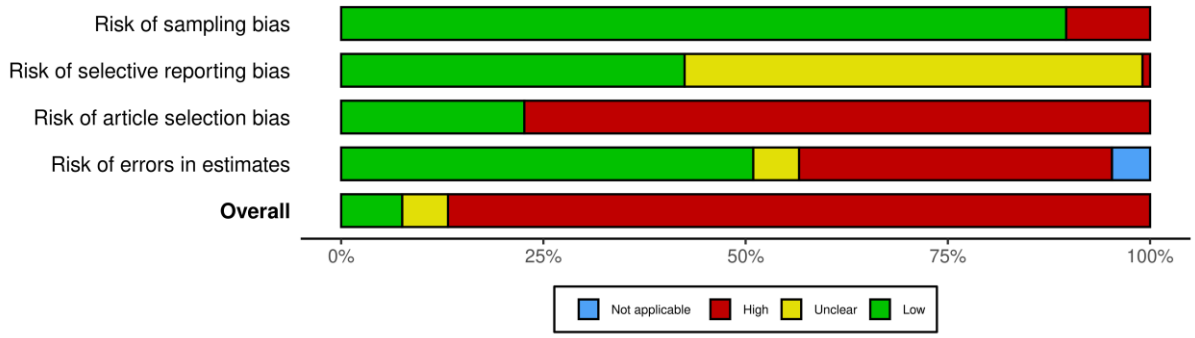
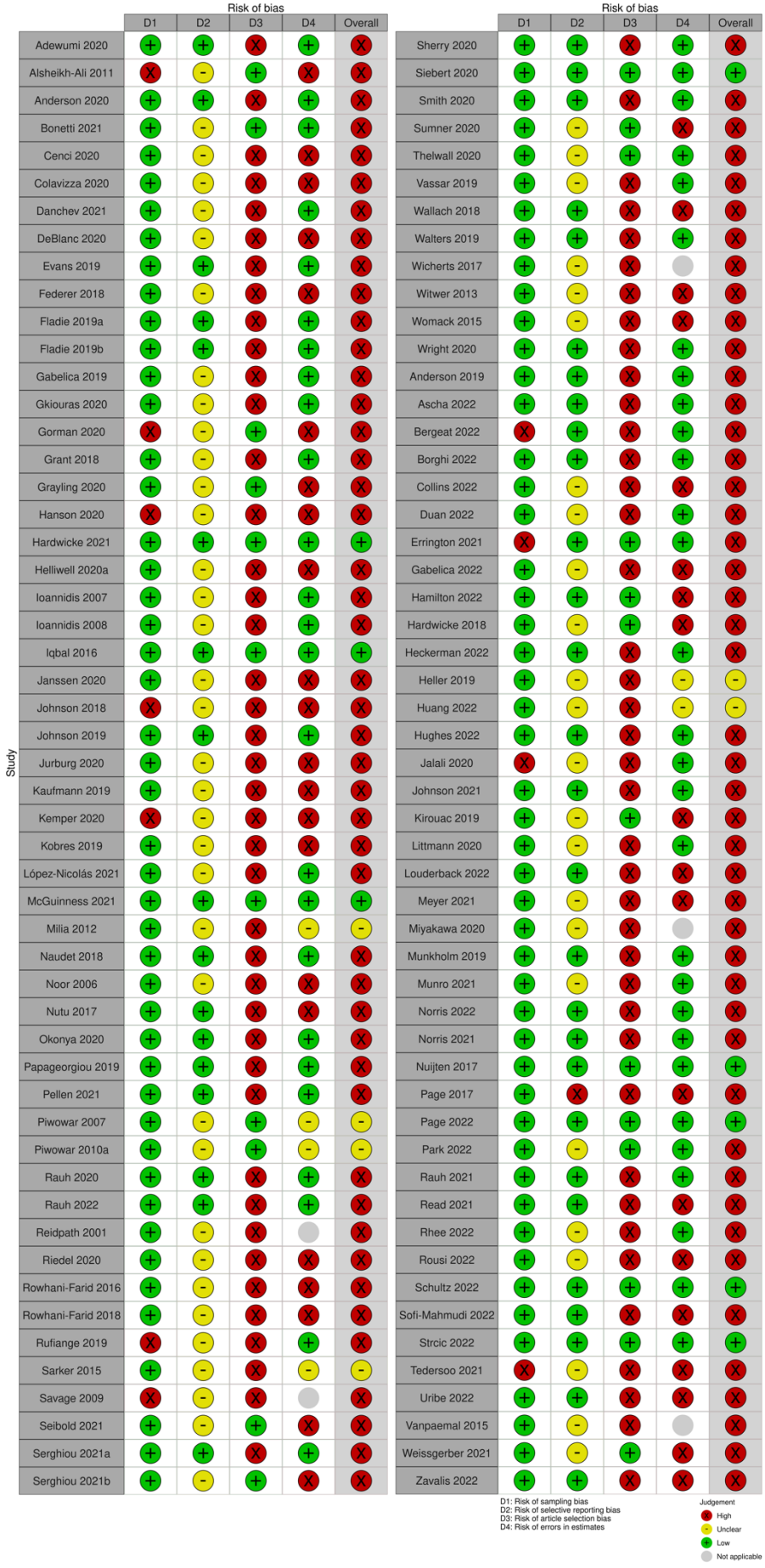
SUPPLEMENTARY INFORMATION

Appendix to: Hamilton DG, Hong K, Fraser H, Rowhani-Farid A, Fidler F & Page MJ. Prevalence and predictors of data and code sharing in the medical and health sciences: A systematic review with meta-analysis of individual participant data. BMJ 2023; 382: e075767. DOI: 10.1136/bmj-2023-075767.

