Reproducible Research Practices and Barriers to Reproducible Research in Geography: Insights from a Survey

Peter Kedron, Joseph Holler & Sarah Bardin

Published online: 01 Dec 2023.

Submit your article to this journal ⬚

Article views: 586

View related articles ⬚

View Crossmark data ⬚

OPEN ACCESS

Check for updates

Reproducible Research Practices and Barriers to Reproducible Research in Geography: Insights from a Survey

Peter Kedron,[a] Joseph Holler,[b] and Sarah Bardin[c]

[a]Department of Geography, University of California Santa Barbara, USA, and School of Geographical Sciences and Urban Planning, Arizona State University, USA; [b]Department of Geography, Middlebury College, USA; [c]School of Geographical Sciences and Urban Planning and Spatial Analysis Research Center (SPARC), Arizona State University, USA

The number of reproduction and replication studies undertaken across the sciences continues to rise, but such studies have not yet become commonplace in geography. Existing attempts to reproduce geographic research suggest that many studies cannot be fully reproduced, or are simply missing components needed to attempt a reproduction. Despite this suggestive evidence, a systematic assessment of geographers' perceptions of reproducibility and use of reproducible research practices remains absent from the literature, as does an identification of the factors that keep geographers from conducting reproduction studies. We address each of these needs by surveying active geographic researchers selected using probability sampling techniques from a rigorously constructed sampling frame. We identify a clear division in perceptions of reproducibility among geographic subfields. We also find varying levels of familiarity with reproducible research practices and a perceived lack of incentives to attempt and publish reproduction studies. Despite many barriers to reproducibility and divisions between subfields, we also find common foundations for examining and expanding reproducibility in the field. These include interest in publishing transparent and reproducible methods, and in reproducing other researchers' studies for a variety of motivations including learning, assessing the internal validity of a study, or extending prior work. *Key Words: epistemology, geographic research methods, open science, reproducibility, researcher survey.*

Reproducible research publicly discloses the evidence used to support claims made in prior work, facilitates the independent verification of those claims, and enables the extension of that work by the broader research community (Schmidt 2009; Nosek, Spies, and Motyl 2012; Earp and Trafimow 2015). Following the National Academies of Science, Engineering, and Medicine (NASEM 2019), reproducibility refers to the ability to independently re-create the results of a study using the same materials, procedures, and conditions of analysis. Although reproducibility is not a guarantee of scientific or practical usefulness, it does provide a strong basis for the collective evaluation of ideas. Nonetheless, systematic reviews of published research papers reveal a lack reproducibility (Collberg et al. 2014; Iqbal et al. 2016); furthermore, attempts to reproduce published research papers frequently fail (Chang and Li 2015; Raghupathi, Raghupathi, and Ren 2022). Previous studies consistently link the irreproducibility of research to inadequate recordkeeping, opaque reporting, the inaccessibility of research components, and a lack of incentives to share research details or to attempt reproduction studies (Ranstam et al. 2000; Anderson, Martinson, and De Vries 2007; NASEM 2019). Surveys of researchers find that few researchers are attempting to independently reproduce the work of others (Baker 2016; Boulbes et al. 2018). At the same time, a sizable portion of survey respondents report knowing of instances in which researchers engaged in questionable or biased research practices tied to the publication of irreproducible results (Fanelli 2009; Fraser et al. 2018).

Despite these concerns, the available literature currently provides insufficient evidence to conclusively evaluate the reproducibility of research generally, or disciplinary research specifically. This knowledge gap exists in part because few reproduction studies have

been published in many fields of research, which limits the quantity of empirical evidence available to make judgments about reproducibility. Any judgments made based on the currently available set of reproduction studies are likely to be limited in scope because existing reproductions typically focus on re-creating the results of a small number of studies selected based on topical interest or researcher familiarity (Open Science Collaboration 2015; Camerer et al. 2016; Camerer et al. 2018). Another approach to assessing the reproducibility of a field is to draw samples of research papers and check the availability and completeness of the research components required for a reproduction. Assessments of this type have narrowly sampled from conference paper series, specific journals, or disciplinary repositories (Stodden et al. 2016; Byrne 2017; Gundersen and Kjensmo 2018; Stodden, Krafczyk, and Bhaskar 2018). Similarly, surveys of researchers asking participants about their use of reproducible research practices have commonly sampled authors from specific journals, members of professional associations, or conference attendees (Baker 2016; NASEM 2019). Surveys have also commonly failed to systematically report the methodological details (e.g., response rate) needed to assess and address potential bias in survey response. Furthermore, reproduction attempts, assessments, and surveys have all typically focused on evaluating the computational components of studies, such as data and code availability, rather than all aspects of research design and execution. In combination, the small number of reproduction studies, reliance on convenience samples of publications and survey participants, and a tendency to focus on computation constrains the scope and generalizability of reproducibility evaluations.

Reproducibility surveys and reproduction attempts in the geographic literature face these same challenges. The few available reproducibility surveys in geography have relied on convenience samples drawn from specialist conferences and have only focused on computationally intensive forms of geographic research (Ostermann and Granell 2017; Nüst et al. 2018; Konkol, Kray, and Pfeiffer 2019; Balz and Rocca 2020). The small number of published attempts to reproduce geographic research have similarly focused on the computational reproducibility of conference papers (Nüst et al. 2018; Ostermann et al. 2021; Nüst et al. 2023) or on specific topics such as COVID-19 (Paez 2021; Kedron, Bardin, et al. 2022; Holler et al. 2023; Kedron,

Bardin, et al. 2023). More recent reproduction attempts by Kedron, Bardin, et al. (2023) show that the factors hindering the re-creation of results and the evaluation of claims likely extend beyond computation into the conceptualization and design of geographic research.

Geographers continue to debate the role of reproduction studies in the discipline (Brunsdon 2016; Singleton, Spielman, and Brunsdon 2016; Goodchild et al. 2021; Kedron et al. 2021; Sui and Kedron 2021; Wainwright 2021; Kedron and Holler 2022), examine the reproducibility of individual studies (Ostermann et al. 2021; Nüst et al. 2023), and build the infrastructure needed to support reproducible research (Nüst and Hinz 2019; Yin et al. 2019; Wilson et al. 2021; Kedron, Bardin, et al. 2022). We have yet to systematically assess the use of reproducible research practices across the discipline's diverse research traditions, or identify the factors that have hindered geographers from adopting reproducible research practices and conducting reproduction studies. Without a systematic assessment of these issues, it is unclear which actions geographers should take if they wish to improve the reproducibility of work in the discipline.

To address this gap in our collective knowledge, we surveyed geographic researchers about their understanding of reproducibility, perception of reproducibility in their subfields, familiarity and use of reproducible research practices, and barriers to reproducibility. To support generalization, we designed a sampling frame to capture researchers from across disciplinary subfields and methodological approaches, and draw survey participants from that frame using a probability sampling scheme. In the remainder of this article, we first present the design of our survey, sampling strategy, and analytical approach. We then present our results, focusing on researcher perceptions and use of reproducible research practices, and then analyzing researcher experiences of attempts to reproduce prior work. Finally, we discuss the implications of our survey results and limitations of our work and conclude by proposing where geography might go from here and how the discipline can contribute to reproducibility across the sciences.

Data and Methods

Complete documentation of the procedures, survey instrument, and other materials used in this study are available through the Survey of Reproducibility in

Geographic Research project (Kedron, Holler, et al. 2023; see—https://osf.io/5yeq8/) hosted by the Open Science Framework (OSF). The OSF project connects to a GitHub repository that hosts the anonymized data set and code used to create all results and supplemental materials along with a complete history of their development. All of the results presented in this article can be independently reproduced using the materials in that repository. The repository links to an interactive visualization of the survey results, which allows users to examine additional cross-tabulations and statistical summaries of the survey data. We encourage interested readers to critically evaluate and build on these materials. Before the start of data collection, we registered a preanalysis plan for the survey with OSF Registries (Kedron, Holler, and Bardin 2022; see—https://osf.io/6zjcp). The survey was conducted under the approval and supervision of the Arizona State Institutional Review Board (STUDY00014232).

Sampling Frame

Our target population of interest is researchers who have recently published in the field of geography. We followed a four-step procedure to create a sampling frame for our survey that captures this diverse population of researchers and the approaches they use when studying geography.

