4%) and *Social Forces* (5.2%). Second, even among the small minority of articles that share materials, more than half of the packages are not verifiable in practice because they lack data, code, or sufficient documentation. Third, when reproduction can be attempted, success rates are broadly comparable to those reported in economics and psychology. However, the overall yield remains small when measured as the share of fully reproducible articles in the original sample. In short, on the limited evidence available from verifiable packages, sociology does not appear unusually error-prone in the analyses we could check. The more visible gap is that very little is shared in a form that allows checking at all.

5.2 Implications for Journal Policy

The cross-journal pattern in our results points to a clear policy lever. ASR, AJS, and SF are considered top general-interest journals, yet their posting rates differ by an order of magnitude. The most plausible explanation is the policy environment. ASR's submission guidelines follow the ASA Code and signal an expectation of deposit, whereas AJS treats sharing as encouraged but optional, and SF allows authors to opt out with an "available upon request" statement. Evidence from other disciplines supports this interpretation: posting rates rise sharply where mandates are in place and enforced, and remain low where they are not (Vilhuber, 2020; Rainey et al., 2025; Matthews and Rantanen, 2025).

The behavior of SF after its 2023 data availability statement requirement provides the most direct evidence in our data, though we treat it as suggestive given the short post-period and the absence of a comparison group. The policy does not require authors to deposit materials; it simply asks them to state their position. Even so, posting rates more than quadrupled in the following year, rising from 3.3% to 13.6%, despite the most common statement being "available upon request." Although this level remains well below what mandate-and-verify policies achieve in economics, the relative increase mirrors a broader pattern in the policy literature: even soft transparency requirements that make non-sharing visible can shift behavior (Hardwicke et al., 2018; Christensen and Miguel, 2018). Stronger requirements, such as deposit prior to publication, verification by a data editor, and ineligibility for award consideration without a complete package, would likely produce larger effects, as they have in neighboring fields. Whether such policies can be designed to fit the diversity of work published in sociology journals is the more difficult question, to which we now turn.

5.3 What the Reproducibility Findings Imply

Two patterns from our reproduction exercise are worth emphasizing, with appropriate caution given the small verifiable subsample. First, no article in our sample is wholly non-reproducible at the article level: when reproduction failed, it affected only a subset of results, and we did not observe cases where a headline finding reversed direction. A stricter protocol that re-executed data-cleaning pipelines would likely yield lower reproduction rates than ours. Second, when authors provided complete materials, reproduction succeeded at rates broadly in line with recent audits in economics (Brodeur et al., 2026) and management (Fišar et al., 2024). We do not interpret this as evidence that sociology is well-positioned on reproducibility. The more cautious reading is that the posting and verifiability gap, rather than analytic error in posted analyses, is where the discipline lags neighboring fields.

These findings suggest that framing sociology's situation as a reproducibility crisis of the kind that has driven reform efforts in psychology and economics is misleading. The core issue is that few articles share materials, and more than half of those that do are not verifiable in practice. These are distinct problems with distinct remedies. The sharing gap is primarily a policy issue: stronger journal requirements can raise deposit rates, as evidence from other fields shows. The verifiability gap is about documentation and infrastructure: deposit alone is not enough, and even shared packages often lack the data, organization, or instructions needed to run the analysis.

The policy implication is that deposit requirements, while necessary, are not sufficient to close the gap between sociology and neighboring fields. Among the 72 packages in our sample, each of which meets a minimal threshold for public sharing, 56% are still unverifiable. Policies that distinguish between deposit and verification, such as pre-publication checks by data editors used at leading economics journals (Vilhuber, 2020), address both problems at

once. By contrast, policies that require only that authors post materials address the first problem but leave the second largely untouched. Our results suggest that the latter is at least as consequential, since unverifiability remains the most common outcome even among articles that do share materials.

5.4 Transparency for a Methodologically Pluralist Discipline

Sociology's leading journals publish a broader range of empirical work than those in any neighboring discipline that has been systematically audited. This methodological pluralism complicates cross-field comparisons: when sociology's sharing rates fall below those of economics or political science, it is not straightforward to determine how much of the gap reflects slower progress toward open science and how much reflects compositional differences in the kinds of research being produced. Our finding that no qualitative article in the sample includes a replication package can be interpreted in two ways. On the one hand, it represents a striking absence. On the other hand, it reflects well-established professional norms recognizing that ethnographic field notes, interview transcripts, and similar materials often cannot be shared publicly without violating confidentiality commitments. The ASA Code of Ethics explicitly permits such restrictions. For qualitative research, then, the key question is not whether "data and code" should be posted in the same sense as in quantitative analysis, but whether alternative forms of transparency might serve comparable goals. These could include documenting analytic decisions, clearly linking claims to their evidence base, or enabling re-analysis by researchers with appropriate credentials. A growing body of work has begun to develop such approaches (Murphy, Jerolmack and Smith, 2021; Khan, Hirsch and Zeltzer-Zubida, 2024), and some journals, for example, *Administrative Science Quarterly*, provide guidelines to improve transparency in qualitative research (see Online Appendix A). Our results provide an empirical baseline for tracking their adoption.

For the quantitative and experimental studies that constitute the majority of articles in our sample, the case for stronger sharing requirements is more straightforward, and cross-disciplinary evidence on their effects is now substantial. Two design choices are particularly important. First, policies should distinguish between the mere deposit of materials and their verifiability. Our finding that 56% of posted packages are not verifiable suggests that posting alone, without verifying that the materials reproduce the published analysis, leaves much of the potential benefit unrealized. The use of dedicated data editors in economics journals offers one concrete model. Second, policies should account for restricted-use and proprietary data, which constitute a significant share of sociological data sources. Rather than requiring full deposit of all source materials, policies can specify what should be shared, such as cleaned analysis datasets, code, and clear documentation of how the data can be accessed. Existing standards, such as the Data and Code Availability Standard (Koren et al., 2022), provide useful templates.

A third design choice concerns mixed-methods research. Articles that combine quantitative and qualitative approaches often fall between existing policy categories. Their quantitative components can typically be shared using standard practices, whereas their qualitative components may not. Our results suggest that the current outcome is often that neither is shared. A practical compromise would be to require the deposit of quantitative components under the same standards applied to purely quantitative work, while allowing qualitative components to follow whatever transparency practices the field develops for that type of research.

5.5 Limitations

This study has several limitations. First, our sample is limited to three general-interest journals over five and a half years; patterns may differ in specialized subfield journals or at other tiers. The broad pattern we identify, with substantial gaps relative to neighboring fields, alongside upward trends and a clear policy signal in the SF case, is likely to generalize, but the precise levels may not. Second, our reproduction protocol is intentionally permissive: we correct minor errors but do not re-execute data-cleaning pipelines. A stricter "push-button" standard would likely yield lower reproducibility rates. Third, our reproduction attempts focus on results reported in the main text; reproducibility may differ for supplementary analyses or robustness checks. Fourth, our coding of qualitative articles shows that they do not provide the types of replication packages we audit, but it does not imply that their findings are unverifiable in a broader sense. Appropriate transparency standards for qualitative research remain an open question that our methodology cannot resolve. Finally, audits like this one capture observable artifacts, namely what authors share and what can be reproduced, rather than underlying behavior, such as what authors would be willing to share if asked or why they choose not to. Survey evidence (Young and Horvath, 2015; Ferguson et al., 2023) offers complementary insight into these issues, and integrating these approaches would be a valuable direction for future research.

6 Conclusion

The findings reported here provide the first systematic, package-level baseline for tracking sociology's progress on data sharing and computational reproducibility. Three patterns stand out. First, sociology shares far less than its quantitative neighbors: only 9.9% of articles in our sample provide a publicly accessible replication package. This rate falls well below those in political science, demography, and psychology, and is an order of magnitude lower than in top economics journals. Second, sharing is uneven both within and across sociology, with one journal, ASR, accounting for roughly 61.1% of all packages we identified. Third, among the small minority of articles that do share materials, more than half of the packages cannot support an independent reproduction attempt.

The picture is not entirely negative. No article in our sample is wholly non-reproducible at the article level, and when authors provide complete materials, reproduction succeeds at rates comparable to those reported in economics and management. The central issue in sociology is not that its analyses fail to reproduce, but that very few analyses are made checkable in the first place, and that those that are often lack the documentation needed to run them.

These two problems, low rates of sharing and low verifiability, call for different solutions. The sharing problem responds to the journal policy. The experience of *Social Forces* following its 2023 data availability statement, where posting more than quadrupled despite the absence of a deposit requirement, suggests that even relatively soft policies can shift behavior. Stronger requirements, such as a mandatory deposit prior to publication and verification by a data editor, would likely have a larger effect. The verifiability problem, by contrast, cannot be solved by deposit alone. It requires that replication packages include working code, the necessary data, and clear documentation linking scripts to published results. The data editor model, now standard in leading economics journals, addresses both challenges simultaneously.