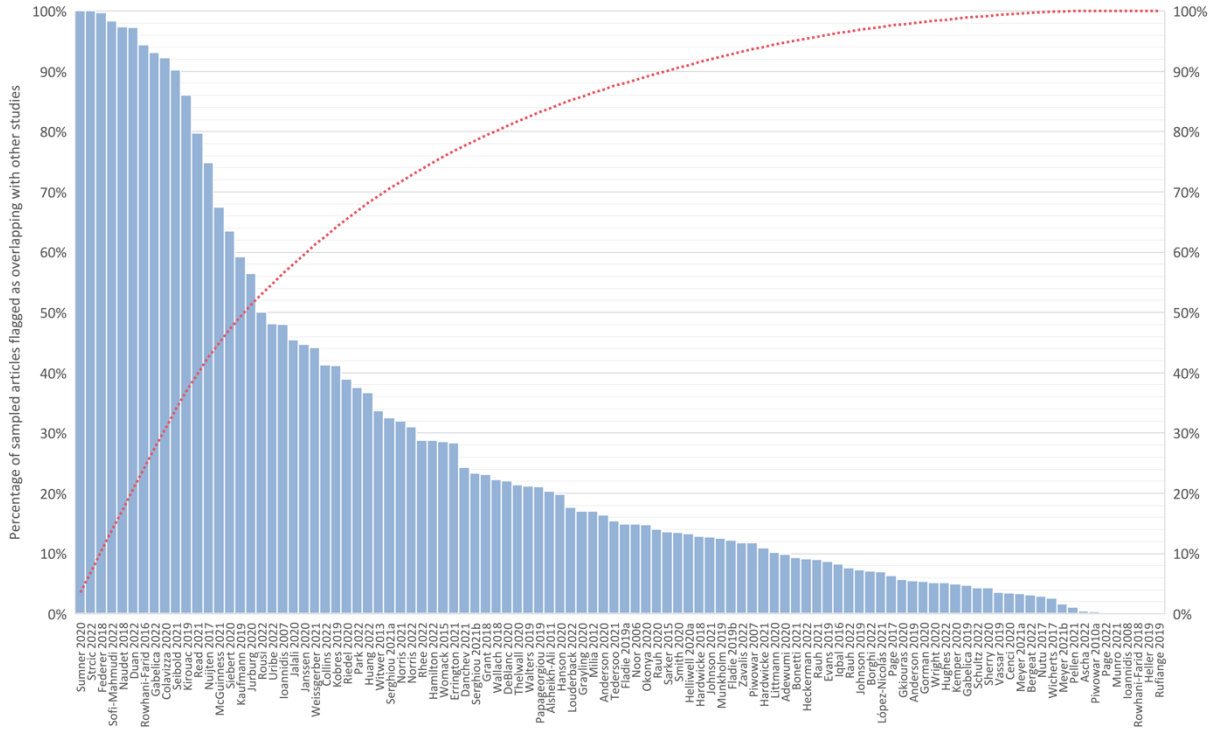
SUPPLEMENTARY FIGURE 1. SUMMARY OF THE RESULTS OF THE RISK OF BIAS ASSESSMENTS.



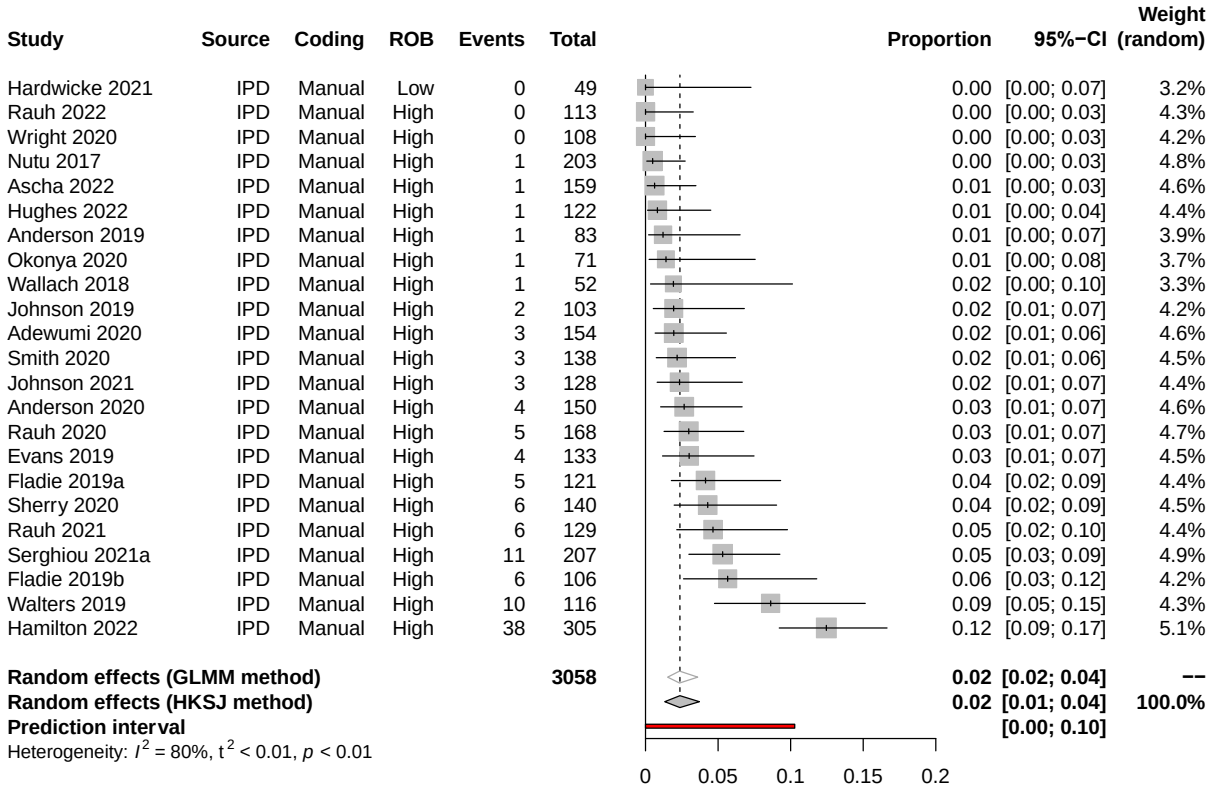
SUPPLEMENTARY FIGURE 2. RISK OF BIAS TRAFFIC LIGHT PLOT.



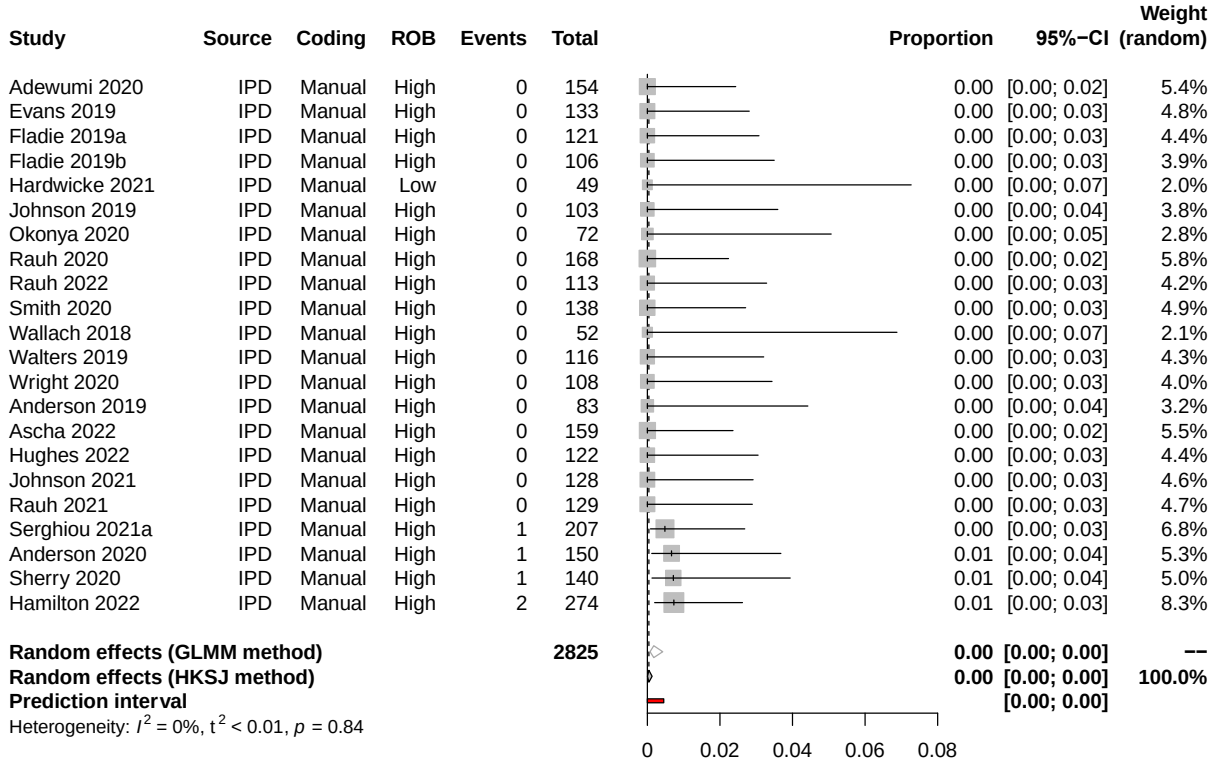
SUPPLEMENTARY FIGURE 3. PROPORTION OF PRIMARY ARTICLES ASSESSMENTS FLAGGED AS POSSIBLY REDUNDANT.



