Prevalence and predictors of data and code sharing in the medical and health sciences: systematic review with meta-analysis of individual participant data

Daniel G Hamilton,[1,2] Kyungwan Hong,[3] Hannah Fraser,[1] Anisa Rowhani-Farid,[3] Fiona Fidler,[1,4] Matthew J Page[5]

[1]MetaMelb Research Group, School of BioSciences, University of Melbourne, Melbourne, VIC, Australia

[2]Melbourne Medical School, Faculty of Medicine, Dentistry, and Health Sciences, University of Melbourne, Melbourne, VIC, Australia

[3]Department of Practice, Sciences, and Health Outcomes Research, University of Maryland School of Pharmacy, Baltimore, MD, USA

[4]School of Historical and Philosophical Studies, University of Melbourne, Melbourne, VIC, Australia

[5]Methods in Evidence Synthesis Unit, School of Public Health and Preventive Medicine, Monash University, Melbourne, VIC, Australia

Correspondence to: D G Hamilton
The University of Melbourne, Melbourne, Victoria 3010, Australia
hamilton.d@unimelb.edu.au (or @danielghamilton on Twitter; ORCID 0000-0001-8104-474X)

Additional material is published online only. To view please visit the journal online.

Cite this as: *BMJ* 2023;382:e075767
http://dx.doi.org/10.1136/bmj-2023-075767

Accepted: 07 June 2023

ABSTRACT

OBJECTIVES

To synthesise research investigating data and code sharing in medicine and health to establish an accurate representation of the prevalence of sharing, how this frequency has changed over time, and what factors influence availability.

DESIGN

Systematic review with meta-analysis of individual participant data.

DATA SOURCES

Ovid Medline, Ovid Embase, and the preprint servers medRxiv, bioRxiv, and MetaArXiv were searched from inception to 1 July 2021. Forward citation searches were also performed on 30 August 2022.

REVIEW METHODS

Meta-research studies that investigated data or code sharing across a sample of scientific articles presenting original medical and health research were identified. Two authors screened records, assessed the risk of bias, and extracted summary data from study reports when individual participant data could not be retrieved. Key outcomes of interest were the prevalence of statements that declared that data or code were publicly or privately available (declared availability) and the success rates of retrieving these products (actual availability). The associations between data and code availability and several factors (eg, journal policy, type of data, trial design, and human participants) were also examined. A two stage approach to meta-analysis of individual participant data was performed, with proportions and risk ratios pooled with the Hartung-Knapp-Sidik-Jonkman method for random effects meta-analysis.

RESULTS

The review included 105 meta-research studies examining 2 121 580 articles across 31 specialties. Eligible studies examined a median of 195 primary articles (interquartile range 113-475), with a median publication year of 2015 (interquartile range 2012-2018). Only eight studies (8%) were classified as having a low risk of bias. Meta-analyses showed a prevalence of declared and actual public data availability of 8% (95% confidence interval 5% to 11%) and 2% (1% to 3%), respectively, between 2016 and 2021. For public code sharing, both the prevalence of declared and actual availability were estimated to be <0.5% since 2016. Meta-regressions indicated that only declared public data sharing prevalence estimates have increased over time. Compliance with mandatory data sharing policies ranged from 0% to 100% across journals and varied by type of data. In contrast, success in privately obtaining data and code from authors historically ranged between 0% and 37% and 0% and 23%, respectively.

CONCLUSIONS

The review found that public code sharing was persistently low across medical research. Declarations of data sharing were also low, increasing over time, but did not always correspond to actual sharing of data. The effectiveness of mandatory data sharing policies varied substantially by journal and type of data, a finding that might be informative for policy makers when designing policies and allocating resources to audit compliance.

SYSTEMATIC REVIEW REGISTRATION

Open Science Framework doi:10.17605/OSF.IO/7SX8U.

WHAT IS ALREADY KNOWN ON THIS TOPIC

In recognition of the benefits of data sharing, key research stakeholders have been increasing the pressure on medical researchers to maximise the availability of their data and code

Many meta-research studies have examined the prevalence of data and code sharing in medicine and health, but most have been narrow in scope and modest in size

WHAT THIS STUDY ADDS

The findings showed that the prevalence of data and code sharing was low in medical research

Statements declaring that data are publicly available have increased over time, but declared availability did not guarantee actual availability

Compliance with mandatory data sharing policies varied among journals, as well as according to the type of data generated

Introduction

Data collection, analysis, and curation have integral roles in the research lifecycle of most scholarly fields, including medicine and health. That research products, like raw data and analytic code, are valuable commodities to the broader medical research community is also well recognised. Greater access to raw data, analytic code, and other materials that underpin published research provides researchers with opportunities to strengthen their methods, validate discovered findings, answer questions not originally considered by the data creators, accelerate research through the synthesis of existing datasets, and educate new generations of medical researchers.[1] Although many challenges with sharing research materials remain (particularly navigating privacy considerations, and time and resource burdens), in

recognition of the benefits of this practice, funders and publishers of medical research have been increasing the pressure on medical researchers over the past two decades to maximise the availability of these products for other researchers.[2-6] Recent examples include the US government advising its federal funding agencies to update their public access policies before the end of 2025, requiring publications and supporting data to be freely and immediately available.[7]

Although policy changes have increased optimism that data and code sharing rates in medicine will rise, questions remain about the current culture of sharing, how it has evolved over time, how successful stakeholder policies are at instigating sharing, and when researchers do share, how often are useful data made available. Many meta-research studies in medicine have looked at these questions, but most have been limited in size and scope, focusing on specific research participants, data types, and outcomes. Therefore, the objectives of this review were to synthesise this research to establish an accurate representation of the prevalence of data and code sharing in medical and health research, assess compliance with stakeholder policies on data and code availability, and explore how other relevant factors influence availability. We anticipate that the findings of this review will highlight several areas for future policy making and meta-research activities.

Methods

We registered our systematic review on 28 May 2021 on the Open Science Framework,[8] and subsequently prepared a detailed review protocol.[9] Supplementary table 1 shows seven deviations from the protocol. The findings of this review are reported in accordance with the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) 2020 statement[10] and its individual participant data extension.[11] We summarise key aspects of the methods below. The supplementary information and review protocol provide further information.[9]

Eligibility criteria

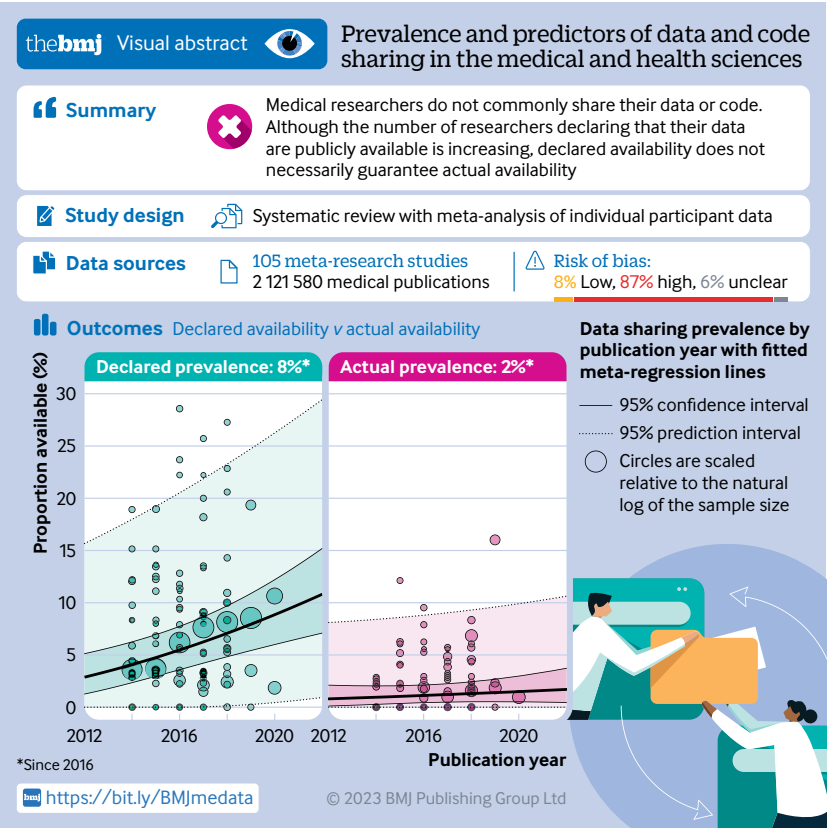
Studies where researchers investigated the prevalence of, or factors associated with, data or code sharing (termed meta-research studies) across a sample of published scientific articles presenting original medical or health related research findings (termed primary articles) were eligible for inclusion in the review. We included studies that used manual or automated methods to assess data and code sharing if they involved examination of the body text of the sampled primary articles. Exclusion criteria for this review included meta-research studies that investigated data or code sharing: as a routine part of a systematic review and meta-analysis of individual participant data; in scientific articles outside of medicine and health; or from sources other than journal articles (eg, clinical trial registries).

Search strategy and study selection

On 1 July 2021, we searched Ovid Medline, Ovid Embase, and the preprint servers medRxiv, bioRxiv, and MetaArXiv to identify potentially relevant studies indexed from database inception up to the search date. On the same date, other preprint servers and relevant online resources (see supplementary methods in supplementary information) were searched to locate other published, unpublished, and registered studies of relevance to the review. Backward and forward citation searches of meta-research studies meeting the inclusion criteria were also performed with citationchaser on 30 August 2022.[12] No language restrictions were imposed on any of the searches. Results from all of the searches of the main databases and preprint servers were imported into Covidence (Covidence systematic review software, Veritas Health Innovation, Melbourne, Australia) and deduplicated. All titles, abstracts, and full text articles were then screened for eligibility by DGH and another author (HF, AR-F, or KH) independently, with disagreements resolved by discussion between authors, or by a third author if necessary (MJP). All literature identified by the additional preprint, online, and citation searches were screened against the eligibility criteria by one of the authors (DGH).

Data collection and processing

When a meta-research study was found to be eligible, one of the authors (DGH) determined whether sufficiently unprocessed article level individual participant data and article identifiers were publicly available. For meta-research studies where complete individual participant data were not available, the corresponding author was contacted and asked if they



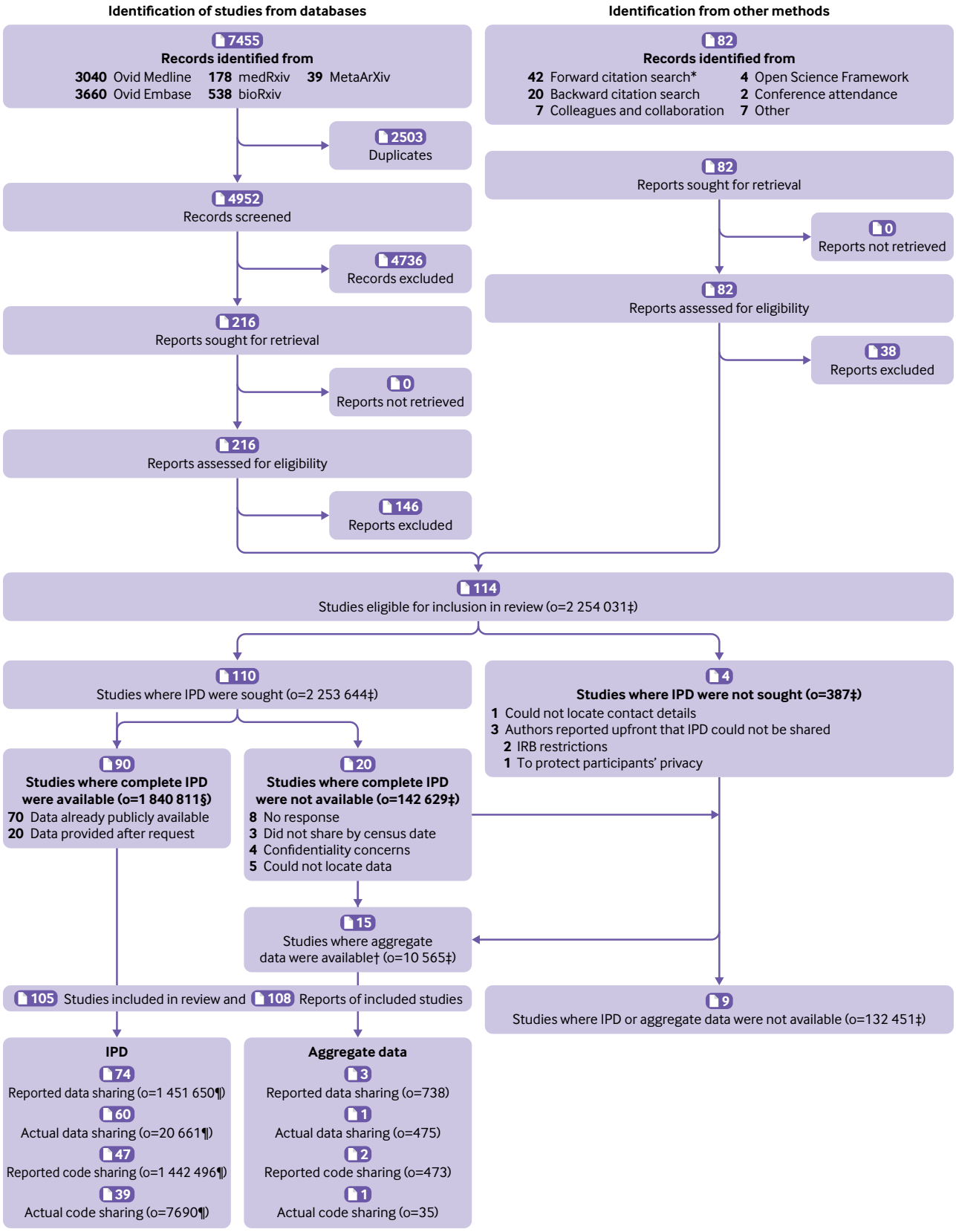


Fig 1 | Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) 2020 and PRISMA-individual participant data (IPD) flow diagram. *Forward citation search was performed on 30 August 2022. †Aggregate data were derived from partial IPD, reports, or authors. ‡Number of observations does not account for the potential presence of duplicate or non-medical primary articles. §Number of observations accounts for non-medical primary articles and duplicate primary articles within, but not between, meta-research studies. ¶Number of observations accounts for non-medical primary articles and both duplicate primary articles within and between meta-research studies. IRB=institutional review board; o=number of primary articles.