Designing standards that accommodate sociology's methodological diversity is the key task moving forward. The ASA Code of Ethics already recognizes that some forms of data cannot be shared in conventional ways, and our finding that no qualitative article in the sample provides a replication package reflects, in part, professional norms that prioritize confidentiality. The risk is that this recognition becomes a default exemption rather than a principled boundary. Effective standards would distinguish between quantitative and qualitative components, specify what should be shared under which conditions, and align policy tools with the problems they are intended to address. Such an approach would align sociology's transparency practices with the demands of its empirical claims, without imposing uniform requirements on a discipline whose strength lies in its diversity.

Sociologists have debated transparency for at least two decades (Winship, 2007; Freese, 2007; Freese and Peterson, 2017; Weeden, 2023). The question is no longer whether transparency matters, but how to implement it in practice. The findings reported here clarify both the scale of the gap and the policy levers most likely to close it. Whether the discipline acts on this evidence is now a decision for its leading journals.

References

American Economic Association. 2024. "Data and Code Availability Policy." https://www.aeaweb.org/journals/data/data-code-policy (accessed May 1, 2026).

American Sociological Association. 2018. "Code of Ethics." https://www.asanet.org/wp-content/uploads/asa_code_of_ethics-june2018.pdf (accessed May 1, 2026).

Auspurg, Katrin, and Josef Brüderl. 2022. "How To Increase Reproducibility and Credibility of Sociological Research." In *Handbook of Sociological Science*. 512–527. Edward Elgar Publishing.

Basellini, Ugofilippo. 2024. "Open Science Practices in Demographic Research: An Appraisal." *Demographic Research*, 50: 1265–1280.

Breznau, Nate, Eike Mark Rinke, Alexander Wuttke, Muna Adem, Jule Adriaans, Esra Akdeniz, Amalia Alvarez-Benjumea, Henrik K. Andersen, Daniel Auer, Flavio Azevedo, et al. 2025. "The Reliability of Replications: A Study in Computational Reproductions." *Royal Society Open Science*, 12(3): 241038.

Brodeur, Abel, Anna Dreber, Fernando Hoces de la Guardia, and Edward Miguel. 2023. "Replication Games: How to Make Reproducibility Research More Systematic." *Nature*, 621(7980): 684–686.

Brodeur, Abel, Derek Mikola, Nikolai Cook, Lenka Fiala, Thomas Brailey, Ryan Briggs, Alexandra De Gendre, Yannick Dupraz, Jacopo Gabani, Romain Gauriot, et al. 2026. "Reproducibility and Robustness of Economics and Political Science Research." *Nature*, 652(8108): 151–156.

Camerer, Colin F., Anna Dreber, Eskil Forsell, Teck-Hua Ho, Jürgen Huber, Magnus Johannesson, Michael Kirchler, Johan Almenberg, Adam Altmejd, Taizan Chan, Taisuke Imai, et al. 2016. "Evaluating Replicability of Laboratory Experiments in Economics." *Science*, 351(6280): 1433–1436.

Camerer, Colin F., Anna Dreber, Felix Holzmeister, Teck-Hua Ho, Jürgen Huber, Magnus Johannesson, Michael Kirchler, Gideon Nave, Brian A. Nosek, Thomas Pfeiffer, et al. 2018. "Evaluating the Replicability of Social Science Experiments in *Nature* and *Science* between 2010 and 2015." *Nature Human Behaviour*, 2(9): 637–644.

Chin, Jason M., Justin T. Pickett, Simine Vazire, and Alex O. Holcombe. 2023. "Questionable Research Practices and Open Science in Quantitative Criminology." *Journal of Quantitative Criminology*, 39(1): 21–51.

Christensen, Garret, and Edward Miguel. 2018. "Transparency, Reproducibility, and the Credibility of Economics Research." *Journal of Economic Literature*, 56(3): 920–980.

Dreber, Anna, and Magnus Johannesson. 2025. "A Framework for Evaluating Reproducibility and Replicability in Economics." *Economic Inquiry*, 63(2): 338–356.

Ferguson, Joel, Rebecca Littman, Garret Christensen, Elizabeth Levy Paluck, Nicholas Swanson, Zenan Wang, Edward Miguel, David Birke, and John-Henry Pezzuto. 2023. "Survey of Open Science Practices and Attitudes in the Social Sciences." *Nature Communications*, 14(1): 5401.

Firebaugh, Glenn. 2007. "Replication Data Sets and Favored-Hypothesis Bias: Comment on Jeremy Freese (2007) and Gary King (2007)." *Sociological Methods & Research*, 36(2): 200–209.

Fišar, Miloš, Ben Greiner, Christoph Huber, Elena Katok, Ali I. Ozkes, and Management Science Reproducibility Collaboration. 2024. "Reproducibility in Management Science." *Management Science*, 70(3): 1343–1356.

Freese, Jeremy. 2007. "Replication Standards for Quantitative Social Science: Why Not Sociology?" *Sociological Methods & Research*, 36(2): 153–172.

Freese, Jeremy, and David Peterson. 2017. "Replication in Social Science." *Annual Review of Sociology*, 43: 147–165.

Greenspan, Rachel Leigh, Logan Baggett, and Brian B. Boutwell. 2024. "Open Science Practices in Criminology and Criminal Justice Journals." *Journal of Experimental Criminology*, 1–21.

Hardwicke, Tom E., Joshua D. Wallach, Mallory C. Kidwell, Theiss Bendixen, Sophia

Crüwell, and John Ioannidis. 2020. "An Empirical Assessment of Transparency and Reproducibility-Related Research Practices in the Social Sciences (2014–2017)." *Royal Society Open Science*, 7(2): 190806.

Hardwicke, Tom E., Manuel Bohn, Kyle MacDonald, Emily Hembacher, Michèle B. Nuijten, Benjamin N. Peloquin, Benjamin E. Demayo, Bria Long, Erica J. Yoon, and Michael C. Frank. 2021. "Analytic Reproducibility in Articles Receiving Open Data Badges at the Journal Psychological Science: An Observational Study." *Royal Society Open Science*, 8(1): 201494.

Hardwicke, Tom E., Maya B. Mathur, Kyle MacDonald, Gustav Nilsonne, George C. Banks, Mallory C. Kidwell, Alicia Hofelich Mohr, Elizabeth Clayton, Erica J. Yoon, Michael Henry Tessler, et al. 2018. "Data Availability, Reusability, and Analytic Reproducibility: Evaluating the Impact of a Mandatory Open Data Policy at the Journal Cognition." *Royal Society Open Science*, 5(8): 180448.

Hardwicke, Tom E., Robert T. Thibault, Beth Clarke, Nicholas Moodie, Sophia Crüwell, Sarah R. Schiavone, Sarah A. Handcock, Khanh An Nghiem, Fallon Mody, Tuomas Eerola, et al. 2024. "Prevalence of Transparent Research Practices in Psychology: A Cross-Sectional Study of Empirical Articles Published in 2022." *Advances in Methods and Practices in Psychological Science*, 7(4): 25152459241283477.

Holzmeister, Felix, Magnus Johannesson, Colin F. Camerer, Yiling Chen, Teck-Hua Ho, Suzanne Hoogeveen, Juergen Huber, Noriko Imai, Taisuke Imai, Lawrence Jin, et al. 2025. "Examining the Replicability of Online Experiments Selected by a Decision Market." *Nature Human Behaviour*, 9(2): 316–330.

Khan, Shamus, Jennifer S. Hirsch, and Ohad Zeltzer-Zubida. 2024. "A Dataset Without a Code Book: Ethnography and Open Science." *Frontiers in Sociology*, 9: 1308029.

Koren, Miklós, Marie Connolly, Joan Llull, and Lars Vilhuber. 2022. "Data and Code Availability Standard [Version 1.0]." https://doi.org/10.5281/zenodo.7436134 (accessed May 1, 2026).

Lucas, Jeffrey W., Kevin Morrell, and Marek Posard. 2013. "Considerations on the 'Replication Problem' in Sociology." *The American Sociologist*, 44: 217–232.

Matthews, Abigail A., and Jason Rantanen. 2025. "Legal Research as a Collective Enterprise: An Examination of Data Availability in Empirical Legal Scholarship." *The Journal of Law, Economics, and Organization*, 41(2): 570–603.

Miguel, Edward. 2021. "Evidence on Research Transparency in Economics." *Journal of Economic Perspectives*, 35(3): 193–214.

Miske, Olivia, Anna Lou Abatayo, Mason Daley, Mirka Dirzo, Nicholas Fox, Noah Haber, Krystal M. Hahn, Melissa Kline Struhl, Brinna Mawhinney, Priya Silverstein, et al. 2026. "Investigating the Reproducibility of the Social and Behavioural Sciences." *Nature*, 652(8108): 126–134.

