The Stark realities of reproducible statistically orientated sociological research

Original Article

Methodological Innovations
1–15
© The Author(s) 2022
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/20597991221111681
journals.sagepub.com/home/mio

The Stark realities of reproducible statistically orientated sociological research: Some newer rules of the sociological method

Vernon Gayle and Roxanne Connelly

Abstract

There is increasing concern that research is not transparent and that empirical results are often impossible to reproduce. Guidelines for undertaking reproducible research have been proposed in a number of academic areas (e.g. computational economics, psychology and medical research), however currently there are no guidelines for sociological research. This methodological paper provides guidance for undertaking reproducible statistically orientated sociological research. We provide an extended demonstration of the issues associated with reproducing results and undertaking transparent analyses. We draw on suitable concepts and techniques from open research, e-research and computing. We propose a set of Newer Rules of the Sociological Method, for undertaking transparent statistically orientated sociological research that supports reproducibility.

Keywords

Reproducibility, transparency, data analysis, Jupyter

Introduction

Sociologists use both a wide spectrum of forms of data and a broad range of methods to study important social problems and societal issues (Schwemmer and Wieczorek, 2020). Empirical findings that flow from sociological inquiries have important implications for social understanding, and for the development of policies and practices (Haux, 2019). It has been recognised that sociology has a reproducibility problem (Freese, 2007). In this paper, our theoretical point of departure is the claim that improving research transparency and making research more easily reproducible will play a central role in demonstrating the rigour of sociological studies and establishing the quality and reliability of their empirical findings.

Across a wide range of academic disciplines there is increasing concern that research findings cannot be reproduced (i.e. consistently repeated), and therefore it is impossible to verify empirical results (Baker, 2016; Christensen et al., 2019; Janz and Freese, 2021; Nature, 2016; Stark, 2018; Yale, 2010). A number of reproducibility guidelines have been proposed. Baiocchi (2007) proposed guidelines for computational economics, Hofner et al. (2016) for biometrics, Begley and Ioannidis (2015) for medical research, Sandve et al. (2013) for computational research, and both Nosek et al. (2012) and Obels et al. (2020) for psychology. At the current time there is an absence of practical guidance for sociological researchers. The motivation for this methodological paper is to engage with current debates and practices associated with transparent and reproducible research, and to provide practical guidance that is suitable for sociological research.

The concept of transparency is integral to reproducible research (see Barba, 2016, 2018; Barba and Thiruvathukal, 2017; Donoho et al., 2009; Leek and Peng, 2015; Peng, 2011; Stodden et al., 2013a, 2013b). The multifariousness of the

University of Edinburgh, Edinburgh, UK

Corresponding author:
Vernon Gayle, University of Edinburgh, George Square, Edinburgh EH8 9LD, UK.
Email: vernon.gayle@ed.ac.uk

research process means that empirical results are much more likely to be reproduced if the work that created the original results is transparent, that is, all the steps of the research process are adequately spelt out. Transparency involves making visible both the empirical foundation of the research, and the logic of the inquiry. In essence transparent research documents both what steps were undertaken, and why they were undertaken.

The plurality of data and analytical techniques within sociology, lead us to conjecture that a single set of guidance for undertaking reproducible sociological research is unlikely to be adequate. The focus of this paper is sociological research that employs statistical techniques for the analysis of large-scale and complex datasets, such as social surveys, Census records, administrative data, and big data resources. We envisage that sociologists working with other forms of data, and employing other data analytical techniques, will produce complementary or interrelated materials that are more pertinent to other types of data and analytical methods. We note that there is promising early work on this issue in relation to qualitative research methods (Aguinis and Solarino, 2019; Moravcsik, 2014; Sukumar and Metoyer, 2019).

The US Statistician and reproducibility advocate Professor Philip Stark, has produced a checklist for undertaking reproducible statistical analyses that outlines 14 reproducibility points on which an analysis can fail (Stark, 2015). In this article we use Stark's guidelines as a framework to explore the concept of undertaking transparent and reproducible sociological analyses of large-scale datasets using statistical techniques.

We address two methodological challenges. First, we attempt to duplicate an existing set of published results that use large-scale nationally representative survey data. Second, we attempt to extend, that is replicate, the original published results using an additional technique and an alternative measure. In this study we will use an electronic notebook-based approach. We will also draw on suitable concepts and techniques from other areas of open research. The paper aims to provide methodological guidance to assist sociologists undertaking transparent and reproducible statistically orientated analyses of large-scale and complex datasets, and to encourage discussion of how best to apply open research principles.

The methodological challenge of reproducibility in statistically orientated sociological research

The methods sections of traditional paper-based academic journals provide insufficient space for authors to comprehensively document the level of detail necessary for statistical analyses of large-scale datasets to be duplicated. It is common to make the popular statement, often contained in a footnote, that further information is 'available by request'. Dr Cristobal Young posted a blog entry after conducting a small field experiment as part of a graduate course in statistical analysis. Students selected sociological articles that they admired and wanted to learn from and asked the authors for a 'replication package' with suitable information in order to 'look under the hood', and to see exactly how the results were produced (Young, 2015). The exercise was relatively small in scale ($n=53$), but it was insightful that only a minority of sociologists provided the data or the research code that produced the published results. This is evidently an inadequate solution to the methodological challenge of reproducible sociological research.

Janz (2016) usefully partitions 'reproducibility' into two interrelated concepts. The first concept is 'duplication'. Results are duplicated when they are consistently reproduced using the same research data and data analytical techniques. We argue that a sociological analysis can be described as being 'duplicated' when a third party that is unconnected with the original analysis can produce results that are identical (e.g. they match published results).

The second concept is 'replication'. A replication study extends the original work. We theorise that a replication study will extend the original sociological analysis because it will either:

(i) include additional measures (e.g. additional explanatory variables)
(ii) include alternative measures (e.g. a different socio-economic measure)
(iii) analyse new data (e.g. a more recent sweep of a survey or data from a more recent cohort)
(iv) employ different data analytical techniques.

A replication study may also combine any of these four methodological extensions.[1]

The methodological challenge associated with duplicating results in statistically orientated sociological research, for example in studies that analyse social survey data, is relatively easy to conceptualise. Within the limits of most conventional publications, for example paper-based academic journal articles, it is impracticable for researchers to provide sufficient details on how the analytical research data were organised. It is also difficult for authors to report the specific details of comprehensive data analysis techniques.

The intricate nature of the steps that are required to organise and enable large-scale datasets prior to statistical analysis (a process that is increasing becoming known as 'data wrangling') and the specificity of statistical data analysis techniques, mean that it is usually infeasible to reverse-engineer analytical datasets from the results that are commonly reported in published papers (e.g. tables of statistical modelling results). This restricts both the capacity to verify empirical results through duplication, and the capability to incrementally build on sociological findings through the extensions offered by replication studies. These issues are not unique to statistically orientated sociology however, and they pervade other research disciplines (see Schwab et al., 2000).

Sociological analyses of large-scale datasets usually begin with a ‘raw’ (i.e. only minimally processed) dataset. Typically, this dataset has been downloaded from a national archive. In practice, a great deal of data wrangling (or data enabling) tasks are required to prepare the ‘raw’ dataset and transform it into an ‘analytical’ research dataset that is suitable for statistical analyses (Long, 2009).

The data wrangling work will include operations such as appropriately coding missing values, and re-coding variables into a format that is suitable for the specific sociological analysis. Usually, in this phase the researchers must select appropriate measures and decide how to operationalise them. These choices will be guided both by theoretical considerations and practical requirements. Variable selection is often overlooked within the research methods literature. Variable selection is not a trivial activity when using pre-existing large-scale data resources. Research datasets often contain a wide range of variables, and they routinely contain different versions of key measures such as income, social class and education.

Many large-scale studies have coverage of a general population, and they are purposely designed to be infrastructural research resources that can facilitate a wide spectrum of social and economic research. For example, using British Household Panel Survey data, Boyle et al. (2009) studied the effects of family migration on women’s employment using data on women aged 16–64 living in married or cohabiting partnerships; Bartley et al. (2004) studied participants who reported their health as being excellent or good with no limiting illness; and Ermisch and Jenkins (1999) examined the extent and determinants of residential mobility in people aged 55 and over.