Table 1 | Characteristics of included studies (n=105)

Primary articles examined	No (%)
Median (IQR) No of primary articles examined	195 (113-475)
Median (IQR) primary article publication year	2015 (2012-2018)
Medical specialty (top 7):	
Multidisciplinary	17 (16)
Biomedicine	10 (10)
Infectious diseases	10 (10)
General medicine	9 (9)
Addiction medicine	5 (5)
Clinical psychology	5 (5)
Oncology	5 (5)
Outcome of interest:	
Data sharing only	46 (44)
Code sharing only	5 (5)
Data and code sharing	54 (51)
Coding method:	
Manual	82 (78)
Automated	8 (8)
Both manual and automated	3 (3)
Unclear	7 (7)
Data restrictions:*	
No restrictions	63 (59)
Trial data	16 (15)
Sequence data	6 (6)
Systematic review data	5 (5)
Gene expression data	5 (5)
Other	6 (6)
Not applicable	5 (5)
Journal restrictions:	
No restrictions	56 (53)
High impact	17 (16)
One journal	10 (10)
Hand selected	7 (7)
Preprint servers	5 (5)
Other	10 (10)

Data are count (%) unless specified otherwise. Underlying data are at https://osf.io/ca89e. IQR=interquartile range.
*Data do not sum to 105 because one study assessed two types of data.

would provide the complete or remaining individual participant data. If individual participant data were not provided by 31 December 2022, when available, summary data were independently extracted from study reports by two of the authors (DGH and MJP), with discrepancies resolved by discussion.

When complete individual participant data were assembled, one of the authors (DGH) performed data integrity checks: evaluation of the completeness of the dataset; check of the validity of the dataset; and check that the sharing prevalence for data or code, or both, as stated in the report, could be exactly reproduced. In instances where any of these checks failed, clarification was sought from the meta-research authors. We also checked for, and removed, non-medical articles within individual participant data, as well as redundant assessments across assembled individual participant data (supplementary methods in supplementary information). When the individual participant data checks were complete, one of the authors (DGH) manually extracted and reclassified the required data in line with the review's codebook.

Outcomes of interest

Four prespecified outcome measures were of primary interest to the review: prevalence of primary articles where authors declared that their data or code were publicly available (declared public availability); prevalence of primary articles where meta-researchers verified that data or code were in fact publicly available (actual public availability); prevalence of primary articles where authors declared that their data or code were privately available (declared private availability); and prevalence of primary articles where meta-researchers verified that study data or code were released in response to a private request (actual private availability).

Actual public availability represented the results of the most intensive investigation of an availability statement by meta-researchers. We required data to be immediately available to be classified as actually publicly available. The review protocol and supplementary information provide further details on how we defined actual availability as well as all of our other outcome measures.[9 13] We also included eight secondary outcome measures on the association between journals' policies on sharing, study design, research participants, and data sharing, as well as the association between data and code sharing.

Assessments of risk of bias

The risk of bias in the included meta-research studies was assessed with a tool that was designed based on methods used in previous Cochrane methodology reviews.[14 15] The tool included four domains: sampling bias, selective reporting bias, article selection bias, and risk of errors in the accuracy of reported estimates (supplementary table 2). Each meta-research article was independently assessed by DGH and one other author (KH or AR-F), with discrepancies resolved by discussion, or by a third author (MJP), if necessary. Because the purpose of the tool was to differentiate between studies at a high risk of bias from those with a low risk of bias, a study was only classified as low risk of bias if all of the criteria were assessed as low risk. We did not assess the likelihood of publication bias affecting the findings of the review (eg, with a funnel plot) or the certainty in the body of evidence, because the available methods are not well suited for methodology reviews such as ours.

Statistical analysis

We used a two stage approach to the meta-analysis of individual participant data, where summary statistics were computed from available individual participant data, abstracted from the included study reports, or obtained directly from meta-research authors, and then pooled with conventional meta-analysis techniques. We calculated proportions and 95% confidence intervals for all prevalence outcomes. Where possible, we calculated risk ratios with 95% confidence intervals for all association outcomes. For primary outcome measures, to ensure that the studies were sufficiently clinically and methodologically similar, we only pooled studies that did not use non-random sampling methods, did not restrict primary articles by publication location, funder, institution,

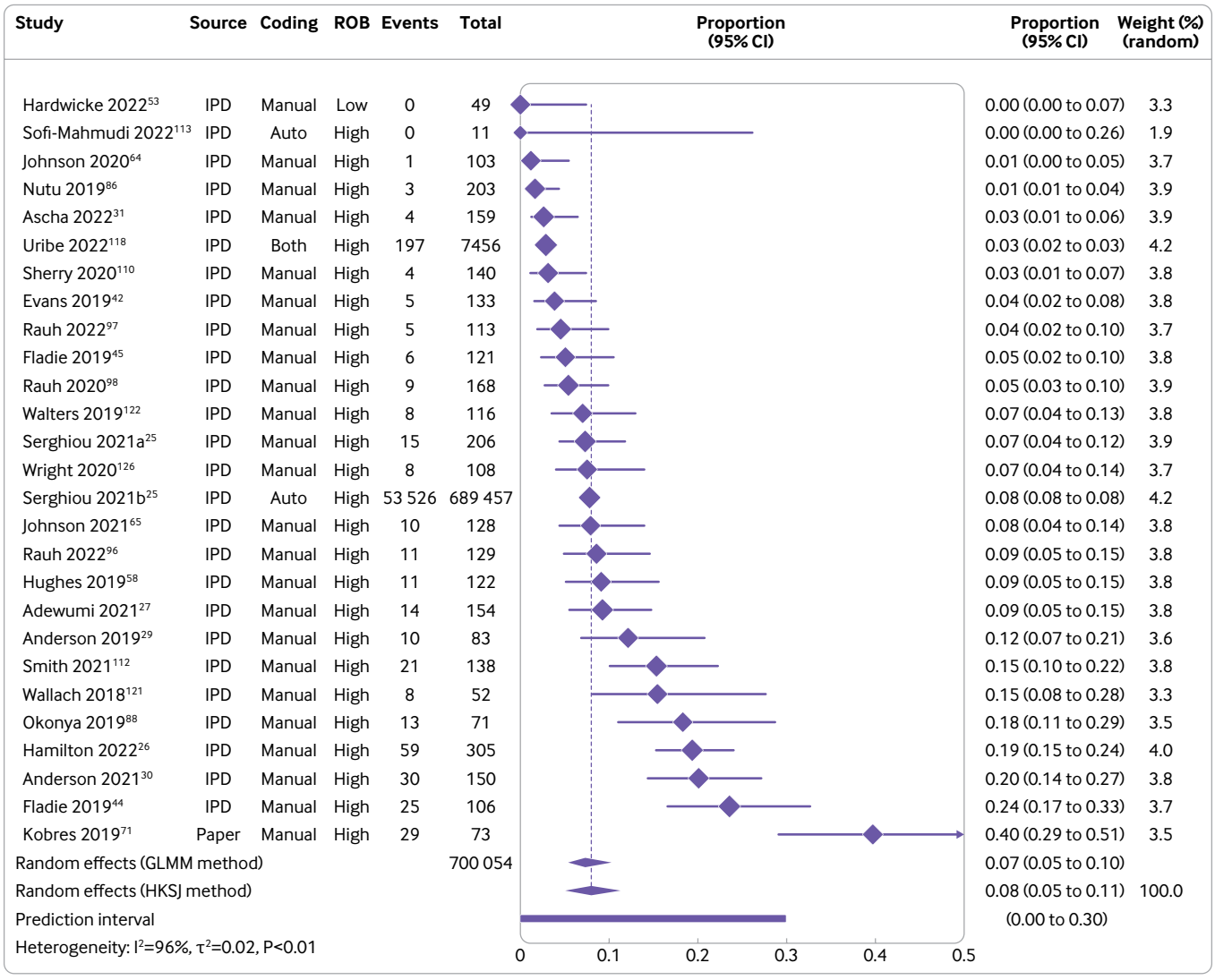


Fig 2 | Prevalence of declared public data sharing between 2016 and 2021. ROB=risk of bias; GLMM=generalised linear mixed model; HKSJ=Hartung-Knapp-Sidik-Jonkman; IPD=individual participant data. Serghiou 2021a and 2021b refer to the manual and automated assessments, respectively, reported in Serghiou et al[25]

or data type, and reported outcome data on primary articles published after 2016.

We pooled prevalence estimates by first stabilising the variances of the raw proportions with arcsine square root transformations, and then applied random effects models with the Hartung-Knapp-Sidik-Jonkman method.[16] The same approach was also used for meta-analyses of risk ratios; however, no transformations were used, and the treatment arm continuity correction proposed by Sweeting and colleagues[17] was applied to studies reporting zero events in one group (double zero cell events were excluded from the main analysis). Statistical heterogeneity was assessed by visual inspection of forest plots, the size of the I^2 statistics, and their 95% confidence intervals, and by 95% prediction intervals where more than four studies were included. Data cleaning, deduplication, analysis, and visualisation were performed in R (R Foundation for Statistical Computing, Vienna, Austria, version 4.2.1).

Subgroup and sensitivity analyses

Our aim was to conduct subgroup analyses to investigate whether estimates of the prevalence of public data sharing differed depending on the data type, or whether primary articles were subject to any mandatory sharing policies by the funders of the research or had posted a preprint before publication. Furthermore, we investigated the influence of publication year on the prevalence of data and code sharing by fitting three level mixed effects meta-regression models on arcsine transformed proportions. We also performed sensitivity analyses to assess changes in pooled estimates when excluding meta-research studies that: were rated as high or had an unclear risk of bias; did not provide individual participant data for the review; were at high risk of overlap with other meta-research studies; did not assess whether publicly available data were also compliant with the FAIR (findability, accessibility, interoperability, and reusability) principles[18]; did not manually assess primary articles; and did not examine research related

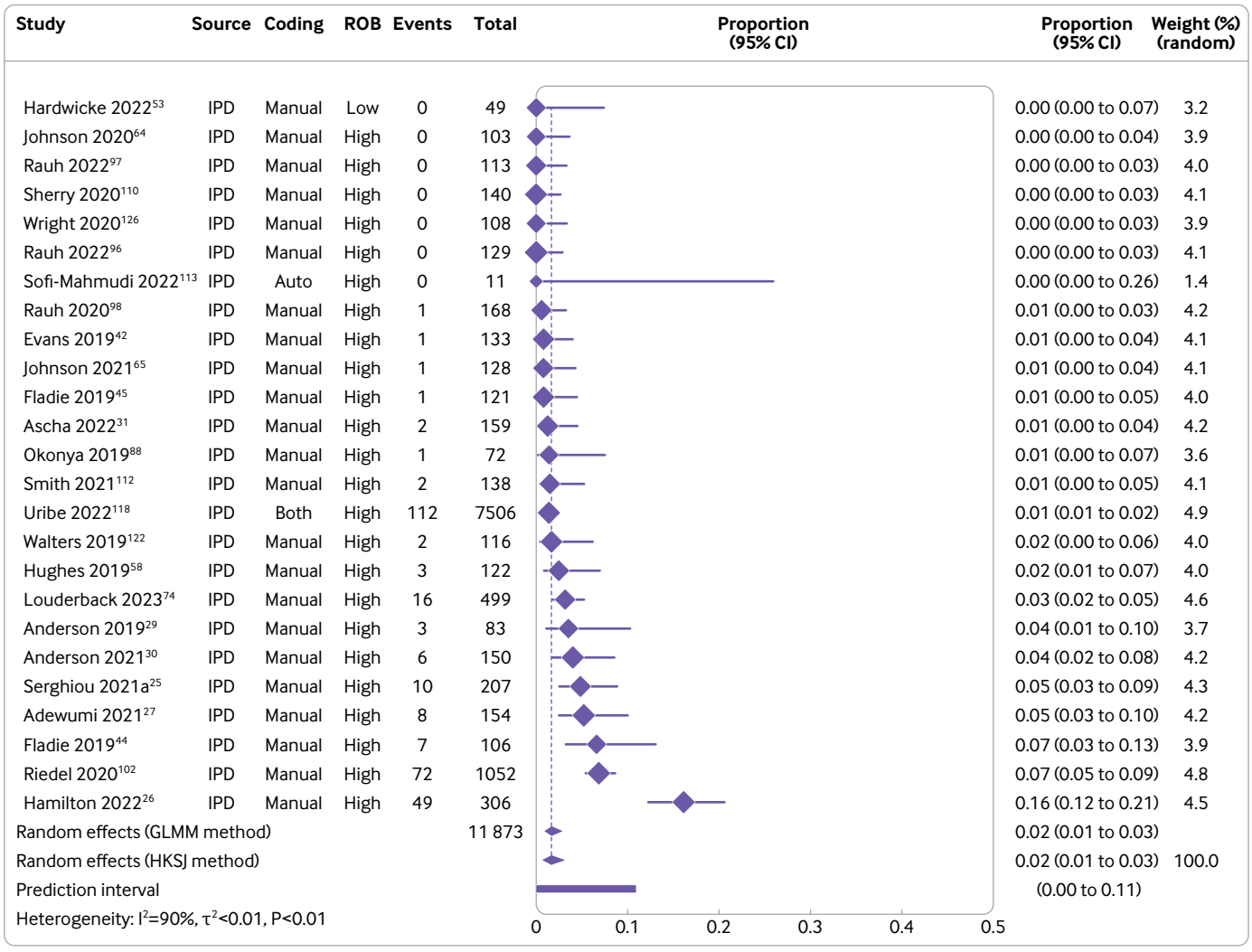


Fig 3 | Prevalence of actual public data sharing between 2016 and 2021. ROB=risk of bias; GLMM=generalised linear mixed model; HKSJ=Hartung-Knapp-Sidik-Jonkman; IPD=individual participant data. Serghiou 2021a refers to the manual assessments reported in Serghiou et al[25]

to covid-19. Also, we examined differences in pooled proportions and risk ratios when generalised linear mixed models were used to aggregate findings.

Patient and public involvement

No patients or members of the public were involved in the conception, development, analysis, or interpretation of the results of the review. A former cancer patient, a cancer patient advocate, and a practising clinician, however, reviewed the manuscript to ensure maximum comprehension to both a lay and general medical audience.

Results

Study selection and individual participant data retrieval

The search of Ovid Medline, Ovid Embase, and the medRxiv, bioRxiv, and MetaArXiv preprint servers, once deduplicated, identified 4952 potentially eligible articles for the review; 4736 articles were excluded after the titles and abstracts were screened. Of the remaining 216 articles, full text articles were retrieved for all papers, and 70 were found to be eligible for the review. More literature searches identified another 44 eligible reports for inclusion, giving a total of 114 eligible meta-research studies examining a combined total of 2 254 031 primary articles for the review.[19-135] After confirmation of eligibility, we searched for publicly available individual participant data for the 114 meta-research studies. Of these studies, complete individual participant data were publicly available in 70 (61%), 20 had published partial individual participant data (18%), and 24 had not shared any individual participant data publicly (21%). Of the 44 studies that did not share complete individual participant data publicly, we retrieved the required individual participant data for 20 studies privately.