Moody, James W., Lisa A. Keister, and Maria C. Ramos. 2022. "Reproducibility in the Social Sciences." *Annual Review of Sociology*, 48: 65–85.

Murphy, Alexandra K., Colin Jerolmack, and DeAnna Smith. 2021. "Ethnography, Data Transparency, and the Information Age." *Annual Review of Sociology*, 47(1): 41–61.

Nosek, Brian A., George Alter, George C. Banks, Denny Borsboom, Sara D. Bowman, Steven J. Breckler, Stuart Buck, Christopher D. Chambers, Gilbert Chin, Garret Christensen, et al. 2015. "Promoting an Open Research Culture." *Science*, 348(6242): 1422–1425.

Nosek, Brian A., Timothy M. Errington, Noah Haber, Theresa Stankov, and Andrew H. Tyner. 2026. "A Brief Glossary of Terms about Repeatability: Replicability, Robustness, and Reproducibility." MetaArXiv. https://doi.org/10.31222/osf.io/mqfp4_v1.

Open Science Collaboration. 2015. "Estimating the Reproducibility of Psychological Science." *Science*, 349(6251): aac4716.

Pérignon, Christophe, Olivier Akmansoy, Christophe Hurlin, Anna Dreber, Felix Holzmeister, Jürgen Huber, Magnus Johannesson, Michael Kirchler, Albert J. Menkveld, Michael Razen, et al. 2024. "Computational Reproducibility in Finance: Evidence from 1,000 Tests." *Review of Financial Studies*, 37(11): 3558–3593.

Prosser, Annayah M. B., Olivia Brown, Grace Augustine, and David Ellis. 2024. "It's Time to Join the Conversation: Visions of the Future for Qualitative Transparency and Openness in Management and Organisation Studies." SocArXiv. https://osf.io/preprints/socarxiv/ntf73_v1.

Rainey, Carlisle, Harley Roe, Qing Wang, and Hao Zhou. 2025. "Data and Code Availability in Political Science Publications from 1995 to 2022." *PS: Political Science & Politics*, 58(2): 339–345.

Scoggins, Bermond, and Matthew P. Robertson. 2024. "Measuring Transparency in the Social Sciences: Political Science and International Relations." *Royal Society Open Science*, 11(7): 240313.

Small, Mario Luis. 2011. "How to Conduct a Mixed Methods Study: Recent Trends in a Rapidly Growing Literature." *Annual Review of Sociology*, 37(1): 57–86.

Stockemer, Daniel, Sebastian Koehler, and Tobias Lentz. 2018. "Data Access, Transparency, and Replication: New Insights from the Political Behavior Literature." *PS: Political Science & Politics*, 51(4): 799–803.

Trisovic, Ana, Matthew K. Lau, Thomas Pasquier, and Mercè Crosas. 2022. "A Large-Scale Study on Research Code Quality and Execution." *Scientific Data*, 9(1): 60.

Vilhuber, Lars. 2020. "Reproducibility and Replicability in Economics." *Harvard Data Science Review*, 2(4): 1–39.

Weeden, Kim A. 2023. "Crisis? What Crisis? Sociology's Slow Progress Toward Scientific Transparency." *Harvard Data Science Review*, 5(4).

Winship, Christopher. 2007. "Introduction to the Special Section on Replication and Data Access." *Sociological Methods & Research*, 36(2): 151–152.

Young, Cristobal, and Aaron Horvath. 2015. "Sociologists Need To Be Better at Replication." https://orgtheory.wordpress.com/2015/08/11/sociologists-need-to-be-better-at-replication-a-guest-post-by-cristobal-young/ (accessed May 1, 2026).

Zenk-Möltgen, Wolfgang, and Greta Lepthien. 2014. "Data Sharing in Sociology Journals." *Online Information Review*, 38(6): 709–722.

Supplementary Information

An Empirical Assessment of Data Sharing and Computational Reproducibility in Sociology

Rio Ikeuchi Taisuke Imai

Contents

A Journal Policies

We reviewed the submission policies of 16 leading sociology journals listed in Table A.1 to identify any references to data/code sharing and pre-registration (May 2026). Our sample builds on the set of journals examined by Weeden (2023), which includes four general-interest sociology journals (1–4) and seven journals in subfields or interdisciplinary areas with substantial overlap with sociology (7–13). We expanded this list by adding five additional high-prestige journals, selected from the top of a widely cited reputational ranking of generalist and specialist sociology journals, excluding two theory journals (McDonnell and Stoltz, 2020).

Table A.1: Policies on data and code sharing and preregistration among top sociology journals.

		Data and code sharing	Pre-registration
Generalist			
1	*American Sociological Review*	encouraged	-
2	*American Journal of Sociology*	encouraged	informational
3	*Social Forces*	encouraged*	-
4	*European Sociological Review*	encouraged*	-
5	*Sociological Science*	required	encouraged*
6	*Social Science Research*	encouraged	-
Specialist			
7	*Demography*	encouraged*	-
8	*Administrative Science Quarterly*	required	encouraged
9	*Criminology*	-	-
10	*Gender & Society*	encouraged	-
11	*Social Problems*	encouraged	-
12	*Sociology of Education*	encouraged	-
13	*Social Networks*	encouraged	-
14	*Journal of Marriage and Family*	informational	-
15	*Sociological Methods & Research*	required	encouraged*
16	*Sociological Methodology*	encouraged	-

Notes: Journals labeled `encouraged*` encourage data sharing or pre-registration but require only a data availability statement or disclosure. *Sociological Research & Methodology* (15) will require pre-registration beginning in July 2026. Journals were classified as `informational` when they mention open science practices without actively encouraging them.

1. American Sociological Review. The "Submission Guidelines" (accessed May 1, 2026) include following statement.

> Ethics: Submission of a manuscript to another professional journal while it is under review by the ASR is regarded by the ASA as unethical. Significant findings or contributions that have already appeared (or will appear) elsewhere must be clearly identified. All persons who publish in ASA journals are required to abide by ASA guidelines and ethics codes regarding plagiarism and other ethical issues. This requirement includes adhering to ASA's stated policy on data-sharing: "As a regular practice, sociologists share data and pertinent documentation as an integral part of a research plan. Sociologists generally make their data available after completion of a project or its major publications, except where proprietary agreements with employers, contractors, or clients preclude such accessibility or when it is impossible to share data and protect the confidentiality of the research participants (e.g., field notes or detailed information from ethnographic interviews)" (ASA Code of Ethics, 2018).

2. American Journal of Sociology. The "Manuscript Submission Information" (accessed May 1, 2026) includes the following statement.

> **Data Access and Reproducibility:** It is expected that papers that involve numerical data analysis will share the analytic code with the journal upon acceptance or at the request of reviewers, and will make a statement about whether or not they make their data, in whole or in part, sharable. This does not apply to data produced from in-depth interviews or observational studies.
>
> ◦ For more information, please see Further Guides for Submission.

> 4. **Open Science I: Code.** It is not now a requirement that articles that do formal (statistical) data analysis share their code and data. However, we now consider it expected that code be archived in the Dataverse account associated with the AJS—even if the data allowing reproduction is not. We do understand that not all researchers, especially early-career ones, proceed in a sufficiently mechanical fashion to make this always appropriate, but we think that it is in almost all cases consonant with current best practices that researchers preserve a log that allows them to share the code. You do not need to comment it or make it useful to others (especially if you cannot share your data); simply allowing others to confirm that there is no disagreement between what you say you did and what you did will be a big step forward. It will be expected, but not required, that such code be given upon submission of the final draft as a con-

dition of acceptance. If you do not wish to share your code, you should say upon first submission that you will not share your code and explain why. A great paper that has a good reason for not sharing code can certainly be published. But, all other things being equal, we will tend to prefer submissions that share the code. You will retain copyright over the code.

5. **Open Science II: Data.** There are four possibilities for the type of data used in a standard quantitative paper: (1) public data to which all have, with minimal effort, access (e.g., the GSS, ANES, and so on); (2) shared data with some restriction (e.g., more sensitive information from GSS or Add-Health); (3) proprietary data collected by individuals or small groups who have agreed to share it with you (as the author); (4) data you have collected yourself. In possibility (1), you need not share your data, and in possibility (2), you cannot share it. For (3), you are of course beholden to those who generated the data, but we strongly ask you to try to get permission from them to make the data, even if in simplified form, public. We are always concerned at the use of data that neither you, nor your reviewers, knows inside and out. For (4), it is not necessary that you share the data you have gathered—and this definitely does not apply to ethnographic data or in-depth interview data, in which sharing the data may identify respondents and is not in keeping with standard practices. However, we will tend to prefer, and trust, papers based on data that you are willing to share. Further, you might consider that the AJS is the appropriate paper for the capstone piece of your project, something that would be published just around the time when you are ready to share the data with others.