The selection of cases in an analysis may be regarded as being inconsequential, but in practice nuanced judgements are usually exercised when analytical datasets are being developed. Even in the seemingly innocuous example of a study of married women, the data analyst might be faced with making a choice between including all women whose ‘current legal status is married’, or only those women who presently are ‘cohabiting with their legal husbands’. Unless specific information about how cases are selected is made available it is difficult to reconstruct analytical datasets. This prevents results being duplicated by third parties that are unconnected with the original research, and ultimately precludes work being extended through replication.

In the analysis of more complex data resources, the intricacies of data wrangling are amplified. For example, sociologists analysing the British Household Panel Survey (BHPS) will have access to data which is supplied in over 180 files (see Taylor et al., 2010). Data wrangling operations routinely involve matching together individual-level and household-level data at different time points. Activities such as matching data collected from other individuals in the study, for example spouses, children, full siblings, half-siblings, step-siblings and other household sharers are additionally intricate (see Longhi and Nandi, 2014).

In addition to the problem of insufficient information on how the analytical data were wrangled prior to the analyses, there is also the issue of the lack of precise details on how the statistical data analysis techniques were applied. In our experience, duplicating rudimentary analyses is feasible when standard statistical data analysis techniques are employed (e.g. bivariate tests, correlations, linear regression, and logistic regression). In these circumstances, as long as the same statistical techniques are used to analyse an identical copy of the analytical dataset, the subsequent results will be indistinguishable. This is not the case when more advanced methods are employed, for example when models with alternative estimation procedures are reported or when multiple imputation techniques are used to address missing data issues (Playford et al., 2020). In the next sections we will provide a detailed example of the challenges associated with duplicating published results. We will also illustrate how sociological research can be extended by replication.

The sociological example

The sociological example presented in this article involves real research data, rather than a ‘textbook example’, in order to provide a genuine illustration of the methodological issues associated with undertaking transparent and reproducible research. The first analytical task is duplicating a statistical model reported in table 5, page 20 of Connolly (2006). The analysis chosen is a logistic regression model analysing data from the Youth Cohort Study of England and Wales (YCS), which was a major British longitudinal study that began in the mid-1980s (see Finch et al., 2004).

The second analytical task is a replication that extends the logistic regression model reported in Connolly (2006) first by incorporating quasi-variance based estimation to better interpret the effects of ethnicity (Gayle and Lambert, 2007); and then augmenting the model with the addition of an alternative social class measure, the UK National Statistics Socio-Economic Classification (NS-SEC). The overall methodological challenge is to undertake this work within a transparent and reproducible framework using Philip Stark’s checklist (see Table 1).

Duplication

The process and the results of the duplication of the logistic regression model reported in table 5, page 20 of Connolly (2006) was successfully achieved. A replication that extends the model by incorporating quasi-variance based estimation, and the addition of an alternative social class measure, the UK National Statistics Socio-Economic Classification (NS-SEC) were also successful.

Table 1. Stark's reproducibility checklist.

1. If you relied on Microsoft Excel for computations, fail.
2. If you did not script your analysis, including data cleaning and munging, fail.
3. If you did not document your code so that others can read and understand it, fail.
4. If you did not record and report the versions of the software you used (including library dependencies), fail.
5. If you did not write tests for your code, fail.
6. If you did not check the code coverage of your tests, fail.
7. If you used proprietary software that does not have an open-source equivalent without a really good reason, fail.
8. If you did not report all the analyses you tried (transformations, tests, selections of variables, models, etc.) before arriving at the one you chose to emphasise, fail.
9. If you did not make your code (including tests) available, fail.
10. If you did not make your data available (and a law like FERPA or HIPPA doesn't prevent it), fail.
11. If you did not record and report the data format, fail.
12. If there is no open source tool for reading data in that format, fail.
13. If you did not provide an adequate data dictionary, fail.
14. If you published in a journal with a paywall and no open-access policy, fail.

See http://www.bitss.org/science-is-show-me-not-trust-me/ accessed 30.05.22 or https://web.archive.org/web/20220530112231/http://www.bitss.org/science-is-show-me-not-trust-me/ archived 30.05.22.

An innovative aspect of this work is that we go beyond providing the usual supplementary material and make the complete workflow openly available using a Jupyter notebook (see Kluyver et al., 2016). Jupyter notebooks are an open source web-based application that enables researchers to author documents that include live code (e.g. R or Stata code), alongside data analysis outputs (e.g. modelling results, plots etc.), and narrative text describing and detailing the workflow. Jupyter notebooks are free to use and can be downloaded at https://jupyter.org/. The complete workflow and the results of the duplication and replication activities are rendered fully transparent in the Jupyter notebook which accompanies this study. The Jupyter notebook is available as a supplementary-material and can be downloaded from: https://osf.io/8nwvu/. We have also made the Jupyter notebook available in portable document format (pdf) this is an alternative non-interactive version of notebook, which can be accessed without Jupyter.

In this example the duplication required about 70 cells of code and commentary to produce the results, even though it was a standard logistic regression model with only three explanatory variables (see Jupyter notebook sections 7, 8, 9.1 and 9.2). Working out how missing data were handled and the specific form of survey weighting that was used to provide the published results took some additional detective work (see Jupyter notebook section 7). The length and the intricacy of the work involved illustrates that reverse-engineering analytical datasets from the results that are commonly reported in published papers (e.g. tables of statistical modelling results) in order to duplicate findings is not a straightforward process. In the case of more complex analyses it is likely that duplication would be infeasible.

It is notable that we probably benefitted from having previously analysed the Youth Cohort Study data (Gayle et al., 2002, 2016). In situations where either the dataset is less familiar, the criteria for the inclusion in the analytical sample is more opaque, more variables are incorporated into the analysis, or non-standard data analysis or modelling techniques are applied (or any combination of these four methodological situations), the likelihood of being able to duplicate published results will be substantially reduced.

Replication

The replication work extended the original published analyses by incorporating quasi-variance based estimation (see Jupyter Notebook sections 9.3 and 9.4) and reparametrizing the explanatory variables in the model (see Jupyter Notebook sections 9.5 and 9.6). This part of the workflow illustrated how additional work can build incrementally. In this example the reformulation of the work exposed an important alternative understanding of the substantive effects of ethnicity on pupils gaining school qualifications. The inclusion of the official UK National Statistics Socio-Economic Classification (NS-SEC) rendered the original work more comparable with other British analyses of social class and education (see Jupyter Notebook sections 9.7 and 9.8). These methodological extensions were beneficial but would have been much more easily achievable if the original work had been rendered transparent and reproducible.

Stark's reproducibility checklist in action: A critical reflection

The workflow for this sociological study did not fail on any of Stark's 14 point checklist for reproducible research. Stark's checklist provided an insightful framework for an assay of reproducible working, and we will now reflect on each of the points.

1. **If you relied on Microsoft Excel for computations, fail.**

Using an Excel spreadsheet for reproducible statistically orientated sociological analysis can never be justified as it is impossible to provide and document a clear audit trail. The now well-known case of the errors in the spreadsheet-based calculations made in Reinhart and Rogoff (2010), which were reported by Herndon et al. (2014), should serve as a stern warning against using spreadsheets in social science data analyses.[2] Similar issues have been highlighted in genetics research (Ziemann et al., 2016).

2. **If you did not script your analysis, including data cleaning and munging, fail.**
3. **If you did not document your code so that others can read and understand it, fail.**

Points 2 and 3 are interconnected. The narrative must detail *what was undertaken*, *how it was undertaken* and *why it was undertaken*, in order to provide sufficient information for a third party unconnected with the original research to reproduce the results.

4. **If you did not record and report the versions of the software you used (including library dependencies), fail.**

The problem of preserving detailed information to recreate the entire computational environment is recognised in scientific computing (see Howe, 2012). Reporting libraries and dependencies is especially critical in the R and Python open source ecosystems (see McKinney, 2010; Plakidas et al., 2016), but it is also important to appropriately document the use of auxiliary resources such as user-written *.ado* files in Stata (see Baum, 2009).

