

Attempts to Reproduce Health Sciences Research Are Uncommon and May Be Declining Over Time: A Systematic Review and Meta-Analysis

Niklas Bobrovitz

n.jhbobro@ucalgary.ca

University of Calgary

Harriet Ware

University of Calgary

Zara Stelfox

University of Calgary

Corson Johnstone

University of Calgary

Juliane Butterfield

University of Calgary

Stephana J Moss

Dalhousie University

Liam Whalen-Browne

University of Calgary

Danial J Niven

University of Calgary

Tom Stelfox

University of Alberta

Anmol Shahid

University of Calgary

Research Article

Keywords: Reproducibility, meta-science, meta-epidemiology, systematic review, meta-analysis, reproducibility crisis

Posted Date: October 14th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10976193/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Background

There are widespread concerns about irreproducible research driving ineffective use of resources, patient harm, and loss of public trust in science. Given these concerns, research reproducibility should be routinely and consistently monitored. However, efforts to quantify research reproducibility are not well characterized.

Objectives

We aimed to synthesize evidence on the prevalence of attempts to reproduce research in the health sciences and evaluate temporal trends in these efforts.

Methods

We searched MEDLINE, Embase, Web of Science, the Cochrane Controlled Trials Registry and Database of Systematic Reviews, and grey literature from January 1, 2000 to May 31, 2023. We included English-language studies published after 1999 that estimated the prevalence of reproducibility attempts in the health sciences. Reproducibility attempts were classified as studies attempting to exactly repeat another study's methods (methodological), compare results yielded from using the same or similar methods as another study (results), or compare conclusions yielded from an attempt to reproduce a study including re-analyses of the original data (inferential). We meta-analyzed prevalence estimates using random effects models. The statistical significance threshold was $p < 0.01$.

Results

We included 41 studies from biomedical/preclinical research ($n = 10$), clinical research ($n = 26$), population health research ($n = 2$), and multiple research domains ($n = 3$), representing 30 Clarivate research areas. Between 2000 and 2023, the pooled estimated prevalence of reproducibility attempts in the health sciences was 1.6% (0.6–4.2%, $n = 52$). There were significant differences in estimates between research domains, research areas, types of reproducibility being evaluated, and the methods to identify reproducibility attempts. Reproducibility attempts were identified in three ways: (1) inference based on study language, (2) inference based on the comparability of two or more studies, and (3) explicit reporting that a study was a reproducibility attempt. Meta-regression identified a non-significant increase over time in the estimated prevalence of reproducibility attempts between 2000 and 2015 (increase in odds of attempts per year 49.8% [4.2% to 115.4%], $p = 0.03$) followed by a significant decrease between 2016 and 2023 (decrease in odds of attempts per year – 50.2% [-31.4% to -63.8%], $p < 0.001$). This decline was not significant in sensitivity analysis using the earliest, median, and latest publication dates of the underlying study data.

Conclusions

Reproducibility attempts constitute a small proportion of the literature. Their prevalence varies by research domain and area, as well as by the method of classification. Since 2016 attempts to reproduce research may have

declined. There is a need to increase attempts to reproduce research and standardize the classification and reporting of reproduction attempts.

INTRODUCTION

Reproducing research is a key element of sequential discovery in health sciences research and essential for informed implementation of findings in pre-clinical research, clinical medicine, health services, and public health. However, an estimated 85% of study methods and 50% of study results and conclusions in the health sciences cannot be reproduced.^{1–14} There is widespread concern about irreproducible research in health sciences disciplines.¹⁵ The consequences of irreproducible research are substantial including approximately \$28 billion wasted annually pursuing false leads in pre-clinical laboratories in the United States¹⁶, exposure of patients to ineffective or harmful clinical practices that were initially shown to be beneficial and later overturned¹⁷, and eroding public trust in science by popularizing findings that are rescinded.¹⁸

Reproducibility studies are a potential safeguard against these issues as they evaluate whether study methods can be reliably repeated and built upon in future research, offer evidence regarding the trustworthiness and generalizability of findings, and identify sources of aberrant or unique results that may prompt reflection on scientific practices, helping to identify challenges and limitations in designing, executing, and reporting research.^{1,19,20}

Reproducibility studies serve to repeat the methods of previous studies (methodological), can corroborate results using the same or similar methods (results), and evaluate whether similar conclusions can be achieved (inferential).^{21–24} Some reproducibility efforts are explicit, while others are implicit and may only become apparent through comparisons between studies addressing the same topic.

Studies suggest that reproducibility is low across health research, including biomedical and pre-clinical research, clinical medicine, health and social care services, and population health. For example, an estimated 15% of health sciences research has reproducible methods, while half of studies demonstrate reproducible results and inferences.²⁵ However, it is unknown how commonly the scientific community quantifies and monitors reproducibility through dedicated reproduction efforts (i.e., reproducibility studies). Moreover, it remains unclear whether the prevalence of these efforts has changed over time in response to growing concerns regarding research reproducibility. Understanding the prevalence of reproducibility studies may provide important insight into the extent to which the health sciences are actively responding to reproducibility concerns and investing in mechanisms to evaluate the trustworthiness of scientific findings.

To address this knowledge gap, we conducted a systematic review with meta-analysis to summarize: (1) the proportion of published health research studies that explicitly or implicitly attempt to reproduce prior research (i.e., prevalence of reproducibility attempts), (2) how attempts to reproduce research are identified, and (3) temporal trends in reproducibility attempts over time.

METHODS

Protocol and registration

This study followed guidelines from the Cochrane handbook²⁶ and is reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) (Supplementary File 1).²⁷ The study protocol was registered and published *a priori* with the PROSPERO.¹

Search strategy

The search strategy (Supplementary File 2) was developed by a health sciences librarian (DL) and validated by a second librarian (ZP) using the Peer Review of Electronic Search Strategies (PRESS) checklist.²⁸ MEDLINE, Embase, Web of Science, the Cochrane Controlled Trials Registry, and the Cochrane Database of Systematic Reviews were searched from 1 January 2000 to 31 May 2023. Two supplemental searches were conducted: (1) keyword searching of high impact journals in general science and medicine (i.e., New England Journal of Medicine, Lancet, Journal of the American Medical Association, Nature, Science, Cell) for journal guidelines, articles, and webpages on research reproducibility and (2) searching websites of scientific institutions for reports on research reproducibility.