If you are indeed willing to share the data, you should say so upon first submission, as the reviewers and/or editorial board may take this into account. "Sharing the data" does not mean that you will have a footnote saying that readers can contact you for the data—it means placing it in the Dataverse archives that are associated with the AJS, and including a link/reference in your article.

6. **Open Science III: Pre-registration.** In no way is pre-registration required or even in all cases favored. However, particularly in experimental work where effect heterogeneity is being explored, pre-registration helps reviewers tell the difference between a very well fleshed out theoretical understanding (on the one hand) and post-hoc explanations of the failure of the effect to hold more generally (on the other). Post-hoc explanation is not necessarily a bad explanation, but statistics from multiple tests are different from statistics for single tests, and reviewers and readers need to know which they have before them. And pre-registration is not about guessing the right lotto numbers: If you pre-register a hypothesis and then find the opposite, this may turn out to

be a very successful project and produce an interesting paper.

3. Social Forces. The "Author guidelines" (accessed May 1, 2026) include the following statement.

> **Availability of data and materials**
> Where ethically feasible, the Journal strongly encourages authors to make all data and software code on which the conclusions of the paper rely available to readers. Authors are required to include a data availability statement in their paper. When data and software underlying the research article are available in an online source, authors should include a full citation in their reference list. For details of the minimum information to be included in data and software citations see the OUP guidance on citing research data and software.
>
> Whenever possible, data should be presented in the main manuscript or additional supporting files or deposited in a public repository. Visit OUP's Research data page for information on general repositories for all data types, and resources for selecting repositories by subject area.
>
> **Data availability statement**
> The inclusion of a data availability statement is a requirement for papers published in the Journal. Data availability statements provide a standardized format for readers to understand the availability of original and third-party data underlying the research results described in the paper. The statement should describe and provide means of access, where possible, by linking to the data or providing the required unique identifier.
>
> More information and example data availability statements.

4. European Sociological Review. The "Information for Authors" (accessed May 1, 2026) states the following.

> **Availability of Data and Materials**
> Where ethically feasible, European Sociological Review strongly encourages authors to make all data and software code on which the conclusions of the paper rely available to readers. Authors are required to include a Data Availability Statement in their article. This policy applies to all papers submitted to the journal on or after 5 September 2022.

> We suggest that data be presented in the main manuscript or additional supporting files or deposited in a public repository whenever possible. Information on general repositories for all data types, and a list of recommended repositories by subject area, is available here.
>
> **Data Availability Statement**
> The inclusion of a Data Availability Statement is a requirement for articles published in European Sociological Review. Data Availability Statements provide a standardised format for readers to understand the availability of data underlying the research results described in the article. The statement may refer to original data generated in the course of the study or to third-party data analysed in the article. The statement should describe and provide means of access, where possible, by linking to the data or providing the required unique identifier.
>
> The Data Availability Statement should be included in the endmatter of your article under the heading 'Data availability'.
>
> More information and example Data Availability Statements can be found here.

5. Sociological Science. The "Reproducibility Policy" (accessed May 1, 2026) states the following.

> Over the last decade, we have witnessed a crisis in science in which many admired research studies have been overturned or found non-replicable. Researchers increasingly recognize that publication itself does not imply that findings are robust, and the public has questioned the credibility of social science research. In order to advance the credibility of sociological research, *Sociological Science* has adopted a reproducibility policy.
>
> Starting with submissions received after April 1, 2023, authors of articles relying on statistical or computational methods will be required to deposit replication packages as a condition of publication in Sociological Science. Replication packages must contain both the statistical code and — when legally and ethically possible — the data required to fully reproduce the reported results. With this policy, Sociological Science hopes other high-impact journals in Sociology will follow suit in setting standards for reproducibility of published work.
>
> In addition to depositing replication packages, papers relying on experimental methods must adhere to the disclosure and pre-registration requirements outlined in the

journal's Policy on Findings from Experimental Data below.

Under many legitimate circumstances, data cannot legally or ethically be made available to readers. When authors cannot make their data available, they must explain why in the main text of the paper. In such cases, making code and other materials available is still required, unless doing so would violate legal or ethical constraints.

Researchers using qualitative data, such as interviews or participant observation data, are not required to submit a replication package. We encourage authors to make qualitative data available when possible, and urge them to consider whether materials such as interview protocols or coding schemes can be shared.

Replication packages are not required at the time of submission and are not part of the editorial review. The packages are only required if a paper is accepted, and shall be deposited when authors return their publication agreement. However, authors must clearly state in their submission that they either meet this standard or are claiming an exemption. Sociological Science considers this transparency statement to be an important element of the article.

Policy on Findings from Experimental Data Published in Sociological Science
These policies for findings from data from experiments are in addition to those for the provision of data and code discussed in the general policy for the transparency and credibility of work published in Sociological Science.

(1) Disclosure expectations
For results from experiments, Sociological Science expects that the discussion of methods, results, or an appendix includes mention of (a) all outcome measures collected and analyzed for the article's target research question; (b) all independent variables or experimental manipulations analyzed for the research questions, whether successful or not; and (c) the total number of observations for which experimental data was collected but are excluded from analyses, along with the reasons for exclusion.

(2) Pre-registration
Sociological Science requires authors of papers that include originally collected experimental data and present findings from those experiments as testing hypotheses (i.e., as "confirmatory" rather than "exploratory" research) to indicate explicitly whether those experiments were pre-registered.

For an experiment to be pre-registered, it must have been provided to an independent online registry prior to data collection, and it must include specific information about the planned analyses in addition to the experimental design and hypotheses. The Open Science Framework provides both an example of a recommended registry and a recommended template for the information to be included in a pre-registration.

Manuscripts reporting pre-registered experiments must report all preregistered analyses in the body of the manuscript or in an appendix, note all deviations from the preregistered analysis plan, and clearly distinguish analyses that were preregistered from those that were not.

An anonymized version of this pre-registration must be included at the time of submission as additional materials available to reviewers. Apart from anonymization, the pre-registration must be the same as that which was provided to the registry prior to data collection.

After acceptance, the version of a study with pre-registered experiments should include a link to the online pre-registration in the version of the paper to be published.

This policy regarding pre-registration does not apply to secondary analyses of already-existing experimental data, such as analyses of the various experimental items available in different waves of the General Social Survey, so long as it is clear from the text that the data were independently collected and already existed at the time of the study.

6. Social Science Research. The "About the journal" section (accessed May 1, 2026) states the following.

Research data

We are committed to supporting the storage of, access to and discovery of research data, and our research data policy sets out the principles guiding how we work with the research community to support a more efficient and transparent research process.

Research data refers to the results of observations or experimentation that validate research findings, which may also include software, code, models, algorithms, protocols, methods and other useful materials related to the project.

Please read our guidelines on sharing research data for more information on depositing, sharing and using research data and other relevant research materials.

Research data deposit, citation and linking

For this journal, Option B instructions from our research data guidelines apply.
This means that you are **encouraged** to:

- Deposit your research data in a relevant data repository.
- Cite and link to this dataset in your article.
- If this is not possible, make a statement explaining why research data cannot be shared.

Data statement
To foster transparency, you are encouraged to state the availability of any data at submission.

Ensuring data is available may be a requirement of your funding body or institution. If your data is unavailable to access or unsuitable to post, you can state the reason why (e.g., your research data includes sensitive or confidential information such as patient data) during the submission process. This statement will appear with your published article on ScienceDirect.

Read more about the importance and benefits of providing a data statement.

Data linking
Linking to the data underlying your work increases your exposure and may lead to new collaborations. It also provides readers with a better understanding of the described research.

If your research data has been made available in a data repository there are a number of ways your article can be linked directly to the dataset.

Read more about the importance and benefits of providing a data statement.

- Provide a link to your dataset when prompted during the online submission process.
- For some data repositories, a repository banner will automatically appear next to your published article on ScienceDirect.
- You can also link relevant data or entities within the text of your article through the use of identifiers. Use the following format: Database: 12345 (e.g. TAIR: AT1G01020; CCDC: 734053; PDB: 1XFN).

Learn more about linking research data and research articles in ScienceDirect.

7. Demography. The "Ethics & Disclosures" section (accessed May 1, 2026) includes the following statement.

Reproducibility and Data Availability Policy

Demography values transparency and reproducibility in scientific research. The journal asks authors of **accepted manuscripts** to direct readers to an accessible repository for the available data sources and implementation code. While *Demography* does not maintain a repository for data or code, authors are asked to provide URL links with persistent identifiers (e.g., DOIs or accession numbers) directing readers to their source materials. In the following, we provide guidance for authors.

All accepted manuscripts must include a data availability statement confirming whether and how data have been shared. For data that are publicly available, this statement should describe how the data can be accessed and include a persistent identifier, such as a DOI or repository accession number. For data that are not publicly available, authors must provide a description of the conditions limiting access to the data. These conditions may include legal or ethical restrictions or proprietary discretion.