5. **If you did not write tests for your code, fail.**

Writing tests for your code involves checking that your statistical software is operating as intended. We compared the results of two methods, which were used in the analysis, against existing published results. Most sociological analyses employ common or routine statistically orientated data analysis methods, therefore this requirement may be too stringent for every single sociological analysis. Stark suggests that you should test your software every time you change it. This is sensible and is a reasonable precaution to safeguard against software bugs etc. This issue would be a particular concern if using user-written commands (e.g. in R or Python), but in curated statistical software such as Stata users can be more confident that built in commands have been thoroughly pre-tested.

6. **If you did not check the code coverage of your tests, fail.**

Code coverage is a measure of how many lines of source code (e.g. the underlying software code which runs a particular test) have been validated during a test of your code. Very few sociological researchers develop new statistical tests or need to implement statistical tests within new software routines. Therefore, this requirement is irrelevant to most mainstream sociological analyses. Researchers who are developing new tests or constructing new routines should test the coverage of their code, and publicly documenting it is an essential course of action.

7. **If you used proprietary software that does not have an open-source equivalent without a really good reason, fail.**

The *sine qua non*, of reproducible research is that third parties unconnected with the original work can duplicate results and verify findings, and then can incrementally develop further research. Using an esoteric statistical analysis software or programming language will not assist in this overall goal. At the current time cursory observations of published research suggest that the majority of statistically orientated sociological data analysis is undertaken using SPSS, Stata or R (cf Lambert et al., 2015; Treiman, 2009). SAS now appears to be less ubiquitous,[3] and currently few analyses appear to have been undertaken using Python.

There is a long history of evaluations of different statistical software packages (see Carpenter and Morganstein, 1986; Platt and Platt, 1981). There are now numerous internet posts that evaluate (usually unsystematically) the pros and cons of different data analysis software packages.[4] In actuality, despite a few minor differences which arguably do not amount to major strengths or weaknesses, the four most popular statistical data analysis tools (SPSS, Stata, R and SAS) can be used to undertake the majority of mainstream techniques that are routinely used in sociological research (see Ward, 2013: Table 1).

Free and open-source software is licenced to use, copy and change, and the source code is openly shared so that users can voluntarily add to (or improve) the software. It is simple to gain access to free statistical software such as R and Python. It is easy to presume that these free languages chime with the principles of reproducible research, but the topic is worthy of consideration beyond this initial presumption. The openness of source code for example when undertaking checks on statistical functions or routines (i.e. ‘looking under the hood’) and for development work (e.g. programming new functions), can be considered as a benefit of open-source software. These are not routine activities in mainstream sociological research. Furthermore, some commercial data analysis packages such as Stata also allow researchers to access the source code for statistical functions and routines. Therefore Stata is no different to an open source language such as R or Python in this respect (for an accessible introduction to Stata programming see Gould, 2018).

Stata is a commercial package but allows community contributed commands and routines, distributed as ado-files (.ado), to be straightforwardly downloaded and incorporated into the software. After appropriate checking, certification, and documentation, some user-written commands have later been adopted by StataCorp to become part of a subsequent official release. Stata combines the extensibility that is more often associated with open-source packages with features usually associated with commercial packages such as software testing and verification, technical support and professional documentation. An especially attractive feature of Stata is that from the very first version it was further developed to support reproducibility through forward and reverse compatibility. The *version* command ensures that users can be confident that in the future files will continue to work, even after new versions of the software have been installed. This special feature is unavailable in other data analysis packages (e.g. R and Python).

In blunt comparison with SPSS, SAS and Stata, at the current time R and Python have less developed help and on-line support material, and fewer resources that contain relevant sociological examples. Stata is generally orientated to social science research with a particular focus on survey data analysis (see StataCorp, 2019). Presently, when using either R or Python to analyse large-scale social science datasets (e.g. surveys) it is difficult to effectively combine the numeric codes for variables along with both their value and variable labels. This means that it is difficult to effectively exploit helpful meta-information on measures, and this presents an obstacle both for data analysis and for reproducing research.

The UK Data Service currently provide data from large-scale resources such as Understanding Society (the UK Household Longitudinal Study) in SPSS and Stata format.[5] The UK Data Service also provides data in a tab-delimited format which is more software package agnostic. In our experience, working with datasets in this format is challenging because of the lack of meta-information on measures in the dataset. Actions such as detailed checking after matching cases or merging information from other household members is much more awkward when data are in this format.

As a result of these practical exigencies our current view is that sociologists can complete reproducible analyses with either free or commercial statistical packages. In order to maximise reproducibility, it is prudent to use the current mainstream statistical tools (i.e. SPSS, Stata, R or SAS) unless there are good reasons for using another software. Any of the current mainstream statistical tools can be used to undertake reproducible analysis, as long as the data wrangling and data analyses are suitably organised and then rendered appropriately transparent within a detailed narrative.

8. **If you did not report all the analyses you tried (transformations, tests, selections of variables, models, etc.) before arriving at the one you chose to emphasise, fail.**

Providing access to the complete workflow is an indispensable aspect of rendering sociological analysis transparent and reproducible. It may make a contribution to limiting negative research practices and provides extra safeguards against nefarious activities (e.g. p-hacking and HARKing). This will ultimately improve confidence in results within and beyond the academic community.

9. **If you did not make your code (including tests) available, fail.**

Stark states that your code should also state how it is licenced. This is a new departure in sociological research. There are a series of licences that could be appropriate to the current activity that would chime with the wider academic ideas about attribution. In this present work we have chosen to use the MIT Licence, because it appears most suitable.[6]

10. **If you did not make your data available (and a law like FERPA or HIPPA doesn't prevent it), fail.**
11. **If you did not record and report the data format, fail.**
12. **If there is no open source tool for reading data in that format, fail.**

Access to data is an integral part of transparent and reproducible sociological research. Making data accessible is not possible for most sociologists working with large-scale datasets that are supplied by national archives because these data do not belong to the researcher, and they are usually provided under some form of 'end user licence' that prevents data sharing.[7] It is therefore extremely important that researchers cite their data using a precise and persistent mechanism, for example a digital object identifier (DOI) (see Paskin, 2010).

13. **If you did not provide an adequate data dictionary, fail.**

Providing an adequate data dictionary is a relatively easy task but it is not currently a ubiquitous sociological practice. The purpose of the data dictionary is to inform third parties unconnected with the original project. Therefore, the acid test of a data dictionary is how easily it can be read and understood.

14. **If you published in a journal with a paywall and no open-access policy, fail.**

Wide access to published results is critical to reproducible research. Scholarly communication is currently in a state of flux (see Whiteley, 2019). In the UK, policy changes from the research funding councils,[8] audit requirements for the Research Excellence Framework[9] and wider policies directed towards encouraging greater knowledge exchange[10] are likely to increase access to published research. In UK higher education research, the move to Green open access which involves publishing in a traditional subscription journal, but also 'self-archiving' in a repository (e.g. a university archive or external subject-based repository) and providing free access is likely to serve to meet this requirement (although free access might be after an embargo period set by the publisher). Under most circumstances researchers can also publish pre-prints of their work on a server such as SocArXiv,[11] making the work openly accessible.

Discussion

Increasing transparency in sociological research is intrinsically attractive for a number of reasons. Greater transparency will (i) increase the capacity to understand how the research was conducted, (ii) help other scholars evaluate the analyses undertaken, (iii) aid the detection of errors and inconsistencies, (iv) facilitate the incremental development of work, (v) contribute to limiting negative research practices, (vi) provide extra safeguards against nefarious practices and (vii) improve confidence in results within and beyond the academic community.

We are not so naive as to imagine that sociologists working in the field of the statistical analysis of large-scale data will adopt transparent and reproducible working overnight. We have delivered a number of seminars and workshop presentations on this topic (Gayle, 2016, 2017, 2018; Gayle et al., 2019a, 2019b; Playford at al., 2020). In question and answer sessions we have received both plaudits and brickbats. A striking feature of these discussions is that they have been sharply split by career stage. In general, researchers who are younger or at an earlier stage in their careers have been much more positive and receptive to these ideas, compared with older and more established researchers.

A number of barriers to undertaking transparent and reproducible statistical analyses of large scale datasets have routinely been suggested by researchers who are critical of this approach. We will reflect on the limiting nature of these obstacles, which can be corralled under the following topic headings; fear of mistakes, increased time and effort required, changes to work practices, no direct academic credit, fear of the scoop, data sharing issues, lack of support from journals, lack of tools and infrastructure, lack of training, and the absence of standards.