SUPPLEMENTARY FIGURE 4. PREVALENCE OF DECLARED PRIVATE DATA SHARING SINCE 2016.

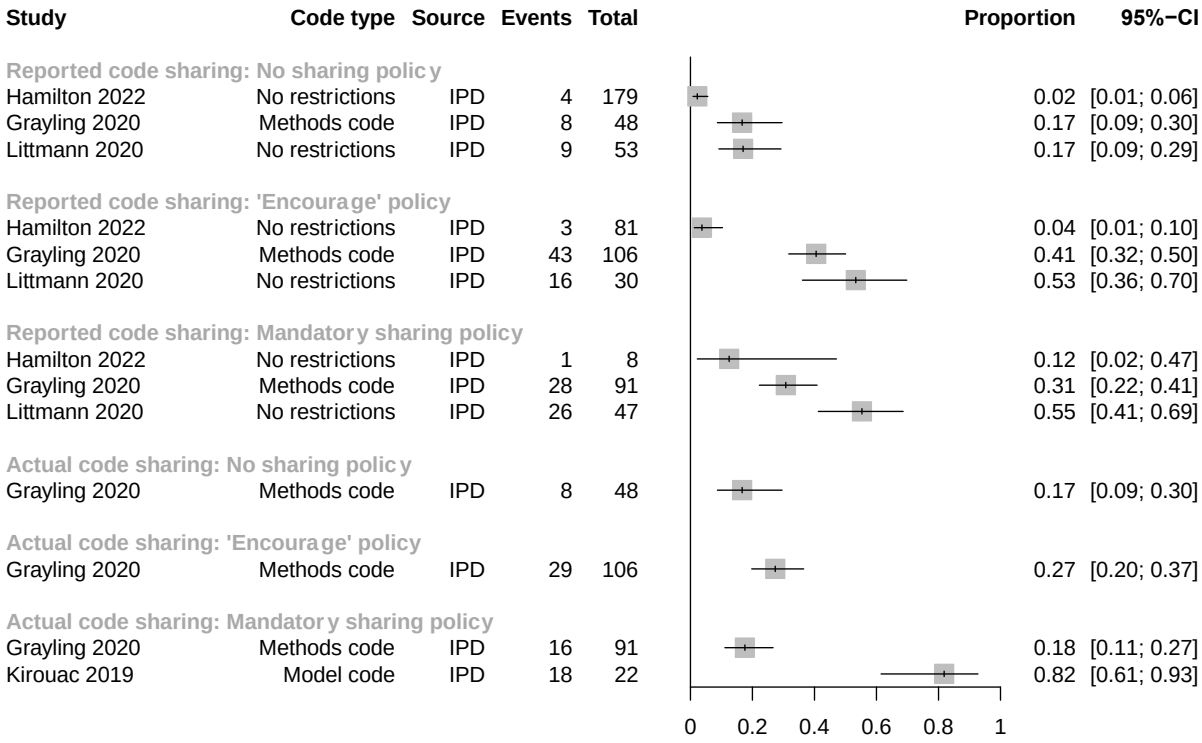


SUPPLEMENTARY FIGURE 5. PREVALENCE OF DECLARED PRIVATE CODE SHARING SINCE 2016.

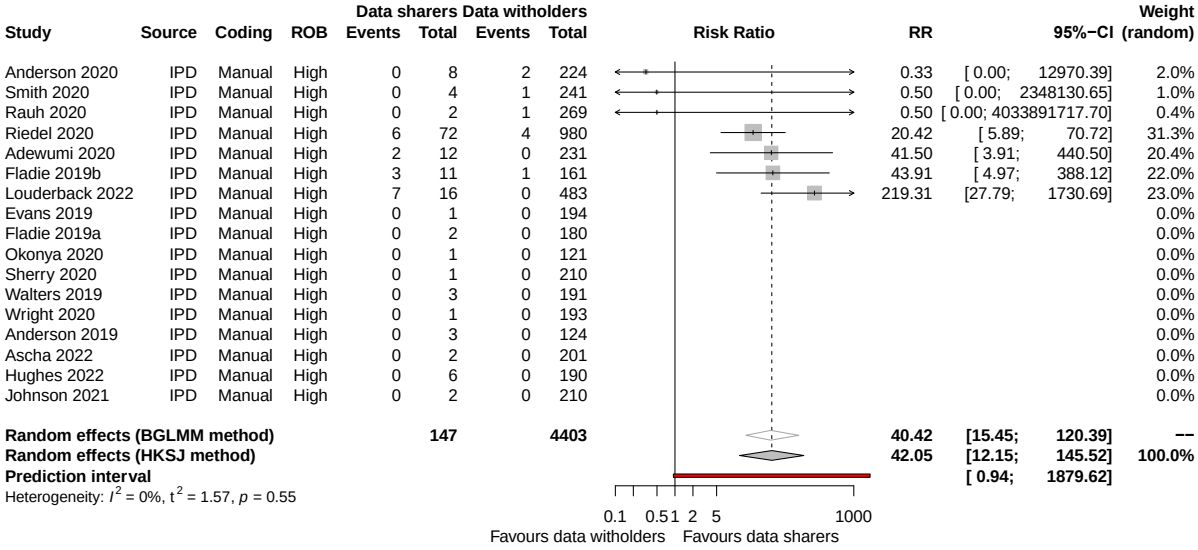


SUPPLEMENTARY FIGURE 6. PREVALENCE OF SUCCESSFUL RESPONSES TO PRIVATE REQUESTS FOR DATA FROM PUBLISHED MEDICAL RESEARCH BY DECLARATION TYPE.