Selection criteria

Two authors independently screened abstracts and full texts (NB, CJ, JB, SM, LWB). Disagreement was resolved by consensus and discussion, or arbitration by a third author (NB, TS).

Articles were included if they were published in English and reported on the proportion of published research articles dedicated to reproducing prior or concurrent research (reproducibility attempts) in the health sciences. We defined health science as research examining any aspect of health, disease, or healthcare with the aim of generating knowledge, interventions, or technologies to improve human health or well-being.^{29,30} We operationalized reproducibility attempts into three categories: (1) methodological (i.e., attempting to exactly repeat another study's methods), (2) results reproducibility (i.e., attempting to reproduce the results of a study using the same or similar methods), and (3) inferential reproducibility (i.e., evaluating whether the same scientific conclusions are reached by repeating a study or part of the study including re-analyses of the original study data).²¹⁻²⁴ Results reproducibility was further classified as approximate (i.e., similar methods but new population, setting, exposure, or analyses), re-test (i.e., same methods but new sample), or unspecified.

Studies were excluded if more than half of their data were unrelated to the health sciences, they were published in non-English language articles or were published prior to the year 2000 to ensure feasibility of completing the review.

The inclusion and exclusion criteria were designed to capture estimates with potential for generalizability while ensuring the review would be feasible. As such, we included studies that evaluated reproducibility across multiple "PECO" (Population, Exposure, Comparison, Outcome) research questions. This definition excluded studies focused on reproducing a single research question. Those are studies typically conducted by investigators focused on answering a single research question. Instead, our definition aimed to capture studies seeking to broadly evaluate reproducibility. PECO questions were considered unique if they differed in their exposures or comparators and populations or outcomes.

Supplementary File 2 contains further details about the inclusion and exclusion criteria.

Data extraction and management

Data were extracted using an electronic spreadsheet (Airtable) developed through a pilot extraction of twenty randomly selected full-text articles (NB, CJ, JB, SM). The remaining articles were extracted by one reviewer and verified by a second (NB, CJ, JB). Disagreements were resolved through discussion and consensus, or by consultation with a third reviewer when required (TS).

Data were extracted on study characteristics including study design, domain of health science the study focused on (i.e., biomedical/pre-clinical, clinical, health/social care services, population health)^{31,32}, specific research area according to the Web of Science Research Area classification (e.g., oncology)³³, overall aim of the article, funding source, and type of reproducibility (i.e., methodological, results, inferential) being targeted by the reproducibility attempts.

Data analysis

Analyses were completed using STATA 18 and R (4.1.2) (HW). Meta and metafor packages were used. All underlying code and dataset are publicly available via the Open Science Framework.³⁴

Each overarching research domains (e.g., clinical) was subcategorized by research area (e.g., oncology) to created unique research domain-areas (e.g., clinical-oncology).³³

We meta-analyzed estimates of the proportion of studies that are reproducibility attempts using mixed effects logistic regression. Estimates were weighted by sample size. For each pooled estimate, 95% confidence intervals and 95% prediction intervals were calculated. Prediction intervals were calculated from the estimated random-effects model (metaprop) and back-transformed to the prevalence scale. To explore variation in estimates we conducted univariable and multivariable meta-regression by fitting generalized linear mixed models for binomial data.^{35,36} Independent predictors were defined a priori as publication year, research domain, research area, metric used to measure reproducibility attempts, and risk of bias. Statistical heterogeneity was assessed using Cochrane's Q.

To evaluate temporal trends in reproducibility attempts we conducted separate meta-analysis for each publication year and plotted point estimates with 95% confidence intervals (95% CI). Meta-regression with time-varying covariates was used to identify statistically significant trends over time. Given that many estimates in the meta-analyses were derived from underlying data spanning multiple years we conducted a sensitivity analysis using the earliest, median, and latest publication dates of the underlying study data. The threshold for statistical significance was $p < 0.01$.

Content analysis was used to categorize the methods described by studies for conducting reproducibility attempts and metrics used by the studies to quantify the proportion studies considered reproducibility attempts.

Estimate risk of bias was assessed using a modified Joanna Briggs Institute Checklist (Supplementary File 2).^{37,38}

Funding

The work was funded by the Canadian Institutes of Health Research. The funding body had no role in the design of the study or in the collection, analysis, or interpretation of data, or in writing the manuscript.

Patient and Public Involvement

Patients or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

RESULTS

Characteristics of included studies

The search yielded 55,218 articles for screening of which 41 were included in the review (Figure 1). Supplementary Table 1 and 2 list references for included and excluded studies.

Summary characteristics of the included studies are reported in Table 1 with individual level characteristics reported in Supplementary Table 3. Articles reported on biomedical/preclinical (n=10) and clinical (n=26) research domains, with few articles from population health (n=2), and multiple domains (n=3). No studies were identified in the health/social care services domain. All included studies reported on attempts to reproduce research results. Across studies, 30 distinct research areas were represented.