Fear of mistakes

Researchers who make their entire research compendium available, risk having their work inspected more closely and perhaps increase the chance of errors being found (Stodden, 2010). The Turing Way Community et al. (2019) succinctly summarise the issue and state that making public materials that could potentially include errors is revealing and potentially intimidating and stressful. They further state that it is likely to motivate researchers to be careful and to conduct analyses to a higher standard. Boettiger (2015) pointed out that cultural and behavioural factors in many fields are a far more extensive barrier to reproducibility than technical barriers. In our view, what is required is a cultural shift moving social scientists away from the view that publishing non-transparent results is more acceptable than exposing work to scrutiny that might reveal errors. A ‘gotcha culture’ in which a common goal is catching researchers out will strongly dissuade researchers from undertaking transparent analyses (Donoho, 2010; Janz and Freese, 2021). What is required is a shift to a culture that is committed to the respectful identification of mistakes, coupled with an ethos of appropriately correcting errors. This would contribute to the incremental and cumulative development of social science research.

Increased time and effort required

The stages involved in undertaking sociological research that employs statistical techniques to analyse large-scale and complex datasets are intricate, and routinely large amounts of time and effort are required (Treiman, 2009). Gayle (2021) warns that it is infeasible to undertake multivariate statistical analyses of large-scale datasets without using a computer, and specialist software or a programming language. He further asserts that whilst software can be operated in different ways the complexity of large-scale datasets means that researchers undertaking anything other than very basic analysis will write out software commands using a syntactical or programming format. These commands form the backbone of the audit trail. As Oliveira and Stewart (2006) state, if computational work is not correct nothing else matters, for without accuracy useful results cannot be expected.

In our experience, a large outlay of time and effort is routinely required in sociological studies analysing large-scale datasets. Undertaking statistically oriented sociological research involves working in an organised and systematic manner. It does not necessarily require researchers to invest additional time or extra effort to render high quality work transparent and reproducible. Therefore, we regard the perceived barrier of the investment of additional time and effort as subterfuge.

Moreover, in our experience further analyses are often required before formal publication is secured, for example addressing comments from an intransigent ‘Reviewer 2’.[12] In many instances requests for further work will arrive more than 6 months after the original manuscript has been submitted to an academic journal. By this stage in the lifecycle of the project it is not uncommon to be beyond the formal end date, and research staff funded by grants often have moved on to other appointments. Therefore, having a planned, organised and literate workflow pays major dividends because it greatly assists researchers in rapidly navigating back through the research process in order to make suitable amendments and undertake additional analyses. Indeed, the Turing Way community suggest that any additional time costs associated with transparent and reproducible working are more than compensated at the end of projects (The Turing Way Community et al., 2019).

Changes in working practices

The initial idea of working within a transparent and reproducible framework may seem daunting for some researchers. The inclusion of comments explaining the logic of the data analysis workflow is a hallmark of good practice in statistically orientated sociological analyses (Long, 2009). In reality most researchers already work systematically even though they may not consciously emphasise this aspect of their work. This means that the move to deliberately developing a narrative of the workflow may, in practice, be a small step rather than a giant leap. The narrative workflow must include details of the research code, that is, what action was undertaken, and a commentary explaining why the action was undertaken. This is conceptualised more formally in computer science. Knuth (1984) introduced the paradigm of 'literate programming', in which a computer programme is given an explanation of its logic in plain language (e.g. English) alongside the computer code (see also Knuth, 1992; Schulte et al., 2012).

The paradigm of 'literate' computation directly appeals to the aim of rendering the social science statistical data analysis workflow transparent. Jupyter notebook pioneer Dr Fernado Perez describes a literate workflow as the weaving of a narrative directly into live computation, interleaving text with code and results in order to construct a complete narrative that relies equally on the textual explanations and the computational components, with the ultimate goal of communicating empirical results (Perez, 2013). A literate workflow will help the original researchers keep track and understand all of the elements of their work during the research process, but is also beneficial, for example when returning to make revisions or when checking proofs. Providing a literate workflow will be revolutionary in providing journal reviewers, PhD examiners and research stakeholders more widely, with access to the entire research process.

The concept of 'peer code review' is the manual inspection of code by groups of researchers rather than by a single author, and it is recognised as a valuable tool for reducing defects and improving the quality of software (see Bacchelli and Bird, 2013). 'Pair programming' is a software development technique in which two colleagues collaborate at one workstation, one working as the driver and the other working as the observer. It is a tactical attempt to limit faults and ensure that there is a strategic overview of the code writing process (see Williams, 2001). The authors have adapted both of these techniques in their reproducible analyses, and it has proved feasible and fruitful in this current study.

No direct academic credit

There is a hierarchy in the employment structure within universities. Schimanski and Alperin (2018) provide an evaluation of scholarship in academic promotion and tenure processes in North America, Parker (2008) studied promotion criteria in the British universities, Smith et al. (2014) investigated academic promotion in Australia, and Macfarlane (2007) investigated defining and rewarding academic citizenship. Buttliere (2014) suggested that the problem is that an ineffective reward system makes prosocial action less favourable because it impacts upon the production of work that counts for promotion.

During our discussions on undertaking transparent and reproducible social science research, we have observed that detractors exhibit the ability to set free their 'inner economist' and make the general appeal that they are only motivated by rewards. Over the course of our careers we have observed that academics routinely undertake a number of tasks that do not return direct academic credit. Most notably such tasks include commenting on colleagues' work, refereeing academic journal articles, reviewing grant proposals, writing references for both staff members and students, and sitting on committees. Participation in tasks that are not explicitly rewarded have also been documented in formal studies of academic work (see Hamilton, 2019). Indeed, Lam (2011) asserts that academics are often motivated by the intrinsic satisfaction of solving a puzzle, rather than by financial, employment or reputational rewards. After much reflection, our position is that good researchers should primarily be driven by the desire to undertake high quality work, and not by extrinsic rewards (see Merton, 1973).

As Nosek et al. (2012) assert the solution requires making incentives for publishing reproducible work competitive with non-reproducible research. Discussions of incentives for reproducibility are already underway in other disciplines such as translational medicine (see Rosenblatt, 2016). Stodden (2014) suggested that data and code citation practices should also be recognised and expected in research, and tenure and promotion could reward the production of computational elements (e.g. software and data analytical code). Academic promotion committees are charged with ensuring that scholarly work is concordant with the culture and policies of their institutions, and is meritorious and consistent with scientific standards (Cabrera et al., 2017). Given the increasing concern that empirical findings cannot be reproduced and verified, we envisage that in future the transparency of research will have a more prominent role in the evaluation of research quality and that credit be awarded to studies that adopt good practices. Munafò et al. (2017) conclude that changing the incentives requires a coordinated effort by all stakeholders to alter reward structures.

Fear of the scoop

Laine (2017) states that the risk of scooping is often used as a counter argument for open science and especially for open data. Being scooped means being beaten in an attempt to be first to publish a new research finding or discovery. The race to publish first might be part of an academic or intellectual competition, however taking another researcher's work and passing it off as your own is simply an illegitimate act and constitutes academic misconduct (Steneck, 2007).

We are aware that many scientific fields have grand challenges which can result in researchers, in different universities or research groups, pursuing similar or overlapping research questions. Indeed, British Mathematician and Abel Prize Laureate Professor Sir Andrew Wiles has admitted in interviews that he worked on his proof of Fermat's last theorem in secret to mitigate against the risk of another mathematician solving the problem first (Plusmathsorg, 2016). The Millennium Grand Challenge in Mathematics offers one million US dollar prizes as an incentive (Jaffe, 2006). In such a situation it is easy to see why researchers might be secretive about their work in progress. It is difficult to envisage any situation in sociological research that is an analogue without entering the realms of fantasy. The long running and widespread practice of delivering conference and seminar presentations openly advertises work in progress and appears not to have incited the practice of scooping. We are sceptical and cannot easily envisage incidences in sociological research where legitimate scooping would occur. A simple defence against scooping, is for the researcher to organise their workflow in a literate and transparent manner but to only make it public once publication has been secured.