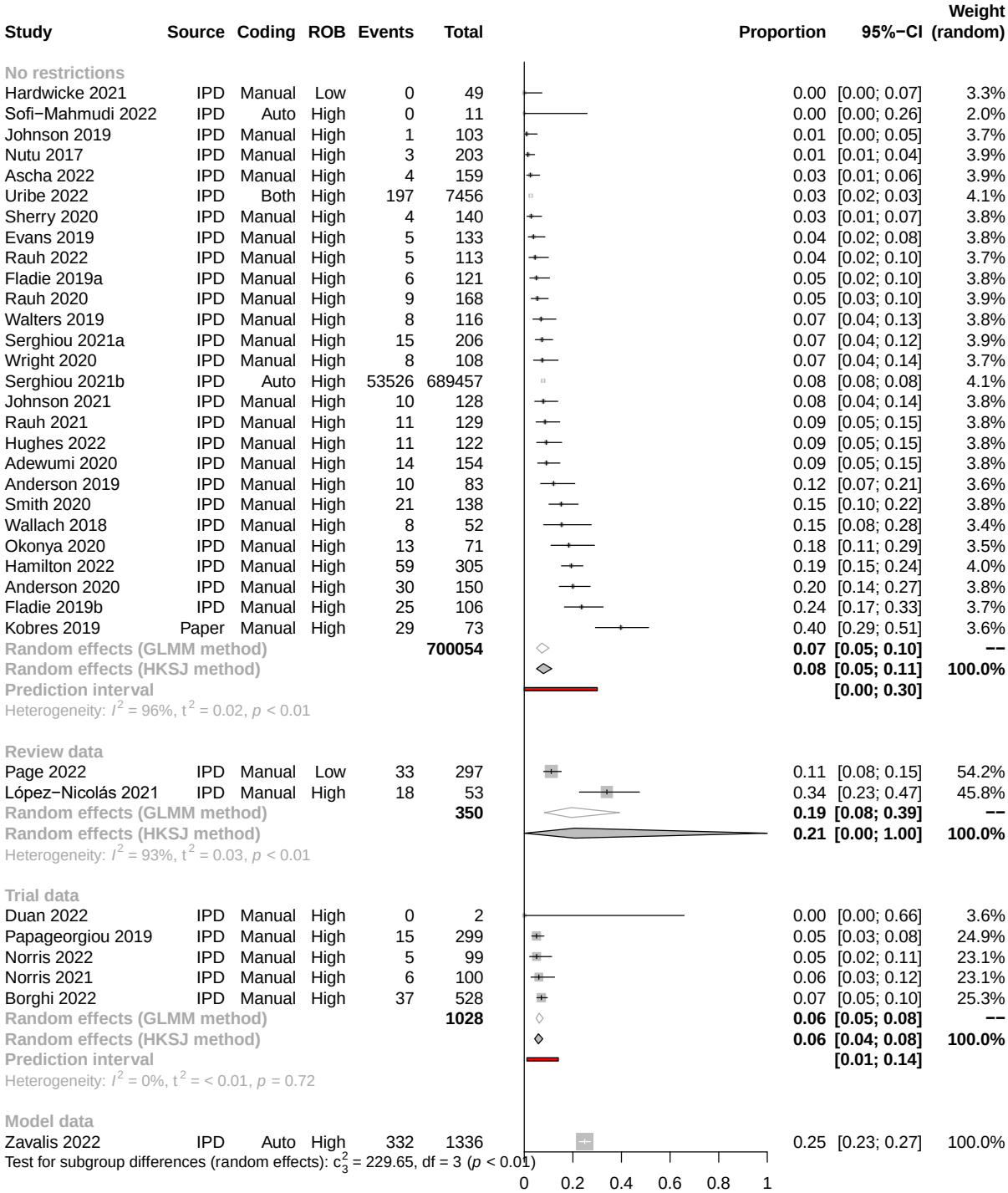
SUPPLEMENTARY FIGURE 7. PREVALENCE OF DECLARED AND ACTUAL PUBLIC CODE SHARING BY JOURNAL CODE SHARING POLICY.



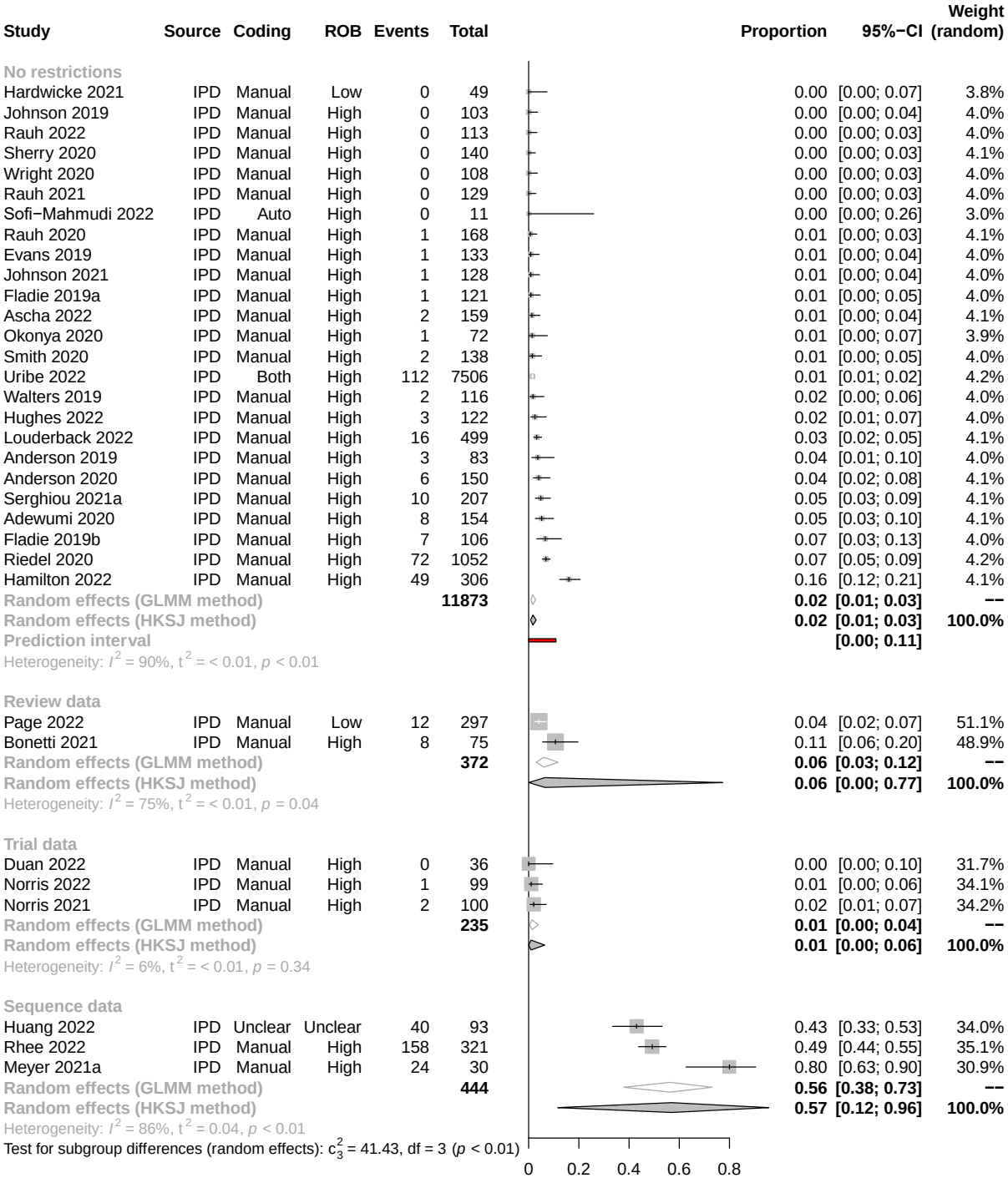
SUPPLEMENTARY FIGURE 8. ASSOCIATION BETWEEN DATA SHARING AND CODE SHARING (ACTUAL AVAILABILITY).