First, beginning at the publication level, we identified journals indexed as either geography or physical geography by the Web of Science's Journal Citation Reports (Clarivate 2023) that also had a five-year impact factor greater than 1.5. From those journals, we created a database of all articles published between 2017 and 2021.

Second, we used the Arizona State University institutional subscription to Scopus (2023) to extract journal information (e.g., subject area), article information (e.g., citation counts), and author information (e.g., corresponding status) for each publication. Because our intention was to capture individuals actively publishing new geographic research, we retained publications indexed by Scopus as document type = "Article" and removed all other publication types (e.g., editorials) from our article database. We also removed articles with missing authorship information.

Third, we created a list of researchers and their published articles, focusing on corresponding authors for two reasons. First, corresponding authorship is one indicator of the level of involvement an individual had in a given work. Although imperfect, it was the best available indicator in the Scopus database as across journals there is no commonly adopted policy for declarations of author work (e.g., CRediT Statements). Second, Scopus maintains e-mail contact information for all corresponding authors, which gave us a means of contacting researchers in our sampling frame. Scopus also maintains a unique identifier for each author (author-id) across time, which allowed us to identify authors across publications.

Fourth, we determined uniqueness by grouping researchers by their author-id, and we determined the most recent contact information by selecting records associated with the most recent year of publication. For 383 researchers who had two or more distinct e-mail addresses in the latest year of publication, we removed noninstitutional personal e-mail addresses and then selected one of the remaining institutional e-mail addresses.

Applying these criteria yielded a sampling frame of 29,828 researchers. On average, these authors published 2.7 articles in geography journals meeting our criteria between 2017 and 2021. Roughly one third (33.0 percent) were most recently a corresponding author for an article published in a general geography journal. A similar proportion (32.0 percent) were most recently a corresponding author for an article published in an earth sciences journal, and smaller proportions published in the social sciences and cultural geography (20.0 percent and 16.0 percent, respectively).

Survey Instrument

The survey first established eligibility based on age and geographic research activity in the past five years and asked researchers to report their primary subfield and methodology. We asked each participant to assess their familiarity with the term *reproducibility* and to provide their own definition. We then provided a definition based on NASEM (2019) to establish a common understanding of reproducibility for the remainder of the survey. Specifically, we defined reproducibility as, "whether research results can be re-created by an independent researcher using the same materials, procedures, and conditions of analysis that were used in the original study." Remaining questions assessed familiarity and

use of reproducible research practices (twenty-two questions), perceptions of the reproducibility of geographic research (two questions), and beliefs about reproducibility with regard to its significance (seventeen questions) and barriers (thirteen questions). For researchers who reported attempting reproductions, we asked them to elaborate on their motivations and outcomes (nine questions).

We developed the survey questions following a review of prior reproducibility surveys (e.g., Fanelli 2009; Baker 2016; Konkol, Kray, and Pfeiffer 2019) and our own reading of recurring issues in the reproducibility literature. We pilot tested the survey instrument with nineteen graduate students and geography faculty with differing levels of experience, disciplinary subfields, and methodological background. After pilot testing, we removed these individuals from our sampling frame to ensure they would not be included in our final sample.

Data Collection

We used a digital form of the tailored design method (Dillman, Smyth, and Christian 2014) to survey geographic researchers between 17 May and 10 June 2022. A simple random sample of 2,000 researchers was drawn without replacement from our sampling frame, and those researchers were invited via e-mail to participate in the online survey. Researchers received their initial invitation on 17 May 2022. Two reminder e-mails were sent to researchers who had not yet completed the survey on 26 May and 31 May 2022.

The online survey was administered through Qualtrics. Participation in the survey was entirely voluntary. Each researcher that opted to participate in the survey was provided with consent documentation approved by institutional review board and linked to the Internet survey instrument. Participants were also given the option to provide an e-mail address for eligibility for one of three prizes of US$90, selected randomly after the data collection period. Participating researchers had the option to exit and reenter the survey and were also able to review and change their answers using a back button as they progressed through the survey. At the end of the data collection period, responses were checked for completeness and coded using the reporting standards of the American Association for Public Opinion Research (AAPOR 2023). Responses were downloaded from Qualtrics, anonymized, and stored in a public, deidentified database in the research compendium.

Analytical Approach

We conducted two statistical analyses of the survey responses. First, we analyzed researcher perspectives on reproducibility following three themes: (1) how geographic researchers define reproducibility, (2) familiarity and experience with reproducible research practices, and (3) perceived barriers to reproducibility. Second, we analyzed researchers' experiences reproducing prior studies including their motivations and experience of successes and barriers. For both analyses, we produced and analyzed descriptive statistical summaries of participant responses to Likert scale questions designed to assess those themes and experiences. We also coded qualitative text responses to selected themes and created quantitative summaries of these themes for each participant. To examine variation among our participants, we cross-tabulated all statistical summaries by disciplinary subfield and methodological approach and compared response frequencies across these subgroups.

Analyzing Researcher Perspectives on Reproducibility

For our first set of analyses, we examined the full set of survey responses. In addition to the examination of statistical summaries of individual Likert scale questions, we created four aggregate measures that summarize participant perceptions and experiences with our four main themes. Complete details about our coding scheme, procedure, and derived data are available in a version-controlled digital compendium that accompanies this publication. The computational code that creates statistical summaries of these variables is similarly available, which makes our entire analysis completely reproducible.

Defining Reproducibility. We coded participants' qualitative definitions of "reproducibility" (1) to assess the similarity between each of the provided definitions and the definition adopted by NASEM (2019), and (2) to determine what participants identified as the motivation for making work reproducible. First, we measured the similarity of each provided definition to the definition adopted by

NASEM (2019). NASEM defines reproducible research as having four characteristics—same data, same procedure, same results, and same conditions. To make this comparison, the authors independently coded each respondent definition for the presence or absence of each of the four characteristics. These assessments were then compiled in a single spreadsheet, which was used to identify disagreements in the independent coding. Disagreements in the assignment of codes were resolved through discussion among the three authors. We created an aggregate measure of definition similarity for the final coded response for each participant by counting the presence of each NASEM definition characteristic, resulting in a measure with the domain [0, 4]. Definitions that received a score of zero did not share any characteristics with the definition provided by NASEM, whereas those that received a score of four included all of the characteristics identified by NASEM.

Second, we also coded each definition to one of four motivations for ensuring the reproducibility of a study: (1) to facilitate the assessment of prior work, (2) to assess experimental research, (3) to improve transparency and facilitate further extension of work, and (4) to improve the transparency and consistency of data collection. We derived this coding from common themes in the responses and our own reading of the reproducibility literature. As earlier, each definition was independently coded by each author before code assignments across authors were compared with disagreements resolved through discussion.

Familiarity and Experience. We measured participant familiarity and experience with five reproducibility-enhancing research practices: (1) the adoption of open source software, (2) the use of research notebooks, (3) data sharing, (4) code and procedure sharing, and (5) research plan preregistration. We assessed familiarity by asking participants to identify whether they were "not at all," "very little," "somewhat," or "to a great extent" familiar with each of the five practices. Participants who identified as being familiar "somewhat" or "to a great extent" with a practice were coded as familiar with that practice. For each participant, we then created an aggregate measure of familiarity with reproducibility-enhancing research practices by counting the number of practices with which they were familiar. This procedure resulted in a familiarity measure with domain [0, 5], where zero indicates a lack of

familiarity with any of the practices assessed and five indicates familiarity with all of the practices assessed.

We followed a similar procedure to construct an aggregate measure of participant experience using reproducibility-enhancing research practices. We assessed researcher experience with each practice by asking participants to identify whether they "never," "rarely," "some of the time," "most of the time," or "always" used that practice in their research. Participants who reported using a practice most of the time or always were coded as having experience with that practice. To create an aggregate measure of experience for each participant, we then counted the number of practices they regularly used. This procedure created an experience score with domain [0, 5], where zero indicates a lack of experience with any of the practices assessed and five indicates experience with all of the practices assessed.

Barriers. Finally, we constructed a measure of participant perceptions of the barriers that hinder reproducibility. We asked participants to identify how frequently they believe twelve different factors contributed to a lack of reproducibility in their subfield. Participants were asked whether they believed each practice "never," "rarely," "occasionally," or "frequently" contributed to a lack of reproducibility. Participants who responded that a factor occasionally or frequently hindered the reproducibility of research were coded as identifying that factor as a barrier. From those responses, we created an aggregate measure of perceived barriers for each participant by counting the number of factors they identified as barriers. This procedure resulted in a measure of barriers with domain [0, 12], where zero indicates a participant identified no barriers to reproducibility.