Supplementary results in the supplementary information provides further information on the individual participant data retrieval process, as well as the outcomes of the data integrity checks. In total, 108 reports of 105 meta-research studies assessing a total of 2 121 580 primary articles were included in the quantitative analysis,[19-126] with complete

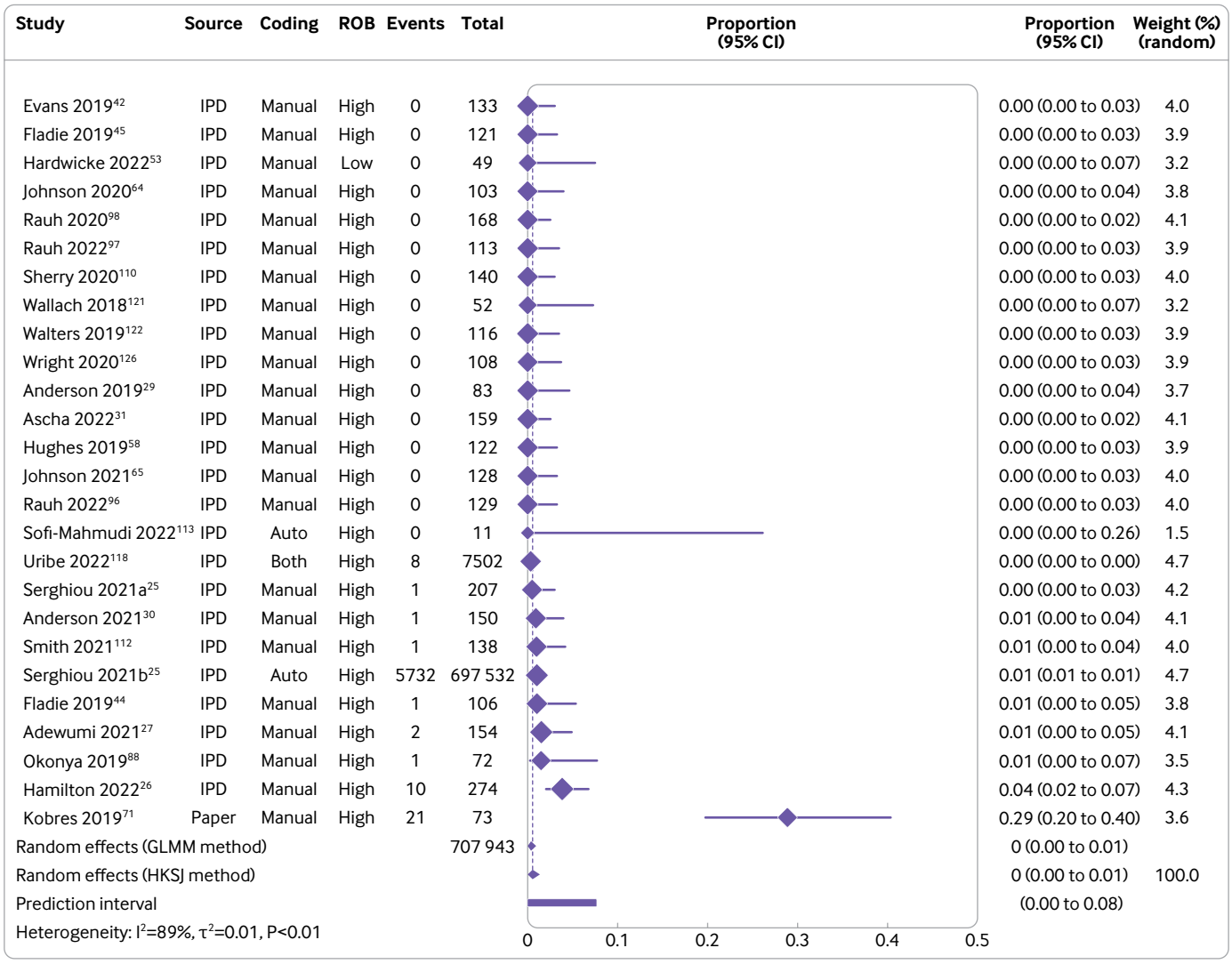


Fig 4 | Prevalence of declared public code sharing between 2016 and 2021. ROB=risk of bias; GLMM=generalised linear mixed model; HKSJ=Hartung-Knapp-Sidik-Jonkman; IPD=individual participant data. Serghiou 2021a and 2021b refer to the manual and automated assessments, respectively, reported in Serghiou et al[25]

individual participant data available for 90 studies, a combination of partial individual participant data and summary data for 10 studies, and summary data only for five studies. Figure 1 shows the full PRISMA flow diagram and supplementary table 3 lists details of the nine studies that were eligible for the review but could not be included in the quantitative analysis.

Study characteristics

Table 1 outlines the summary information for the 105 meta-research studies included in the quantitative analysis of this review. Eligible meta-research studies examined a median of 195 primary articles (interquartile range 113-475; sample size range 10-1 475 401), with a median publication year of 2015 (interquartile range 2012-2018; publication date range 1781-2022). Meta-research studies assessed data and code sharing across 31 specialties. Most commonly, studies were interdisciplinary, examining several medical fields simultaneously (n=17, 16%), followed by biomedicine and infectious disease (each n=10, 10%), general medicine (n=9, 9%), and addiction medicine, clinical psychology, and oncology (each n=5, 5%). Eleven studies (10%) examined articles related to covid-19 disease.

Most meta-research studies did not set any restrictions for type of data (n=63, 59%) or journals of interest (n=56, 53%). When data restrictions were imposed, however, they were most often limited to trial data (n=16, 15%), sequence data (n=6, 6%), and gene expression data and review data (each n=5, 5%). Of the 105 meta-research studies, 31 and four also evaluated compliance with journal data and code sharing policies, respectively. None of the meta-research studies examined compliance with policies instituted by medical research funders or institutions. In total, 95 and 58 meta-research studies, respectively, examined the prevalence of public data and code sharing in primary articles, with five studies examining compliance of publicly shared data with the FAIR principles. In contrast, 10, four, and two studies, respectively, assessed whether study data, code, or

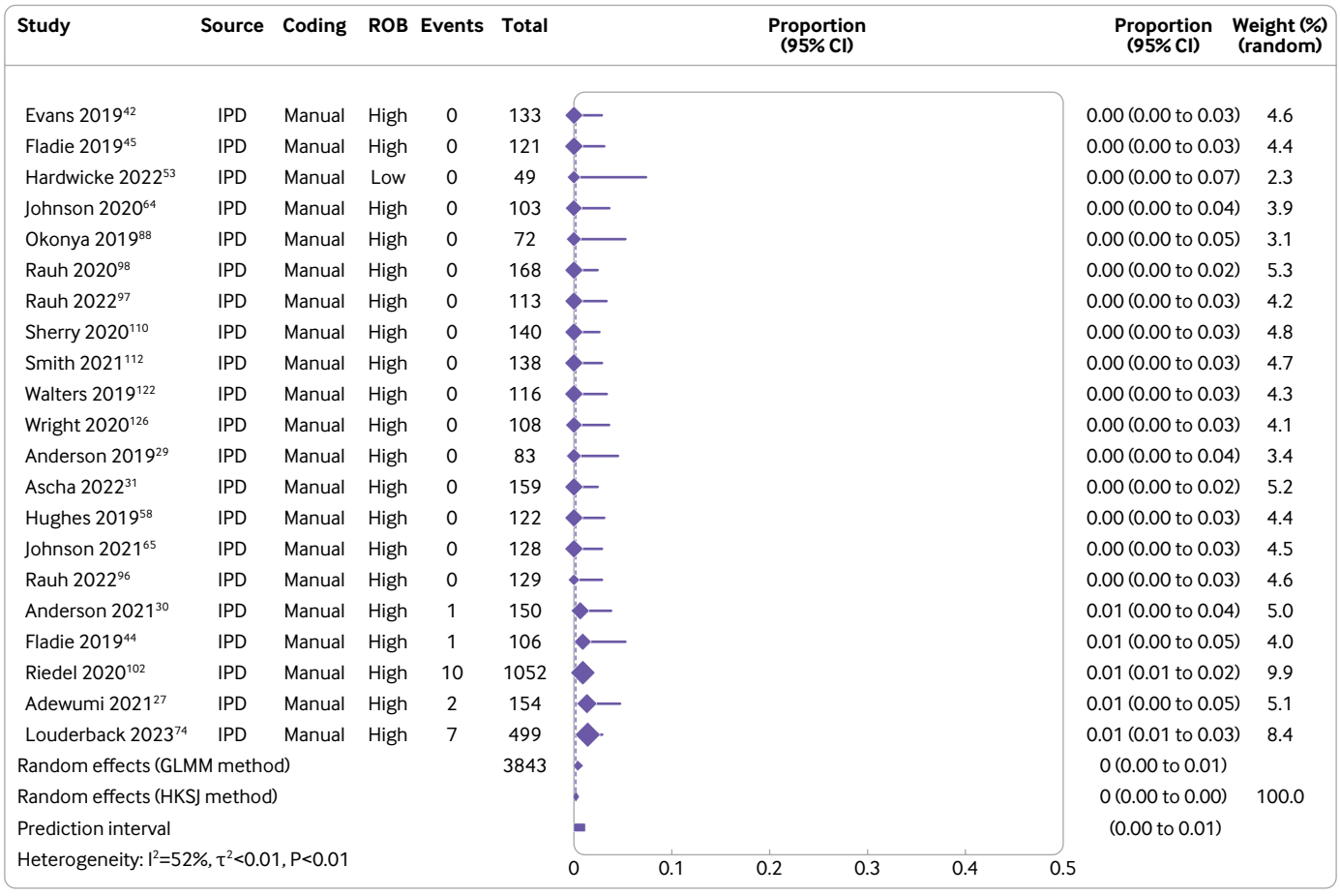


Fig 5 | Prevalence of actual public code sharing between 2016 and 2021. ROB=risk of bias; GLMM=generalised linear mixed model; HKSJ=Hartung-Knapp-Sidik-Jonkman; IPD=individual participant data

both data and code could be retrieved in response to a private request (ie, actual private availability).

Risk of bias assessment

Supplementary figures 1 and 2 show the overall and individual results of the risk of bias assessments. Most eligible meta-research studies were judged favourably on the first risk of bias domain (sampling bias), having randomly sampled primary articles from populations of interest, or assessed all eligible articles identified by their literature searches (n=95, 90%). In contrast, some meta-research studies were judged to be at low risk of selective reporting bias (n=45, 42%) and article selection bias (n=24, 23%). Similarly, only half of the meta-research studies (n=54, 51%) were judged to have used a primary article coding strategy considered to be at low risk of errors. Only eight studies (8%) were classified as low risk of bias for all four domains.

Public data and code sharing

The combination of studies considered sufficiently similar (both clinically and methodologically) in a random effects meta-analysis suggested that 8% of medical articles published between 2016 and 2021 reported that the data were publicly available (95% confidence interval 5% to 11%, k=27 studies, I^2=96%) and 2% actually shared the data (1% to 3%, k=25, I^2=90%) (fig 2 and fig 3). For public code sharing, the prevalence of declared and actual code sharing between 2016 and 2021 were estimated to be 0.3% (0% to 1%, k=26, I^2=89%) and 0.1% (0% to 0.3%, k=21, I^2=52%), respectively (fig 4 and fig 5). Despite the included meta-research studies following similar methodologies, we found high I^2 values for all analyses. Because of the consistency of the point estimates, and the narrow width of the prediction intervals, however, we do not believe this finding indicates concerning levels of variability.

Private data and code sharing

In contrast with declarations of public availability, available on request declarations were less common for data (2%, 95% confidence interval 1% to 4%, k=23, I^2=80%) and code (0%, 0% to 0.1%, k=22, I^2=0%) between 2016 and 2021 (supplementary figs 4 and 5). For actual private data and code availability, we could not combine the findings of relevant meta-research studies because of methodological differences, particularly journal restrictions (ie, policy differences), as well as the type of data being requested. Overall, however, we found that success in privately obtaining data and code from authors of published medical research on request historically ranged between 0% and 37% (k=12) and 0% and 23% (k=5), respectively (fig 6).

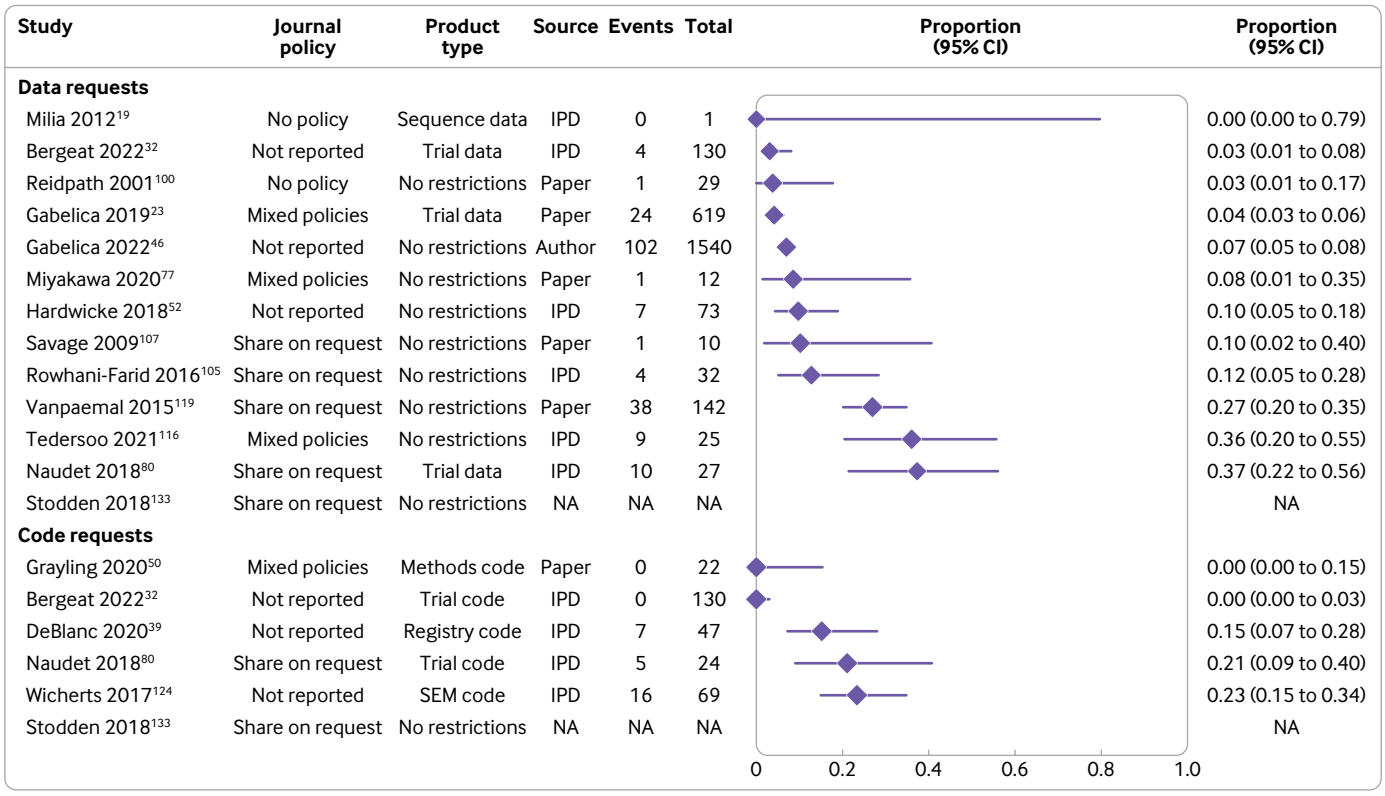


Fig 6 | Prevalence of successful responses to private requests for data and code from published medical research. SEM=structural equation modelling; NA=summary data not available; IPD=individual participant data

We also found that when authors who declared that their data and code were available on request were asked for these products by meta-researchers, success in obtaining data and code improved to 0-100% (k=7) and 0-43% (k=4), respectively. In contrast, when requests for data and code were made to authors who did not include a statement concerning availability, success in obtaining data and code decreased to 0-30% (k=7) and 0-12% (k=3), respectively. Lastly, we found that attempts to obtain data from authors who explicitly declared that their data were unavailable were associated with a 0% sharing rate (k=2). Supplementary figure 6 shows the full results.