Table 1. Summary characteristics of included articles

Characteristic	No. (%) of 41 articles
Study design^a	
Evidence synthesis	
Scoping, systematic, or non-systematic review (cross-sectional)	33 (80.5%)
Scoping, systematic, or non-systematic review (cohort)	7 (17.1%)
Observational	
Secondary dataset analysis (cross-sectional)	1 (2.4%)
Year of publication^a	
2000-2009	1 (2.4%)
2010-2019	13 (31.7%)
2020-2023	27 (65.9%)
Research domain	
Biomedical/Preclinical	10 (24.4%)
Clinical	26 (63.4%)
Health/Social care services	0
Population health	2 (4.9%)
Multiple ^b	3 (7.3%)
Research area^c	
Multiple research areas (Biomedical/Preclinical)	3 (7.3%)
Multiple research areas (Clinical)	3 (7.3%)
Psychiatry (Clinical)	3 (7.3%)
Surgery (Clinical)	3 (7.3%)
Cardiovascular system and cardiology (Clinical)	2 (4.9%)
Sports sciences (Clinical)	2 (4.9%)
Type of reproducibility	
Results	41 (100%)
Results - re-test reproduction	6 (14.6%)
Results - approximate reproduction	10 (29.4%)
Results - not specified	28 (62.3%)
Funding Source^a	
Granting agency or private foundation	14 (34.2%)

Departmental or home institution	8 (19.5%)
Government	5 (12.2%)
None	7 (17.1%)
Not reported	14 (34.2%)

^aPercentages may not add up to 100 due to rounding, missing information, or non-mutually exclusive categories. Study design, type of reproducibility examined, and funding did not contain mutually exclusive categories. ^bArticles reporting relevance to health sciences but not specifying a research domain within health sciences. ^cResearch areas reported in 2 or more articles.

Identifying reproducibility attempts

Investigators used three approaches to determine whether a given study was a reproducibility attempt: 1) inferred based on the language used to describe a study; 2) inferred based on the comparability of design elements for two or more studies; 3) the study explicitly reported being a reproducibility attempt. These three approaches were operationalized in 12 unique ways (Supplementary Table 5).

Study designs used to quantify the prevalence of reproducibility attempts

Three study-designs were identified that estimated the prevalence of reproducibility attempts. These included cross-sectional evidence syntheses (n=33), cohort-style evidence syntheses (n=7), and one secondary dataset analysis. These study designs were operationalized using nine unique approaches (Supplementary Table 4).

Estimates of the prevalence of reproducibility attempts

There were 52 estimates of the prevalence of reproducibility attempts (Supplementary Table 7), including 10 from biomedical/preclinical research, 32 from clinical research, eight for population health research, and two spanning multiple research domains. There were estimates for 30 research areas. Most estimates had a moderate risk of bias (n=48) while a minority were high risk (n=4). (Supplementary Table 6). Estimates were at increased risk of bias primarily due to low sample sizes and lack of adjustment for potentially confounding factors.

The pooled estimated prevalence of reproducibility attempts in the health sciences was 1.6% (95% CI 0.6-4.2%, 95% PI 0.0-91.2%, n=52) (Figure 2). Prevalence estimates differed by subtype of results reproducibility (p=0.004). Attempts classified as approximate reproduction were more prevalent than re-test or unspecified results reproducibility attempts: 10.6% (95% CI 3.4-28.3%, 95% PI 0.1-95.7%), 0.5% (95% CI 0.0-7.2%, 95% PI 0.0-85.9%), 0.7% (95% CI 0.2-2.7%, 95% PI 0.0-86.7%), respectfully (Supplementary Figure 1). There were significant differences between research domains (p<0.001). Reproducibility attempts were most prevalence in population health at 3.5% (95% CI 0.60-18.2%, 95% PI 0.0-93.1%), followed by biomedical/preclinical at 3.0% (95% CI 0.5-17.1%. 95% PI 0.0-96.5%), clinical at 1.3% (95% CI 0.3-4.9%, 95% PI 0.0-95.0%) and multiple research domains at 0.1% (95% CI 0.1-0.1%, 95% PI 0.0-0.5%) (Supplementary Figure 2). Estimates also differed between research areas (p<0.001) with a range of 0.0% to 77.8% (Table 2). However, estimates for 22 of the 30 research areas were based on single studies.

Table 2. Estimates of the prevalence of attempts to reproduce health sciences research

Research domain		
Research area	Biomedical/Preclinical	Clinical
Anesthesiology		0.0% (95% CI 0.0-2.5%, n=1)
Cardiovascular System and Cardiology		8.2% (95% CI 0.6-58.8%, 95% PI 0.0-100.0%, n=2)
Critical Care Medicine		41.9% (95% CI 34.2-49.9%, n=1)
Emergency Medicine		0.0% (95% CI 0.0-1.3%, n=1)
Endocrinology and Metabolism		0.9% (95% CI 0.2-3.6%, n=1)
General Internal Medicine		28.2% (95% CI 20.7-37.1%, n=1)
Genetics and Heredity	60.5% (95% CI 54.7-66.1%, n=1)	
Integrative and Complementary Medicine	77.8% (95% CI 69.3-84.5%, n=1)	
Medical Informatics		0.3% (95% CI 0.2-0.5%, n=1)
Neurosciences and Neurology	12.6% (95% CI 7.2-21.0%, n=1)	0.0% (95% CI 0.0-1.7%, n=1)
Nursing		3.0% (95% CI 0.1-49.1%, 95% PI 0.0-100.0%, n=3)
Obstetrics and Gynecology		0.0% (95% CI 0.0-1.6%, n=1)
Oncology	0.5% (95% CI 0.0-3.3%, n=1)	78.1% (95% CI 59.6-90.1%, n=1)
Otorhinolaryngology	0.0% (95% CI 0.0-3.1%, n=1)	
Pediatrics		17.5% (95% CI 11.8-25.1%, n=1)
Psychiatry		1.7% (95% CI 0.1-17.1%, 95% PI 0.0-94.2%, n=6)
Psychology		1.0% (95% CI 0.1-6.2%, n=1)
Radiology, Nuclear Medicine, and Medical Imaging		7.1% (95% CI 4.5-11.0%, n=1)
Respiratory System	0.4% (95% CI 0.0-2.6%, n=1)	
Sports Science		0.0% (95% CI 0.0-100.0%, 95% PI 0.0-100.0%, n=2)
Surgery		0.0% (95% CI 0.0-100.0%, 95% PI 0.0-100.0%, n=3)
Urology and Nephrology	0.0% (95% CI 0.0-2.7%, n=1)	0.3% (95% CI 0.0-2.2%, n=1)
Multiple Research Areas	4.4% (95% CI 1.8-10.5%, 95% PI 0.1-65.5%, n=3)	57.8% (95% CI 31.5-80.3%, 95% PI 1.3-99.3%, n=3)

n=denotes the number of contributing estimates. There was one estimate in the “Population Health” domain for the research area “Public, Environmental, and Occupational Health” [3.5% (95% CI 0.6-18.2%, 95% PI 0.0-93.1%, n=8)]. There were two estimates in the “Multiple domains” category for the research areas “Neurosciences and Neurology” [0.1% (95% CI 0.1-0.1%, n=1)] and “Substance Abuse” [0.4% (95% CI 0.0-2.6%, n=1)].