Data sharing issues

Access to data is an integral part of transparent and reproducible sociological research. In fields such as political science there have been longstanding discussions of data sharing (see Bueno de Mesquita et al., 2003). Within these discussions there are both advocates of data sharing such as Meier (1995), and critics such as Sieber (1991) and Gibson (1995). In practice, making data accessible is not possible for most sociologists working with large-scale datasets that are supplied by national archives because these data do not belong to the researcher, and they are usually provided under some form of 'end user licence' that prevents data sharing.[13] It is therefore extremely important that researchers cite their data using a precise and persistent mechanism, for example a digital object identifier (DOI) (see Paskin, 2010). This practice enables third parties who are unconnected with the original research to accurately and definitively locate the correct version of the data that was used in the original study.

Grahe (2018) wrote in a blog post 'I think most reservations about sharing data and materials reflect inertia and lack of understanding rather than actual conflict with the idea of openness'. Data sharing should be practiced in situations where the sociologist is the data collector and has guardianship of the data. Researchers that are unconnected with the original study should be given access to data that facilitates the duplication of studies. Data sharing must follow legal requirements and suitable ethical safeguards must be in place. Shared data should conform to standards that avoid the identification of participants and prevent the unwarranted disclosure of information.

A small number of studies have found positive relationships between data sharing and increased citations in a range scientific fields (e.g. Dorch et al., 2015; Henneken and Accomazzi, 2011; Piwowar et al., 2007). Furthermore, Gleditsch et al. (2003) studied the *Journal of Peace Research*, and concluded that making data available seemed to serve authors well in terms of citations. Whilst we do not wish to place too much emphasis on this small and disparate set of studies, they are nonetheless encouraging.

Lack of support from academic journals

Currently over 5000 journals across all disciplines are signatories of the Transparency and Openness Promotion (TOP) Guidelines,[14] which require data and code sharing standards.[15] Unfortunately, the current list of TOP signatories includes few journals from Sociology and allied disciplines. If more sociology journals signed up to the TOP Guidelines, this would result in a dramatic change in the research landscape. This is not to suggest that new publication guidelines are a panacea, indeed some evaluations for example Herndon and O'Reilly (2016) suggest that while journals are adopting new policies they are doing so in an incomplete and varied manner.

Despite current low levels of support from sociology journals we do not feel that researchers should be constrained by existing journal policies. Even if a journal does not explicitly require research transparency, most journals allow the publication of additional online supplementary materials alongside published articles (see Connelly and Gayle, 2019). Additional materials could also be made available via platforms such as the Open Science Framework (OSF).

Lack of tools and infrastructure

Lowering the technical barriers to undertaking reproducible research is likely to be beneficial (Boettiger, 2015). In this present study and in Connelly and Gayle (2019) we have demonstrated that there are tools and resources that are routinely used in areas such data science and e-research that are valuable for undertaking transparent and reproducible sociological research using large-scale datasets, which we will now reflect upon.

Electronic research notebooks. The electronic research notebook is a convincing methodological tool for rendering the entire workflow transparent. Jupyter notebooks provide an elegant way of combing executable research code, commentary, and results in a unified narrative. Electronic notebooks can easily be made available as part of online supplementary materials and deposited in a repository (we will expand on the use of electronic repositories below). This approach has already been successfully deployed in sociological research (see Connelly and Gayle, 2019), and is widely used in other disciplines (see Abbott et al., 2021).

We are aware that moving to using electronic research notebooks would be a big leap within sociology. A less elegant, but more achievable step will be providing documented files containing research code and outputs. At its simplest, these might take the form of annotated .do files and log files of outputs for Stata users; Command files in .sps format and output files in .spv format for SPSS users; and .R script files and .txt output files for sociologists using R. More sophisticated approaches would include the use of R Markdown and Stata Dynamic Documents.

Despite the benefits, Jupyter notebooks are not without problems however (see Perkel, 2018). In our experience the initial installation can be fiddly.[16] Installing kernels to run different software and computer languages is often not a smooth process, and from time to time there are perplexing errors and inconsistencies with some libraries and dependencies. These issues are likely to present sizable obstacles for many sociologists, especially those with more restricted computing skills. These issues are likely to be evanescent given the speed of development in the areas of e-research and computational science however. Other notebook applications may emerge as potential alternatives to Jupyter, and may be more readily suited to statistically orientated sociological research (Perkel, 2021).

Research objects. Within data science the concept of a Research Object (RO) describes an artefact that packages up research outputs (e.g. data, metadata, code, results, documentation and papers) (Bechhofer et al., 2013; Sefton et al., 2019). An innovative way of conceptualising the production of publicly accessible workflows is to consider them as Research Objects. A practical guide to developing research objects when undertaking reproducible statistically orientated social science research has been developed by Gayle (2021). Sociological Research Objects should be produced under the FAIR principles, this means that they should be **F**indable, **A**ccessible, **I**nteroperable and **R**eusable (see Wilkinson et al., 2016).

Repositories. The use of repositories or archives is fundamental to achieving transparency in reproducible research (Freese, 2007). In this present study we have used both GitHub and the Open Science Framework (OSF). The functionality of GitHub lends itself well to developing public repositories of social science workflows. We have also used the Open Science Framework (OSF) which provides a specialist platform where research code can be shared alongside further project related materials such as conference presentations and preprints (Foster and Deardorff, 2017). The OSF platform shows promising signs that it could emerge as a dominant eco-system for transparent and reproducible sociology. The re3data project is a registry of research data repositories.[17]

The development of reproducible sociological workflows will be greatly assisted by the routine use of version controlling protocols to keep track of files within the workflow, and by the use of repositories (see Blischak et al., 2016). Integral to GitHub is the concept of version control, which allows teams to efficiently collaborate when developing, and editing code (see Playford et al., 2016).

Lack of training

Specific training in undertaking transparent and reproducible social science is in its infancy. Pownall et al. (2022) provide the first comprehensive review of how integrating open and reproducible scholarship into teaching and learning may impact students, using a large-scale, collaborative, team-science approach. A range of possible tools and infrastructural resources are introduced in Kitzes et al. (2018), and Christensen et al. (2019) is a textbook on transparent and reproducible social science research.

A number of social science research methods organisations provide training courses on reproducible research.[18] Data Carpentry develops and teaches workshops on the fundamental data skills needed to conduct research.[19] Sullivan et al. (2019) provide authors with step-by-step instructions for using the free and open source Open Science Framework (OSF) to create a data management plan, preregister their study, use version control, share data and other research materials, or post a preprint for quick and easy dissemination. Integrating training in research transparency into undergraduate and postgraduate research methods curriculums is likely to encourage improvements in reproducibility in the work of future generations of researchers.

Absence of standards

It is disingenuous to suggest that there is an absence of standards for reproducible research in statistically orientated sociology. This is because over a decade ago Freese (2007) suggested a set of standards whereby, at the time of publication, sociologists would use online archives to deposit the resources required for duplicating published results along with the maximum amount of information about the study. Furthermore, as we mentioned above, guidelines have been produced in a range of other research fields that offer insights (e.g. Baiocchi, 2007; Begley and Ioannidis, 2015; Hofner et al., 2016; Nosek et al., 2012; Obels et al., 2020; Sandve et al., 2013).

The motivation for this present study has been to engage with current debates associated with transparent and reproducible research and, by using Stark's guidelines as a framework, to explore the concept of undertaking transparent and reproducible sociological analyses of large-scale datasets using statistical techniques. We have sought to provide practical guidance that is suitable for sociological research and in the next section we further contribute to establishing practicable standards for sociological research.

Conclusion

Stark's checklist provided an insightful framework for exploring the methodological challenges associated with undertaking reproducible statistically oriented research, and worked well as a 'sensitising device' (see Giddens, 1984). Although we achieved all of the 14 items on the checklist in this research enterprise, on reflection we consider that Stark's list is primarily orientated towards technical aspects of statistical work rather than to statistically orientated social science data analysis. In our view the list does not neatly dovetail with the activities that commonly constitute the workflow when undertaking statistically orientated sociological analyses, and we do not feel they provide a suitable blueprint for sociologists to undertake reproducible research.