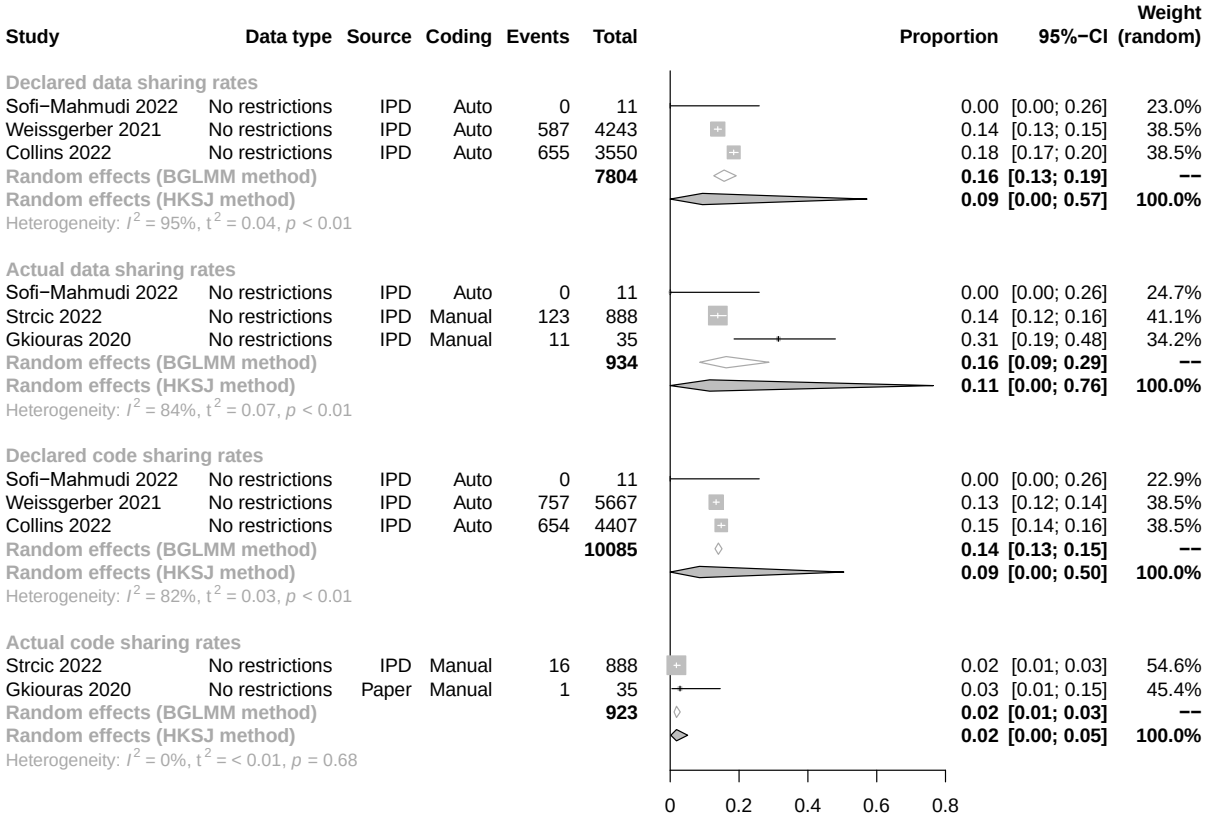
SUPPLEMENTARY FIGURE 9. PREVALENCE OF DECLARED PUBLIC DATA SHARING SINCE 2016 BY DATA TYPE.



SUPPLEMENTARY FIGURE 10. PREVALENCE OF ACTUAL PUBLIC DATA SHARING SINCE 2016 BY DATA TYPE.



SUPPLEMENTARY FIGURE 11. PREVALENCE OF PUBLIC DATA AND CODE SHARING AMONG STUDIES INVESTIGATING COVID-19.



SUPPLEMENTARY TABLE 1. DEVIATIONS FROM THE ORIGINAL REVIEW PROTOCOL.

Original plan	Revised plan	Reason for modification
Please refer to Table 1 in the review protocol (https://f1000research.com/articles/10-491/v2 - T1).	Two extra options were added to the 'risk of sampling bias' ("Sampled whole population of interest") and 'risk of article selection bias' categories ("No article screening performed"). Both items were considered low risk of bias.	The two options were added to account for meta-research articles that assessed all articles within a population of interest (e.g., assessed all articles indexed in PubMed Central a la Serghiou et al. (2021)) or did not perform article screening respectively.
Prevalence estimates will be transformed using the Freeman-Tukey double arcsine transformation and combined using standard inverse variance methods.	We pooled prevalence estimates by first stabilising the variances of the raw proportions using arcsine square root transformations, then applied random-effects models using the Hartung-Knapp-Sidik-Jonkman method.	Due to large sample size imbalances between included studies, and the negative influence such skewed ranges of sample sizes can have on the harmonic mean which is used to back-transform meta-analytic estimates transformed with the Freeman-Tukey double arcsine method (Schwarzer 2019), we decided to use the arcsine square root transformation instead.
We did not plan to conduct sensitivity analyses to investigate differences in pooled risk ratios when using generalized linear mixed models.	We examined differences in pooled risk ratios when using generalised linear mixed models to aggregate findings [32,33].	We decided post-hoc to check the robustness of the meta-analyses of risk ratios when using bivariate generalised linear mixed effects models as proposed by Chu et al. (2012) [33]. Like the meta-analyses of proportions, we chose this method as it has been specifically recommended in situations when the event of interest is rare, and individual study sample sizes are small and circumvent the need to add arbitrary continuity corrections allowing the analysis of both single-zero and double-zero events [27; 165].
We did not plan to collect information on data type, nor perform a subgroup analysis to explore its effects on the study's findings.	We collected data on data type and conducted a sub-group analysis to investigate whether prevalence estimates of public data sharing differed depending on the data type	We decided to collect data on data type and perform a subgroup analysis exploring its effects on reported findings based on feedback from colleagues on the protocol.