Secondary outcomes

Insufficient data were available to evaluate five of our eight secondary outcome measures (supplementary results in supplementary information has more information). We also could not directly compare outcomes for mandatory versus non-mandatory journal sharing policies. For articles subject to mandatory data sharing policies of journals, however, we estimated that 65% of primary articles (95% confidence interval 36% to 88%, k=5, I^2=99%) declare that the data are publicly available and 33% actually share the data (5% to 69%, k=3, I^2=93%). One stage analysis of available individual participant data also showed that prevalence of actual data sharing among journals with mandatory data sharing policies ranged between 0% and 100% (median 40%, interquartile range 20-60%). In contrast, we estimated that private requests for data from authors whose journals have "must share on request" policies will be successful 21% of the time (95% confidence interval 4% to 47%, k=3, I^2=30%). For comparison, declared and actual data sharing prevalence estimates for journals with "encourage sharing" policies were 17% (0% to 62%, k=6, I^2=98%) and 8% (0% to 48%, k=3, I^2=90%), respectively. Similarly, we estimated that the prevalence of declared and actual data sharing for articles published in journals with no data sharing policy are 17% (0% to 59%, k=4, I^2=95%) and 4% (0% to 95%, k=2, I^2=83%), respectively. Supplementary figure 7 shows the prevalence estimates for declared and actual public code sharing according to journal policies.

Furthermore, our data suggested that triallists are 31% less likely to declare that the data are publicly available than non-triallists (risk ratio 0.69, 95% confidence interval 0.45 to 1.07, k=23, I^2=0%). When we examined actual data sharing, however, neither group seemed more or less likely to share their data (0.96, 0.53 to 1.72, k=19, I^2=0%) (fig 7 and fig 8). For data derived from human participants, researchers were estimated to be 35% less likely to declare that their data are publicly available than researchers working with non-human participants (0.65, 0.42 to 0.99, k=19, I^2=57%). This finding became more pronounced when we examined actual data sharing prevalence estimates (0.44, 0.24 to 0.81, k=16, I^2=28%) (fig 9 and fig 10). Lastly, we estimated that researchers who declare that their data are publicly available are eight times more likely to declare code to

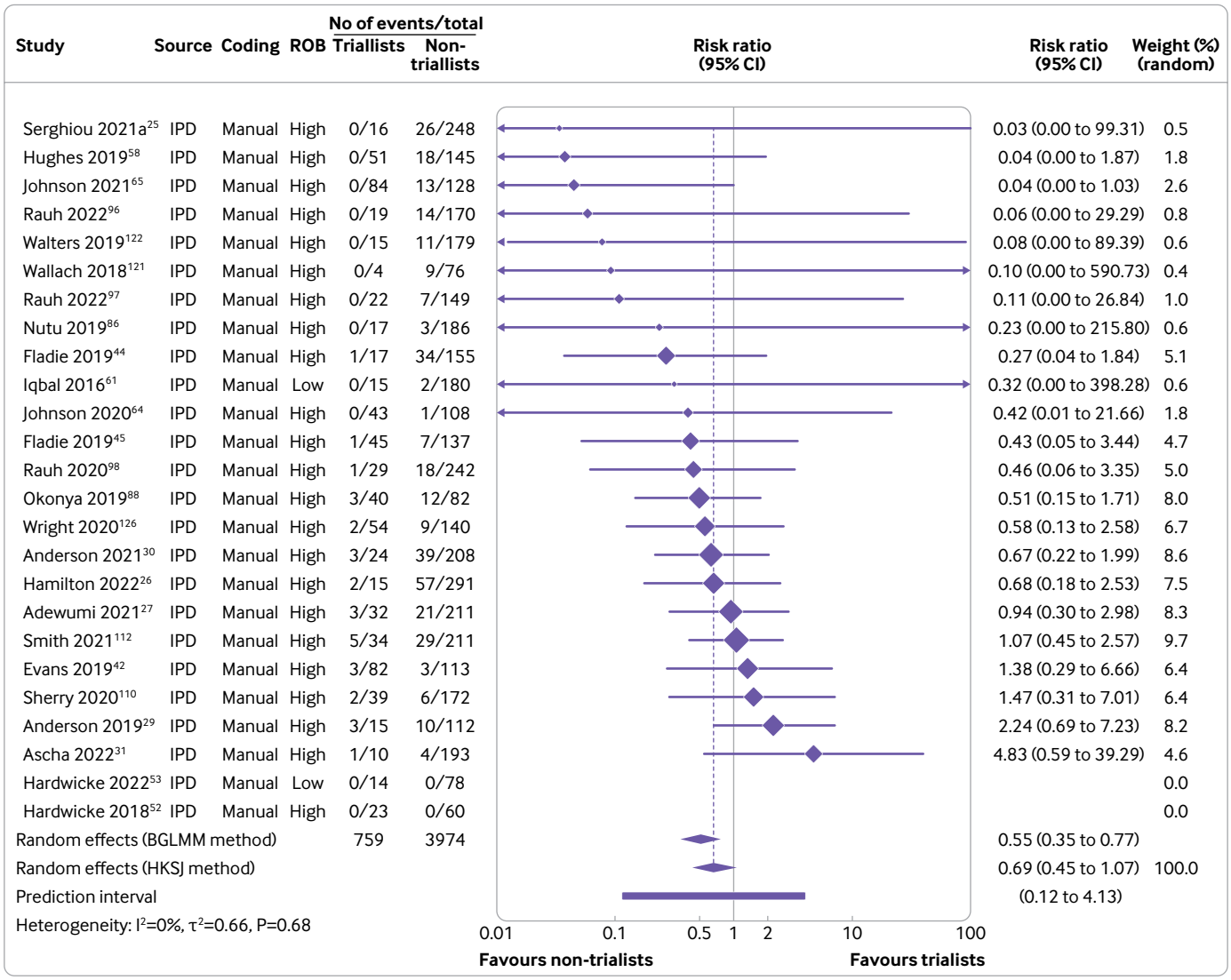


Fig 7 | Association between trial design and prevalence of declared public data sharing. BGLMM=bivariate generalised linear mixed model; HKSJ=Hartung-Knapp-Sidik-Jonkman; IPD=individual participant data. Serghiou 2021a refers to the manual assessments reported in Serghiou et al[25]

be available also (8.03, 2.86 to 22.53, k=12, I^2=32%). Furthermore, researchers who were verified to have made data available were estimated to be 42 times more likely than researchers who withheld data to also share code (42.05, 12.15 to 145.52, k=7, I^2=0%) (supplementary fig 8).

Subgroup and sensitivity analyses

Several subgroup analyses were performed (supplementary results in the supplementary information has the full results). We found that rates for both declared and actual public data sharing significantly differed according to type of data (P<0.01), with the highest estimate of actual data sharing occurring among authors working with sequence data (57%, 95% confidence interval 12% to 96%, k=3, I^2=86%), systematic review data (6%, 0% to 77%, k=2, I^2=75%), and then trial data (1%, 0% to 6%, k=3, I^2=6%) (supplementary figs 9 and 10). We also found substantial differences in compliance rates with journal policies depending on the type of data (table 2). Publication year was also found to be a significant moderator of the prevalence of declared data sharing (β=0.017, 95% confidence interval 0.008 to 0.025, P<0.001) but not actual data sharing (β=0.004, −0.005 to 0.013, P=0.36) (fig 11 and supplementary table 4). Specifically, we found an estimated rise in the prevalence of declared data sharing from 4% in 2014 (95% confidence interval 2% to 6%; 95% prediction interval 0% to 18%) to 9% in 2020 (6% to 12%; 0% to 26%). Both declared and actual code sharing prevalence estimates did not seem to have meaningfully increased over time.

Supplementary tables 5 and 6 show the full results of the sensitivity analyses for the primary and secondary outcomes, respectively. We found that meta-analyses of prevalence estimates were similar when generalised linear mixed models or standard inverse variance aggregation methods were used. Also, limiting the analyses to meta-research studies where authors

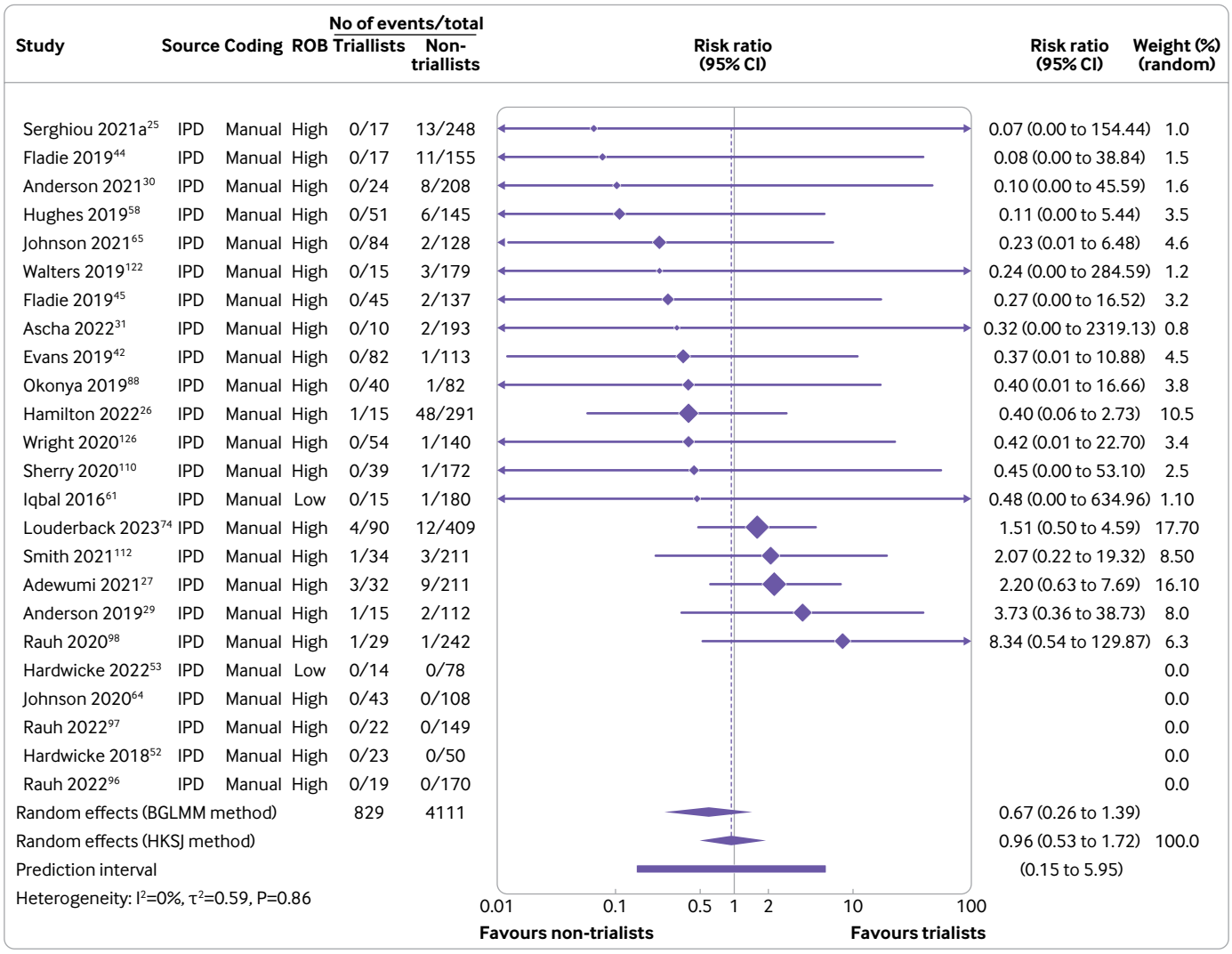


Fig 8 | Association between trial design and prevalence of actual public data sharing. BGLMM=bivariate generalised linear mixed model; HKSJ=Hartung-Knapp-Sidik-Jonkman; IPD=individual participant data. Serghiou 2021a refers to the manual assessments reported in Serghiou et al[25]

manually coded articles did not meaningfully change the results. Lastly, we estimated that the prevalence of declared and actual public data sharing for studies investigating covid-19 (including preprints and peer reviewed publications) were 9% (95% confidence interval 0% to 57%, k=3, o=7804, I^2=95%) and 11% (0% to 76%, k=3, o=934, I^2=84%), respectively.

Discussion

Principal findings of the review

In this systematic review and meta-analysis of individual participant data, we used multiple data sources and methods to investigate public and private availability of data and code in the medical and health literature. We also examined several factors associated with sharing. Table 3 shows a summary of the main findings of the review. Aggregation of the findings of sufficiently similar studies suggested that on average, 8% of medical papers published between 2016 and 2021 declared that their data were publicly available and 2% actually shared their data publicly. Pooled prevalence estimates for declared and actual code sharing since 2016 were even lower, with both estimated to be <0.5%, with little change over time. The prediction intervals from our analyses were also relatively precise, suggesting that we now have good estimates for the prevalence of data and code sharing for medical and health research between 2016 and 2021.