Analyzing Researcher Reproduction Attempts

For our second set of analyses, we examined only the responses of researchers who reported attempting a reproduction in the past two years to understand what motivated reproduction attempts, how successful those attempts were, and what factors hindered success.

Motivations. To assess what motivated researchers to attempt reproductions, two of the authors independently coded qualitative text responses to the question, "What made you decide to attempt the reproduction(s)?" Each response was categorized as

one of four types of motivation, which we derived from recurring themes in participants' responses and from our review of the reproducibility literature. The four motivation types were to (1) verify or check published research, (2) learn from published research for extension or teaching, (3) internally check their own research to verify their work or increase the transparency of their work, and (4) replicate a study with new data. After each response was coded independently by the two authors, we identified disagreements in motivation assignments across authors. Disagreements were resolved through a discussion that was moderated by the third author. After a review of the coded responses, we chose to use the first and second motivations as a filter to narrow our sample to participants who attempted reproductions that matched the definition of reproducibility in presented by NASEM. We chose to remove participants reporting that they attempted to reproduce their own work because it is unlikely these respondents would encounter the same barriers as researchers attempting to reproduce the work of others, and because a core component of epistemological function of reproducibility is that it acts as an independent check of prior claims. We chose to remove participants who reported replicating a study because the collection of new data changes the purpose and experience of re-creating a study.

Success and Barriers. After narrowing our sample to participants attempting independent reproductions of the work of others, we analyzed participant responses to a set of questions related to the experience making those attempts. To analyze participant success, we created statistical summaries for a series of questions that asked researchers to identify whether they were able to partially or completely re-create some or all results of the target study. We similarly analyzed barriers to the reproduction of

results by creating statistical summaries of participants' ability to access key study artifacts (e.g., data, procedural information, and code).

Results

A total of $n = 218$ of the authors we contacted completed the online survey with information sufficient for analysis. The contact rate for the survey was 13.9 percent and the cooperation rate was 78.7 percent, yielding an overall response rate of 10.9 percent. The refusal rate was 2.9 percent.[1] Another forty authors started the survey but did not complete enough of the survey to be included in the analysis.

Respondents were predominantly male (65.1 percent) and between the ages of thirty-five and fifty-five (62.4 percent). The majority of respondents were academics, and they were balanced across career levels from graduate students to full professors with no one career level comprising more than 30 percent of the sample. Respondents identified with each of the four major disciplinary subfields—physical geography (29.8 percent), geographic methods and GIScience (28.0 percent), nature and society (10.1 percent), and human geography (30.7 percent)—and three major methodological approaches—quantitative (42.2 percent), mixed methods (39.0 percent), and qualitative (18.3 percent).

Table 1 summarizes how researchers define reproducibility, their familiarity with reproducible research practices and experience using them, and the factors they see as barriers to reproducibility in geography. Table 1 presents the mean and standard deviation of the summary measures we created for each of these four themes. Each row in the table captures one of those four themes. The columns of the table separate the statistical summaries associated with those themes by subfield and methodological

Table 1. Descriptive summary of researcher perceptions and experiences with reproducibility

Measure	Subfield				Approach			Overall	n	Missing
	PH	MT	NS	HU	QN	MX	QL			
Definition	1.73 (1.11)	2.00 (1.09)	1.86 (1.06)	1.84 (1.14)	1.91 (1.13)	1.82 (1.04)	1.69 (1.26)	1.83 (1.12)	181	37
Familiarity	3.71 (0.98)	3.97 (0.86)	2.82 (1.76)	2.36 (1.36)	3.75 (0.98)	3.46 (1.22)	1.75 (1.30)	3.26 (1.36)	218	0
Experience	2.06 (1.18)	1.85 (1.38)	1.18 (0.96)	0.60 (0.80)	1.63 (1.32)	1.69 (1.24)	0.48 (0.68)	1.44 (1.28)	218	0
Barriers	8.31 (3.00)	9.31 (2.79)	7.44 (3.20)	7.32 (3.57)	8.78 (2.80)	8.58 (2.96)	5.97 (3.83)	8.20 (3.24)	182	36

Note. Each data cell contains the mean and the standard deviation in parentheses of aggregate measures of researcher definitions, familiarity, experience, and barriers. The domain of each measure is: Definition [0, 4]; Familiarity [0, 5]; Experience [0, 5]; Barriers [0; 12]. PH = physical geography; MT = GIScience and methods; NS = nature and society; HU = human geography; QN = quantitative; MX = mixed methods; QL = qualitative.

approach. For example, the first entry in the overall column of the table indicates that respondents on average included 1.83 of the four components of the NASEM definition in their own definitions of reproducibility, with a standard deviation of 1.12 components. Moving down to the familiarity and experience rows of the same column, these entries indicate that respondents were on average familiar with 3.26 of the five reproducible research practices we survey but indicated having experience using only 1.44 of those same practices. The barriers entry from the same column indicates that respondents identified an average of 8.20 of the 12 factors we surveyed as hindering reproducibility in the discipline. In contrast, the entry summarizing qualitative researchers' perceptions of barriers to reproducibility indicates that these researchers identified an average of 5.97 of 12 barriers to reproducibility.

In aggregate, the data reveal consistent trends in definition, familiarity, experience, and barriers of reproducibility between the subdisciplines and methodological approaches. Respondents who self-identified as specializing in physical geography and geographic methods consistently reported greater familiarity with reproducibility than those working in nature and society and human geography. Similarly, respondents who identified as primarily using quantitative and mixed-methods approaches consistently report greater familiarity of reproducibility than those using qualitative methods. The following subsections present detailed results for each of these topics, highlighting the principal sources of difference between subfields and methodological approaches.

Researcher Perspectives on Reproducibility

Reproducibility is on the minds of geographic researchers. Nearly all researchers reported being at least somewhat familiar with the term reproducibility (89.0 percent), with half reporting being very familiar with the term (53.6 percent). More pointedly, the majority of survey respondents reported thinking about the reproducibility of their own research (80.7 percent), discussing reproducibility with a colleague (70.6 percent), and questioning the reproducibility of published work (57.3 percent) in the past two years. More than half of the researchers we surveyed (52.8 percent) also reported considering reproducibility while peer reviewing a grant proposal or

publication during the same time frame. Researchers, however, estimated that only 50.6 percent of the results published in the discipline were reproducible, albeit with a large standard deviation of 24.7 percent that suggests a great deal of uncertainty about the true value. Few respondents reported attempting to reproduce the work of other researchers (14.7 percent) with fewer still attempting to publish those reproduction studies (6.8 percent).

In total, 58.0 percent of respondents agreed with the statement, "Reproducibility is incompatible with the epistemologies within my subfield," 28.0 percent disagreed with the statement, and 13.0 percent indicated that they did not know. About half of the respondents specializing in human geography (49.3 percent) and nature and society (50.0 percent) indicated that reproducibility was incompatible with the epistemologies of their subfields. Respondents conducting primarily qualitative research were similarly skeptical of the epistemological role of reproducibility in their subfield. Seventy-five percent of qualitative researchers indicated that reproducibility was epistemologically incompatible with their subfield.

Definitions and Importance of Reproducibility. A total of 181 (83.0 percent) of our survey respondents provided an interpretable definition of reproducibility. Geographic researchers provided definitions of reproducibility that explicitly included an average of 1.83 of the four characteristics from the NASEM definition. The availability and use of the same research procedures (80.7 percent) and results (74.0 percent) were the characteristics of reproducibility most frequently identified by researchers. Less than half of respondents explicitly included use of the same data (38.1 percent) or the need to work in the same context (17.7 percent) in their definitions. The pattern of similarity to the NASEM definition and each of its components was consistent across subfields and methodological approaches, with a slightly greater emphasis on data and procedural availability among quantitative and geographic methods and GIScience researchers. The lower inclusion of data and context in definitions might be explained by researchers conceptualizing reproducibility as the formal NASEM definition of replicability, which emphasizes the testing of similar questions and procedures in new contexts with new data. For example, one respondent defined reproducibility as "the extent to which the research design can be replicated in different

geographical contexts." We observed this alternative definition of reproducibility in 20.4 percent of respondents' definitions.