SUPPLEMENTARY TABLE 1 (CONTINUED). DEVIATIONS FROM THE ORIGINAL REVIEW PROTOCOL.

Original plan	Revised plan	Reason for modification
We did not plan to conduct sensitivity analyses to investigate differences in pooled prevalence estimates when excluding studies that used automated coding strategies.	We conducted sensitivity analyses to investigate differences in pooled prevalence estimates when excluding studies that used automated coding strategies.	We decided to include this sensitivity analysis to evaluate whether pooled prevalence estimates deviated when the results of meta-research studies which used automated coding strategies (methods have been shown to have inferior accuracy, positive predictive value (PPV), and negative predictive value (NPV) when compared to manual coding strategies) were removed.
We originally planned to perform subgroup analyses to explore differences in public data sharing frequencies between primary articles reporting the results of clinical trials or not, as well as articles reporting the results of studies using human participants versus not.	We conducted analyses of association for these outcomes rather than a subgroup analysis comparing pooled proportions between groups. Consequently, these two subgroup analyses have been included as secondary outcomes.	We changed the analysis plan for these outcomes due to the availability of data that allowed us to directly explore associations between both of these factors (i.e. calculation of risk ratios)
We originally planned to perform a subgroup analysis to explore differences in public data sharing frequencies between primary articles studying COVID-19 or not.	We conducted a sensitivity analysis to examine how data sharing frequencies changed when analyses were restricted to meta-research studies examining COVID-19.	We decided not to compare data and code sharing rates between COVID and non-COVID research because of the large amount of methodological heterogeneity in meta-research studies examining non-COVID research.

SUPPLEMENTARY TABLE 2. RISK OF BIAS CRITERIA.

Item	Low risk of bias	High risk of bias	Unclear risk of bias
Risk of sampling bias	The meta-research study evaluated a random sample of primary articles or sampled the entire population of interest.	The meta-research study included a non- or pseudorandom sample of primary articles.	The sampling frame for the sample of primary articles was unclear.
Risk of selective reporting bias	Eligible outcomes and associations reported in the protocol for the meta-research study were fully reported in the results section of the publication.	Not all eligible outcomes and associations reported in the protocol for the meta-research study were reported in the results section of the publication.	It was unclear if all eligible outcomes and associations were fully reported in the results section of the publication (e.g., because a study protocol for the meta-research study was unavailable).
Risk of article selection bias	Details about which studies were excluded from the study and why have been shared and match the criteria described in the methods, or no article screening needed to be performed (e.g., because all articles identified by a literature search were analysed)	Details about which studies were excluded and why were not reported.	Details about the eligibility criteria and study selection process was unclear.
Risk of errors in the accuracy of reported estimates	All outcome data were either manually coded by at least two people independently in parallel or coded by one person and checked in full by another.	Outcome data were manually coded by one researcher, an automated algorithm, or according to another methodology different from that outlined in the Low Risk category.	The method used to extract data from the included primary studies was unclear.

SUPPLEMENTARY TABLE 3. FINDINGS OF ELIGIBLE META-RESEARCH STUDIES WHERE SUMMARY DATA WERE NOT AVAILABLE FOR THE REVIEW (N=9).

Study	Year	Discipline	Journals examined	Primary study date range	Data types	Sample size	IPD available	Exclusion reason
Helliwell 2020b	2020	COVID-19, MERS	Multiple	2019-2020, 2018-2019	Any	398, 55	Partial	Reported prevalence estimates could not be coded in accordance with the study codebook
Hemkens 2016	2016	General Medical	Multiple	2012	Clinical data	124	No	Reported prevalence estimates could not be coded in accordance with the study codebook
Jiao 2022	2022	Multidisciplinary	PLOS One	2014-2020	Any	127,935	No	Prevalence estimates not reported separately for medical articles
McDonald 2017	2017	General Medical	BMJ	2015-2017	Clinical data	237	Partial	Reported prevalence estimates could not be coded in accordance with the study codebook
Ramke 2018	2018	Ophthalmology	Multiple	2000-2014	Clinical data	153	No	Prevalence estimates not reported
Rustici 2021	2021	Biomedicine	Multiple	2009-2013, 2012	RNA-Seq, Microarray	1,114, 347	No	Prevalence estimates not reported separately for medical articles
Stodden 2018	2018	Multidisciplinary	Science	2009-2010	Any	204	No	Reported prevalence estimates could not be coded in accordance with the study codebook and are also not reported separately for medical articles
Towse 202	2020	Clinical Psychology	Multiple	2014-2017	Any	1,900	Partial	Prevalence estimates not reported separately for medical articles
Zhao 2017	2017	Multidisciplinary	PLOS One	2014-2015	Any	50	No	Prevalence estimates not reported separately for medical articles

SUPPLEMENTARY TABLE 4. META-REGRESSION RESULTS.

	Model Coefficients						Level 3 (Between-study)			Level 2 (Within-study)				
	Intercept	SE	β*	SE	95% CI	*p*	τ^2	I^2	k	τ^2	I^2	o	AIC	BIC
Declared data sharing														
Three-level (All)	-32.9763	8.4334	0.0165	0.0042	0.0082-0.0248	0.0001	0.0122	90.67%	27	0.0012	9.06%	117	-166.64	-155.66
Three-level (Manual)	-32.8099	14.1371	0.0164	0.007	0.0025-0.0303	0.0213	0.012	55.63%	25	0.0037	17.15%	105	-127.34	-116.8
Actual data sharing														
Three-level (All)	-8.2121	9.0308	0.0041	0.0045	-0.0048-0.013	0.3589	0.0089	74.66%	26	0.0004	3.16%	111	-199.79	-189.02
Three-level (Manual)	-16.8115	11.1834	0.0084	0.0055	-0.0026-0.0194	0.1334	0.0088	62.28%	25	0.0006	4.01%	106	-180.64	-170.07
Declared code sharing														
Three-level (All)	-22.7095	1.7921	0.0113	0.0009	0.0095-0.013	<0.0001	0.0011	95.31	24	0	1.54%	114	-294.16	-283.28
Three-level (Manual)	-1.5145	10.6961	0.0008	0.0053	-0.0098-0.0113	0.8852	0.0014	18.54	22	0	0%	102	-236.2	-225.78
Actual code sharing														
Three-level (All)	-5.1997	10.5109	0.0026	0.0052	-0.0078-0.0129	0.6199	0.0009	15.15	21	0	0	99	-242.96	-232.66
Three-level (Manual)	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA

*Beta coefficients represent the arcsine-transformed change in availability rate each year, All = all eligible studies, Manual = studies that only manually assessed sampled primary articles

SUPPLEMENTARY TABLE 5. SENSITIVITY ANALYSES FOR PRIMARY OUTCOMES.