Therefore, we conclude by suggesting six *Newer Rules of the Sociological Method* which provide a more suitable and practicable set of guidelines for advancing reproducible statistically orientated sociological research. The six rules are cognisant of the present research culture in sociology, and this includes the accessibility of large-scale data, the statistical methods that are routinely used, the general level of computational skills, the current low levels of access to knowledge and skills in data science and e-research, and the forms of research methods training that are currently available.

Newer rules of the sociological method

(i) Use established data analysis tools (e.g. Stata, SPSS, R or SAS), and clearly state the version, and all the libraries, dependencies and plugins that are used.

(ii) Clearly identify the version of the dataset and its origins (i.e. where and when it was obtained) using a persistent identifier such as a digital object identifier (DOI).

(iii) Construct a data dictionary in a literate format that can easily be understood by someone unconnected with the project.

(iv) Write down all of the code for how the data were prepared for analysis, in a literate format that can easily be understood by someone unconnected with the project.

(v) Write down all of the code for all of the analyses undertaken, and not just the analyses that are presented in the published work, in a literate format that can easily be understood by someone unconnected with the project.

(vi) Archive the project materials, bundled-up as a research object in a findable and accessible location, and endeavour to make them interoperable and re-useable in the future.

The six points provide guidance for undertaking reproducible statistically orientated sociological research. They are not intended to be set in stone, and it is likely that as thinking unfolds, and new protocols and tools emerge from neighbouring computational fields, revisions will be required. These *Newer Rules of the Sociological Method*, provide a clarion call for sociologists who are engaged in statistical analyses to go beyond the boundaries of the current research norms and move towards a transparent research culture that supports reproducibility.

Acknowledgement

We would like to thank Professor Philip Stark for his work which greatly helped to crystalize our thoughts on reproducibility and the data analysis workflow. Vernon would also like to thank Philip for introducing him to Fernado Perez who kindly invited him to BIDS which proved to be very insightful. Min RK deserves a special mention for his initial help installing the original Stata kernel into the Jupyter ecosystem. We would also like to thank Chris Playford, Yuji Shimohira Calvo, Kevin Ralston, Alasdair Gray and Audrey Thomas.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Vernon Gayle https://orcid.org/0000-0002-1929-5983

Notes

[1] Other authors adopt different terminology. For example Nosek and Errington (2020) suggest the term *computational reproducibility* for retesting using the same data and analysis, and *robustness* for investigating the same data with different analyses. Alternatively, Goodman et al. (2016) employ the term *methods reproducibility* when the same data and tools are used, and *results reproducibility* to describe the production of corroborating results. A historical account of the terminology is provided by Plesser (2018).

[2] In addition Philip Stark points to the more general problems of bugs in spreadsheet software see http://eusprig.org/horror-stories.htm accessed 30.06.22.

[3] An interesting blog post entitled 'The Popularity of Data Science Software' was posted by Robert A. Muenchen which indicates that the use of SAS might be in decline more generally http://r4stats.com/articles/popularity/ accessed 30.05.22.

[4] For example see http://r4stats.com/articles/popularity/ accessed 30.05.22; https://www.thoughtco.com/quantitative-analysis-software-review-3026539 accessed 30.05.22; https://www.r-bloggers.com/whats-the-best-statistical-software-a-comparison-of-r-python-sas-spss-and-stata/ accessed 30.05.22; https://www.r-bloggers.com/python-r-vs-spss-sas/ accessed 30.05.22; http://fmwww.bc.edu/GStat/docs/StataVSPSS.html accessed 30.05.22; https://www.guru99.com/sas-versus-r.html accessed 30.05.22;

https://www.quora.com/What-are-the-major-differences-between-Python-and-R-for-data-science accessed 30.05.22.

5. See UK Data Service Study Number SN6614 Understanding Society: Waves 1-9, 2009-2018 and Harmonised BHPS: Waves 1-18, 1991-2009.
6. For an introduction to licence types see https://choosealicense.com accessed 30.05.22.
7. See https://www.ukdataservice.ac.uk/conditions.aspx accessed 30.05.22 for detailed information on data supplied by the UK Data Service.
8. See https://www.ukri.org/files/legacy/documents/rcukopenaccesspolicy-pdf/ accessed 30.05.22.
9. See https://www.ref.ac.uk/media/1228/open_access_summary__v1_0.pdf accessed 30.05.22.
10. See https://www.jisc.ac.uk/guides/an-introduction-to-open-access accessed 30.05.22.
11. See https://osf.io/preprints/socarxiv accessed 30.05.22.
12. For an explanation of this meme in social media see Watling et al. (2021).
13. See https://www.ukdataservice.ac.uk/conditions.aspx accessed 30.05.22 for detailed information on data supplied by the UK Data Service.
14. See https://www.cos.io/ accessed 30.05.22.
15. See https://osf.io/9f6gx/ accessed 30.05.22.
16. Practical information on installation is available here: https://osf.io/8nwvu/ accessed 13.07.22.
17. See https://www.re3data.org/ accessed 30.05.22.
18. See https://www.ncrm.ac.uk/ accessed 30.05.22 and https://www.aqmen.ac.uk/ accessed 30.05.22.
19. See https://datacarpentry.org/ accessed 11.04.22.

Supplemental material

The supplementary materials associated with this article are also available here: https://osf.io/8nwvu/

References

Abbott R, Abbott TD, Abraham S, et al. (2021) Open data from the first and second observing runs of Advanced LIGO and Advanced Virgo. *SoftwareX* 13: 100658.

Aguinis H and Solarino AM (2019) Transparency and replicability in qualitative research: The case of interviews with elite informants. *Strategic Management Journal* 40: 1291–1315.

Bacchelli A and Bird C (2013) Expectations, outcomes, and challenges of modern code review. In: *2013 35th International conference on software engineering (ICSE)*, San Francisco, CA, 18–26 May 2013, pp.712–721. New York, NY: IEEE.

Baiocchi G (2007) Reproducible research in computational economics: Guidelines, integrated approaches, and open source software. *Computational Economics* 30: 19–40.

Baker M (2016) 1,500 scientists lift the lid on reproducibility. *Nature* 533: 452–454.

Barba LA (2016) The hard road to reproducibility. *Science* 354: 142–142.

Barba LA (2018) Terminologies for reproducible research. Available at: https://arxiv.org/pdf/1802.03311.pdf (accessed 13 July 2022).

Barba LA and Thiruvathukal GK (2017) Reproducible research for computing in science & engineering. *Computer Science and Engineering* 19: 85–87.

Bartley M, Sacker A and Clarke P (2004) Employment status, employment conditions, and limiting illness: Prospective evidence from the British household panel survey 1991–2001. *Journal of Epidemiology and Community Health* 58: 501–506.

Baum CF (2009) *An Introduction to Stata Programming*. College Station, TX: Stata Press.

Bechhofer S, Buchan I, De Roure D, et al. (2013) Why linked data is not enough for scientists. *Future Generation Computer Systems* 29: 599–611.

Begley CG and Ioannidis JP (2015) Reproducibility in science: Improving the standard for basic and preclinical research. *Circulation Research* 116: 116–126.

Blischak JD, Davenport ER and Wilson G (2016) A quick introduction to version control with Git and GitHub. *PLoS Computational Biology* 12: e1004668.

Boettiger C (2015) An introduction to Docker for reproducible research. *ACM SIGOPS Operating Systems Review* 49: 71–79.

Boyle P, Feng Z and Gayle V (2009) A new look at family migration and women's employment status. *Journal of Marriage and Family* 71: 417–431.

Bueno de Mesquita B, Gleditsch NP, James P, et al. (2003) Symposium on replication in international studies research. *International Studies Perspectives* 4: 72–107.

Buttliere BT (2014) Using science and psychology to improve the dissemination and evaluation of scientific work. *Frontiers in Computational Neuroscience* 8: 82.

Cabrera D, Vartabedian BS, Spinner RJ, et al. (2017) More than likes and tweets: creating social media portfolios for academic promotion and tenure. *Journal of Graduate Medical Education* 9: 421–425.

Carpenter J and Morganstein D (1986) Comparing statistical software for microcomputers. *Computers & Operations Research* 13: 185–196.