Demography encourages authors to adhere to FAIR data-sharing principles, which recognize that access to shared data may be restricted to protect confidential or proprietary information, while still promoting data sharing in ways that are as open as possible without compromising data use agreements.

Sample data availability statements illustrating typical scenarios are available upon request from the *Demography* editorial office, including examples for administrative or remote data center (RDC) data, proprietary commercial data, and survey data held by authors but not deposited in repositories.

Demography also encourages authors to provide, upon acceptance, reproducible code to readers. When code is made available, the code statement should specify the software version used to produce the results. The code itself should be clear and should include informative comments explaining the purpose of each command or formula and how these relate to the paper's results.

Authors may direct readers to accessible repositories or archived URLs with persistent identifiers for both data and code. Because *Demography* does not maintain its own repository, authors are responsible for selecting appropriate repositories. Authors are

> invited to submit datasets to an appropriate discipline-specific or generalist repository. Authors can find additional guidance and a list of repositories by exploring repository registries such as FAIRsharing.org and re3data.org. This information is intended to assist authors in meeting data- and code-sharing expectations as transparently and straightforwardly as possible.
>
> *Demography* does not edit or validate data or code. All data and code availability statements will be included in published manuscripts following the Acknowledgments. When materials are provided, the availability statement will include the following note (as appropriate): "The author has provided data and code for replicating results."

8. Administrative Science Quarterly. The "Data and method transparency" section (accessed May 1, 2026) states the following.

> Ensuring the credibility of our research is central to the academic mission, and many authors are searching for ways to demonstrate their willingness and desire to be transparent about their research process. This is a collective responsibility. Our efforts must begin with doctoral education and teaching students to do rigorous, careful work with the goal of developing deep knowledge of organizations and organizational processes. To advance our collective goal of deeper understanding, we all must work together to uphold standards of excellence, rigor, and credibility. The following policy was designed with this mission in mind. More transparency in our articles helps all scholars do better research. If our research is to be useful and built upon, it is essential for other scholars to know how that research has been conducted and for readers to have confidence in the findings.
>
> All authors who submit to ASQ are required to explain how they will share sufficient information regarding their data and methods to maximize readers' understanding and trust of their research efforts while adhering to practices that are appropriate for the types(s) of data and method(s) the authors employ. Before a paper is accepted for publication in ASQ, authors are expected to ensure the proposed transparency plan has been implemented. The following sets of guidelines are intended to communicate the expectations for the policy as it applies to different types of data and to help authors in developing their proposed transparency plan.
>
> We seek to balance two broad objectives: to support authors and to support transparency in the research we publish. Recognizing that transparency can be achieved in various ways, and to reflect the diversity of methods employed in ASQ articles, we hope the following policy and guidelines for quantitative and qualitative data help au-

thors be transparent and help readers understand how research was performed. Valuable knowledge accumulates more reliably and efficiently when others can fully understand, trust, and build on prior work. Authors are required to adhere to current best practices for transparent data description as appropriate for their specific methodology. This involves clearly documenting and communicating the nature, sources, and limitations of the data used.

Policy and Guidelines for Quantitative Data

Policy for Data and Code Sharing

We require that authors of accepted ASQ papers publicly share their data and code for studies employing statistical or computational methods unless there is a compelling reason not to do so. All quantitative data and code necessary to reproduce the results reported in an article must be permanently archived in a secure, third-party data repository, such as Borealis, Dataverse, openICPSR, OSF, ResearchBox, or Zenodo. It is the responsibility of authors to ensure that the data and code they post remains permanently accessible. The code should be annotated so that readers can understand the analyses. If original data were collected through experiments or surveys, the relevant experimental stimuli or survey instruments should also be included. Prior to sharing data, authors should remove any individually identifying or confidential information.

Authors are not required to provide these materials at the time of submission but must post them publicly before final acceptance. At the time of submission, authors will be required to respond to ASQ data transparency questions in order to indicate how they will satisfy our policy and guidelines. These responses can be updated, if necessary, when a revision is submitted. When data disclosure is inappropriate or infeasible, authors should outline an alternative approach; consider these examples of alternative approaches. Options might include sharing a subset of the data, providing a detailed enough description of the data collection steps to enable other researchers to construct a comparable dataset, or supplying precise instructions for accessing licensed data, among other possible strategies. We recognize that ethical or legal considerations might prohibit the dissemination of data. Ultimately, the paper should uphold the principles of transparency; the final plan will be determined in conversation with the handling editor.

Authors might propose a reasonable delay in the public posting of data to have more time to benefit from their data collection effort. However, they must still share the data with the editorial team prior to final acceptance. All parties downloading the data must agree to use the data only to reproduce and explore the results in the paper.

ASQ is not opposed to readers using shared data for other research purposes if and when the original author expressly consents to such uses.

Guidelines for Pre-Registration

Our expectations about pre-registration take into account the diversity of methods and types of data used in ASQ articles. While we do not require pre-registration, we encourage it for experiments and original surveys, as pre-registration can strengthen other scholars' inferences about the research. For studies that test hypotheses using data that has already been collected, we value pre-registration of analytical plans before analyses are conducted.

Pre-registrations are useful for identifying what aspects of a study should be considered deductive and what parts are not. ASQ is open to manuscripts that test theory when designs permit but also engage in informed speculation based on unanticipated results. Pre-registration helps authors and readers distinguish these intellectual objectives. ASQ's editorial team believes that we can advance knowledge by sharing positive, negative, and null results. We are, therefore, open to publishing studies based on pre-registered analyses that produce either supported or unsupported hypotheses. We ask authors to be transparent about which parts of a manuscript are consistent with their pre-registered research design and analyses and which parts are not. At the time of submission, you may share pre-registration by including a link in the manuscript to anonymous materials housed on a site such as OSF or aspredicted.org.

When authors want or need to deviate from pre-registrations, they should disclose this in the submission, keeping in mind that deviations do not make the study inadequate or necessarily invalidate the pre-registration. An editor or reviewers may request follow-up action based on such information, such as asking to see results from both pre-registered and exploratory analyses.

Policy and Guidelines for Qualitative Research

Policy for Qualitative Research

It is important for authors to explain why the setting chosen is appropriate to answer their research question, the process by which data were collected, the steps taken to analyze the data, and how their findings and the theory generated emerged from the data they collected.

We require that authors state clearly in the methods section where the study took place, when the data were collected (i.e., months/years in which different sources of

data were collected), which types of data were collected and for what purposes, and who collected the data. Authors should communicate what population of study the findings are based on, as well as the reasoning for choosing that population. When presenting data gathered from different people, occupations, groups, or organizations, the use of confidential identifiers or pseudonyms provides data transparency without compromising confidentiality. Such identifiers help the reader trace the distribution or preponderance of evidence throughout the manuscript and assure readers that the data in the manuscript are not elicited from a single place, person, day, or field visit. When data collection involves interviews, we expect an appendix with an interview protocol and an anonymized description or table of interview.

Guidelines for Transparency in Data Collection and Analysis

We do not advocate any particular approach to conducting qualitative research or analyzing qualitative data; the suggestions below are intended as broad guidelines for improving transparency. Transparency invites your editor and reviewers along with you on the process of discovery and can simplify the reviewing process.

There are many approaches to the analysis of qualitative data. Be clear about the approach you chose and why it is appropriate given your data and research question. In the methods section, share information about how you organized, interpreted, and coded the data using examples of how your coding framework emerged from the data. Communicate honestly and concisely with journal readers to make your theorizing process transparent in terms of the moves taken to advance from data analyses to developing theory. Such an explicit account might explain how you developed insights or made discoveries and what drove the process of theorizing from the data analyzed. The explanation of the analytic steps taken should have a relationship to how the findings are presented – so that readers can map the methodological approach described to the findings presented.

Tables and figures can also be helpful to support your detailed textual descriptions of data collection and analysis but are not required. Supporting tables that organize and document the different types of data gathered can be useful for readers, especially for studies with complex data collection efforts with different sources over time. Creative visualizations can help display your data and enhance the transparency of analysis.

Be as clear as possible about how the stated findings are supported by the data gathered. Your claims regarding findings should all be supported by data, and you should make clear to the reader how the raw data produced the interpretation presented. There are many ways to achieve this. We encourage sharing rich data in the manuscript

text, such as full quotations and descriptions of events observed. Any supporting tables should offer additional but not duplicative examples. If mixed methods are used, it is important to specify which methods address which research questions and ensure that the rationale for using multiple methods is clear and justified.

9. Criminology. The "Author Guidelines" (accessed May 1, 2026) do not specify any policy regarding data and code sharing.