	Declared public sharing					Actual public sharing				
	%	95% CI	95% PI	k	I^2	%	95% CI	95% PI	k	I^2
Public data sharing										
Low ROB	-	-	-	-	-	-	-	-	-	-
Provided IPD	7%	5-10%	0-25%	26	96%	2%	1-3%	0-11%	25	90%
Assessed FAIR	5%	0-50%	NA	3	98%	4%	0-43%	NA	3	98%
Manually coded primary articles	9%	5-12%	0-31%	24	90%	2%	1-3%	0-12%	23	90%
COVID-19	9%	0-57%	NA	3	95%	11%	0-76%	NA	3	84%
HKSJ method	8%	5-11%	0-30%	27	96%	2%	1-3%	0-11%	25	90%
GLMM method	7%	5-10%	1-34%	27	95%	2%	1-3%	0-15%	25	91%
Public code sharing										
Low ROB	-	-	-	-	-	-	-	-	-	-
Provided IPD	0.2%	0-0.4%	0-2%	25	86%	0.1%	0-0.3%	0-1%	21	52%
Assessed FAIR	0.9%	0-12%	NA	3	93%	-	-	-	-	-
Manually coded primary articles	0.3%	0-1%	0-9%	23	82%	0.1%	0-0.3%	0-1%	21	52%
COVID-19	9%	0-50%	NA	3	82%	2%	1-3%	NA	2	0%
HKSJ method	0.3%	0-1%	0-8%	26	89%	0.1%	0-0.3%	0-1%	21	52%
GLMM method	0.2%	0-1%	0-12%	26	91%	0.2%	0-0.9%	0-3%	21	0%

CI – confidence interval, PI – prediction interval, k – number of included studies, HKSJ – Hartung-Knapp-Sidik-Jonkman method for random-effects meta-analysis, GLMM – generalised linear mixed-models, ROB – risk of bias, IPD – individual participant data

SUPPLEMENTARY TABLE 5 (CONTINUED). SENSITIVITY ANALYSES FOR PRIMARY OUTCOMES (CONTINUED).

	Declared public sharing					Actual public sharing				
	%	95% CI	95% PI	k	I^2	%	95% CI	95% PI	k	I^2
Private data sharing										
Low ROB	0%	0-2%	NA	1	NA	-	-	-	-	-
Provided IPD	2%	1-4%	0-10%	23	80%	-	-	-	-	-
Assessed FAIR	12%	9-16%	NA	1	NA	-	-	-	-	-
Manually coded primary articles	2%	1-4%	0-10%	23	80%	-	-	-	-	-
HKSJ method	2%	1-4%	0-10%	23	80%	-	-	-	-	-
GLMM method	2%	2-4%	0-12%	23	67%	-	-	-	-	-
Private code sharing										
Low ROB	0%	0-2%	NA	1	NA	-	-	-	-	-
Provided IPD	0%	0-0.1%	0-0.5%	22	0%	-	-	-	-	-
Assessed FAIR	0.7%	0-2%	NA	1	NA	-	-	-	-	-
Manually coded primary articles	0%	0-0.1%	0-0.5%	22	0%	-	-	-	-	-
HKSJ method	0%	0-0.1%	0-0.5%	22	0%	-	-	-	-	-
GLMM method	0.1%	0-0.4%	0-0.5%	22	0%	-	-	-	-	-

CI – confidence interval, PI – prediction interval, k – number of included studies, HKSJ – Hartung-Knapp-Sidik-Jonkman method for random-effects meta-analysis, GLMM – generalised linear mixed-models, ROB – risk of bias, IPD – individual participant data

SUPPLEMENTARY TABLE 6. SENSITIVITY ANALYSES FOR SECONDARY OUTCOMES AND SUBGROUP ANALYSES.

	Declared public sharing					Actual public sharing				
	RR	**95% CI**	**95% PI**	**k**	**I^2**	**RR**	**95% CI**	**95% PI**	**k**	**I^2**
Association between data and code sharing										
HKSJ method	8.03	2.86-22.53	0.33-194.43	12	32%	42.05	12.15-145.52	0.94-1879.62	7	0%
BGLMM method (SZC)	7.88	2.44-18.01	NA	12	NA	40.42	15.45-120.39	NA	7	NA
BGLMM method (DZC)	10.51	3.00-18.01	NA	23	NA	52.85	9.46-132.52	NA	17	NA
Low ROB	-	-	-	-	-	-	-	-	-	-
FAIR studies	11.84	0-1.33x10^7	NA	2	82%	-	-	-	-	-
IPD only	-	-	-	-	-	-	-	-	-	-
Manual coding	4.52	1.38-14.86	0.20-101.05	10	0%	-	-	-	-	-
Trial versus non-trial										
HKSJ method	0.69	0.45-1.07	0.12-4.13	23	0%	0.96	0.53-1.72	0.15-5.95	19	0%
BGLMM method (SZC)	0.55	0.35-0.77	NA	23	NA	0.67	0.26-1.39	NA	19	NA
BGLMM method (DZC)	0.56	0.37-0.79	NA	25	NA	0.69	0.27-1.52	NA	24	NA
Human versus non-human										
HKSJ method	0.65	0.42-0.99	0.12-3.61	19	57%	0.44	0.24-0.81	0.05-3.57	16	28%
BGLMM method (SZC)	0.69	0.46-1.01	NA	19	NA	0.58	0.29-1.00	NA	16	NA
BGLMM method (DZC)	0.69	0.48-1.00	NA	20	NA	0.59	0.30-0.97	NA	20	NA

SZC – Included studies with no events in one group but not studies with no events in both groups in analyses, DZC – Included both studies with no events in one or both groups in analyses

SUPPLEMENTARY METHODS.

Protocol and registration

We registered our systematic review on May 28th, 2021 on the Open Science Framework (OSF), prior to commencing the literature search [1], and subsequently prepared a detailed review protocol [2]. We report seven deviations from the protocol in Supplementary Table 1. As the research subjects of interest were scientific publications, ethics approval was not required for this research. The findings of this review are reported in accordance with the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) 2020 statement [3] and its IPD extension [4]. We summarise key aspects of the methods below; for further details, please refer to the review protocol [2].

Eligibility criteria

Any study in which 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") was eligible for inclusion in the review. No restrictions were placed on the publication location (e.g., preprint server, peer-reviewed journal) or the format (e.g., conference abstract, research letter) of either group. Nor were restrictions placed on the strategy used to identify and select primary articles, the type of data assessed (e.g., trial data, review data) or the level of sharing assessed (e.g., partial versus complete sharing). Furthermore, we included studies that used either manual or automated methods to assess data and code sharing provided it involved some examination of the body text of 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 IPD meta-analysis; among scientific articles outside of medicine and health; or via avenues other than journal articles (e.g., clinical trial registries).