Christensen G, Freese J and Miguel E (2019) *Transparent and Reproducible Social Science Research: How To Do Open Science*. Berkeley: University of California Press.

Connelly R and Gayle V (2019) An investigation of social class inequalities in general cognitive ability in two British birth cohorts. *The British Journal of Sociology* 70(1): 90–108.

Connolly P (2006) The effects of social class and ethnicity on gender differences in GCSE attainment: A secondary analysis of the Youth Cohort Study of England and Wales 1997–2001. *British Educational Research Journal* 32: 3–21.

Donoho DL (2010) An invitation to reproducible computational research. *Biostatistics* 11: 385–388.

Donoho DL, Maleki A, Rahman IU, et al. (2009) Reproducible research in computational harmonic analysis. *Computer Science and Engineering* 11: 8–18.

Dorch BF, Drachen TM and Ellegaard O (2015) The data sharing advantage in astrophysics. *Proceedings of the International Astronomical Union* 11: 172–175.

Ermisch JF and Jenkins SP (1999) Retirement and housing adjustment in later life: Evidence from the British Household Panel Survey. *Labour Economics* 6: 311–333.

Finch SA, La Valle I, McAleese I, et al. (2004) *Youth Cohort Study of England and Wales, 1998-2000* [data collection], 5th edn. Colchester: UK Data Service. SN: 4009. DOI: 10.5255/UKDA-SN-4009-1.

Foster ED and Deardorff A (2017) Open science framework (OSF). *Journal of the Medical Library Association JMLA* 105: 203–206.

Freese J (2007) Replication standards for quantitative social science: Why not sociology? *Sociological Methods & Research* 36: 153–172.

Gayle V (2016) Is the Paper Just a Palimpsest? An appeal for reproducible social stratification research. In: *Social stratification research seminar*, Cambridge.

Gayle V (2017) Some newer rules of the sociological method: Reproducible stratification research. In: *Social stratification research seminar*, University of Edinburgh, Edinburgh.

Gayle V (2018) *Transparent and Reproducible Social Science Research*. Luxembourg: University of Luxembourg.

Gayle V (2021). A practical guide to developing research objects when undertaking reproducible statistically orientated social science research during COVID-19. National Centre for Research Methods. Available at: https://eprints.ncrm.ac.uk/id/eprint/4465/1/research_objects_practical%20guide.pdf (accessed 13 July 2022).

Gayle V, Berridge D and Davies R (2002) Young people's entry into higher education: Quantifying influential factors. *Oxford Review of Education* 28: 5–20.

Gayle V, Connelly R and Playford C (2019a) Then we wrote a gazillion comments - Methodological reflections on collaborative transparent and reproducible stratification research. In: *Social stratification research seminar*, Vrije Universiteit Amsterdam.

Gayle V, Connelly R and Playford C (2019b) *Transparent and Reproducible Social Data Science – A Reflection on Tools and Techniques Suitable for Lifecourse and Equality Research*. Exeter: University of Exeter.

Gayle V and Lambert P (2007) Using quasi-variance to communicate sociological results from statistical models. *Sociology* 41: 1191–1208.

Gayle V, Murray S and Connelly R (2016) Young people and school General Certificate of Secondary Education attainment: Looking for the 'missing middle'. *British Journal of Sociology of Education* 37: 350–370.

Gibson JL (1995) Cautious reflections on a data-archiving policy for political science. *PS: Political Science & Politics* 28: 473–476.

Giddens A (1984) *The Constitution of Society: Outline of the Theory of Structuration*. Berkeley, CA: University of California Press.

Gleditsch NP, Metelits C and Strand H (2003) Posting your data: Will you be scooped or will you be famous. *International Studies Perspectives* 4: 89–97.

Goodman SN, Fanelli D and Ioannidis JP (2016) What does research reproducibility mean? *Science Translational Medicine* 8: 341ps12.

Gould WW (2018) *The Mata Book: A Book for Serious Programmers and Those Who Want to be*. College Station, TX: Stata Press.

Grahe J (2018) *Should You Share Your Data? The Opportunities and Challenges of Data Sharing*. Available at: https://authorservices.taylorandfrancis.com/should-you-share-your-data/ (accessed 13 July 2022).

Hamilton JE (2019) Cash or kudos: Addressing the effort-reward imbalance for academic employees. *International Journal of Stress Management* 26: 193–203.

Haux T (2019) *Dimensions of Impact in the Social Sciences: The Case of Social Policy, Sociology and Political Science Research*. Bristol: Policy Press.

Henneken EA and Accomazzi A (2011) Linking to data-effect on citation rates in astronomy. arXiv preprint arXiv:1111.3618.

Herndon J and O'Reilly R (2016) Data sharing policies in social sciences academic journals: Evolving expectations of data sharing as a form of scholarly communication. In: Kellam L and Thompson K (eds) *Databrarianship: The Academic Data Librarian in Theory and Practice*. Chicago, IL: Association of College and Research Libraries, pp.219–243.

Herndon T, Ash M and Pollin R (2014) Does high public debt consistently stifle economic growth? A critique of Reinhart and Rogoff. *Cambridge Journal of Economics* 38: 257–279.

Hofner B, Schmid M and Edler L (2016) Reproducible research in statistics: A review and guidelines for the Biometrical Journal. *Biometrical journal* 58: 416–427.

Howe B (2012) Virtual appliances, cloud computing, and reproducible research. *Computer Science and Engineering* 14: 36–41.

Jaffe AM (2006) The millennium grand challenge in mathematics. *Notices of the AMS* 53: 652–660.

Janz N (2016) Bringing the gold standard into the classroom: Replication in university teaching. *International Studies Perspectives* 17: 392–407.

Janz N and Freese J (2021) Replicate others as you would like to be replicated yourself. *PS: Political Science & Politics* 54: 305–308.

Kitzes J, Turek D and Deniz F (2018) *The Practice of Reproducible Research: Case Studies and Lessons for the Data-Intensive Sciences*. Oakland, CA: University of California Press.

Kluyver T, Ragan-Kelley B, Pérez F, et al. (2016) Jupyter notebooks - a publishing format for reproducible computational workflows. In: Loizides F and Scmidt B (eds) *Positioning and Power in Academic Publishing: Players, Agents and Agendas*. Amsterdam: IOS Press, pp.87–90.

Knuth DE (1984) Literate programming. *The Computer Journal* 27: 97–111.

Knuth DE (1992) *Literate Programming*. Stanford, CA: Addison-Wesley Publishing Company.

Laine H (2017) Afraid of scooping – Case study on researcher strategies against fear of scooping in the context of open science. *Data Science Journal* 16: 29.

Lam A (2011) What motivates academic scientists to engage in research commercialization: 'Gold', 'ribbon' or 'puzzle'? *Research Policy* 40: 1354–1368.

Lambert P, Browne W and Michaelides D (2015) Contemporary developments in statistical software for social scientists. In: Procter R and Halfpenny P (eds) *Innovations in Digital Research Method*. London: SAGE, pp.143–160.

Leek JT and Peng RD (2015) Opinion: Reproducible research can still be wrong: Adopting a prevention approach. *Proceedings of the National Academy of Sciences* 112: 1645–1646.

Longhi S and Nandi A (2014) *A Practical Guide to Using Panel Data*. London: SAGE.

Long JS (2009) *The Workflow of Data Analysis Using Stata*. College Station, TX: Stata Press.

Macfarlane B (2007) Defining and rewarding academic citizenship: The implications for university promotions policy. *Journal of Higher Education Policy and Management* 29: 261–273.

McKinney W (2010) Data structures for statistical computing in python. In: *Proceedings of the 9th Python in science conference*, Austin, TX, vol. 445, pp.51–56.

Meier KJ (1995) Replication: A view from the streets. *PS: Political Science & Politics* 28: 456–459.

Merton RK (1973) The Normative Structure of Science. In: Storer NW (ed.) *The Sociology of Science - Theoretical and Empirical Investigations*. Chicago, IL: University of Chicago Press, pp.267–280.

Moravcsik A (2014) Transparency: The revolution in qualitative research. *PS: Political Science & Politics* 47: 48–53.

Munafò MR, Nosek BA, Bishop DVM, et al. (2017) A manifesto for reproducible science. *Nature Human Behaviour* 1: 1–9.