The approaches used to identify reproducibility attempts were associated with significant differences in estimated prevalence ($p=0.001$). Studies inferred to be reproducibility attempts based on the language used in the study were most prevalent at 12.6% (95% CI 5.1%-28.2%, 95% PI 0.3-88.7%), followed by those inferred to be reproducibility attempts based on the comparability of study design elements to previous research at 11.9% (95% CI 2.4%-42.8%, 95% PI 0.0-98.5%), and studies that explicitly reported being reproducibility attempts at 0.3% (95% CI 0.1-1.0%, 95% PI 0.0-45.6%) (Supplementary Figure 3).

Prevalence estimates did not differ by risk of bias (moderate 1.7% [95% CI 0.7-4.4%, 95% PI 0.0-90.5%]), high 0.7% [95% CI 0.0-56.0%, 95% PI 0.0-100.0%], $p=0.75$) (Supplementary Figure 4).

Temporal trends in reproducibility attempts

Piece-wise meta-regression identified a non-significant increase over time in the prevalence of reproducibility attempts between 2000 and 2015 (increase in odds of attempts per year 49.8% [4.2% to 115.4%], $p=0.0291$) followed by a significant decrease between 2016 and 2023 (decrease in odds of attempts per year -50.2% [-31.4% to -63.8%], $p<0.001$) (Figure 3, Supplementary Table 8).

Sensitivity analysis showed similar trends however, the magnitude and statistical significance of those trends varied by analytic assumption. Analysis without a piece-wise cut point showed a non-significant decline from 2000 to 2023 in the odds of attempts per year (-18.3% [-2.1% to -31.7%], $p=0.028$). Analysis excluding the year 2000, after which there was an 11-year gap in data, showed a significant decline from 2012 to 2023 in the odds of attempts per year (-29.6% [-14.6% to -41.9%], $p<0.001$) (Supplementary Figure 5-6).

Evidence for a temporal decline was sensitive to how calendar year was assigned to estimates derived from multi-year data. Analysis using different assumptions about publication dates for the underlying study data (earliest, median, and latest year of data contributing to the estimates) showed non-significant decreases in odds of attempts per year: earliest year of data -0.02% [-0.05% to 0.03%], $p=0.46$; median year of data -0.04% [-0.11% to 0.04%], $p=0.32$; latest year of data -0.06% [-0.19% to 0.07%], $p=0.34$. (Supplementary Figure 7, Table 9-11).

Black solid lines represent 95% confidence intervals. Black dotted lines are the slope of the piece-wise meta-regression time-varying coefficients using 2015 as a cut-point.

Meta-regression analysis for heterogeneity

Meta-regression showed that moderate to large amounts of variation between estimates were accounted for by the research area ($R^2=76.2\%$) and the approaches used to identify reproducibility attempts ($R^2=42.3\%$). Smaller amounts of variation were accounted for by type of reproducibility being assessed ($R^2=17.6\%$) and the combination of research domain and publication year ($R^2=17.2\%$). Multi-variable models containing the variables research area and study design used to estimate reproducibility attempts did not converge due to low sample sizes (Supplementary Table 12).

DISCUSSION

This systematic review and meta-analysis found 41 studies that estimated the prevalence of reproducibility attempts in the health sciences (i.e. proportion of studies that attempted to reproduce prior research). The results suggest that attempts to reproduce health sciences research are uncommon (1.6% of the health sciences literature). However, estimates varied substantially between research domains, research areas, and methods used to identify reproducibility attempts. The prevalence of reproducibility attempts declined from 2016 to 2023 however, evidence for a temporal decline was sensitive to analytic assumptions.

Few studies have examined the prevalence of reproducibility attempts across multiple research domains and topics. Cobey et al. conducted a scoping review in which they screened 11,224 articles published in 2018/19 in psychology, health sciences, biomedicine, economics, and education journals. They found 42 (0.37%) reproducibility attempts in the health science and six (0.05%) in biomedical science.³⁹ Our estimate was higher, perhaps because we included data from a different time-period (2000-2023) and used meta-analysis to pool a collection of primary estimates, as opposed to identifying a sample of studies and determining the proportion of reproducibility attempts in the sample.

The findings of this study have important implications. First, greater investment in reproducibility attempts may be warranted. Despite recognition of the importance of reproducing research the proportion of literature dedicated to reproducing studies is low and appears to have declined since 2016. In some areas of research less than 1% of the literature were attempts to reproduce research. Conducting reproducibility studies is necessary to identify irreproducible science, build the evidence-base for a particular research question, and address issues with scientific practice that may be contributing to reproducibility issues.

There is no consensus on the optimal number of reproducibility attempts for a given topic, an appropriate ratio of reproducibility attempts relative to novel studies, or the proportion of funding that should be dedicated to reproducibility studies. These targets may be selected by scientists, clinicians, policymakers and patients as all are affected by the consequences of irreproducible research. There are active funding and policy initiatives that will support reproducibility attempts including the National Institutes of Health agency wide strategy to create infrastructure to embed reproducibility into NIH research, incentivise reproducibility studies, and further support existing reproducibility focused policies, resources, and programs.⁴⁰ The impact of these types of initiatives should be evaluated.