10. Gender & Society. The "Submission guidelines" (accessed May 1, 2026) state the following.

Data availability

The Journal is committed to facilitating openness, transparency and reproducibility of research, and has the following research data sharing policy. For more information, including FAQs please visit the Sage Research Data policy pages. Subject to appropriate ethical and legal considerations, authors are encouraged to:

- Share your research data in a relevant public data repository
- Include a data availability statement linking to your data. If it is not possible to share your data, use the statement to confirm why it cannot be shared.
- Cite this data in your research

11. Social Problems. The "Instructions to Authors" (accessed May 1, 2026) state the following.

Availability of Data and Materials

Where ethically feasible, Social Problems strongly encourages authors to make all data and software code on which the conclusions of the paper rely available to readers. We suggest that data be presented in the main manuscript or additional supporting files or deposited in a public repository whenever possible. Information on general repositories for all data types, and a list of recommended repositories by subject area.

Data and Software Citation

Social Problems supports the Force 11 Data Citation Principles and the recommendations of the FORCE11 Software Citation Implementation Group. When data and software underlying the research article are available in an online source, authors should include a full citation in their reference list.

For details of the minimum information to be included in data and software citations see the guidance on Citing research data and software.

12. Sociology of Education. The "Submission guidelines" (accessed May 1, 2026) state the following.

Ethics: Submission of a manuscript to another professional journal while it is under review by SOE is regarded by the ASA as unethical. Significant findings or contributions that have already appeared (or will appear) elsewhere must be clearly identified. All persons who publish in ASA journals are required to abide by ASA guidelines and ethics codes regarding plagiarism and other ethical issues. This requirement includes adhering to ASA's stated policy on data-sharing: "As a regular practice, sociologists share data and pertinent documentation as an integral part of a research plan. Sociologists generally make their data available after completion of a project or its major publications, except where proprietary agreements with employers, contractors, or clients preclude such accessibility or when it is impossible to share data and protect the confidentiality of the research participants (e.g., field notes or detailed information from ethnographic interviews)" (ASA Code of Ethics, 2018).

13. Social Network. The "Guide for authors"(accessed May 1, 2026) states the following.

Research data

We are committed to supporting the storage of, access to and discovery of research data, and our research data policy sets out the principles guiding how we work with the research community to support a more efficient and transparent research process.

Research data refers to the results of observations or experimentation that validate research findings, which may also include software, code, models, algorithms, protocols, methods and other useful materials related to the project.

Please read our guidelines on sharing research data for more information on depositing, sharing and using research data and other relevant research materials.

Research data deposit, citation and linking

For this journal, Option B instructions from our research data guidelines apply.
This means that you are **encouraged** to:

- Deposit your research data in a relevant data repository.
- Cite and link to this dataset in your article.
- If this is not possible, make a statement explaining why research data cannot be shared.

Research data

This journal encourages and enables you to share data that supports your research publication where appropriate, and enables you to interlink the data with your published articles. Research data refers to the results of observations or experimentation that validate research findings, which may also include software, code, models, algorithms, protocols, methods and other useful materials related to the project.

Below are a number of ways in which you can associate data with your article or make a statement about the availability of your data when submitting your manuscript. If you are sharing data in one of these ways, you are encouraged to cite the data in your manuscript and reference list. Please refer to the "References" section for more information about data citation. For more information on depositing, sharing and using research data and other relevant research materials, visit the research data page.

Data linking

If you have made your research data available in a data repository, you can link your article directly to the dataset. Elsevier collaborates with a number of repositories to link articles on ScienceDirect with relevant repositories, giving readers access to underlying data that gives them a better understanding of the research described.

There are different ways to link your datasets to your article. When available, you can directly link your dataset to your article by providing the relevant information in the submission system. For more information, visit the database linking page.

For supported data repositories a repository banner will automatically appear next to your published article on ScienceDirect.

In addition, you can link to relevant data or entities through identifiers within the text of your manuscript, using the following format: Database: xxxx (e.g., TAIR: AT1G01020; CCDC: 734053; PDB: 1XFN).

Research elements

This journal enables you to publish research objects related to your original research

– such as data, methods, protocols, software and hardware – as an additional paper in a Research Elements journal.

Research Elements is a suite of peer-reviewed, open access journals which make your research objects findable, accessible and reusable. Articles place research objects into context by providing detailed descriptions of objects and their application, and linking to the associated original research articles. Research Elements articles can be prepared by you, or by one of your collaborators.

During submission, you will be alerted to the opportunity to prepare and submit a manuscript to one of the Research Elements journals.

More information can be found on the Research Elements page.

Data statement
To foster transparency, we encourage you to state the availability of your data in your submission. This may be a requirement of your funding body or institution. If your data is unavailable to access or unsuitable to post, you will have the opportunity to indicate why during the submission process, for example by stating that the research data is confidential. The statement will appear with your published article on ScienceDirect. For more information, visit the Data Statement page.

14. Journal of Marriage and Family. The "Journal of Marriage and Family (JMF) Style Guide" (accessed May 1, 2026) states the following.

Online Supporting Materials

Analyses (or other information, including data and/or code) that are not essential to the printed article, but that are necessary to provide useful information about the dataset and measures or support the rationale for the analytic approach, can be submitted as appendixes and published in the online version of the article. The online availability of supporting material should be stated in the manuscript. Author guidelines for preparing supporting material, including acceptable formats and file sizes, are available online at Wiley Author Services. The online supplemental appendix should be no more than 10 pages in length.

15. Sociological Methods & Research. The "Submission guidelines" (accessed May 1, 2026) state the following.

Open Science Policy

SMR is a signatory to the TOP Guidelines for open science, designed to promote research transparency and increase public trust in science.

SMR's Open Science Policy applies to all manuscripts submitted after June 1, 2025.

Data, code, and materials sharing

SMR requires that the data, code, and materials used to produce the manuscript's empirical, simulation, and methodological results are clearly documented and openly accessible. Typically, authors will share a well-documented replication package (containing their data, code, and materials and a README file) openly in a trusted public repository. Openly sharing data, code, and materials improves the transparency of and trust in sociological science and enables the validation, reproduction, replication, reanalysis, and reinterpretation of sociological studies.

Exceptions to this requirement are noted below.

At submission

At submission, authors must provide their anonymized code and materials for peer review. Optionally, authors may also share their data at submission.

Reviewers will receive author's code/materials/data only after agreeing to review the manuscript and confirming that they will access code/materials/data confidentially and solely for purposes of review.

Authors must include an availability statement disclosing where anonymized code and materials (and, optionally, data) can be accessed for peer review. If data are not shared at submission, the availability statement must also disclose how data will be shared if the paper is accepted and/or justify clearly if exceptions apply.

Acceptable ways to share code, materials, and (optionally) data at submission:

- Authors deposit code and materials in a repository that allows for anonymous link sharing (see recommended repositories), then include the anonymous link to the code/materials in the submitted manuscript. This is the preferred method.
- Authors include an Anonymous GitHub link to the code/materials in the manuscript submission.*
- Authors include code/materials by uploading them as supplementary material with the manuscript submission during Step 2 of the submission process. Only .pdf, .txt, or .docx files are acceptable.*

* If the paper is accepted, code and materials must be moved to a trusted data repository prior to publication. If shared via GitHub, they may be archived via Zenodo prior to publication.

Unacceptable ways to submit code, materials, and data at submission:

- Authors include supplementary material as files other than .pdf, .txt, or .docx with their manuscript submission.
- Authors include a link to a personal, institutional, unanonymized GitHub, CRAN, or any other web page that contains personal or institutional information about any of the authors.

Before acceptance

Prior to acceptance for publication, authors must make their data, code, and materials openly available via a trusted public repository (see recommended repositories). SMR does not generally accept that data, code, or materials are exclusively available "upon request from the authors," or are exclusively shared via authors' GitHub or personal webpages. However, authors may permanently archive their GitHub repository on Zenodo.

Prior to acceptance, all manuscripts must update their availability statement to disclose how the study data, code, and materials can be permanently accessed. If exceptions apply, the reasons must be stated clearly in the availability statement, along with instructions and/or conditions for access.

Exceptions

Data, code, and materials should be shared as openly as possible while respecting legal and ethical constraints. Exceptions should be applied minimally. Exceptions will often apply only to part of the data, while the rest of the data, code, and materials can be shared openly.

If exceptions apply, the manuscript's availability statement must (a) clearly state the reasons and (b) include instructions and/or conditions for accessing the data, code, and materials to the fullest extent possible.