Nature (2016) Reality check on reproducibility. *Nature* 533: 437.

Nosek BA and Errington TM (2020) What is replication? *PLoS Biology* 18: e3000691.

Nosek BA, Spies JR and Motyl M (2012) Scientific Utopia: II. Restructuring incentives and practices to promote truth over publishability. *Perspectives on Psychological Science* 7: 615–631.

Obels P, Lakens D, Coles NA, et al. (2020) Analysis of Open Data and computational reproducibility in registered reports in psychology. *Advances in Methods and Practices in Psychological Science* 3: 229–237.

Oliveira S and Stewart DE (2006) *Writing Scientific Software: A Guide to Good Style*. Cambridge: Cambridge University Press.

Parker J (2008) Comparing research and teaching in university promotion criteria. *Higher Education Quarterly* 62: 237–251.

Paskin N (2010) Digital object identifier (DOI®) system. *Encyclopedia of library and information sciences* 3: 1586–1592.

Peng RD (2011) Reproducible research in computational science. *Science* 334: 1226–1227.

Perez F (2013) Literate computing" and computational reproducibility: IPython in the age of data-driven journalism. Available at: http://blog.fperez.org/2013/04/literate-computing-and-computational.html (accessed 13 July 2022).

Perkel JM (2018) Why Jupyter is data scientists' computational notebook of choice. *Nature* 563: 145–146.

Perkel JM (2021) Reactive, reproducible, collaborative: computational notebooks evolve. *Nature* 593: 156–157.

Piwowar HA, Day RS and Fridsma DB (2007) Sharing detailed research data is associated with increased citation rate. *PLoS One* 2: e308.

Plakidas K, Stevanetic S, Schall D, et al. (2016) How do software ecosystems evolve? A quantitative assessment of the R ecosystem. In: *Proceedings of the 20th international systems and software product line conference*, pp.89–98.

Platt W and Platt C (1981) Microcomputer statistical packages and regression performances. In: *1981 Proceedings of the statistical computing section*, pp.309–316. Boston, MA: American Statistical Association.

Playford CJ, Gayle V, Connelly R, et al. (2016) Administrative social science data: The challenge of reproducible research. *Big Data & Society* 3: 2053951716684143.

Playford CJ, Connelly R and Gayle V (2020) The Challenges of Reproducible Research and Teaching it. In: *Love your code event UK data service and office for national statistics*, London.

Plesser HE (2018) Reproducibility vs. replicability: A brief history of a confused terminology. *Frontiers in Neuroinformatics* 11: 76. Epub ahead of print 18 January 2018. Doi: 10.3389/fninf.2017.00076

Plusmathsorg (2016) Andrew Wiles - beginnings, working in secret and saying goodbye to an old friend. Available at: https://www.youtube.com/watch?v=aGko3eEPq8o (accessed 13 July 2022).

Pownall M, Azevedo F, König LM, et al. (2022) The impact of open and reproducible scholarship on students' scientific literacy, engagement, and attitudes towards science: A review and synthesis of the evidence. Available at: https://osf.io/preprints/metaarxiv/9e526 (accessed 13 July 2022).

Reinhart CM and Rogoff KS (2010) Growth in a time of debt. *American Economic Review* 100: 573–578.

Rosenblatt M (2016) *An Incentive-Based Approach for Improving Data Reproducibility*. Washington, DC: American Association for the Advancement of Science, p.336ed335.

Sandve GK, Nekrutenko A, Taylor J, et al. (2013) Ten simple rules for reproducible computational research. *PLoS Computational Biology* 9: e1003285.

Schimanski LA and Alperin JP (2018) The evaluation of scholarship in academic promotion and tenure processes: Past, present, and future. *F1000Research* 7: 1605.

Schulte E, Davison D, Dye T, et al. (2012) A multi-language computing environment for literate programming and reproducible research. *Journal of Statistical Software* 46: 1–24.

Schwab M, Karrenbach N and Claerbout J (2000) Making scientific computations reproducible. *Computer Science and Engineering* 2: 61–67.

Schwemmer C and Wieczorek O (2020) The methodological divide of sociology: Evidence from two decades of Journal Publications. *Sociology* 54: 3–21.

Sefton P, Carragáin EÓ, Goble C, et al. (2019) *Introducing RO-Crate: Research Object Data Packaging*. Brisbane: eResearch Australia.

Sieber JE (1991) *Sharing Social Science Data*. London: SAGE.

Smith KM, Else F and Crookes PA (2014) Engagement and academic promotion: A review of the literature. *Higher Education Research & Development* 33: 836–847.

Stark PB (2015) Science is "show me," not "trust me." Available at: https://www.bitss.org/science-is-show-me-not-trust-me/ (accessed 13 July 2022).

Stark PB (2018) Before reproducibility must come preproducibility. *Nature* 557: 613–614.

StataCorp (2019) *Stata 16 Base Reference Manual*. College Station, TX: Stata Press.

Steneck NH (2007) *Introduction to the Responsible Conduct of Research*. Washington, DC: US Government Printing Office.

Stodden V (2010) *The scientific method in practice: Reproducibility in the computational sciences*. MIT Sloan Research Paper No. 4773-10, Available at: https://ssrn.com/abstract=1550193 (accessed 13 July 2022).

Stodden V, Borwein J and Bailey DH (2013a) Setting the default to reproducible. *computational science research. SIAM News* 46: 4–6.

Stodden V, Guo P and Ma Z (2013b) Toward reproducible computational research: An empirical analysis of data and code policy adoption by journals. *PLoS One* 8: e67111.

Stodden V (2014) The reproducible research movement in statistics. *Statistical Journal of the IAOS* 30: 91–93.

Sukumar PT and Metoyer R (2019) *Replication and Transparency of Qualitative Research From a Constructivist Perspective*. Open Science Framework. Epub ahead of print 2019. DOI: 10.31219/osf.io/6efvp.

Sullivan I, DeHaven A and Mellor D (2019) Open and reproducible research on open science framework. *Current Protocols Essential Laboratory Techniques* 18: e32.

Taylor MF, Brice J, Buck N, et al. (2010) *British Household Panel Survey User Manual: Volume A*. Colchester: ISER.

The Turing Way Community, Arnold B, Bowler L, et al. (2019) *The Turing Way: A Handbook for Reproducible Data Science* (v0.0.4). Zenodo.

Treiman DJ (2009) *Quantitative Data Analysis: Doing Social Research to Test Ideas*. San Francisco, CA: John Wiley & Sons.

Ward BW (2013) What's Better—R, SAS®, SPSS®, or Stata®? Thoughts for instructors of statistics and Research Methods Courses. *Journal of Applied Social Science* 7: 115–120.

Watling C, Ginsburg S and Lingard L (2021) Don't be reviewer 2! Reflections on writing effective peer review comments. *Perspectives on Medical Education* 10: 299–303.

Whiteley AM (2019) *From access to praxis: The case for open access in the humanities and social sciences and the public good*. Unpublished Doctoral Thesis, University of Calgary, Calgary.

Wilkinson MD, Dumontier M, Aalbersberg IJ, et al. (2016) The FAIR guiding principles for scientific data management and stewardship. *Scientific data* 3: 160018.

Williams L (2001) Integrating pair programming into a software development process. In: *Proceedings 14th conference on software engineering education and training. 'In search of a software engineering profession' (Cat. No. PR01059)*, pp.27–36. New York, NY: IEEE.

Yale (2010) Law school roundtable on data and code sharing reproducible research. *Computer Science and Engineering* 12: 8–13.

Young C (2015) Sociologists need to be better at replication. Available at: https://orgtheory.wordpress.com/2015/08/11/sociologists-need-to-be-better-at-replication-a-guest-post-by-cristobal-young/ (accessed 13 July 2022).

Ziemann M, Eren Y and El-Osta A (2016) Gene name errors are widespread in the scientific literature. *Genome Biology* 17(1): 177.

Author biographies

Vernon Gayle is Professor of Sociology and Social Statistics at the University of Edinburgh and a Co-Director of the ESRC National Centre for Research Methods.

Roxanne Connelly is a senior lecturer in Sociology and Quantitative Methods at the University of Edinburgh. She works in the fields of social stratification and the sociology of education.