Instruction on how to conduct reproducibility studies has been proposed to be included as part of standard scientific training with the goal of increasing capacity for conducting reproducibility attempts.^{41,42} Some argue pursuing novel work should take priority over repeating what has already been done. However, the two are not mutually exclusive. Conducting reproducibility studies can provide novel findings. Investigating why the methods, results, or conclusions of a study may not have been reproduced may have useful insights relevant to future studies on the same topic as well as other scientific questions.

Second, estimates of prevalence differed by the approach investigators used to identify reproducibility attempts. This suggests that standardization of methods to identify reproducibly attempts are needed. For example, studies that explicitly self-identify as reproducibility attempts may be considered deliberate efforts to reproduce prior research, as opposed to studies only deemed to be reproducibility attempts because their research question and methods overlapped with another similar study, whether knowingly or unknowingly. These latter studies may be incidental reproducibility attempts and may yield different insights than deliberate attempts. Future researchers should make this distinction when classifying reproducibility studies.

Third, some areas of research had significantly higher prevalence of reproducibility attempts than others. Some of these differences may be attributable to how reproducibility attempts were defined. However, there may be an opportunity to evaluate whether the prevalence of reproduction attempts in an area of research is related to the reproducibility of research methods, findings, and inferences. For example, a research area with more reproduction attempts may have more reproducible research as the number of attempts may reflect a disciplinary interest and focus on robust science. Conversely, a research area that has low levels of reproducibility may trigger more reproduction attempts. To the best of our knowledge this relationship has not been evaluated at scale. Future work could also evaluate associations between the prevalence of reproducibility attempts and other meta-science metrics of scientific robustness including the proportion of studies at high risk of bias, low GRADE certainty of evidence, adherence to methodological and reporting checklists, and proportion of retracted articles.

Fourth, there is a need for primary studies to quantify the prevalence of methodological and inferential reproducibility attempts. We found no estimates of the prevalence of types of studies. There have been many large scale and well-funded interventions focused on improving methodological reproducibility including protocol registries (e.g. ClinicalTrials.gov, Open Science Framework, PROSPERO), reporting guidelines and guideline networks (e.g. MIAME, ARRIVE, PRISMA, EQUATOR), and code and data repositories. As such, it would be prudent to monitor the prevalence of methodological reproducibility as this serves to determine whether these interventions are having the intended effect. The lack of inferential reproducibility estimates should also be addressed in future studies. Assessments of inferential reproducibility are crucial to understanding the potential consequences of reproducibility issues given that irreproducible methods and results may not necessarily translate into different conclusions and as such may not impact applied scientific and clinical decision making. Inferential reproducibility merits dedicated measurement and may deserve attention comparable to that given to results reproducibility. It is currently unknown whether this is the case.

This review had several limitations. First, for feasibility purposes focused on articles reporting the prevalence of attempts for multiple research questions to provide insight into trends within and between research domains and areas. It is unclear if including articles that reported the prevalence of reproduction attempts for a single research question would have impacted our estimates. Second, there were only enough data to create pooled estimates for eight research areas. The other 22 research areas were represented by estimates from single studies. As additional data become available further pooled analyses may provide insight into heterogeneity between research areas and improve the robustness of prevalence estimates. Third, assigning a single publication year to the prevalence estimates may have masked trends within the data. We conducted sensitivity analysis to explore different assumptions about publication years which showed declines over time but there were smaller in magnitude and non-significant. As such, the trends over time that we report should be interpreted with caution. Future primary studies may consider reporting the prevalence of reproducibility attempts by year so that trends are clearer and allow synthesis of data to occur by year. Fourth, differences in how primary studies identified reproducibility attempts may have introduced measurement inconsistency, making it difficult to distinguish true between-field differences from differences in ascertainment. Fifth, we restricted inclusion to English-language studies published from 2000 onward, which may limit generalizability to earlier or non-English meta-research.

CONCLUSION

Reproducibility attempts constituted a small proportion of the literature. Attempts to reproduce increased from 2000 to 2015 then declined from 2016 to 2023. The prevalence of attempts varied by research domain and area, as well as by the method of classification. Research reproducibility is not being routinely and consistently monitored.

As such there is a need to increase attempts to reproduce research and to standardize the classification and reporting of reproduction attempts.

Declarations

ACKNOWLEDGMENTS

We would like to acknowledge Dr. Diane Lorenzetti and Dr. Zahra Premji, information specialists that reviewed the search strategy and completed the PRESS checklist.

COMPETING INTEREST DECLARATION

The authors declare no competing interests.

AUTHOR STATEMENT

N.B., DJ.N., and T.S., conceived the idea and designed the study. N.B., C.J., J.K., L.W.B., F.K., R.P., A.D., and B.F. screened articles, extracted data, and verified data. N.B. and H.W. analysed the data. N.B., H.W., DJ.N., and T.S. interpreted the data. N.B. and S.J.M. completed risk of bias assessments. N.B., Z.S., T.S., and A.S. prepared the manuscript with substantive input from all the co-authors. All authors approved the submitted version of the manuscript.

DATA AVAILABILITY AND CODE AVAILABILITY

The full dataset, analysis approach, and statistical code underlying all analysis and figures are available in the Open Science Framework project repository entitled “The Reproducibility of Health Research”, 10.17605/OSF.IO/3FHD9, along with a readme file that describes each file (2026 readme systematic review on research reproducibility attempts).