SMR permits the following exceptions:

- **Publicly available data, code, and materials**
  - Many commonly used source datasets, code packages, and study materials

are already permanently archived by another provider (e.g., PSID data and codebooks, R packages on CRAN).
  - Authors using publicly available source data, code, and materials must provide detailed instructions for access in their availability statement.
  - Even if source data are publicly available, authors should still post their derived analytic dataset, and authors must provide the complete code that derived the analytic dataset from the source data.
  - Some public datasets, such as the PSID or Population Registers, have existing policies and procedures regarding data sharing that restrict authors' ability to share the source data or derived analytic dataset openly in a public repository. Authors should comply with the sharing policy of the data provider.
- **Proprietary data, code, and materials**
  - When proprietary data, code, and materials may not legally be shared by the authors, the authors must provide all information necessary to apply for access in their availability statement.
  - Authors must share openly in a public repository any subset of proprietary data, code, and materials that they are allowed to share.
- **Sensitive or restricted data, code, and materials**
  - If sensitive data, code, and materials can be fully de-identified or anonymized with reasonable effort, then the de-identified or anonymized version should be shared with open access via a public repository, barring other exceptions.
  - If sensitive data, code, or materials cannot reasonably be de-identified or anonymized, authors should:
    1. Share the sensitive portion of the data, code, or materials in a public repository with restricted access if that is possible (see recommended repositories).
       * Authors must note access restrictions in their availability statement and provide detailed instructions for access.
       * When sharing data, code, or materials with restricted access, access should be managed by a third party (typically the repository) rather than the authors. SMR does not generally accept that data, code, or materials are exclusively "available upon request from the authors."
    2. Openly share any non-sensitive portions of the data in a trusted repository.

       For example: Consider a qualitative manuscript that presents verba-

tim quotations as evidence, but full field notes or interview transcripts cannot legally or ethically be shared. The authors should share longer passages around the direct quotations via a trusted repository to provide additional context.

3. If data, code, or materials cannot be shared via methods (a) or (b) (e.g., if the consent process for a study, IRB conditions, or data use agreements prohibit sharing in a public repository with or without restricted access), then the availability statement must explain this, along with where, how, and under what conditions access can be gained, if at all possible. (For example, access to a replication package that contains data from national population registers and is hosted in a secure enclave may be available only to researchers affiliated with a university in a particular country, upon application to a specific office.)

- Inquiries about exceptions to the data, code, and materials sharing policy, other than those listed above, must be communicated and explained to the editors at the time of submission via the cover letter and will be evaluated on a case-by-case basis.

Inquiries about exceptions to the data, code, and materials sharing policy, other than those listed above, must be communicated and explained to the editors at the time of submission via the cover letter and will be evaluated on a case-by-case basis.

Data format and permissions

The data may be shared in any standard format (e.g., .dta, .txt, .R, .csv), preferably readable by free (open source) software. SMR encourages authors to use field- and method-specific guidelines for preparing data and code for sharing. See the FAIRsharing website for guidance.

Authors are responsible for ensuring that their replication package is usable and well-documented. The Econometrics Society provides some guidance and best practices for preparing your data and code in an organized and well-documented replication package. Social Science Data Editors provides an excellent template README for replication packages.

Authors are responsible for having appropriate rights to (re)distribute data.

Examples of availability statements for data, code, and materials

Data, code, and materials availability statements may be combined (e.g. if the data, code, and materials are all shared via one repository link).

Examples of data availability statements:

- The datasets generated and/or analyzed during this study are available in the [REPOSITORY NAME] repository, [PERSISTENT WEB LINK TO DATASETS].
- The datasets generated and/or analyzed during this study are not publicly available due to [REASON WHY DATA ARE NOT PUBLIC]. However, they can be accessed in the [REPOSITORY NAME] repository, [PERSISTENT WEB LINK TO DATASETS], with the following restrictions: [LIST OF RESTRICTIONS].
- The data were used under license from [DATA PROVIDER NAME] for the current study and are not publicly available. However, the data can be accessed using the following instructions: [DETAILED INSTRUCTIONS FOR ACCESSING DATA INCL. WEBSITE/EMAIL].
- Data sharing is not applicable to this article as no data were generated or analyzed during this study.
- All data generated or analyzed during this study are included in this published article.

Examples of code availability statements:

- The code used during this study and documentation for the code are available in the [REPOSITORY NAME] repository, [PERSISTENT WEB LINK TO CODE].
- All code used during this study is included in this published article.
- The code was used under license from [THIRD PARTY NAME] for the current study and is not publicly available. However, the code can be accessed using the following instructions: [DETAILED INSTRUCTIONS FOR ACCESSING CODE].

Examples of materials availability statements:

- The [NAME OF STUDY MATERIAL] used during the current study is available in the [REPOSITORY NAME] repository, [PERSISTENT WEB LINK TO DATASETS].
- All materials used during this study are included in this published article.
- The [NAME OF STUDY MATERIAL] was used under license from [THIRD PARTY NAME] for the current study and are not publicly available. However, the [NAME OF STUDY MATERIAL] can be accessed using the following instructions: [DETAILED INSTRUCTIONS FOR ACCESSING STUDY MATERIAL].

Recommended repositories

Trusted public repositories adhere to policies that make data discoverable, accessible, and usable. They also provide for long-term preservation and assign unique and persistent identifiers (DOIs) (see TOP Guidelines, Lin et al. 2020)

Here is a list of recommended trusted public repositories. Authors can search for other data repositories at the FAIRsharing website or the Registry of Research Data Repositories.

Data citation

Authors must formally cite datasets in the body of the text and include them in the References section. Dataset citation ensures proper credit for data collectors and recognizes the value of datasets as first class research objects similar to scholarly publications. Dataset references should follow the Datacite format and include a permanent DOI identifier, repository name, and dataset version number (if applicable). For example:

Freese, Jeremy (2018): 2014-18 NCAA Women's Lacrosse Scores. figshare. Dataset. https://doi.org/10.6084/m9.figshare.6160451.v1

Lundberg, Ian (2023): Class exercise: Predicting income mobility in PSID. PSID Public Data Extract Repository at the Inter-university Consortium for Political and Social Research. Dataset. https://doi.org/10.3886/E185941V2.

Citations to author-generated datasets should follow SMR's self-citation policy.

Preregistration

SMR encourages authors of confirmatory empirical research to preregister their study and/or analysis plan with an independent institutional registry.

All manuscripts must include a preregistration statement, placed immediately before the Data, Code, and Materials availability statement, that states whether or not the study was preregistered.

SMR publishes both confirmatory (i.e., hypothesis-testing) and exploratory or inductive empirical analyses—often within the same manuscript. Preregistration helps distinguish confirmatory from exploratory analyses and can strengthen the credibility of

confirmatory claims. Preregistration intends to curtail the misrepresentation of post-hoc hypotheses as ex-ante.

Preregistration involves registering the study design, hypotheses, coding of key variables, sample inclusion/exclusion criteria, and planned analyses prior to collecting or receiving the data and conducting the research.

Preregistration is especially helpful for confirmatory research that collects new data or uses administrative or restricted data that provide a verifiable record of time of first access. SMR recognizes the difficulty of credibly preregistering analyses of publicly available secondary data.

Major preregistration registries include ClinicalTrials.gov, the American Economic Association's registry of randomized clinical trials, the Open Science Framework registry, Evidence in Governance and Politics' registry, and the Registry for International Development Impact Evaluation.

Manuscripts presenting preregistered analyses may also include non-preregistered and exploratory analyses if these are transparently labeled.

Starting July 2026, SMR will require preregistration for all newly submitted manuscripts reporting randomized controlled trials, including field, lab, and survey experiments.

Preregistration statements and disclosures

- *For manuscripts not reporting empirical results:*
  - The preregistration statement states that the manuscript does not contain empirical results.
- *For manuscripts reporting empirical results:*
  - The preregistration statement states whether or not the study and/or analysis plan were preregistered.
  - Manuscripts without preregistration may explain why, for example by noting that the presented empirical analyses are exploratory, inductive, or illustrative, or use publicly available data that already existed before the study was begun.
  - *For preregistered studies:*
    * Manuscripts must report all preregistered analyses in the body of the manuscript or in an appendix, note all deviations from the preregistered analysis plan, and clearly distinguish analyses that were prereg-

istered from those that were not.

  - *At submission:*
    - The preregistration statement must state when the study was preregistered (i.e. prior to the start of data collection, prior to the receipt of any data, or prior to the receipt of outcomes data).
    - Authors must provide an anonymized version of the preregistration materials as a supplementary document in the online submission system or via an anonymized link to an institutional registry.
  - *Prior to acceptance:*
    - The preregistration statement must include a link to the time-stamped registration at the institutional registry.

Examples of preregistration statements

- This study/analysis plan was preregistered with [NAME OF REGISTRY] prior to the start of data collection/prior to the receipt of any data/prior to the receipt of outcomes data. The registration can be viewed at [LINK TO TIME-STAMPED REGISTRATION]. All analyses reported in the paper that were not preregistered have been clearly noted as such.
- This study was not preregistered.
- This study was not preregistered because it does not report empirical results.

Starting July 2026, SMR will require preregistration for all newly submitted manuscripts reporting randomized controlled trials, including field, lab, and survey experiments.