In contrast with public data and code availability prevalence estimates, the overall success in privately obtaining data and code from authors of published medical research ranged between 0% and 37% and 0% and 23%, respectively. These findings are consistent with similar research conducted in disciplines outside of medicine.[133 136-140] We found that these ranges differed, however, according to the type of data and code being requested, the policy of the journal, and whether authors declared that the products were available on request. Finally, although data were not available to assess compliance with the data sharing policies of funders and institutions, we found varying

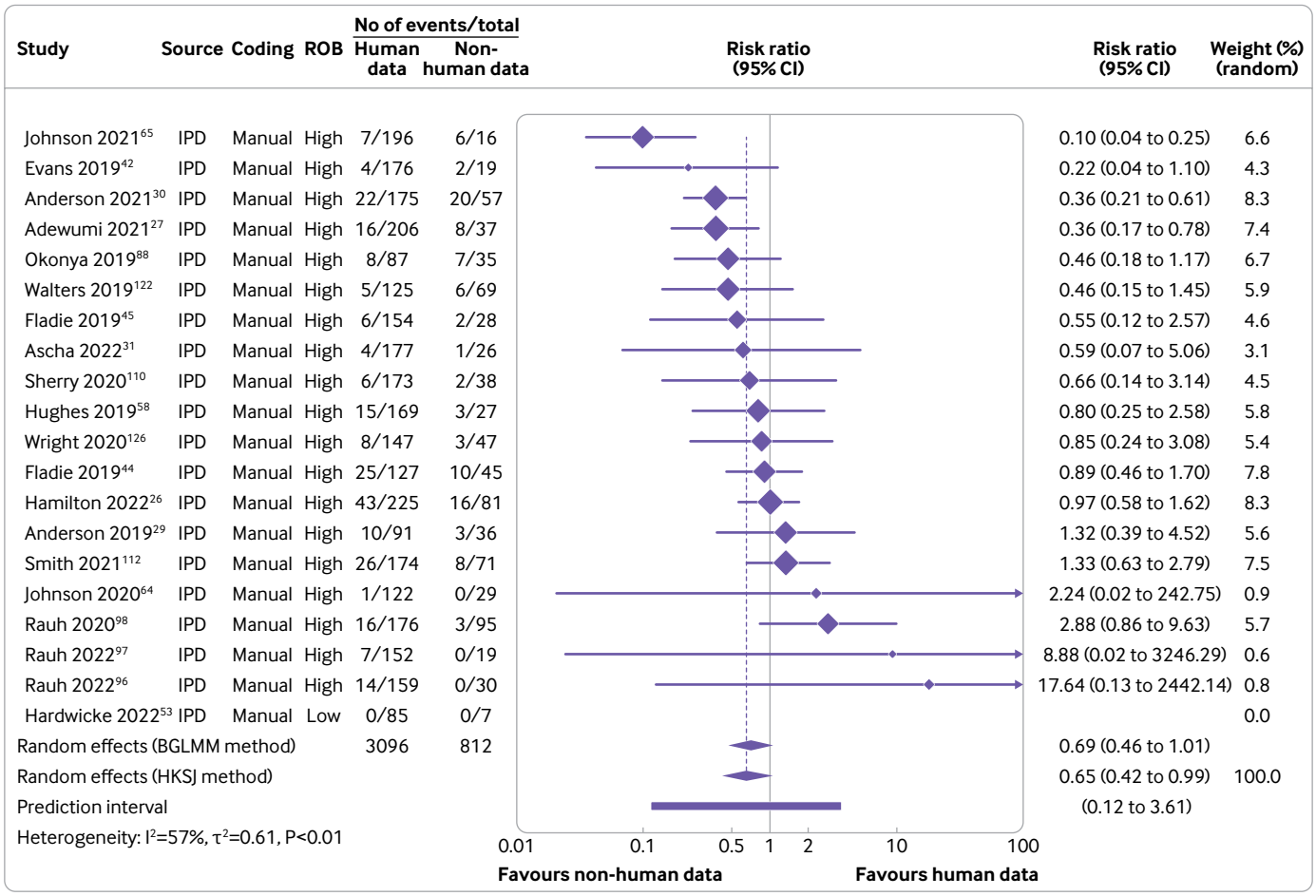


Fig 9 | Association between type of research participant and prevalence of declared public data sharing. ROB=risk of bias; BGLMM=bivariate generalised linear mixed model; HKSJ=Hartung-Knapp-Sidik-Jonkman; IPD=individual participant data

compliance with the data sharing policies of journals, particularly depending on the type of data.

Review findings in context

When we examined similar research conducted in other scientific areas, declared data sharing prevalence estimates in medicine seemed to be higher than in some disciplines (eg, humanities, earth sciences, and engineering[25]), but lower than in others (eg, experimental biology[141] and hydrology[142]). The low prevalence of code sharing in our review was consistent with other disciplines, except for ecology[143] and computer sciences,[25] despite evidence suggesting that most medical researchers use data analysis software capable of exporting syntax or files that preserve analytic decisions.[26]

One explanation for interdisciplinary differences in the prevalence of data sharing is that researchers in areas outside of the medical, health, behavioural, and social sciences are more likely to make data available because typically they do not need to navigate privacy protections associated with the collection and sharing of data from human participants.[144] For example, national and international protection laws, like the US Health Insurance Portability and Accountability Act (HIPAA) and the European Union's 2018 General Data Protection Regulation (GDPR), impose strong restrictions on the processing of personal medical data.[145 146] Our results support this notion. We found that medical researchers studying data from human participants were 56% less likely to actually make their data publicly available than those who had used data derived from non-human participants.

We also see other researchers point to differences in data sharing rates between medical and non-medical researchers working with the same human derived data types, which could indicate possible cultural differences. For example, for mitochondrial and Y chromosomal data, Anagnostou and colleagues[147] found a substantially different prevalence of data sharing between medical (64%) and forensic (90%) genetics researchers. Follow-up work by the same authors suggested that discrepancies between sharing estimates likely reflected differences in how these disciplines value openness and transparency, in contrast with the burdens associated with navigating privacy constraints.[147]

Potential implications of our findings

Our findings raise important implications for researchers and policy makers. For policy makers, our findings suggest that because of substantial variability

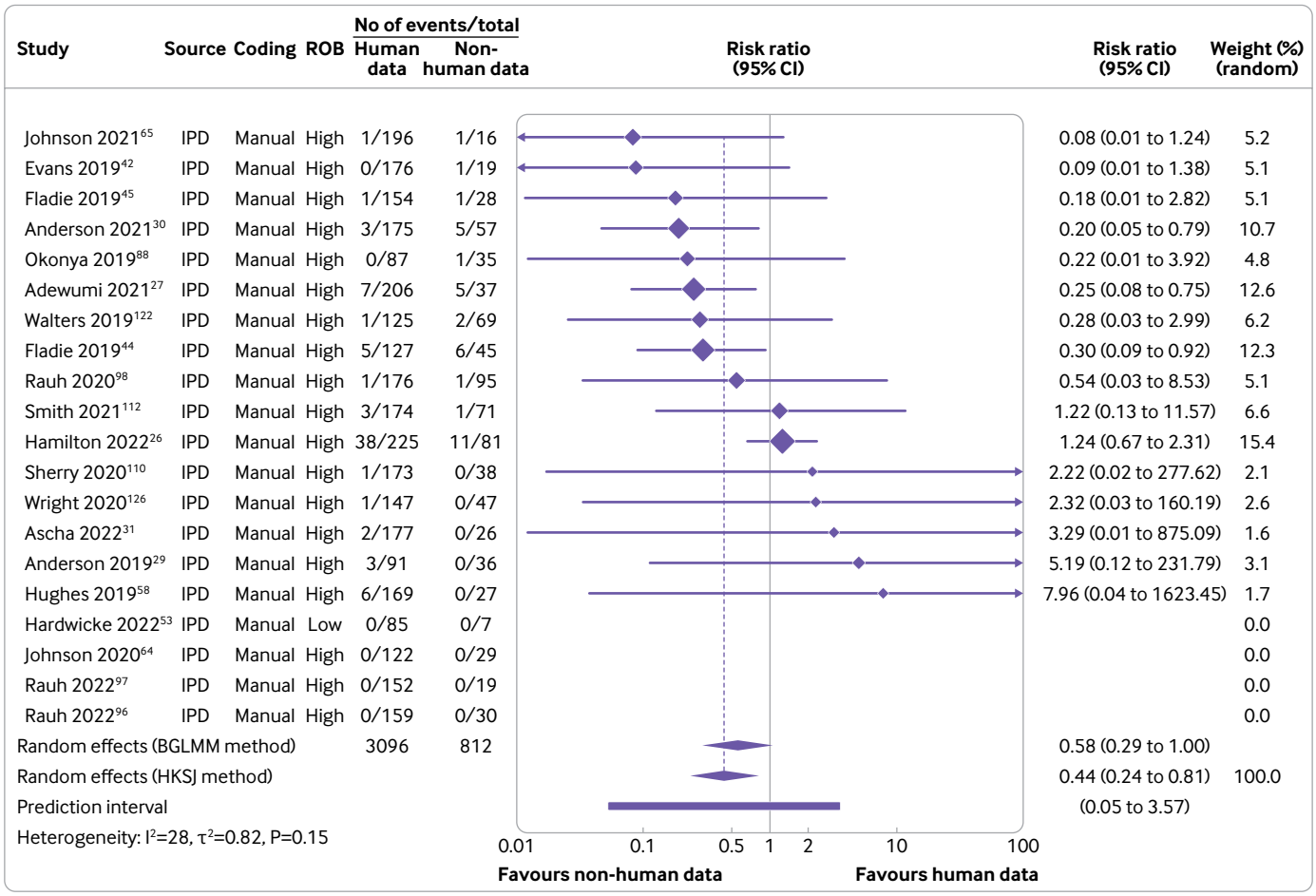


Fig 10 | Association between type of research participant and prevalence of actual public data sharing. ROB=risk of bias; BGLMM=bivariate generalised linear mixed model; HKSJ=Hartung-Knapp-Sidik-Jonkman; IPD=individual participant data

in compliance among the journals that we studied, average compliance with mandatory data sharing policies in medicine and health were lower than those reported in other disciplines.[19 139 148] However, these policies seem to be more effective than the studied alternatives (eg, "must share on request" and "encourage sharing" policies). Furthermore, these policies might vary in their effectiveness according to the type of data. Based on the large variability in compliance, we recommend that policy makers periodically audit compliance with these policies, possibly triaging audits by type of data, and strengthen policing if substantial non-compliance is detected. Enforcement of policies in this setting could range from simple checks of commonly reported problems (eg, that links are present and functional[149]), to verifying that data can be freely downloaded and are complete, sufficiently unprocessed, and well annotated.

For researchers, the finding that data reported to be available are frequently not accessible or reusable highlights the need for improved research training in how to share and reuse data. Also, because the average prevalence of data sharing is low, the medical research community could consider more incentives to increase the frequency and quality of data sharing. For example, some commonly proposed strategies, beyond implementation of policies mandating sharing, include open science badges, data embargoes, data publications, new altmetrics, as well as changes to funding schemes to allow applicants to budget for data archival costs, and academic hiring and promotion criteria to reward sharing.[80 150 151] Although such strategies have long been suggested by medical research stakeholders, such as the US National Academy of Medicine,[152] as previous research has noted, in medicine, more opinion pieces on the lack of incentives for researchers to share data exist than there are empirical tests of these incentives.[150] Consequently, the effectiveness of most of these strategies in medicine is unclear.

Strengths and limitations of this study

Our review had many methodological advantages over previous research in this area. Firstly, because data and code sharing are relatively rare events, meta-analysis of individual participant data allowed us to bring together many imprecise findings to give more precise estimates. Retrieval of useful individual participant data from 95% of the included studies also allowed us to conduct several data quality checks, identify and remove substantial amounts of redundant assessments, perform subgroup analyses not possible

Table 2 | Subgroup analysis: prevalence of declared and actual public data sharing according to type of data and journal policy

	Declared data sharing					Actual data sharing				
Policy	Sharing prevalence (%)	95% CI (%)	95% PI (%)	k	I^2	Sharing prevalence (%)	95% CI (%)	95% PI (%)	k	I^2 (%)
No data restrictions:										
No policy	17*	0 to 59	NA	4	95	4*	0 to 95	NA	2	83
Encourage policy	17*	0 to 62	0 to 100	6	98	8*	0 to 48	NA	3	90
Mandatory policy	65*	36 to 88	2 to 100	5	99	33*	5 to 69	NA	3	93
Sequence data:										
No policy	—	—	—	—	—	46*	0 to 100	NA	2	94
Encourage policy	—	—	—	—	—	57*	15 to 94	NA	2	0
Mandatory policy	—	—	—	—	—	67*	45 to 86	NA	3	70
Gene expression data:										
No policy	23	11 to 42	NA	1	NA	—	—	—	—	—
Encourage policy	30	19 to 44	NA	1	NA	43	31 to 55	NA	1	NA
Mandatory policy	69*	0 to 100	NA	2	93	43*	0 to 100	NA	2	53
Trial data:										
No policy	0*	0 to 46	NA	2	72	—	—	—	—	—
Encourage policy	0*	0 to 5	NA	3	24	—	—	—	—	—
Mandatory policy	55*	40 to 70	NA	2	0	56	33 to 77	NA	1	NA
Systematic review data:										
No policy	5*	2 to 8	NA	2	0	0	0 to 4	NA	1	NA
Encourage policy	3*	0 to 100	NA	2	87	1	0 to 4	NA	1	NA
Mandatory policy	62*	0 to 100	NA	2	92	28	16 to 44	NA	1	NA

CI=confidence interval; PI=prediction interval; k=number of eligible meta-research studies; NA=not applicable.
*Pooled estimate from random effects meta-analysis.

when conducting a meta-analysis of aggregate data, as well as minimise the risk of data availability biases. Secondly, the meta-analyses of our primary and secondary outcomes included more studies than the average meta-analysis of prevalence and rare events,[153 154] reducing the risk of power issues, and making our review a comprehensive analysis of the prevalence of actual data and code sharing. We also had more than double the recommended number of estimates for each covariate for our meta-regression analyses, minimising the risk of problems such as overfitting.[155] Thirdly, the review included checks for robustness with generalised linear mixed models, which have been recommended over conventional meta-analyses of arcsine transformed proportions.[156]