Researchers' definitions of reproducibility were primarily connected to two epistemic functions. Just over half of respondents (52.5 percent) defined reproducibility as a means of assessing prior work for errors or inconsistencies through comparison of original results to results from an attempted reproduction. These comparisons ranged from rigid bitwise quantitative interpretations, as in the "ability to regenerate exactly the results published based on the data and code provided by the authors," to more flexible interpretations in which "other researchers could use the same or similar methodology without great difficulty and, given similar data, arrive at comparable results." Responses also included definitions with a focus on experimental science, as in "an ability to produce consistent results when an experiment is repeated."

Nearly all other researchers (40.9 percent) tied reproducibility to the need for transparency in research so that others could independently expand on prior studies. For example, a quantitative geographer stated, "The methods should provide sufficient information to be able to reproduce the results. In quantitative science this should, at minimum, provide all the equations and algorithms used for any calculation. In the interest of increasing transparency in science, the practice of sharing the code should be encouraged." For others, open science did not necessarily need to result in identical results, reflected in this definition:

> As a qualitative researcher doing in-depth case study research, my studies cannot be perfectly reproduced. But reproducibility sits in the openness about methods and data collection practices, as well as critical reflection about strengths and weaknesses of my research. When we write about those things in the methods section in our papers and theses, their reproducibility is increased.

The remainder of responses (6.6 percent) emphasized repeatable or reliable data observation over all other dimensions of reproducibility. For example,

> As a historical geographer, working with qualitative research methods, I understand reproducibility more in terms of sources than of methods. I see reproducible research as being that which makes clear the origin and location of its data.

A physical geographer similarly emphasized data observations: "Data/observations of some variable can be recovered repeatedly by different observers/methods."

Responses to related Likert questions from the full survey sample ($n = 218$) support the results from the subsample of 181 qualitative definitions analyzed previously. A majority of researchers identified reproducibility as important for validating (75.2 percent) and establishing the credibility (72.5 percent) of research. Respondents also saw reproducing studies as important to reducing the presence of persistent errors in the discipline (77.5 percent) and to increasing trust in research findings (78.5 percent). In parallel with the need for openness and transparency in science, most respondents agreed with the importance of reproducibility for research efficiency (63.3 percent), communication with academics (68.8 percent) and practitioners (64.7 percent), and training students (75.7 percent).

Despite wide recognition of reproducibility as epistemically important, respondents were cautious about drawing conclusions from a single study or reproduction attempt. Only half of the respondents (50.9 percent) agreed that when researchers do not share their data they have less trust in a study. A smaller percentage (41.7 percent) agreed that inability to reproduce a result detracts from the validity of a study, and an even smaller minority agreed that such inability implies that the result is false (26.2 percent).

Qualitative researchers identified reproducibility as playing a much smaller epistemic role compared to the discipline as a whole. A small percentage of qualitative researchers agreed that reproducibility is important for validating research (25.0 percent) or establishing its credibility (20.0 percent). Qualitative researchers similarly placed less emphasis on reproducibility as a means of increasing the accessibility and extensibility of research. Few qualitative researchers had less trust in a study when researchers did not share their data (27.5 percent) or saw reproducibility as important for sharing research with academics (27.5 percent).

The data from our sample of researchers show that there is broad recognition of reproducibility and its importance in geography with three caveats—conflation of reproducibility and replicability, different perspectives from researchers using qualitative methods, and caution about judging the

trustworthiness or validity of published research based on success or failure of an attempt to reproduce a study. In this context, are individual researchers aware of the research practices needed to enhance reproducibility, and have these practices already been adopted for use in research?

Familiarity and Experience with Reproducible Research Practices. Geographic researchers were familiar with an average of 3.26 different reproducible research practices, but only reported experience using an average of 1.44 of these practices in their own work. Table 2 presents researcher familiarity and use of five different reproducible research practices. More than half of all researchers reported familiarity with data sharing (86.7 percent), open source software (85.3 percent), field and lab notebooks (67.0 percent), and code sharing (59.2 percent). A far smaller number of researchers reported using these "familiar" practices regularly in their own work, however. Less than half of the researchers surveyed reported sharing their data (44.5 percent), using open source software (38.1 percent), using field or lab notebooks to record their work (40.0 percent), or sharing their code (18.8 percent) most or all of the time. Only a small subset of researchers reported familiarity with the preregistration of research designs and protocols (27.5 percent) or regular use of this practice (2.7 percent).

Researcher familiarity and use of reproducible research practices varied by disciplinary subfield and methodological approach. Researchers who identified as physical geographers or methodologists and GIScientists reported being familiar with one to two more reproducible research practices than human geographers and those focused on nature and society. Researcher practices similarly diverged by subfield, but no subset of researchers reported using on average more than two of these practices regularly in their work. Quantitative and mixed-methods researchers reported familiarity with and use of an average of two more reproducible research practices when compared to qualitative researchers.

Differences in researcher familiarity and use of specific reproducible research practices across subfields and approaches was greatest for practices more typical of quantitative workflows. When compared to qualitative researchers, quantitative and mixed-methods researchers reported greater familiarity with all reproducible research practices. For example, just 12.5 percent of qualitative researchers reported familiarity with code sharing, whereas 81.5 percent of quantitative and 57.7 percent of mixed-methods researchers reported familiarity with the same practice. Even among quantitative and mixed-methods researchers, familiarity with reproducible research practices did not translate into regular use of those

Table 2. Researcher familiarity with and use of reproducible research practices

Barrier	Subfield				Approach			Overall	Missing
	PH	MT	NS	HU	QN	MX	QL		
Open source software									
Familiarity with	90.8%	100.0%	72.7%	71.7%	96.7%	89.5%	52.5%	85.3%	0
Use of	47.4%	59.0%	31.8%	13.4%	47.8%	41.1%	10.0%	36.1%	0
Lab and field notebooks									
Familiarity with	90.8%	78.0%	63.7%	35.8%	75.0%	71.8%	40.0%	67.0%	0
Use of	63.0%	44.2%	40.9%	14.9%	40.2%	39.4%	20.0%	40.0%	2
Sharing and archiving data									
Familiarity with	95.4%	95.4%	72.7%	71.7%	93.5%	92.9%	57.5%	86.7%	0
Use of	63.0%	44.2%	40.9%	20.9%	51.1%	51.8%	13.0%	44.5%	0
Sharing of code/scripts									
Familiarity with	64.6%	83.6%	45.4%	38.8%	81.5%	57.7%	12.5%	59.2%	0
Use of	20.0%	32.8%	4.5%	10.5%	22.8%	21.1%	5.0%	18.8%	1
Research preregistration									
Familiarity with	29.2%	27.2%	29.2%	17.9%	28.3%	34.2%	12.5%	27.5%	0
Use of	4.6%	4.9%	0.0%	0.0%	1.1%	5.9%	0.0%	2.7%	0
N	65	61	22	67	92	85	40	218	0

Note. Cells report the percentage of respondents reporting being "somewhat" or "very" familiar with a reproducible research practice or using those practices "most of the time" or "always." PH = physical geography; MT = GIScience and methods; NS = nature and society; HU = human geography; QN = quantitative; MX = mixed methods; QL = qualitative.

practices. Just over half of quantitative and mixed-methods researchers reported regularly sharing their data. Less than a quarter of these same researchers reported regularly sharing their code. These rates were lower for qualitative geographers, of which only a small subset reported using open source software or field notes or sharing data and code most of the time or always.

Perceived Barriers to Reproducible Research. Aligning with the differences we observed between researcher familiarity with reproducibility-enhancing research practices and researcher use of those same practices, respondents to our survey identified an average of 8.20 factors contributing to a lack of reproducible research. Geographic researchers working in different subfields and with different methodological approaches each identified a similar number of barriers to reproducibility, as shown in Table 3. Qualitative geographers might be the one group that deviates from this consistent pattern. These researchers identified fewer barriers to reproducibility on average, but with a greater variance that left us unable to distinguish this group from any other. To examine differences in the specific factors that researchers believe hinder reproducibility in the discipline, we divided the twelve factors into three groups—those related to the research environment, the availability of research artifacts, and study-specific characteristics.