References

1. Khan F, Kennet J, Bobrovitz N, Stelfox T. Prevalence of Research Reproducibility in the Health Sciences; 2023. PROSPERO. <https://www.crd.york.ac.uk/PROSPERO/view/CRD42023399411>
2. Begley CG, Ioannidis JPA. Reproducibility in Science: Improving the Standard for Basic and Preclinical Research. *Circul Res*. 2015;116(1):116–26. 10.1161/CIRCRESAHA.114.303819.
3. Open Science Collaboration. Estimating the reproducibility of psychological science. *Science*. 2015;349(6251):aac4716. 10.1126/science.aac4716.
4. Camerer CF, Dreber A, Forsell E, et al. Evaluating replicability of laboratory experiments in economics. *Science*. 2016;351(6280):1433–6. 10.1126/science.aaf0918.
5. Brandt MJ, IJzerman H, Dijksterhuis A, et al. The Replication Recipe: What makes for a convincing replication? *J Exp Soc Psychol*. 2014;50:217–24. 10.1016/j.jesp.2013.10.005.
6. Baker M, Penny D. Is there a reproducibility crisis? *Nature*. 2016;533(7604):452–4. 10.1038/533452a.
7. Nature. Challenges in irreproducible research. *Nature*. 2018. Accessed March 3, 2022. <https://www.nature.com/collections/prbfkwmwvz/>
8. Ioannidis JPA. Contradicted and initially stronger effects in highly cited clinical research. *JAMA*. 2005;294(2):218–28. 10.1001/jama.294.2.218.

9. Errington TM, Iorns E, Gunn W, Tan FE, Lomax J, Nosek BA. An open investigation of the reproducibility of cancer biology research. *Elife*. 2014;3. 10.7554/eLife.04333.
10. Karp NA. Reproducible preclinical research—Is embracing variability the answer? *PLoS Biol*. 2018;16(3):e2005413. 10.1371/journal.pbio.2005413.
11. Fanelli D, Opinion. Is science really facing a reproducibility crisis, and do we need it to? *Proc Natl Acad Sci U S A*. 2018;115(11):2628–31. 10.1073/pnas.1708272114.
12. Ioannidis JPA. Meta-research: Why research on research matters. *PLoS Biol*. 2018;16(3):e2005468. 10.1371/journal.pbio.2005468.
13. Gall T, Ioannidis JPA, Maniatis Z. The credibility crisis in research: Can economics tools help? *PLoS Biol*. 2017;15(4):e2001846. 10.1371/journal.pbio.2001846.
14. Iqbal SA, Wallach JD, Khoury MJ, Schully SD, Ioannidis JPA. Reproducible Research Practices and Transparency across the Biomedical Literature. *PLoS Biol*. 2016;14(1):e1002333. 10.1371/journal.pbio.1002333.
15. Cobey KD, Ebrahimzadeh S, Page MJ, Thibault RT, Nguyen PY, Abu-Dalfa F, Moher D. Biomedical researchers' perspectives on the reproducibility of research. *PLoS Biol*. 2024;22(11):e3002870. 10.1371/journal.pbio.3002870. . PMID: 39499707; PMCID: PMC11537370.
16. Freedman LP, Cockburn IM, Simcoe TS. The Economics of Reproducibility in Preclinical Research. *PLoS Biol*. 2015;13(6):e1002165. 10.1371/journal.pbio.1002165.
17. Niven DJ, Rubinfeld GD, Kramer AA, Stelfox HT. Effect of Published Scientific Evidence on Glycemic Control in Adult Intensive Care Units. *JAMA Intern Med*. 2015;175(5):801–9. 10.1001/jamainternmed.2015.0157.
18. Davey M. Retracted studies may have damaged public trust in science, top researchers fear. *The Guardian*. <https://www.theguardian.com/science/2020/jun/06/retracted-studies-may-have-damaged-public-trust-in-science-top-researchers-fear>. June 5, 2020.
19. Munafo MR, Nosek BA, Bishop DVM et al. A manifesto for reproducible science. *Nat Hum Behav*. 2017; 1(1).
20. National Academies of Sciences, Engineering, and Medicine; Policy and Global Affairs. Reproducibility and Replicability in Science. National Academies Press (US); 2019. <https://www.ncbi.nlm.nih.gov/books/NBK547537/>. 10.17226/25303.
21. Niven DJ, McCormick TJ, Straus SE, et al. Reproducibility of clinical research in critical care: a scoping review. *BMC Med*. 2018;16(1):26. 10.1186/s12916-018-1018-6.
22. Goodman SN, Fanelli D, Ioannidis JPA. What does research reproducibility mean? *Sci Transl Med*. 2016;8(341):341ps12. 10.1126/scitranslmed.aaf5027.
23. Clemens MA. The Meaning of Failed Replications: A Review and Proposal. *J Economic Surveys*. 2017;31(1):326–42. 10.1111/joes.12139.
24. Plesser HE. Reproducibility vs. Replicability: A Brief History of a Confused Terminology. *Front Neuroinform*. 2017;11:76. 10.3389/fninf.2017.00076.
25. Bobrovitz N, Harriet W, Johnston C, Kennett J, Moss SJ, Whalen-Browne L, Khan F, Fletcher B, Niven DJ, Stelfox T. Reproducibility of Health Sciences Research: A Systematic Review and Meta-Analysis of Prevalence and Metrics. Under-review.
26. Higgins J, Thomas J, Chandler J et al. Cochrane Handbook for Systematic Reviews of Interventions. Accessed September 2, 2022. <https://training.cochrane.org/handbook>

27. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, Shamseer L, Tetzlaff JM, Akl EA, Brennan SE, Chou R, Glanville J, Grimshaw JM, Hróbjartsson A, Lalu MM, Li T, Loder EW, Mayo-Wilson E, McDonald S, McGuinness LA, Stewart LA, Thomas J, Tricco AC, Welch VA, Whiting P, Moher D. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. 2021;372:n71. 10.1136/bmj.n71. . PMID: 33782057; PMCID: PMC8005924.
28. McGowan J, Sampson M, Salzwedel DM, Cogo E, Foerster V, Lefebvre C. PRESS Peer Review of Electronic Search Strategies: 2015 Guideline Statement. *J Clin Epidemiol*. 2016;75:40–6. 10.1016/j.jclinepi.2016.01.021.
29. Medical Research Council UK Research and Innovation. Understanding health research. January 11. 2023. <https://www.ukri.org/councils/mrc/facilities-and-resources/find-an-mrc-facility-or-resource/mrc-regulatory-support-centre/understanding-health-research/what-is-health-research/>
30. Canadian Institutes of Health Research. Health Research in Canada and You. January 1. 2015. <https://cihr-irsc.gc.ca/e/43753.html>
31. Canadian Institutes of Health Research. Glossary of Funding-Related Terms. 2018. Accessed March 3, 2022. <http://www.cihr-irsc.gc.ca/e/34190.html#r5>
32. Medical Research Council. Health Research Classification System. 2018. http://hrcsonline.net/wp-content/uploads/2018/01/HRCS_Main_Handbook_v2_Feb2018.pdf
33. Web of Science. Web of Science Research Areas. Published online 2021. <https://incites.help.clarivate.com/Content/Research-Areas/wos-research-areas.htm#:~:text=The%20Web%20of%20Science%20schema,represented%20by%20smaller%20sub%2Dfields>
34. Khan F, Kennet J, Bobrovitz N, Stelfox T. Prevalence of Research Reproducibility in the Health Sciences: Systematic Review and Meta-Analysis Protocol; 2021. 10.17605/OSF.IO/3FHD9
35. Schwarzer G, Chemaitelly H, Abu-Raddad LJ, Rücker G. Seriously misleading results using inverse of Freeman-Tukey double arcsine transformation in meta-analysis of single proportions. *Res Synth Methods*. 2019;10(3):476–83. 10.1002/jrsm.1348.
36. Lin L, Chu H. Meta-analysis of Proportions Using Generalized Linear Mixed Models. *Epidemiology*. 2020;31(5):713–7. 10.1097/EDE.0000000000001232.
37. Bobrovitz N, Noël K, Li Z, et al. SeroTracker-RoB: A decision rule-based algorithm for reproducible risk of bias assessment of seroprevalence studies. *Res Synth Methods*. 2023;14(3):414–26. 10.1002/jrsm.1620.
38. Joanna Briggs Institute. *The Joanna Briggs Institute*. <https://jbi.global/>.
39. Cobey KD, Fehlmann CA, Christ Franco M, et al. Epidemiological characteristics and prevalence rates of research reproducibility across disciplines: A scoping review of articles published in 2018–2019. *Elife*. 2023;12. 10.7554/eLife.78518.
40. National Institutes of Health. Strengthening Replication and Reproducibility of NIH-funded Research. <https://www.nih.gov/replicationandreproducibility>. Accessed July 31, 2026.
41. Susann Auer NA, Haeltermann TL, Weissgerber JC, Erlich D, Susilaradeya, Reproducibility for Everyone Team. Magdalena Julkowska, Małgorzata Anna Gazda, Benjamin Schwessinger, Nafisa M Jadavji, (2021) Science Forum: A community-led initiative for training in reproducible research *eLife* 10:e64719.
42. Kohrs FE, Auer S, Bannach-Brown A, Fiedler S, Haven TL, Heise V, Holman C, Azevedo F, Bernard R, Bleier A, Bössel N, Cahill BP, Castro LJ, Ehrenhofer A, Eichel K, Frank M, Frick C, Frieze M, Gärtner A, Gierend K, Grüning DJ, Hahn L, Hülsemann M, Ihle M, Illius S, König L, König M, Kulke L, Kutlin A, Lammers F, Mehler DMA, Miehl C, Müller-Alcazar A, Neuendorf C, Niemeyer H, Pargent F, Peikert A, Pfeuffer CU, Reinecke R, Röer JP, Rohmann

JL, Sánchez-Tójar A, Scherbaum S, Sixtus E, Spitzer L, Straßburger VM, Weber M, Whitmire CJ, Zerna J, Zorbek D, Zumstein P, Weissgerber TL. Eleven strategies for making reproducible research and open science training the norm at research institutions. *Elife*. 2023;12:e89736. 10.7554/eLife.89736. . PMID: 37994903; PMCID: PMC10666927.