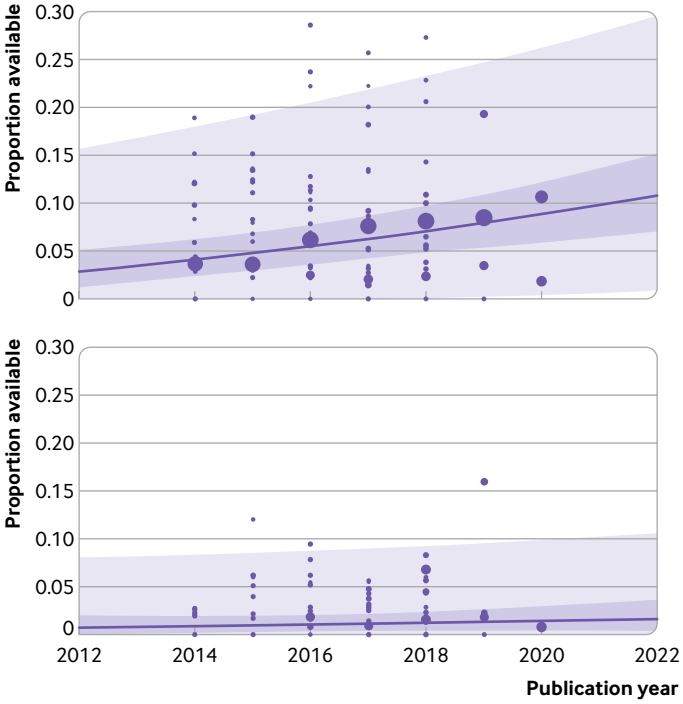


Fig 11 | Bubble plot of the prevalence of declared (top) and actual (bottom) data sharing by publication year with fitted meta-regression lines, 95% confidence intervals (dark purple shaded area), and 95% prediction intervals (light purple shaded area). Circles are scaled relative to the natural log of the sample size

Our review had some limitations. We might have missed relevant literature because of challenges in designing the search strategies (eg, lack of controlled vocabulary, variations in the way studies described themselves) and limiting searches to predominantly English language databases. Also, we could not include the findings of nine studies because we could not source individual participant data or useable summary data. Because 97% of primary articles examined by the excluded studies were at high risk of overlap with studies that were included in the analysis, however, we do not believe that their omission would have substantially changed our findings. We also assumed that authors will always declare in the text when data or code have been made publicly available, which previous studies have shown is not always the case.[85] This practice seems to be uncommon, however, and therefore was unlikely to have substantially affected our results. Most of the primary articles in our meta-analyses of declared availability also originated from two large studies which used automated coding strategies,[25 118] although sensitivity analyses showed that removal of these studies did not change any of the reported findings. Finally, despite efforts to ensure studies were clinically homogeneous, our meta-analyses of proportions showed high levels of statistical inconsistency (I^2). Considering that 75% of published meta-analyses of proportions report I^2 values >90%,[153] however, the statistic's usefulness for assessing heterogeneity in this context is unclear.

Table 3 | Summary of main findings

	Declared availability					Actual availability				
Outcome	Sharing prevalence (%)	95% CI (%)	95% PI (%)	k	o	Sharing prevalence (%)	95% CI (%)	95% PI (%)	k	o
Public data sharing	8*	5 to 11	0 to 30	27	700 054	2*	1 to 3	0 to 11	25	11 873
Public code sharing	0.3*	0 to 1	0 to 8	26	707 943	0.1*	0 to 0.3	0 to 1	21	3 843
Private data sharing	2*	1 to 4	0 to 10	23	3 058	0 to 37†	NA	NA	12	NA
Private code sharing	0*	0 to 0.1	0 to 0.5	22	2 825	0 to 23†	NA	NA	5	NA

CI=confidence interval; PI=prediction interval; k=number of meta-research studies; o=number of primary articles; NA=not applicable.
*Pooled estimate from random effects meta-analysis. †Point estimate range.

Consequently, evidence synthesis researchers have recommended that greater priority should be given to visually inspecting forest plots and prediction interval widths instead.[153] Therefore, although we acknowledge these high I^2 values, because of the consistency of the study methods and reported estimates, as well as the narrow width of the prediction intervals, we do not believe that these values indicate concerning levels of variability in this context.

Conclusion

The results of this review suggest that although increasing numbers of medical and health researchers are stating that their data are publicly available, such declarations are rare, and not all declarations lead to actual availability of the data. In contrast, the prevalence of both declared and actual code sharing are persistently low in medicine. We also found varying levels of success in privately obtaining data and code from authors of published medical research. Although no data were available to evaluate the effectiveness of the data sharing policies of funders and institutions, assessments of journal policies suggested that mandatory sharing policies were more effective than non-mandatory policies, but showed varying compliance depending on the journal and type of data. This finding might be informative for policy makers when designing policies and allocating resources to audit compliance.

We thank Steve McDonald for his assistance with the literature searches, Sue Finch for her advice on the statistical analyses, Rebecca Hamilton and Jason Wasiak for their feedback on the manuscript, and all of the meta-researchers who took the time to prepare data and provide clarifications for the review.

Contributors: DGH conceived and designed the study, collected the data, performed the formal analysis, curated the data, and wrote and prepared the original draft. KH, HF, and AR-F collected the data and validated the study. MJP collected the data, and supervised and validated the study. All authors contributed to the methodology, interpreted the results, contributed to writing the manuscript, approved the final version, and had final responsibility for the decision to submit for publication. DGH is the guarantor. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

Funding: The study received no funding. DGH is supported by an Australian Commonwealth Government Research Training Programme Scholarship. MJP is supported by an Australian Research Council Discovery Early Career Researcher Grant (DE200101618). The Laura and John Arnold Foundation funds the RIAT Support Centre (no grant number), which supports the salaries of AR-F and KH. KH's project was supported by the US Food and Drug Administration (FDA), Department of Health and Human Services (HHS), as part of a financial assistance award U01FD005946 totalling US$5000 with 100% funded by FDA/HHS. The contents are those of the authors and do not necessarily represent the official views of, or an endorsement of, the FDA/HHS or the US government. The funders had no role in considering the study design or in the collection, analysis, interpretation of data, writing of the report, or decision to submit the article for publication.

Competing interests: All authors have completed the ICMJE uniform disclosure form at https://www.icmje.org/disclosure-of-interest/ and declare: support from the Australian Research Council during the conduct of this research for the submitted work; some authors had support from research institutions listed in the funding statement; some authors received financial support to attend an international congress to present preliminary results of this project; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work.

Ethical approval: Not applicable.

Data sharing: Summary level data and the code required to reproduce all of the findings of the review are freely available on the Open Science Framework (OSF) (doi: 10.17605/OSF.IO/U3YRP) under a Creative Commons zero version 1.0 universal (CC0 1.0) license. Harmonised versions of the 70 datasets that were originally posted publicly are also available on the OSF. However, to preserve the rights of data owners, harmonised versions of the remaining individual participant data that were shared privately with the review team will only be released with the permission of the data guarantor of the relevant meta-research study. To request harmonised individual participant data, please follow the instructions on the project's Open Science Framework page (https://osf.io/stnk3).

The lead author affirms that this manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as registered have been explained.

Dissemination to participants and related patient and public communities: The results of this research will be disseminated at national and international conferences, as well as at seminars and workshops aimed at clinicians, researchers, publishers, funders, and other relevant research stakeholders. The results of the review will be circulated to all of the meta-researchers who privately shared their data, and the findings will be posted on social media (eg, Twitter, LinkedIn).

Provenance and peer review: Not commissioned; externally peer reviewed.

1 Shahin MH, Bhattacharya S, Silva D, et al. Open Data Revolution in Clinical Research: Opportunities and Challenges. *Clin Transl Sci* 2020;13:665-74. doi:10.1111/cts.12756
2 Kim J, Kim S, Cho HM, Chang JH, Kim SY. Data sharing policies of journals in life, health, and physical sciences indexed in Journal Citation Reports. *PeerJ* 2020;8:e9924. doi:10.7717/peerj.9924
3 Hamilton DG, Fraser H, Hoekstra R, Fidler F. Journal policies and editors' opinions on peer review. *Elife* 2020;9:e62529. doi:10.7554/eLife.62529

4 Resnik DB, Morales M, Landrum R, et al. Effect of impact factor and discipline on journal data sharing policies. *Account Res* 2019;26:139-56. doi:10.1080/08989621.2019.1591277
5 DeVito NJ, French L, Goldacre B. Noncommercial Funders' Policies on Trial Registration, Access to Summary Results, and Individual Patient Data Availability. *JAMA* 2018;319:1721-3. doi:10.1001/jama.2018.2841
6 Gaba JF, Siebert M, Dupuy A, Moher D, Naudet F. Funders' data-sharing policies in therapeutic research: A survey of commercial and non-commercial funders. *PLoS One* 2020;15:e0237464. doi:10.1371/journal.pone.0237464
7 The White House. OSTP Issues Guidance to Make Federally Funded Research Freely Available Without Delay. 25 August 2022. https://www.whitehouse.gov/wp-content/uploads/2022/08/08-2022-OSTP-Public-Access-Memo.pdf
8 Hamilton DG, Fraser H, Fidler F, Rowhani-Farid A, Hong K, Page MJ. Rates and predictors of data and code sharing in the medical and health sciences: A systematic review and individual participant data meta-analysis. Open Science Framework 2021. doi:10.17605/OSF.IO/7SX8U.
9 Hamilton DG, Fraser H, Fidler F, et al. Rates and predictors of data and code sharing in the medical and health sciences: Protocol for a systematic review and individual participant data meta-analysis[version 2; peer review: 2 approved]. *F1000Res* 2021;10:491. doi:10.12688/f1000research.53874.2.
10 Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Syst Rev* 2021;10:89. doi:10.1186/s13643-021-01626-4
11 Stewart LA, Clarke M, Rovers M, et al, PRISMA-IPD Development Group. Preferred Reporting Items for Systematic Review and Meta-Analyses of individual participant data: the PRISMA-IPD Statement. *JAMA* 2015;313:1657-65. doi:10.1001/jama.2015.3656
12 Haddaway NR, Grainger MJ, Gray CT. citationchaser: An R package and Shiny app for forward and backward citations chasing in academic searching.2021. doi:10.5281/zenodo.4543513.
13 Hamilton DG, Fraser H, Fidler F, et al. *A review of data and code sharing rates in medical and health research.* Open Science Framework, 2023, 10.17605/OSF.IO/H75V4.
14 Hansen C, Bero L, Hróbjartsson A, et al. Conflicts of interest and recommendations in clinical guidelines, opinion pieces, and narrative reviews. *Cochrane Database Syst Rev* 2019;1010.1002/14651858.MR000040.pub2.
15 Page MJ, McKenzie JE, Kirkham J, et al. Bias due to selective inclusion and reporting of outcomes and analyses in systematic reviews of randomised trials of healthcare interventions. *Cochrane Database Syst Rev* 2014;2014:MR000035. doi:10.1002/14651858.MR000035.pub2
16 IntHout J, Ioannidis JP, Borm GF. The Hartung-Knapp-Sidik-Jonkman method for random effects meta-analysis is straightforward and considerably outperforms the standard DerSimonian-Laird method. *BMC Med Res Methodol* 2014;14:25. doi:10.1186/1471-2288-14-25.
17 Sweeting MJ, Sutton AJ, Lambert PC. What to add to nothing? Use and avoidance of continuity corrections in meta-analysis of sparse data. *Stat Med* 2004;23:1351-75. doi:10.1002/sim.1761
18 Wilkinson MD, Dumontier M, Aalbersberg IJ, et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 2016;3:160018. doi:10.1038/sdata.2016.18
19 Milia N, Congiu A, Anagnostou P, et al. Mine, yours, ours? Sharing data on human genetic variation. *PLoS One* 2012;7:e37552. doi:10.1371/journal.pone.0037552.
20 Rufiange M, Rousseau-Blass F, Pang DSJ. Incomplete reporting of experimental studies and items associated with risk of bias in veterinary research. *Vet Rec Open* 2019;6:e000322. doi:10.1136/vetreco-2018-000322.
21 Witwer KW. Data submission and quality in microarray-based microRNA profiling. *Clin Chem* 2013;59:392-400. doi:10.1373/clinchem.2012.193813.
22 Zavalis EA, Ioannidis JPA. A meta-epidemiological assessment of transparency indicators of infectious disease models.*medRxiv* 2022. doi:10.1101/2022.04.11.22273744.
23 Gabelica M, Cavar J, Puljak L. Authors of trials from high-ranking anesthesiology journals were not willing to share raw data. *J Clin Epidemiol* 2019;109:111-6. doi:10.1016/j.jclinepi.2019.01.012.
24 Nguyen PY, Kanukula R, McKenzie JE, et al. Changing patterns in reporting and sharing of review data in systematic reviews with meta-analysis of the effects of interventions: cross sectional meta-research study. *BMJ* 2022;379:e072428. doi:10.1136/bmj-2022-072428.
25 Serghiou S, Contopoulos-Ioannidis DG, Boyack KW, Riedel N, Wallach JD, Ioannidis JPA. Assessment of transparency indicators across the biomedical literature: How open is open?*PLoS Biol* 2021;19:e3001107. doi:10.1371/journal.pbio.3001107.
26 Hamilton DG, Page MJ, Finch S, Everitt S, Fidler F. How often do cancer researchers make their data and code available and what factors are associated with sharing?*BMC Med* 2022;20:438. doi:10.1186/s12916-022-02644-2.
27 Adewumi MT, Vo N, Tritz D, Beaman J, Vassar M. An evaluation of the practice of transparency and reproducibility in addiction medicine literature. *Addict Behav* 2021;112:106560. doi:10.1016/j.addbeh.2020.106560.
28 Alsheikh-Ali AA, Qureshi W, Al-Mallah MH, Ioannidis JPA. Public availability of published research data in high-impact journals. *PLoS One* 2011;6:e24357. doi:10.1371/journal.pone.0024357.
29 Anderson JM, Niemann A, Johnson AL, Cook C, Tritz D, Vassar M. Transparent, Reproducible, and Open Science Practices of Published Literature in Dermatology Journals: Cross-Sectional Analysis. *JMIR Dermatology.* 2019;2:e16078. doi:10.2196/16078.
30 Anderson JM, Wright B, Rauh S, et al. Evaluation of indicators supporting reproducibility and transparency within cardiology literature. *Heart* 2021;107:120-6. doi:10.1136/heartjnl-2020-316519.
31 Ascha M, Katabi L, Stevens E, Gatherwright J, Vassar M. Reproducible Research Practices in the Plastic Surgery Literature. *Plast Reconstr Surg* 2022;149:810e-23e. doi:10.1097/PRS.0000000000008956.
32 Bergeat D, Lombard N, Gasmi A, Le Floch B, Naudet F. Data Sharing and Reanalyses Among Randomized Clinical Trials Published in Surgical Journals Before and After Adoption of a Data Availability and Reproducibility Policy. *JAMA Netw Open* 2022;5:e2215209. doi:10.1001/jamanetworkopen.2022.15209.
33 Bonetti AF, Tonin FS, Lucchetta RC, Pontarolo R, Fernandez-Llimos F. Methodological standards for conducting and reporting meta-analyses: Ensuring the replicability of meta-analyses of pharmacist-led medication review. *Res Social Adm Pharm* 2022;18:2259-68. doi:10.1016/j.sapharm.2021.06.002.
34 Borghi JA, Payne C, Ren L, Woodward AL, Wong C, Stave C. Open Science and COVID-19 Randomized Controlled Trials: Examining Open Access, Preprinting, and Data Sharing-Related Practices During the Pandemic.*medRxiv* 2022. doi:10.1101/2022.08.10.22278643.
35 Cenci MS, Franco MC, Raggio DP, Moher D, Pereira-Cenci T. Transparency in clinical trials: Adding value to paediatric dental research. *Int J Paediatr Dent* 2020;31(Suppl 1):4-13. doi:10.1111/ipd.12769.
36 Colavizza G, Hrynaszkiewicz I, Staden I, Whitaker K, McGillivray B. The citation advantage of linking publications to research data. *PLoS One* 2020;15:e0230416. doi:10.1371/journal.pone.0230416.
37 Collins A, Alexander R. Reproducibility of COVID-19 pre-prints. *Scientometrics* 2022;127:4655-73. doi:10.1007/s11192-022-04418-2.
38 Danchev V, Min Y, Borghi J, Baiocchi M, Ioannidis JPA. Evaluation of Data Sharing After Implementation of the International Committee of Medical Journal Editors Data Sharing Statement Requirement. *JAMA Netw Open* 2021;4:e2033972. doi:10.1001/jamanetworkopen.2020.33972.
39 DeBlanc J, Kay B, Lehrich J, et al. Availability of Statistical Code From Studies Using Medicare Data in General Medical Journals. *JAMA Intern Med* 2020;180:905-7. doi:10.1001/jamainternmed.2020.0671.
40 Duan Y, Luo J, Zhao L, et al. Reporting and data sharing level for COVID-19 vaccine trials: A cross-sectional study. *EBioMedicine* 2022;78:103962. doi:10.1016/j.ebiom.2022.103962.
41 Errington TM, Denis A, Perfito N, Iorns E, Nosek BA. Challenges for assessing replicability in preclinical cancer biology. *Elife* 2021;10:e67995. doi:10.7554/eLife.67995.
42 Evans S, Fladie IA, Anderson JM, Tritz D, Vassar M. Evaluation of Reproducible and Transparent Research Practices in Sports Medicine Research: A Cross-sectional study.*bioRxiv* 2019. doi:10.1101/773473.
43 Federer LM, Belter CW, Joubert DJ, et al. Data sharing in PLOS ONE: An analysis of Data Availability Statements. *PLoS One* 2018;13:e0194768. doi:10.1371/journal.pone.0194768.
44 Fladie IA, Adewumi TM, Vo NH, Tritz DJ, Vassar MB. An Evaluation of Nephrology Literature for Transparency and Reproducibility Indicators: Cross-Sectional Review. *Kidney Int Rep* 2019;5:173-81. doi:10.1016/j.ekir.2019.11.001.
45 Fladie IA, Evans S, Checketts J, Tritz D, Norris B, Vassar M. Can Orthopaedics become the Gold Standard for Reproducibility? A Roadmap to Success.*bioRxiv* 2019. doi:10.1101/715144.
46 Gabelica M, Bojčić R, Puljak L. Many researchers were not compliant with their published data sharing statement: a mixed-methods study. *J Clin Epidemiol* 2022;150:33-41. doi:10.1016/j.jclinepi.2022.05.019.
47 Gkiouras K, Nigdelis MP, Grammatikopoulou MG, Goulis DG. Tracing open data in emergencies: The case of the COVID-19 pandemic. *Eur J Clin Invest* 2020;50:e13323. doi:10.1111/eci.13323.
48 Gorman DM. Availability of Research Data in High-Impact Addiction Journals with Data Sharing Policies. *Sci Eng Ethics* 2020;26:1625-32. doi:10.1007/s11948-020-00203-7.
49 Grant R, Hrynaszkiewicz I. The impact on authors and editors of introducing Data Availability Statements at Nature journals. *Int J Digit Curation* 2018;13:195-203. doi:10.2218/ijdc.v13i1.614.