Geographic researchers identified the incentive structure of the researcher environment as an important barrier to reproducibility. A majority of geographic researchers identified both the pressure to publish original research (71.5 percent) and insufficient oversight of the research process (71.1 percent) as barriers. A minority of qualitative researchers identified both factors as barriers, but a majority of researchers in all other approaches and subfields identified both factors as barriers to reproducibility. Physical, methods-focused, and quantitative researchers identified these factors as barriers in higher numbers. A minority of geographic researchers (28.4 percent) believe that the fabrication of data, the manipulation of research results, and similar forms of fraud are a cause of irreproducibility in the discipline. This percentage is consistent with concerning results from large surveys and meta-analyses of research on scientific fraud across other scientific disciplines (Fanelli 2009; Baker 2016).

Researchers identified the unavailability of research artifacts (e.g., data) as a second barrier to reproducibility, but the importance placed on different artifacts varied by subfield and methodological approach. A higher percentage of physical and methods-focused researchers identified all five of the artifacts we investigated as common barriers to reproducibility as compared to human and nature–society researchers. The largest differences between

Table 3. Barriers to reproducibility

Barrier	Subfield				Approach			Overall	Missing
	PH	MT	NS	HU	QN	MX	QL		
Research environment									
Pressure to publish	83.1%	75.0%	54.5%	61.2%	77.3%	75.3%	47.5%	71.5%	11
Insufficient oversight	81.6%	82.0%	63.6%	56.7%	82.6%	74.2%	40.0%	71.1%	13
Fraud	32.3%	39.3%	9.1%	22.4%	29.4%	31.8%	20.0%	28.4%	13
Artifact availability									
Insufficient metadata	76.9%	78.7%	59.1%	62.6%	79.4%	72.9%	45.0%	80.2%	15
Unavailable data	89.3%	78.6%	59.1%	65.7%	87.0%	81.2%	42.5%	75.2%	15
Unavailable protocol/code	77.0%	86.9%	50.0%	59.7%	87.0%	70.6%	37.5%	71.1%	14
Published incomplete results	73.8%	78.7%	63.7%	56.7%	78.2%	72.9%	37.2%	68.4%	14
Use restricted data/software	56.9%	82.0%	40.9%	44.8%	70.7%	56.5%	32.5%	57.8%	19
Study characteristics									
Geographic variation	80.0%	72.2%	81.8%	61.2%	79.4%	74.1%	40.0%	71.5%	16
Researcher positionality	50.8%	70.5%	72.8%	70.2%	56.5%	65.9%	80.0%	64.2%	19
Chance	66.2%	65.6%	68.2%	55.2%	67.4%	64.4%	52.5%	62.3%	19
Different computation	53.8%	68.9%	31.8%	38.8%	63.1%	56.5%	25.0%	50.9%	18
N	65	61	22	67	92	85	40	218	0

Note. Cells report the percentage of respondents reporting each factor occasionally or frequently contributed to a lack of reproducibility in geographic research. PH = physical geography; MT = GIScience and methods; NS = nature and society; HU = human geography; QN = quantitative; MX = mixed methods; QL = qualitative.

these groups existed in researchers' beliefs about how often the availability of research protocols and code and the use of restricted data or software affected reproducibility. A similar gap existed between qualitative researchers and mixed-methods or quantitative researchers with regard to identifying code availability or the use of restricted data or software as contributing to irreproducibility.

A majority of researchers identified the complexity and variability of a system (71.5 percent), researcher positionality (64.2 percent), and chance (62.3 percent) as study-specific factors limiting the reproducibility of geographic research. Minor variations in the emphasis placed on these factors exist across subfields and approaches. A higher percentage of nature–society and physical researchers emphasized the important role that spatial variation and complexity of geographic processes can play when attempting to reproduce geographic research, but this factor was also recognized by researchers across subfields and approaches. A smaller percentage of physical geographers placed emphasis on the impact researcher positionality could have on reproducibility when compared to all other subfields. Positionality acknowledges that knowledge is embedded in power relations and that researchers' social and cultural positions affect their relations with research subjects and materials, thus necessitating declaration of that researcher position to evaluate findings (Pratt 2009; Qin 2016; Holmes 2020). Researchers declare and reflect on their positions to assess how their own identity and history might influence aspects of the research process, such as data collection and interpretation. Qualitative researchers were the group most likely to identify positionality as a barrier to reproducibility (80.0 percent). Differences between the computational environment (computer hardware and software) used to conduct an original study and a reproduction attempt were generally not seen as a factor contributing to a lack of reproducibility in the discipline. Of all subgroups, only a majority of methods-focused and quantitative researchers were concerned with computational environments, reflecting research practices used in their areas of research.

Attempted Reproductions

A total of 102 of the researchers who responded to our survey (46.8 percent) reported attempting a reproduction study during the past two years. Twenty-three of those researchers, however, were reproducing their own research results, and another thirteen were replicating prior studies in new locations. In the end, only thirty-two (14.7 percent) of all respondents reported attempting to reproduce a study originally conducted by another researcher during the past two years.

This subset of thirty-two participants formed the basis for our analysis of researcher practices and experiences when attempting reproductions of the work of others. Reproduction attempts were predominantly made by geographic researchers who self-identified with the physical geography (43.8 percent) or geographic methods and GIS (37.5 percent) subfields. Respondents attempting reproductions were also focused on quantitative (68.8 percent) and mixed-methods (41.2 percent) approaches. Only eight of the researchers who attempted reproductions reported submitting any of their findings for publication.

Most of the thirty-two researchers who attempted to reproduce a prior study reported at least some success in accessing data and procedures and in reproducing the prior study results. The majority of researchers (87.5 percent) were able to access some of the data used in the original study, but few researchers (12.5 percent) reported access to all of the original data. Researchers also reported the ability to access at least some information about the study procedures (68.8 percent) and computational environment (59.4 percent), but limited ability to access all procedural (9.4 percent) and computational environment information (12.5 percent).

Reproduction attempts might produce results for comparison to some or all of the results in a prior study. A reproduction could be identical by finding the exact same results, or could be partial by finding slightly different results that still support the same conclusions. Nearly all researchers reported at least partially reproducing some results (81.3 percent), but only seven (21.9 percent) reported being able to at least partially reproduce all results. Only three researchers (9.4 percent) were able to identically reproduce all results.

The reproduction attempt rate and success rates we observed are similar to analogous rates reported in other studies of the reproducibility of geographic research. For example, the Konkol, Kray, and Pfeiffer (2019) survey of participants from the European Geosciences Union General Assembly found that 7 percent of respondents reported often or always

attempting to reproduce the results of other studies. The authors also found rates of reproduction success similar to those identified in our survey. Specifically, the authors found 24 percent of the survey respondents reported being able to often or always reproduce results and 38 percent reported being able to sometimes reproduce results. Five years of reproducibility reviews of AGILE and GIScience conference papers conducted by Nüst et al. (2023) and Ostermann et al. (2021) consistently identified low levels of research artifact availability, implying low reproducibility.

Access to prior study data and procedural information appears to affect the ability to reproduce prior study results. When researchers had access to some of the data from the original study, they reported being able to at least partially reproduce all results in six of twenty-four instances. That success rate rose to three of four when researchers reported access to all data. Procedural information and code appears to matter as much as data. When researchers had access to some of the procedural information from the original study, they reported being able to at least partially reproduce all results in six of nineteen instances. That success rate rose to three of three when researchers reported access to all procedural information and all code.

The small number of reproduction study attempts reported in the survey results makes it difficult to draw broad conclusions. The results are internally consistent, however, and intuitively support the importance of available data and procedures for the reproducibility of geographic research.

Discussion

Our survey results indicate that geographic researchers are aware of reproducibility and reproducible research practices but have yet to incorporate many of those practices into their own work. We found that few researchers attempt to independently reproduce the work of others, or to publish the reproduction attempts they do undertake. In alignment with the broader reproducibility literature, geographic researchers identify the lack of methodological transparency and the unavailability of data and procedural information as key barriers to reproducibility in the discipline. These findings align with a small survey of conference participants conducted by Nüst et al. (2018), which found that geographic

researchers understood the importance of reproducibility but identified data restrictions and a lack of time as key barriers to making their own work more reproducible. Our results also suggest the need to change the culture of research, publication, and promotion within the discipline. This new culture would recognize and reward both original research that is reproducible and attempts to conduct and publish reproduction studies. On the whole, some awareness of reproducible research practices and the infrastructure to attempt reproductions and publish reproducible work exist within the discipline, but geographers have yet to make either a regular part of disciplinary practice.