Figures

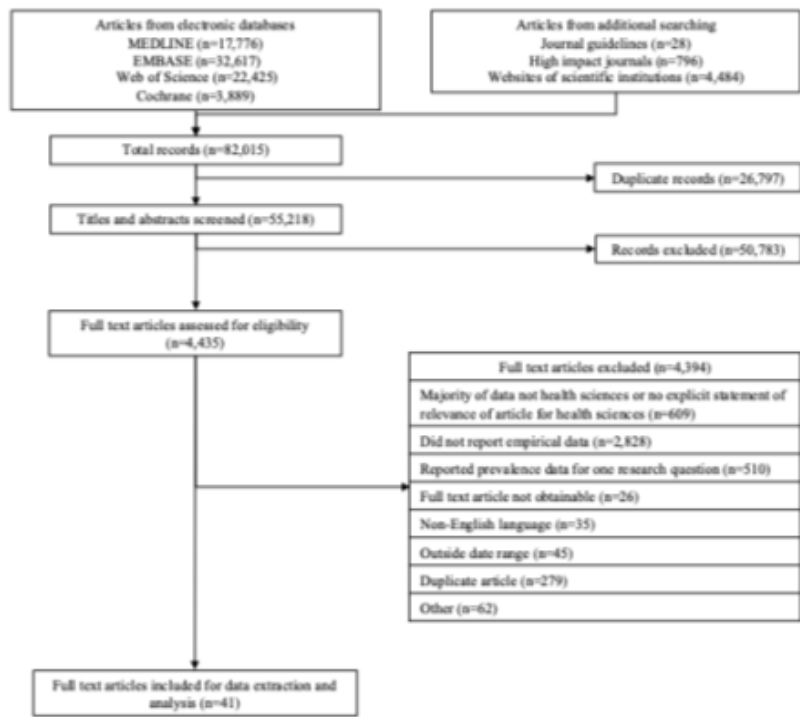


Figure 1

Study selection

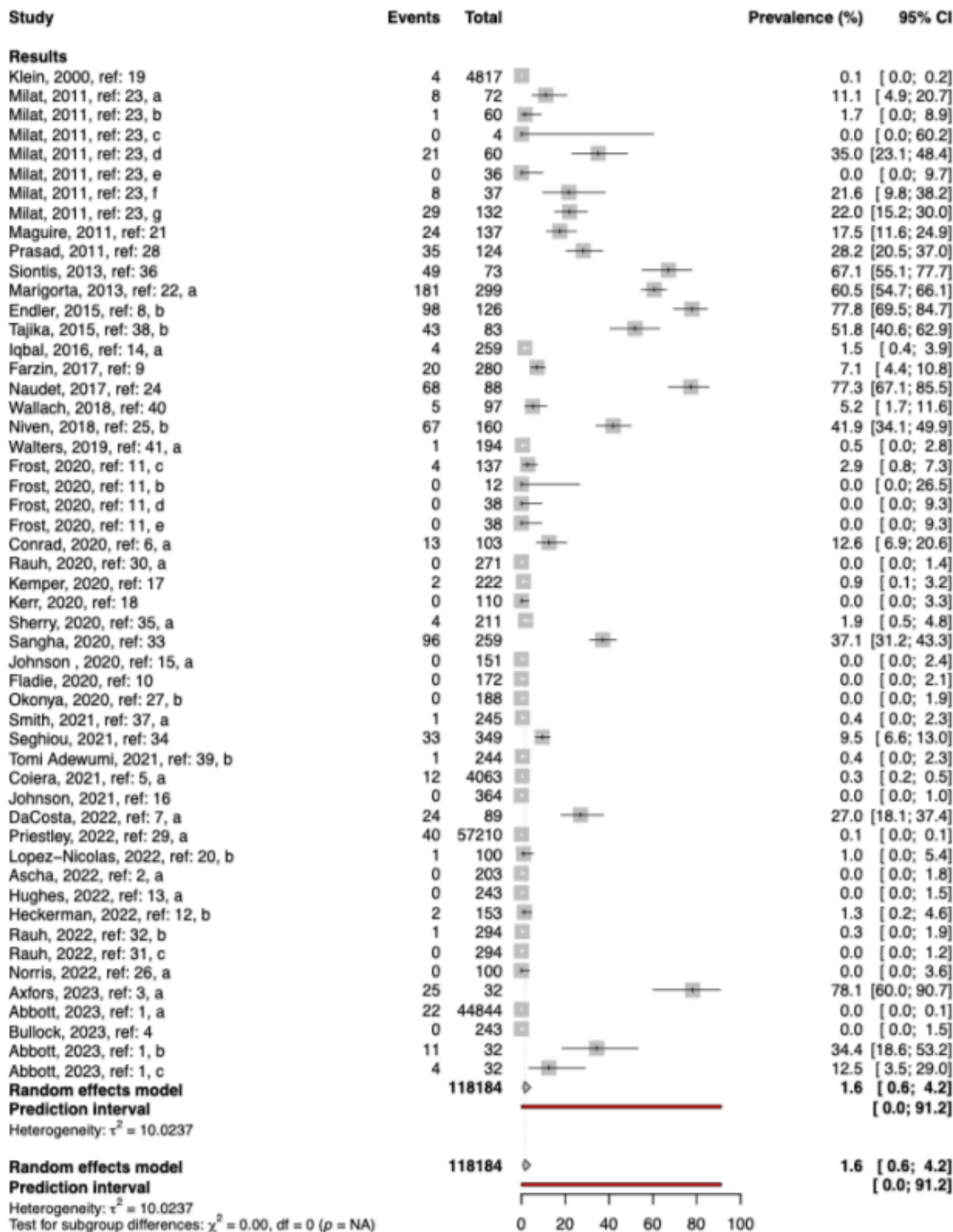


Figure 2

Pooled estimated prevalence of reproducibility attempts

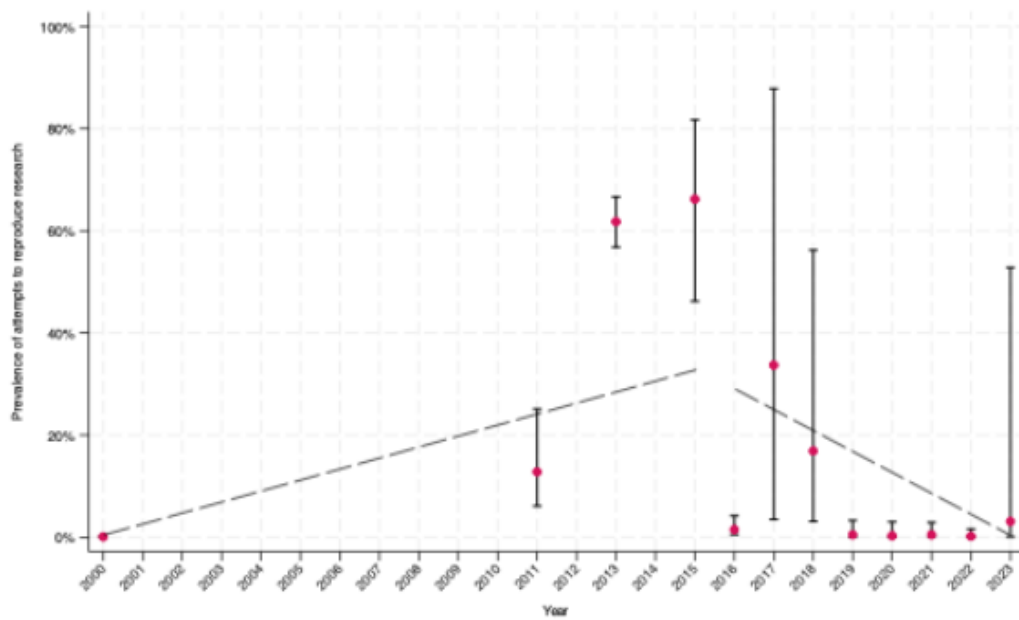


Figure 3

Trend in annual prevalence of attempts to reproduce research. Red dots represent point estimates.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryTables.xlsx](#)
- [SupplementaryFigures.pdf](#)
- [SupplementaryMaterial.docx](#)