50 Grayling MJ, Wheeler GM. A review of available software for adaptive clinical trial design. *Clin Trials* 2020;17:323-31. doi:10.1177/1740774520906398.
51 Hanson KA, Almeida N, Traylor JI, Rajagopalan D, Johnson J. Profile of Data Sharing in the Clinical Neurosciences. *Cureus* 2020;12:e9927. doi:10.7759/cureus.9927.
52 Hardwicke TE, Ioannidis JPA. Populating the Data Ark: An attempt to retrieve, preserve, and liberate data from the most highly-cited psychology and psychiatry articles. *PLoS One* 2018;13:e0201856. doi:10.1371/journal.pone.0201856.
53 Hardwicke TE, Thibault RT, Kosie JE, Wallach JD, Kidwell MC, Ioannidis JPA. Estimating the Prevalence of Transparency and Reproducibility-Related Research Practices in Psychology (2014-2017). *Perspect Psychol Sci* 2022;17:239-51. doi:10.1177/1745691620979806.
54 Heckerman G, Tzng E, Campos-Melendez A, Ekwueme C, Mueller AL. Accessibility and Reproducible Research Practices in Cardiovascular Literature.*bioRxiv* 2022. doi:10.1101/2022.07.06.498942.
55 Heller N, Rickman J, Weight CJ, Papanikolopoulos N. The Role of Publicly Available Data in MICCAI Papers from 2014 to 2018. *arXiv* 2019. doi:10.48550/arXiv.1908.06830.
56 Helliwell JA, Shelton B, Mahmood H, et al. Transparency in surgical randomized clinical trials: cross-sectional observational study. *BJS Open* 2020;4:977-84. doi:10.1002/bjs5.50333.
57 Huang Y-N, Patel NA, Mehta JH, et al. Data Availability of Open T-Cell Receptor Repertoire Data, a Systematic Assessment. *Front Syst Biol* 2022;2. doi:10.3389/fsysb.2022.918792.
58 Hughes T, Niemann A, Tritz D, Boyer K, Robbins H, Vassar M. Transparent and Reproducible Research Practices in the Surgical Literature.*bioRxiv* 2019. doi:10.1101/779702
59 Ioannidis JPA, Allison DB, Ball CA, et al. Repeatability of published microarray gene expression analyses. *Nat Genet* 2009;41:149-55. doi:10.1038/ng.295.
60 Ioannidis JPA, Polyzos NP, Trikalinos TA. Selective discussion and transparency in microarray research findings for cancer outcomes. *Eur J Cancer* 2007;43:1999-2010. doi:10.1016/j.ejca.2007.05.019.
61 Iqbal SA, Wallach JD, Khoury MJ, Schully SD, Ioannidis JPA. Reproducible Research Practices and Transparency across the Biomedical Literature. *PLoS Biol* 2016;14:e1002333. doi:10.1371/journal.pbio.1002333.
62 Jalali MS, DiGennaro C, Sridhar D. Transparency assessment of COVID-19 models. *Lancet Glob Health* 2020;8:e1459-60. doi:10.1016/S2214-109X(20)30447-2.
63 Janssen MA, Pritchard C, Lee A. On code sharing and model documentation of published individual and agent-based models. *Environ Model Softw* 2020;134:104873. doi:10.1016/j.envsoft.2020.104873.
64 Johnson AL, Torgerson T, Skinner M, Hamilton T, Tritz D, Vassar M. An assessment of transparency and reproducibility-related research practices in otolaryngology. *Laryngoscope* 2020;130:1894-901. doi:10.1002/lary.28322.
65 Johnson BS, Rauh S, Tritz D, Schiesel M, Vassar M. Evaluating Reproducibility and Transparency in Emergency Medicine Publications. *West J Emerg Med* 2021;22:963-71. doi:10.5811/westjem.2021.3.50078.
66 Johnson JN, Hanson KA, Jones CA, Grandhi R, Guerrero J, Rodriguez JS. Data Sharing in Neurosurgery and Neurology Journals. *Cureus* 2018;10:e2680. doi:10.7759/cureus.2680.
67 Jurburg SD, Konzack M, Eisenhauer N, Heintz-Buschart A. The archives are half-empty: an assessment of the availability of microbial community sequencing data. *Commun Biol* 2020;3:474. doi:10.1038/s42003-020-01204-9.
68 Kaufmann I, Gattrell WT. Assessment of journal compliance with data sharing guidelines from the International Committee of Medical Journal Editors (ICMJE). *Curr Med Res Opin* 2019;35:31-41. doi:10.1080/03007995.2019.1583496.
69 Kemper JM, Rolnik DL, Mol BWJ, Ioannidis JPA. Reproducible research practices and transparency in reproductive endocrinology and infertility articles. *Fertil Steril* 2020;114:1322-9. doi:10.1016/j.fertnstert.2020.05.020.
70 Kirouac DC, Cicali B, Schmidt S. Reproducibility of Quantitative Systems Pharmacology Models: Current Challenges and Future Opportunities. *CPT Pharmacometrics Syst Pharmacol* 2019;8:205-10. doi:10.1002/psp4.12390.
71 Kobres P-Y, Chretien J-P, Johansson MA, et al. A systematic review and evaluation of Zika virus forecasting and prediction research during a public health emergency of international concern. *PLoS Negl Trop Dis* 2019;13:e0007451. doi:10.1371/journal.pntd.0007451.
72 Littmann M, Selig K, Cohen-Lavi L, et al. Validity of machine learning in biology and medicine increased through collaborations across fields of expertise. *Nat Mach Intell* 2020;2:18-24. doi:10.1038/s42256-019-0139-8.
73 López-Nicolás R, López-López JA, Rubio-Aparicio M, Sánchez-Meca J. A meta-review of transparency and reproducibility-related reporting practices in published meta-analyses on clinical psychological interventions (2000-2020). *Behav Res Methods* 2022;54:334-49. doi:10.3758/s13428-021-01644-z.
74 Louderback ER, Gainsbury SM, Heirene RM, et al. Open Science Practices in Gambling Research Publications (2016-2019): A Scoping Review. *J Gambl Stud* 2023;39:987-1011. doi:10.1007/s10899-022-10120-y.
75 McGuinness LA, Sheppard AL. A descriptive analysis of the data availability statements accompanying medRxiv preprints and a comparison with their published counterparts. *PLoS One* 2021;16:e0250887. doi:10.1371/journal.pone.0250887.
76 Meyer A, Faverjon C, Hostens M, Stegeman A, Cameron A. Systematic review of the status of veterinary epidemiological research in two species regarding the FAIR guiding principles. *BMC Vet Res* 2021;17:270. doi:10.1186/s12917-021-02971-1.
77 Miyakawa T. No raw data, no science: another possible source of the reproducibility crisis. *Mol Brain* 2020;13:24. doi:10.1186/s13041-020-0552-2.
78 Munkholm K, Faurholt-Jepsen M, Ioannidis JPA, Hemkens LG. Consideration of confounding was suboptimal in the reporting of observational studies in psychiatry: a meta-epidemiological study. *J Clin Epidemiol* 2020;119:75-84. doi:10.1016/j.jclinepi.2019.12.002.
79 Munro BA, Bergen P, Pang DSJ. Randomization, blinding, data handling and sample size estimation in papers published in Veterinary Anaesthesia and Analgesia in 2009 and 2019. *Vet Anaesth Analg* 2022;49:18-25. doi:10.1016/j.vaa.2021.09.004.
80 Naudet F, Sakarovitch C, Janiaud P, et al. Data sharing and reanalysis of randomized controlled trials in leading biomedical journals with a full data sharing policy: survey of studies published in *The BMJ* and *PLOS Medicine*. *BMJ* 2018;360:k400. doi:10.1136/bmj.k400.
81 Noor MAF, Zimmerman KJ, Teeter KC. Data sharing: how much doesn't get submitted to GenBank?*PLoS Biol* 2006;4:e228. doi:10.1371/journal.pbio.0040228.
82 Norris E, He Y, Loh R, West R, Michie S. Assessing Markers of Reproducibility and Transparency in Smoking Behaviour Change Intervention Evaluations. *J Smok Cessat* 2021;2021:6694386. doi:10.1155/2021/6694386.
83 Norris E, Sulevani I, Finnerty AN, Castro O. Assessing Open Science practices in physical activity behaviour change intervention evaluations. *BMJ Open Sport Exerc Med* 2022;8:e001282. doi:10.1136/bmjsem-2021-001282.
84 Ntzani EE, Ioannidis JP. Predictive ability of DNA microarrays for cancer outcomes and correlates: an empirical assessment. *Lancet* 2003;362:1439-44. doi:10.1016/S0140-6736(03)14686-7.
85 Nuijten MB, Borghuis J, Veldkamp CLS, Dominguez-Alvarez L, van Assen MALM, Wicherts JM. Journal Data Sharing Policies and Statistical Reporting Inconsistencies in Psychology. *Collabra Psychol* 2017;3. doi:10.1525/collabra.102.
86 Nutu D, Gentili C, Naudet F, Cristea IA. Open science practices in clinical psychology journals: An audit study. *J Abnorm Psychol* 2019;128:510-6. doi:10.1037/abn0000414.
87 Ochsner SA, Steffen DL, Stoeckert CJJr, McKenna NJ. Much room for improvement in deposition rates of expression microarray datasets. *Nat Methods* 2008;5:991. doi:10.1038/nmeth1208-991
88 Okonya O, Rorah D, Tritz D, Umberham BA, Wiley M, Vassar M. Analysis of Practices to Promote Reproducibility and Transparency in Anaesthesiology Research: Are Important Aspects "Hidden Behind the Drapes?"*bioRxiv* 2019. doi:10.1101/729129
89 Page MJ, Altman DG, Shamseer L, et al. Reproducible research practices are underused in systematic reviews of biomedical interventions. *J Clin Epidemiol* 2018;94:8-18. doi:10.1016/j.jclinepi.2017.10.017.
90 Page MJ, Nguyen P-Y, Hamilton DG, et al. Data and code availability statements in systematic reviews of interventions were often missing or inaccurate: a content analysis. *J Clin Epidemiol* 2022;147:1-10. doi:10.1016/j.jclinepi.2022.03.003.
91 Papageorgiou SN, Antonoglou GN, Martin C, Eliades T. Methods, transparency and reporting of clinical trials in orthodontics and periodontics. *J Orthod* 2019;46:101-9. doi:10.1177/1465312519842315.
92 Park HY, Suh CH, Woo S, Kim PH, Kim KW. Quality Reporting of Systematic Review and Meta-Analysis According to PRISMA 2020 Guidelines: Results from Recently Published Papers in the *Korean Journal of Radiology*. *Korean J Radiol* 2022;23:355-69. doi:10.3348/kjr.2021.0808.
93 Pellen C, Caquelin L, Jouvance-Le Bail A, et al. Intent to share Annals of Internal Medicine's trial data was not associated with data re-use. *J Clin Epidemiol* 2021;137:241-9. doi:10.1016/j.jclinepi.2021.04.011.
94 Piwowar HA, Chapman WW. Public sharing of research datasets: a pilot study of associations. *J Informetr* 2010;4:148-56. doi:10.1016/j.joi.2009.11.010.
95 Piwowar HA, Day RS, Fridsma DB. Sharing detailed research data is associated with increased citation rate. *PLoS One* 2007;2:e308. doi:10.1371/journal.pone.0000308.
96 Rauh S, Bowers A, Rorah D, et al. Evaluating the reproducibility of research in obstetrics and gynecology. *Eur J Obstet Gynecol Reprod Biol* 2022;269:24-9. doi:10.1016/j.ejogrb.2021.12.021.