Considerations for qualitative studies

SMR's Open Science Policy applies to all studies. SMR recognizes that data sharing, data citation, and preregistration have a longer history in quantitative research. The Qualitative Data Repository has resources and best practices specific to sharing qualitative data. The Center for Open Science has guidance for the preregistration of qualitative studies.

Definitions

Data

Data refers to the cleaned analytic dataset, text corpus, interview transcripts, and field notes used to produce the results reported in the manuscript. (Alternatively, authors may provide the larger source dataset(s) from which the analytic dataset(s) was de-

rived).

Code
Code refers to all user-written software programs, scripts, and coding schemes used to produce (a) the analytic dataset from the source data, and (b) all empirical, methodological, and simulation results, including all tables and figures, reported in the manuscript.

Materials
Materials refers to data collection and documentation materials, such as survey instruments, codebooks, interview guidelines, etc.

16. Sociological Methodlogy. The "Submission guidelines" (accessed May 1, 2026) state the following.

Ethics: Submission of a manuscript to another professional journal while it is under review by SOE is regarded by the ASA as unethical. Significant findings or contributions that have already appeared (or will appear) elsewhere must be clearly identified. All persons who publish in ASA journals are required to abide by ASA guidelines and ethics codes regarding plagiarism and other ethical issues. This requirement includes adhering to ASA's stated policy on data-sharing: "As a regular practice, sociologists share data and pertinent documentation as an integral part of a research plan. Sociologists generally make their data available after completion of a project or its major publications, except where proprietary agreements with employers, contractors, or clients preclude such accessibility or when it is impossible to share data and protect the confidentiality of the research participants (e.g., field notes or detailed information from ethnographic interviews)" (ASA Code of Ethics, 2018).

B Additional Materials

B.1 Keyword List

Table B.1: Keywords used for the full-text search.

1	␣experiment␣	31	open data
2	␣laboratory␣	32	data is available
3	␣field␣	33	dataverse
4	laboratory experiment	34	openicpsr
5	field experiment	35	figshare
6	online experiment	36	qualitative data repository
7	survey experiment	37	qdr.
8	audit stud	38	analysis plan
9	vignette	39	pre-analysis plan
10	mechanical turk␣	40	␣pap␣
11	prolific␣	41	pap.
12	replicat	42	␣pre regist
13	intervention	43	␣preregist
14	randomized trial	44	␣pre-regist
15	randomized control trial	45	␣register
16	␣rct␣	46	␣registration␣
17	rct.	47	aea rct
18	quantitative	48	aearct
19	analytic strategy	49	socialscienceregistry
20	analytic approach	50	open science framework
21	dependent variable	51	␣osf␣
22	qualitative	52	osf.
23	interview␣	53	aspredicted
24	interview with	54	evidence in governance and politics␣
25	in-depth	55	␣egap
26	field work␣	56	␣ridie
27	fieldwork	57	clinicaltrials
28	ethnograph		
29	mixed method		
30	simulat		

Notes: ␣ denotes a white space.

B.2 Reproduction Procedure

We describe the computational environment in which the replication packages were executed and clarify the scope of the reproducibility checks reported in the main text. All replications were conducted by one of the authors between fall 2024 and winter 2025.

B.2.1 Computing environment

Replications were performed on an Apple Silicon Mac (`aarch64-apple-darwin20 (64-bit)`). We used the software versions listed below.

- `StataNow/SE 18.5` for Mac (Apple Silicon), revision 07 August 2024.
- `R 4.3.1 (2023-06-16) "Beagle Scouts"`
- `RStudio 2023.12.0+369 "Ocean Storm"`

`R` packages required by individual replication packages were installed on demand, using the version available on CRAN at the time of replication.

B.2.2 Scope of the reproducibility checks

Two scope restrictions were applied uniformly across all packages.

Main-text results only. We attempted to reproduce only the figures and tables from the main text of each article. Several articles place additional analyses in an appendix that is bundled with the manuscript file rather than in a separate online supplement. We excluded those appendix-only results from the reproducibility assessment. Our results, therefore, reflect the reproducibility of the headline empirical content, not of every output the authors produced.

Cleaned (analysis-ready) datasets. A number of packages provide both the raw data and the scripts needed to construct an analysis-ready dataset from them, along with the analysis-ready dataset itself. In these cases, we started from the analysis-ready dataset rather than re-running the data construction pipeline. This decision reflects a time constraint and means that any errors introduced during data construction would not be detected by our procedure. We treat data-construction reproducibility as out of scope for the present study and leave it to future work.

No additional downloading. We conducted all reproduction attempts using only the data, code, and materials contained within each replication package. In some cases, README files direct users to download additional external data or resources. Retrieving such materials was beyond our scope, and we did not incorporate them into the reproduction process.

B.3 Classification Criteria: Research Type

- **Observational studies**: Research that uses non-experimental data and quantitative methods. Non-experimental data include large-scale surveys (e.g., the General Social Survey, National Longitudinal Survey of Youth), original author-collected surveys, census data, administrative records, and quantified qualitative data (e.g., text analysis). Quantitative methods are broadly defined as mathematical or statistical techniques. Simulations are classified as observational if they incorporate any of these data sources; simulations based solely on artificial data are placed in the "Other" category.

- **Qualitative studies**: Research that relies on qualitative methods such as interviews, fieldwork, archival work, narrative analysis, interpretation of meaning, or textual analysis. In our sample, all qualitative studies draw on qualitative data, including interview transcripts and ethnographic field notes.

- **Experimental studies**: Research that uses experimental data and quantitative methods. Experimental data involve some form of manipulation in the data-generating process. Examples include online survey experiments, audit studies, and field experiments, many of which rely on randomized interventions.

- **Mixed methods**: Articles are classified as mixed methods when they involve at least two distinct analyses combining quantitative and qualitative approaches (or a mix of observational/experimental and qualitative methods). While mixed methods lack a universal definition (Small, 2011), we adopt this criterion because distinguishing between quantitative and qualitative analyses is essential for evaluating data- and code-sharing practices, which differ substantially across these approaches.

- **Other**: Articles that do not involve empirical analysis (e.g., theoretical papers, reviews, methodological contributions), as well as replication studies, meta-analyses, studies based entirely on simulated data, and any research that does not fit the categories above.

B.4 Classification Criteria: Reproducibility

We primarily follow the classification strategies of Fišar et al. (2024) and Ankel-Peters et al. (2025), classifying reproducibility according to their proposed criteria. Our scope is restricted to reproducibility achievable from the analysis data.

- **Unverifiable**: The figure or result cannot be produced due to missing data or irrecoverable code.
- **Fully reproducible**: All reported numbers or outputs match exactly, aside from non-essential visual differences (e.g., color or line type in figures).
- **Largely reproducible with minor issues**: Small quantitative discrepancies may arise (e.g., from rounding, software versions, random seeds, or typos), but the qualitative conclusions drawn from the table or figure remain unchanged.
- **Largely not reproducible, with major issues**: Substantial quantitative differences occur such that the qualitative conclusions would differ, or important components of the table or figure cannot be produced. For example, some models in a regression table may reproduce while others yield substantially different results or fail to run.
- **Not reproducible**: The reproduced results do not support the paper's conclusions based on that table or figure, either because the outputs differ or the code does not run.
- **Not reproducible but consistent with provided log file**: Direct reproduction using the data and code is not possible, but included log files match the results reported in the paper. This includes cases where numerical results in the logs (even if not plotted) match the figures.
- **Table/Figure not based on data or analysis**: The table or figure does not originate from data analysis or executed code and therefore does not require documentation. Examples include experimental design diagrams, timelines, variable lists, screenshots, illustrations, or conceptual model visualizations.

References

Ankel-Peters, Jörg, Abel Brodeur, Anna Dreber, Magnus Johannesson, Florian Neubauer, and Julian Rose. 2025. "A Protocol for Structured Robustness Reproductions and Replicability Assessments." *Q Open*, 5(3): qoaf004.

Fišar, Miloš, Ben Greiner, Christoph Huber, Elena Katok, Ali I. Ozkes, and Management Science Reproducibility Collaboration. 2024. "Reproducibility in Management Science." *Management Science*, 70(3): 1343–1356.

McDonnell, Erin, and Dustin Stoltz. 2020. "Devilish Details: A Reputational Ranking of Generalist and Specialist Sociology Journals." https://scatter.wordpress.com/2020/01/29/devilish-details-a-reputational-ranking-of-generalist-and-specialist-sociology-journals/ (accessed May 1, 2026).

Small, Mario Luis. 2011. "How to Conduct a Mixed Methods Study: Recent Trends in a Rapidly Growing Literature." *Annual Review of Sociology*, 37(1): 57–86.

Weeden, Kim A. 2023. "Crisis? What Crisis? Sociology's Slow Progress Toward Scientific Transparency." *Harvard Data Science Review*, 5(4).