Our findings also suggest that geographic researchers do not share a single definition of reproducibility. Although researchers share beliefs about the epistemological functions of independent reproductions, they provide definitions that contain different requirements for similarity across studies in terms of data, procedures, results, and context. Moreover, a subset of researchers define reproducibility as what NASEM (2019) defined as replicability—the ability to obtain consistent results across studies designed to answer the same question, each of which has obtained its own data. The interchangeable use of reproducibility and replicability, or the outright reversal of definitions we observed in our sample, has also been documented across the sciences (Plesser 2017; Barba 2018). Given that geography has no established standard use of either term and that many geographic researchers are also trained in other disciplines, it is likely that researchers at least partially inform their definition of reproducibility using concepts prominent in their cognate fields.

The variation in terminology we observed is important for at least two reasons. First, variation in geographic researchers' understanding of reproducibility reflects the discipline's diverse traditions and ways of knowing. Acknowledging this diversity as a strength of the discipline, productive discussions about reproducibility should consider how reproducible research practice fits into different traditions and what common understanding exists across traditions. Second, if researchers lock into a protracted debate about terminology, the community might hinder a more productive discussion about the epistemological role independent reproductions and open science practices can or should play in the discipline.

Our findings point to potentially productive pathways for such a discussion. For example, qualitative geographers are the subset of respondents that most frequently diverged from respondents using other approaches and working in other subfields. This subset of respondents had much less familiarity and use of reproducible research practices and more frequently disagreed that reproducibility was compatible with their epistemological approach. These differences might also explain their lower rates of reporting barriers to reproducibility. Our qualitative respondents, however, did consistently value particular epistemic functions of reproducibility at higher rates than their disagreement with reproducibility on epistemological grounds suggests. This contradiction suggests that qualitative methodologists and reproducibility researchers have yet to meaningfully engage despite sharing some common values. One commonly held value that could serve as a platform for such an engagement is the shared belief in the importance of transparency and precise communication in research.

Qualitative researchers might also have much to contribute to the reproducibility literature owing to their unique perspective and approach to research. For example, qualitative researchers are more concerned than any other group that research positionality is a barrier to reproducibility, but other groups also recognize the impact researcher position and experience can have on research. Perhaps one way forward is to initiate a conversation that highlights how reproducibility is not an absolute standard or determinant of research quality, but instead a means of clarifying for others what was done in a study and why conclusions were drawn as they were. Even if a researcher believes positionality influences data collection and interpretation, using reproducible research practices to control all variables of research design except researcher position might help convey that position and its impact on study results. In other words, reproducibility could open up possibilities for new research questions about researcher positionality and its implications for the evaluation of qualitative and quantitative results. To our knowledge, there is little explicit discussion in the reproducibility literature about how researcher positionality can or should be recorded and conveyed to other researchers. Such work could also move the conversation about reproducibility away from a current focus on the exact re-creation of numerical results, and back to the practice's deeper function—independently assessing the claims of prior research.

Finally, although our results most directly inform quantitative and computational forms of research, we see a number of ways in which the practices examined in our survey can be used to improve the reproducibility of qualitative research in geography. Work in other disciplines can provide a foundation for these improvements. For example, Aguinis and Solarino (2019) developed twelve transparency criteria qualitative researchers studying management can use to catalog and share their data, methods, and overall approach. Roberts, Dowell, and Nie (2019) similarly introduced a methodology to guide reproducible codebook development for thematic analysis. The challenge will be adapting these approaches to geographic analysis. Take the case of using video and audio recordings to capture and share qualitative data. Combining Roberts, Dowell, and Nie's (2019) codebook methodology with version tracking software, a researcher could create a well-documented and detailed record of the interview coding process. When provided with the original recordings, that iterative record could be used to understand, re-create, and assess the final coding. When used to examine geographic phenomena, however, this approach might raise questions about the preservation of participant anonymity. For example, such recordings might require the redaction of not only participant information, but references made in conversations to particular places and times that could identify participants. These redactions both change the amount of information contained in the data, which could make it less useful to a study, and create the additional need to track redaction procedures. Reproducibility might help communicate the rigor of social science research practices and critical social science might help mediate competing values of reproducibility and protection of research subjects. This work remains limited and challenging, however, requiring creative solutions and improvements to public open science infrastructure.

Limitations

To our knowledge, our work is the first systematic attempt to survey a diverse set of geographic researchers about reproducibility. To draw a reliable and generalizable understanding of this issue, we developed a

robust sampling frame representative of the diversity of active geographic researchers. Ideally, we would stratify this set of potential respondents into meaningful subgroups based on their knowledge of reproducibility, and then randomly draw participants from these subgroups. If our resulting sample was imbalanced, we would then use a poststratification procedure to balance the response.

We could not follow this approach for two reasons. First, meaningful stratification and poststratification require knowledge of what predicts differences in response. Given the currently limited understanding of reproducibility within geography, prior to our study we could only speculate about the researcher characteristics predictive of different levels of familiarity and experience with reproducible research (e.g., subfield, methodological approach, or career position). We did not have the knowledge needed to identify reliable predictors. In this respect, our survey lays an initial foundation for examining reproducibility in subsequent studies by providing the first discipline-wide measurement of predictive researcher characteristics. Second, meaningful stratification and poststratification require a population-wide census of key predictors of reproducibility. We are not aware of any census of geographic researchers that contains these data, and we believe that conducting such a census would be difficult given the diversity of the field and the fuzzy boundaries between the discipline's subfields. Given the limitations to stratifying or balancing a survey on reproducibility in geography, our study should be viewed as an exploratory analysis with random sampling and a transparent, reproducible methodology for sample frame construction.

Absent stratification, we have taken steps to reduce several forms of potential bias in our survey. We have worked to eliminate exclusion bias by including in our sampling frame all researchers publishing as corresponding authors in a wide range of geography journals over a five-year period. Although we cannot eliminate the possibility of self-selection bias from our survey, we attempted to quantify potential self-selection by calculating and comparing the completion rates across subfields and approaches. Completion rates for all subfields were between 84.0 percent and 87.0 percent, except slightly higher rates for geographic methods and GIS researchers (96.8 percent). Completion rates were 84.2 percent for mixed methods, 87.0 percent for qualitative methods, and 91.1 percent for quantitative methods. These values suggest

that self-selection was not a significant issue. Finally, we attempted to mitigate the potential for questionnaire bias, which could be caused by partially basing our survey instrument on prior studies that overrepresent perspectives from the computational and experimental sciences. To address this concern, we incorporated into our survey questions from a parallel review of the reproducibility literature available within geography and a review of critiques of positivist science made by social scientists and human geographers. We also included space for text-based qualitative responses in each survey theme, and pilot tested our instrument with a diverse set of geographers.

In light of our finding that participants often provided definitions of reproducibility that only partially matched the NASEM-based definition used in our survey, we cannot be certain which definition respondents had in mind when answering survey questions. We attempted to preemptively address this concern by repeatedly providing the NASEM-based definition during the survey. Although we cannot directly assess which definition each participant used when responding to our questions, we have attempted to indirectly measure this issue and its potential effect on aggregate participant response. Specifically, we examined whether participants who provided reproducibility definitions that shared one or two characteristics with the NASEM definition answered survey questions differently than those that provided definitions that shared three or four characteristics. We found little difference in the responses of researchers in these two groups. For example, 83.8 percent of participants who provided definitions highly similar to the NASEM definition identified unavailable data as a barrier to reproduction, whereas 92.9 percent of participants with low similarity identified this same factor as a barrier. Participants in the two groups also provided similar estimates of the percentage of reproducible results published in the discipline—39.4 percent for the low-similarity group compared to 40.3 percent for the high-similarity group. These levels of similarity were observed across all survey questions. As a final robustness check, we conducted a similar analysis by splitting our sample based on our identification of participant definitions that closely aligned with the NASEM definition of replication. We again found little difference in survey responses between these groups. These checks led us to conclude either that participants used our provided definition when

answering questions or that differences between our provided definition and those of participants were unlikely to have affected their response to our specific set of survey questions.