97 Rauh S, Johnson BS, Bowers A, Tritz D, Vassar BM. A review of reproducible and transparent research practices in urology publications from 2014 to2018. *BMC Urol* 2022;22:102. doi:10.1186/s12894-022-01059-8.
98 Rauh S, Torgerson T, Johnson AL, Pollard J, Tritz D, Vassar M. Reproducible and transparent research practices in published neurology research. *Res Integr Peer Rev* 2020;5:5. doi:10.1186/s41073-020-0091-5.
99 Read KB, Ganshorn H, Rutley S, Scott DR. Data-sharing practices in publications funded by the Canadian Institutes of Health Research: a descriptive analysis. *CMAJ Open* 2021;9:E980-7. doi:10.9778/cmajo.20200303.
100 Reidpath DD, Allotey PA. Data sharing in medical research: an empirical investigation. *Bioethics* 2001;15:125-34. doi:10.1111/1467-8519.00220.
101 Rhee S-Y, Kassaye SG, Jordan MR, Kouamou V, Katzenstein D, Shafer RW. Public availability of HIV-1 drug resistance sequence and treatment data: a systematic review. *Lancet Microbe* 2022;3:e392-8. doi:10.1016/S2666-5247(21)00250-0.
102 Riedel N, Kip M, Bobrov E. ODDPub - a Text-Mining Algorithm to Detect Data Sharing in Biomedical Publications. *Data Sci J* 2020;19:42. doi:10.5334/dsj-2020-042.
103 Rousi AM. Using current research information systems to investigate data acquisition and data sharing practices of computer scientists. *J Librarian Inform Sci* 2022;096100062210930. doi:10.1177/09610006221093049.
104 Rowhani-Farid A, Barnett AG. Badges for sharing data and code at *Biostatistics:* an observational study. *F1000Res* 2018;7:90. doi:10.12688/f1000research.13477.1.
105 Rowhani-Farid A, Barnett AG. Has open data arrived at the British Medical Journal (BMJ)? An observational study. *BMJ Open* 2016;6:e011784. doi:10.1136/bmjopen-2016-011784.
106 Sarker A, Ginn R, Nikfarjam A, et al. Utilizing social media data for pharmacovigilance: A review. *J Biomed Inform* 2015;54:202-12. doi:10.1016/j.jbi.2015.02.004.
107 Savage CJ, Vickers AJ. Empirical study of data sharing by authors publishing in PLoS journals. *PLoS One* 2009;4:e7078. doi:10.1371/journal.pone.0007078.
108 Schulz R, Langen G, Prill R, Cassel M, Weissgerber TL. Reporting and transparent research practices in sports medicine and orthopaedic clinical trials: a meta-research study. *BMJ Open* 2022;12:e059347. doi:10.1136/bmjopen-2021-059347.
109 Seibold H, Czerny S, Decke S, et al. A computational reproducibility study of PLOS ONE articles featuring longitudinal data analyses. *PLoS One* 2021;16:e0251194. doi:10.1371/journal.pone.0251194.
110 Sherry CE, Pollard JZ, Tritz D, Carr BK, Pierce A, Vassar M. Assessment of transparent and reproducible research practices in the psychiatry literature. *Gen Psychiatr* 2020;33:e100149. doi:10.1136/gpsych-2019-100149.
111 Siebert M, Gaba JF, Caquelin L, et al. Data-sharing recommendations in biomedical journals and randomised controlled trials: an audit of journals following the ICMJE recommendations. *BMJ Open* 2020;10:e038887. doi:10.1136/bmjopen-2020-038887.
112 Smith CA, Nolan J, Tritz DJ, et al. Evaluation of reproducible and transparent research practices in pulmonology. *Pulmonology* 2021;27:134-43. doi:10.1016/j.pulmoe.2020.07.001.
113 Sofi-Mahmudi A, Raittio E. Transparency of COVID-19-Related Research in Dental Journals. *Front Oral Health* 2022;3:871033. doi:10.3389/froh.2022.871033.
114 Strcic J, Civljak A, Glozinic T, Pacheco RL, Brkovic T, Puljak L. Open data and data sharing in articles about COVID-19 published in preprint servers medRxiv and bioRxiv. *Scientometrics* 2022;127:2791-802. doi:10.1007/s11192-022-04346-1.
115 Sumner J, Haynes L, Nathan S, Hudson-Vitale C, McIntosh LD. Reproducibility and reporting practices in COVID-19 preprint manuscripts.*medRxiv* . 2020. doi:10.1101/2020.03.24.20042796.
116 Tedersoo L, Küngas R, Oras E, et al. Data sharing practices and data availability upon request differ across scientific disciplines. *Sci Data* 2021;8:192. doi:10.1038/s41597-021-00981-0.
117 Thelwall M, Munafò M, Mas-Bleda A, et al. Is useful research data usually shared? An investigation of genome-wide association study summary statistics. *PLoS One* 2020;15:e0229578. doi:10.1371/journal.pone.0229578.
118 Uribe SE, Sofi-Mahmudi A, Raittio E, Maldupa I, Vilne B. Dental Research Data Availability and Quality According to the FAIR Principles. *J Dent Res* 2022;101:1307-13. doi:10.1177/00220345221101321.
119 Vanpaemel W, Vermorgen M, Deriemaecker L, Storms G. Are We Wasting a Good Crisis? The Availability of Psychological Research Data after the Storm. *Collabra* 2015;1. doi:10.1525/collabra.13.
120 Vassar M, Jellison S, Wendelbo H, Wayant C. Data sharing practices in randomized trials of addiction interventions. *Addict Behav* 2020;102:106193. doi:10.1016/j.addbeh.2019.106193.
121 Wallach JD, Boyack KW, Ioannidis JPA. Reproducible research practices, transparency, and open access data in the biomedical literature, 2015-2017. *PLoS Biol* 2018;16:e2006930. doi:10.1371/journal.pbio.2006930.
122 Walters C, Harter ZJ, Wayant C, et al. Do oncology researchers adhere to reproducible and transparent principles? A cross-sectional survey of published oncology literature. *BMJ Open* 2019;9:e033962. doi:10.1136/bmjopen-2019-033962.
123 Weissgerber T, Riedel N, Kilicoglu H, et al. Automated screening of COVID-19 preprints: can we help authors to improve transparency and reproducibility?*Nat Med* 2021;27:6-7. doi:10.1038/s41591-020-01203-7.
124 Wicherts JM, Crompvoets EAV. The poor availability of syntaxes of structural equation modeling. *Account Res* 2017;24:458-68. doi:10.1080/08989621.2017.1396214.
125 Womack RP. Research Data in Core Journals in Biology, Chemistry, Mathematics, and Physics. *PLoS One* 2015;10:e0143460. doi:10.1371/journal.pone.0143460.
126 Wright BD, Vo N, Nolan J, et al. An analysis of key indicators of reproducibility in radiology. *Insights Imaging* 2020;11:65. doi:10.1186/s13244-020-00870-x.
127 Helliwell JA, Bolton WS, Burke JR, Tiernan JP, Jayne DG, Chapman SJ. Global academic response to COVID-19: Cross-sectional study. *Learn Publ* 2020;33:385-93. doi:10.1002/leap.1317.
128 Hemkens LG, Benchimol EI, Langan SM, et al. The reporting of studies using routinely collected health data was often insufficient. *J Clin Epidemiol* 2016;79:104-11. doi:10.1016/j.jclinepi.2016.06.005.
129 Jiao C, Li K, Fang Z. Data sharing practices across knowledge domains: A dynamic examination of data availability statements in PLOS ONE publications. *J Inf Sci* 2022;016555152211018. doi:10.1177/01655515221101830.
130 McDonald L, Schultze A, Simpson A, Graham S, Wasiak R, Ramagopalan SV. A review of data sharing statements in observational studies published in the BMJ: A cross-sectional study. *F1000Res* 2017;6:1708. doi:10.12688/f1000research.12673.1.
131 Ramke J, Kuper H, Limburg H, et al. Avoidable Waste in Ophthalmic Epidemiology: A Review of Blindness Prevalence Surveys in Low and Middle Income Countries 2000-2014. *Ophthalmic Epidemiol* 2018;25:13-20. doi:10.1080/09286586.2017.1328067.
132 Rustici G, Williams E, Barzine M, et al. Transcriptomics data availability and reusability in the transition from microarray to next-generation sequencing.*bioRxiv* 2021. doi:10.1101/2020.12.31.425022.
133 Stodden V, Seiler J, Ma Z. An empirical analysis of journal policy effectiveness for computational reproducibility. *Proc Natl Acad Sci U S A* 2018;115:2584-9. doi:10.1073/pnas.1708290115.
134 Towse JN, Ellis DA, Towse AS. Opening Pandora's Box: Peeking inside Psychology's data sharing practices, and seven recommendations for change. *Behav Res Methods* 2021;53:1455-68. doi:10.3758/s13428-020-01486-1.
135 Zhao M, Yan E, Li K. Data set mentions and citations: A content analysis of full-text publications. *J Assoc Inf Sci Technol* 2017;69:32-46. doi:10.1002/asi.23919.
136 Krawczyk M, Reuben E. (Un)available upon request: field experiment on researchers' willingness to share supplementary materials. *Account Res* 2012;19:175-86. doi:10.1080/08989621.2012.678688
137 Minocher R, Atmaca S, Bavero C, McElreath R, Beheim B. Estimating the reproducibility of social learning research published between 1955 and 2018. *R Soc Open Sci* 2021;8:210450. doi:10.1098/rsos.210450
138 Wicherts JM, Borsboom D, Kats J, Molenaar D. The poor availability of psychological research data for reanalysis. *Am Psychol* 2006;61:726-8. doi:10.1037/0003-066X.61.7.726
139 Vines TH, Andrew RL, Bock DG, et al. Mandated data archiving greatly improves access to research data. *FASEB J* 2013;27:1304-8. doi:10.1096/fj.12-218164
140 Vines TH, Albert AYK, Andrew RL, et al. The availability of research data declines rapidly with article age. *Curr Biol* 2014;24:94-7. doi:10.1016/j.cub.2013.11.014
141 Roche DG, Raby GD, Norin T, et al. Paths towards greater consensus building in experimental biology. *J Exp Biol* 2022;225(Suppl_1):jeb243559. doi:10.1242/jeb.243559
142 Stagge JH, Rosenberg DE, Abdallah AM, Akbar H, Attallah NA, James R. Assessing data availability and research reproducibility in hydrology and water resources. *Sci Data* 2019;6:190030. doi:10.1038/sdata.2019.30
143 Culina A, van den Berg I, Evans S, Sánchez-Tójar A. Low availability of code in ecology: A call for urgent action. *PLoS Biol* 2020;18:e3000763. doi:10.1371/journal.pbio.3000763
144 Tenopir C, Dalton ED, Allard S, et al. Changes in Data Sharing and Data Reuse Practices and Perceptions among Scientists Worldwide. *PLoS One* 2015;10:e0134826. doi:10.1371/journal.pone.0134826

145 Gourd E. GDPR obstructs cancer research data sharing. *Lancet Oncol* 2021;22:592. doi:10.1016/S1470-2045(21)00207-2
146 Evans BJ, Jarvik GP. Impact of HIPAA's minimum necessary standard on genomic data sharing. *Genet Med* 2018;20:531-5. doi:10.1038/gim.2017.141
147 Anagnostou P, Capocasa M, Milia N, et al. When data sharing gets close to 100%: what human paleogenetics can teach the open science movement. *PLoS One* 2015;10:e0121409. doi:10.1371/journal.pone.0121409
148 Thelwall M, Kousha K. Do journal data sharing mandates work? Life sciences evidence from Dryad. *Aslib J Inf Manag* 2017;69:36-4510.1108/AJIM-09-2016-0159.
149 Rowhani-Farid A, Barnett AG. Badges for sharing data and code at *Biostatistics:* an observational study. *F1000Res* 2018;7:90. doi:10.12688/f1000research.13477.1
150 Rowhani-Farid A, Allen M, Barnett AG. What incentives increase data sharing in health and medical research? A systematic review. *Res Integr Peer Rev* 2017;2:4. doi:10.1186/s41073-017-0028-9
151 Devriendt T, Shabani M, Borry P. Data Sharing in Biomedical Sciences: A Systematic Review of Incentives. *Biopreserv Biobank* 2021;19:219-27. doi:10.1089/bio.2020.0037
152 Institute of Medicine (IOM). *Sharing clinical trial data: Maximizing benefits, minimizing risk.* The National Academies Press, 2015.
153 Migliavaca CB, Stein C, Colpani V, et al, Prevalence Estimates Reviews-Systematic Review Methodology Group (PERSyst). Meta-analysis of prevalence: I^2 statistic and how to deal with heterogeneity. *Res Synth Methods* 2022;13:363-7. doi:10.1002/jrsm.1547
154 Jia P, Lin L, Kwong JSW, Xu C. Many meta-analyses of rare events in the Cochrane Database of Systematic Reviews were underpowered. *J Clin Epidemiol* 2021;131:113-22. doi:10.1016/j.jclinepi.2020.11.017
155 Deeks JJ, Higgins JPT, Altman DG. Analysing data and undertaking meta-analyses. In: Higgins JPT, Thomas J, Chandler J, et al, eds. *Cochrane Handbook for Systematic Reviews of Interventions version 6.3.* Cochrane, 2022.
156 Warton DI, Hui FKC. The arcsine is asinine: the analysis of proportions in ecology. *Ecology* 2011;92:3-10. doi:10.1890/10-0340.1

Web appendix: Supplementary information