Conclusion

In this study, we have provided the first systematic survey of the use of reproducible research practices across geography's diverse research traditions. Our results make clear that geographic researchers are aware of reproducible research practices but lack direct experience using those practices. Academic incentive systems and the inaccessibility of key components of prior research hinder reproducibility, and a small percentage of researchers are attempting to independently reproduce past work.

Arising from the survey results, we see an opportunity for geographers to contribute to the interdisciplinary challenges and debates surrounding reproducibility. There has been a tendency to reduce reproducibility to a matter of sharing computational artifacts (e.g., data and code), and to codify artifact sharing as the narrowed goal of reproducibility through requirements for publishing, funding, or badging. Although these practices might allow independent researchers to more easily reproduce and evaluate some aspects of prior studies, they lose sight of the underlying epistemological functions of reproduction studies. Our results demonstrate that geographic researchers have a more varied understanding of reproducibility. Although the discipline does not agree on the importance of sharing of artifacts for computational reproducibility, there is alignment on the clear, precise, and open communication of research and the use of reproduction studies to evaluate or extend the claims of prior work. This shared understanding provides common ground for a discipline-wide debate about the role reproductions can or might play within different epistemologies and subfields, or in the presence of spatial heterogeneity and unique place-based characteristics.

Our work also creates a foundation for the further empirical investigation of reproducibility within geography and its many disciplinary traditions, and more broadly across the sciences. We have made all the materials used in the development and execution of this research openly available so that others can critique and extend our work. We urge other researchers to reanalyze our data, replicate our study, improve our sampling frame and survey instrument, and

progressively create a deeper understanding of questions we only begin to address in this work. One immediate path would be to use our materials to survey geographic researchers about replicability, as our results show that some researchers appear to see a clearer role for replications over exact reproductions in their subfields, whereas others conflate reproducibility and replicability. Disentangling these concepts and connecting them with the epistemological debates presented here is particularly salient in the context of convergence research addressing the most urgent challenges facing humanity, including climate change, global inequality and poverty, global health, and political conflict.

Acknowledgments

Disclosure Statement

No potential conflict of interest was reported by the authors.

Funding

ORCID

Peter Kedron http://orcid.org/0000-0002-1093-3416
Joseph Holler http://orcid.org/0000-0002-2381-2699
Sarah Bardin http://orcid.org/0000-0001-8657-1725

Note

1. All outcome rates are reported using AAPOR (2023) standards. The outcome rates used were response rate 2, cooperation rate 2, refusal rate 1, and contact rate 1.

References

Aguinis, H., and A. M. Solarino. 2019. Transparency and replicability in qualitative research: The case of interviews with elite informants. *Strategic Management Journal* 40 (8):1291–1315. doi: 10.1002/smj.3015.

American Association for Public Opinion Research (AAPOR). 2023. *2023 standard definitions: Final dispositions of case codes and outcome rates for surveys.* 10th ed. Alexandria, VA: AAPOR. https://www.aapor.org/AAPOR_Main/media/publications/Standard-Definitions20169theditionfinal.pdf.

Anderson, M. S., B. C. Martinson, and R. De Vries. 2007. Normative dissonance in science: Results from a national survey of U.S. scientists. *Journal of Empirical Research on Human Research Ethics: JERHRE* 2 (4):3–14. doi: 10.1525/jer.2007.2.4.3.

Baker, M. 2016. 1,500 scientists lift the lid on reproducibility. *Nature* 533 (7604):452–54. doi: 10.1038/533452A.

Balz, T., and F. Rocca. 2020. Reproducibility and replicability in SAR remote sensing. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 13:3834–43. doi: 10.1109/JSTARS.2020.3005912.

Barba, L. A. 2018. Terminologies for reproducible research. *arXiv Preprint* arXiv:1802.03311. doi: 10.48550/arXiv.1802.03311.

Boulbes, D. R., T. Costello, K. Baggerly, F. Fan, R. Wang, R. Bhattacharya, X. Ye, and L. M. Ellis. 2018. A survey on data reproducibility and the effect of publication process on the ethical reporting of laboratory research. *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research* 24 (14):3447–55. doi: 10.1158/1078-0432.CCR-18-0227.

Brunsdon, C. 2016. Quantitative methods I: Reproducible research and quantitative geography. *Progress in Human Geography* 40 (5):687–96. doi: 10.1177/0309132515599625.

Byrne, M. 2017. Making progress toward open data: Reflections on data sharing at PLoS ONE. https://everyone.plos.org/2017/05/08/making-progress-toward-open-data/.

Camerer, C. F., A. Dreber, E. Forsell, T.-H. Ho, J. Huber, M. Johannesson, M. Kirchler, J. Almenberg, A. Altmejd, T. Chan, et al. 2016. Evaluating replicability of laboratory experiments in economics. *Science* 351 (6280):1433–36. doi: 10.1126/science.aaf0918.

Camerer, C. F., A. Dreber, F. Holzmeister, T.-H. Ho, J. Huber, M. Johannesson, M. Kirchler, G. Nave, B. A. Nosek, T. Pfeiffer, et al. 2018. Evaluating the replicability of social science experiments in nature and science between 2010 and 2015. *Nature Human Behaviour* 2 (9):637–44. doi: 10.1038/s41562-018-0399-z.

Chang, A. C., and P. Li. 2015. Is economics research replicable? Sixty published papers from thirteen journals say "usually not." *Finance and Economics Discussion Series* 2015 (83):1–26. https://www.federalreserve.gov/econres/feds/is-economics-research-replicable-sixty-published-papers-from-thirteen-journals-say-quotusually-notquot.htm. doi: 10.17016/FEDS.2015.083.

Clarivate. 2023. Web of Science journal citation reports. Accessed January 15, 2022. https://clarivate.com.

Collberg, C., T. Proebsting, G. Moraila, A. Shankaran, Z. Shi, and A. M. Warren. 2014. Measuring reproducibility in computer systems research. *Technical Report* 37, Department of Computer Science, University of Arizona. http://reproducibility.cs.arizona.edu/v1/tr.pdf

Dillman, D. A., J. D. Smyth, and L. M. Christian. 2014. *Internet, phone, mail, and mixed-mode surveys: The tailored design method.* Hoboken, NJ: Wiley.

Earp, B. D., and D. Trafimow. 2015. Replication, falsification, and the crisis of confidence in social psychology. *Frontiers in Psychology* 6:621. doi: 10.3389/fpsyg.2015.00621.

Fanelli, D. 2009. How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. *PLoS ONE* 4 (5): e5738. doi: 10.1371/journal.pone.0005738.

Fraser, H., T. Parker, S. Nakagawa, A. Barnett, and F. Fidler. 2018. Questionable research practices in ecology and evolution. *PLoS ONE* 13 (7):e0200303. doi: 10.1371/journal.pone.0200303.

Goodchild, M. F., A. S. Fotheringham, P. Kedron, and W. Li. 2021. Introduction: Forum on reproducibility and replicability in geography. *Annals of the American Association of Geographers* 111 (5):1271–74. doi: 10.1080/24694452.2020.1806030.

Gundersen, O. E., and S. Kjensmo. 2018. State of the art: Reproducibility in artificial intelligence. *Proceedings of the AAAI Conference on Artificial Intelligence* 32 (1):1644–51. doi: 10.1609/aaai.v32i1.11503.

Holler, J., P. Kedron, D. An-Pham, D. Burt, and J. Zhou. 2023. *Reproduction of Chakraborty 2021 distribution of COVID-19 and intra-categorical analysis of people with disabilities.* OSF. osf.io/s5mtq. doi: 10.17605/OSF.IO/S5MTQ.

Holmes, A. G. D. 2020. Researcher positionality—A consideration of its influence and place in qualitative research—A new researcher guide. *Shanlax International Journal of Education* 8 (2):1–9. doi: 10.34293/education.v8i2.1477.

Iqbal, S. A., J. D. Wallach, M. J. Khoury, S. D. Schully, and J. P. Ioannidis. 2016. Reproducible research practices and transparency across the biomedical literature. *PLoS Biology* 14 (1):e1002333. doi: 10.1371/journal.pbio.1002333.

Kedron, P., S. Bardin, T. D. Hoffman, M. Sachdeva, M. Quick, and J. Holler. 2022. A replication of DiMaggio et al. (2020) in Phoenix, AZ. *Annals of Epidemiology* 74:8–14. doi: 10.1016/j.annepidem.2022.05.005.

Kedron, P., S. Bardin, J. Holler, J. Gilman, B. Grady, M. Seeley, X. Wang, and W. Yang. 2023. A framework for moving beyond computational reproducibility: Lessons from three reproductions of geographical analyses of COVID-19. *Geographical Analysis*. doi: 10.1111/gean.12370.

Kedron, P., and J. Holler. 2022. Replication and the search for the laws in the geographic sciences. *Annals of GIS* 28 (1):45–56. doi: 10.1080/19475683.2022.2027011.

Kedron, P., J. Holler, and S. Bardin. 2023. *A survey of reproducibility in geographic research*. OSF. osf.io/5yeq8 doi: 10.17605/OSF.IO/5YEQ8.

Kedron, P., J. Holler, S. Bardin, and Z. Hilgendorf. 2022. *Reproducibility, replicability, and open science practices in the geographical sciences*. OSF. doi: 10.17605/OSF.IO/C5A2R.

Kedron, P., W. Li, S. Fotheringham, and M. Goodchild. 2021. Reproducibility and replicability: Opportunities and challenges for geospatial research. *International Journal of Geographical Information Science* 35 (3):427–45. doi: 10.1080/13658816.2020.1802032.

Konkol, M., C. Kray, and M. Pfeiffer. 2019. Computational reproducibility in geoscientific papers: Insights from a series of studies with geoscientists and a reproduction study. *International Journal of Geographical Information Science* 33 (2):408–29. doi: 10.1080/13658816.2018.1508687.

National Academies of Science, Engineering, and Medicine (NASEM). 2019. *Reproducibility and replicability in science*. Washington, DC: National Academies Press. doi: 10.17226/25303.

Nosek, B. A., J. R. Spies, and M. Motyl. 2012. Scientific utopia: II. Restructuring incentives and practices to promote truth over publishability. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science* 7 (6):615–31. doi: 10.1177/1745691612459058.

Nüst, D., C. Granell, B. Hofer, M. Konkol, F. O. Ostermann, R. Sileryte, and V. Cerutti. 2018. Reproducible research and GIScience: An evaluation using agile conference papers. *PeerJ* 6:e5072. doi: 10.7287/peerj.preprints.26561v1.

Nüst, D., and M. Hinz. 2019. containerit: Generating dockerfiles for reproducible research with R. *Journal of Open Source Software* 4 (40):1603. doi: 10.21105/joss.01603.

Nüst, D., F. O. Ostermann, R. Sileryte, B. Hofer, C. Granell, M. Teperek, A. Graser, K. Broman, K. Hettne, and M. Konkol. 2023. *Reproducible publications at agile conferences*. OSF. osf.io/phmce. doi: 10.17605/OSF.IO/PHMCE.

Open Science Collaboration. 2015. Estimating the reproducibility of psychological science. *Science* 349 (6251):aac4716. doi: 10.1126/science.aac4716.

Ostermann, F. O., and C. Granell. 2017. Advancing science with VGI: Reproducibility and replicability of recent studies using VGI. *Transactions in GIS* 21 (2):224–37. doi: 10.1111/tgis.12195.

Ostermann, F. O., D. Nüst, C. Granell, B. Hofer, and M. Konkol. 2021. Reproducible research and GIScience: An evaluation using GIScience conference papers. In *11th international conference on geographic information science (GISience 2021)*. Vol. 208, Part 2, , ed. K. Janowicz and J. A. Verstegen, 2:1–2:16. Dagstuhl, Germany: Schloss Dagstuhl – Leibniz-Zentrum für Informatik. doi: 10.4230/LIPIcs.GIScience.2021.II.2.

Paez, A. 2021. Reproducibility of research during COVID-19: Examining the case of population density and the basic reproductive rate from the perspective of spatial analysis. *Geographical Analysis* 54 (4):860–80. doi: 10.1111/gean.12307.

Plesser, H. E. 2017. Reproducibility vs. replicability: A brief history of a confused terminology. *Frontiers in Neuroinformatics* 11:76. doi: 10.3389/fninf.2017.00076.

Pratt, G. 2009. "Positionality." In *The Dictionary of Human Geography*, ed. D. Gregory, R. Johnston, G. Pratt, M. Watts, and S. Whatmore, 556–57. Malden, MA: Wiley-Blackwell.

Qin, D. 2016. Positionality. *The Wiley Blackwell Encyclopedia of Gender and Sexuality Studies*, ed. N. A. Naples. London: John Wiley.

Raghupathi, W., V. Raghupathi, and J. Ren. 2022. Reproducibility in computing research: An empirical study. *IEEE Access* 10:29207–23. doi: 10.1109/ACCESS.2022.3158675.

Ranstam, J., M. Buyse, S. L. George, S. Evans, N. L. Geller, B. Scherrer, E. Lesaffre, G. Murray, L. Edler, J. L. Hutton, et al. 2000. Fraud in medical research: An international survey of biostatisticians. *Controlled Clinical Trials* 21 (5):415–27. doi: 10.1016/s0197-2456(00)00069-6.

Roberts, K., A. Dowell, and J.-B. Nie. 2019. Attempting rigour and replicability in thematic analysis of qualitative research data: A case study of codebook development. *BMC Medical Research Methodology* 19 (1):66. doi: 10.1186/s12874-019-0707-y.

Schmidt, S. 2009. Shall we really do it again? The powerful concept of replication is neglected in the social sciences. *Review of General Psychology* 13 (2):90–100. doi: 10.1037/a0015108.

Scopus. 2023. Home page. Accessed January 20, 2022. http://www.scopus.com.

Singleton, A. D., S. Spielman, and C. Brunsdon. 2016. Establishing a framework for open geographic information science. *International Journal of Geographical Information Science* 30 (8):1507–21. doi: 10.1080/13658816.2015.1137579.

Stodden, V., M. S. Krafczyk, and A. Bhaskar. 2018. Enabling the verification of computational results: An empirical evaluation of computational reproducibility. In *Proceedings of the first international workshop on practical reproducible evaluation of computer systems*, 1–5. London and New York: Association for Computing Machinery. doi: 10.1145/3214239.3214242.

Stodden, V., M. McNutt, D. H. Bailey, E. Deelman, Y. Gil, B. Hanson, M. A. Heroux, J. P. A. Ioannidis, and M. Taufer. 2016. Enhancing reproducibility for computational methods. *Science* 354 (6317):1240–41. doi: 10.1126/science.aah6168.

Sui, D., and P. Kedron. 2021. Reproducibility and replicability in the context of the contested identities of geography. *Annals of the American Association of Geographers* 111 (5):1275–83. doi: 10.1080/24694452.2020.1806024.

Wainwright, J. 2021. Is critical human geography research replicable? *Annals of the American Association of Geographers* 111 (5):1284–90. doi: 10.1080/24694452.2020.1806025.

Wilson, J. P., K. Butler, S. Gao, Y. Hu, W. Li, and D. J. Wright. 2021. A five-star guide for achieving replicability and reproducibility when working with GIS software and algorithms. *Annals of the American Association of Geographers* 111 (5):1311–17. doi: 10.1080/24694452.2020.1806026.

Yin, D., Y. Liu, H. Hu, J. Terstriep, X. Hong, A. Padmanabhan, and S. Wang. 2019. CyberGIS-Jupyter for reproducible and scalable geospatial analytics. *Concurrency and Computation: Practice and Experience* 31 (11):e5040. doi: 10.1002/cpe.5040.

PETER KEDRON is an Associate Professor in the Department of Geography at the University of California Santa Barbara, Santa Barbara CA 93106, USA. E-mail: peterkedron@ucsb.edu. His research program develops and uses spatial analytical methods to understand persistently uneven patterns of spatial development. His recent work focuses on improving the production and accumulation of knowledge used to benefit society through policy.

JOSEPH HOLLER is an Assistant Professor of Geography at Middlebury College, Middlebury, VT 05753, USA. E-mail: josephh@middlebury.edu. He is a human geographer and GIScientist with research interests in social vulnerability and adaptation in the context of climate change and environmental degradation. His research intersects with work in political ecology, development geography, human dimensions of global change, and GIScience.

SARAH BARDIN is a Doctoral Student in Geography at the School of Geographical Sciences and Urban Planning at Arizona State University, Tempe, AZ 85381, USA. E-mail: sfbardin@asu.edu. Her research focuses on leveraging spatial modeling and GIS to generate policy-relevant, actionable insights from data.