Information sources and search strategy

On July 1st, 2021, we searched Ovid MEDLINE, Ovid Embase, and the medRxiv, bioRxiv, and MetaArXiv preprint servers to identify potentially relevant studies indexed from database inception up to the search date. The full search strategies, bibliographic citation files, as well as snapshots of the medRxiv and bioRxiv databases are available on the project's OSF page (https://osf.io/jgzsa/). Details on the development of the search strategy are outlined in the review protocol [2]. In addition to the database searches, other preprint servers (PeerJ, Research Square) and relevant online resources (Open Science Framework, aspredicted.org and connectedpapers.com) were searched to locate additional 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 using citationchaser on August 30th, 2022 [5]. Finally, potentially relevant studies recommended by colleagues, discovered through collaborations, and seen at meta-research conferences were also screened for eligibility. No language restrictions were imposed on any of the searches.

Study selection

Results from all main database and preprint server searches were imported into Covidence (Covidence systematic review software, Veritas Health Innovation, Melbourne, Australia) and deduplicated. For the preprint searches, if a version of an eligible meta-research study was discovered in a peer-reviewed journal, it was included in place of the original preprint. All titles, abstracts, and full-text articles were then screened for eligibility in Covidence by DGH and another author (HF, ARF, or KH) independently, with disagreements resolved via discussion between authors, or by a third author if necessary (MJP). All literature identified by the additional preprint and online searches were screened against the eligibility criteria by one author (DGH). When multiple reports on the same dataset were identified, we

used data from the most up-to-date report. A spreadsheet containing all screening decisions is publicly available on the OSF (https://osf.io/6tj87).

Data collection

Once a meta-research study was found to be eligible, one member of the team (DGH) determined whether sufficiently unprocessed article-level IPD and article identifiers (e.g., digital object identifiers (DOIs), PubMed identifiers (PMIDs), article titles) for the included primary articles were publicly available. For meta-research studies where complete IPD were not available (i.e., no data or partial data had been shared), the corresponding author was contacted and asked if they would provide the complete or remaining IPD. If meta-research authors responded that they were either unable or unwilling to share, we then asked whether they would calculate the summary statistics necessary for the review. For meta-research authors who were unable or refused to provide summary data for the review, did not respond, or did not provide the promised IPD by the census date of December 31st, 2022, summary data reported in the meta-research papers were independently extracted by two authors (DGH; MJP), with discrepancies resolved through discussion. A list of all the data that were extracted from each meta-research study for the review can be found on the project's OSF page (https://osf.io/mav89/).

IPD integrity checks and harmonisation

When complete IPD were obtained for a meta-research study, one author (DGH) performed the following integrity checks prior to harmonising the data: i) an evaluation of the completeness of the dataset (e.g., whether any variables or values were missing), ii) a check of the validity of the dataset (e.g., presence of out-of-range values, incorrectly coded values) and iii) a check that the overall sample size and data and/or code sharing prevalence estimates as stated in the report could be exactly reproduced (note that the checks for an included study led by the first author of this review (Hamilton et al 2022 [6]) were performed by another author (HF)). In instances where any of these checks failed, clarification was sought from the meta-research authors. We also checked for, and removed duplicate rows in datasets (i.e., checked if the same primary articles were sampled more than once). Additionally, for meta-research studies that sampled primary articles across multiple scientific disciplines, Digital Science's Dimensions platform (https://app.dimensions.ai) was used to identify which were medical and health-related using their automated 2020 Australia and New Zealand Standard Research Classification (ANZSRC) Fields of Research (FOR) Codes classification service [7]. When primary articles were not indexed in Dimensions, the first author (DGH), who has close to a decade of experience working as an allied health professional, clinical trial coordinator and medical researcher, classified articles as being medical or health-related or not. Furthermore, for meta-research studies with sample sizes less than 500, primary articles not assigned medical FOR codes by the Dimensions platform were manually reviewed and recoded if deemed false negatives.

Once the IPD checks were complete, one author (DGH) then manually extracted and reclassified required data in line with the study's codebook. When all available IPD had been assembled and harmonised, datasets were then merged and the extent of overlapping primary articles between meta-research studies was assessed for each outcome of interest by checking for duplicate DOIs and PMIDs in R (R Foundation for Statistical Computing, Vienna, Austria, v4.2.1) using the duplicated function. We decided to keep data originating from primary articles that were flagged as having been sampled by more than one meta-research study only for the study with the highest score for the fourth risk of bias domain (i.e., lowest risk of errors in the accuracy of reported estimates), or in the event of a tie, the overall lowest risk of bias judgement, or the most recent publication date. More details on the scoring system developed to resolve overlap can be found within the spreadsheet reporting the results of the risk of bias assessments (https://osf.io/a59vj). For eligible meta-research studies where summary data

were only available from study reports, but primary study identifiers were known, information from overlapping primary articles was removed from the meta-research studies that shared complete IPD. For meta-research studies where both primary study identifiers and article-level data were unavailable, we assessed the likelihood of overlap with other meta-research studies by comparing: i) outcome data collected, ii) primary article date range and iii) sampled journals.

Outcomes of interest

The following four pre-specified outcome measures for both research data and code availability were of primary interest to the review:

i) the prevalence of primary articles where authors declared that their data or code are publicly available ('declared public availability');

ii) the prevalence of primary articles in which meta-researchers verified that data or code were indeed publicly available ('actual public availability');

iii) the prevalence of primary articles where authors declared their data or code are privately available (i.e., "available on request" statements) ('declared private availability'), and;

iv) the prevalence of primary articles in which meta-researchers confirmed 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 (e.g., checks that reported URLs were functional, that data could be freely downloaded and opened, that datasets were complete, that reported results could be independently reproduced). We also required data to be immediately available for it to be classified as actually publicly available (i.e., did not accept 'intention to share' and 'under embargo' statements), and took the strictest definition of actual availability when alternatives were available (i.e., if a study assessed both partial and complete sharing, we took the results of the 'full' data availability). Further information on how we defined 'actual availability' as well as all our other outcome measures can be found in the review protocol and the study codebook on the project's OSF page (https://osf.io/u3yrp/).

In addition to the primary outcome measures, we also included eight secondary outcome measures:

i) the prevalence of formalised sections within primary articles dedicated to addressing data and/or code availability;

ii) the association between the presence of a data availability statement and public sharing of data in primary articles;

iii) the association between the presence of a code availability statement and public sharing of research code in primary articles;

iv) the association between a journal's policy on data sharing (any 'mandatory posting' policy versus other policy) and public sharing of research data in primary articles;

v) the association between a journal's policy on data sharing ('make available on request' policy versus other non-mandatory policy) and private sharing of research data in primary articles;

vi) the association between study design (clinical trial versus non-trial) and public sharing of data in primary articles;

vii) the association between the subjects of the research (human participants versus non-human participants) and public sharing of data in primary articles, and;

viii) the association between public sharing of research data and the sharing of code in primary articles.

Assessments of risk of bias

The risk of bias of included meta-research studies was assessed using a tool designed based on methods used in previous Cochrane Methodology reviews [8, 9]. The tool included four domains: i) sampling bias, ii) selective reporting bias, iii) article selection bias, and iv) the 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 ARF), with discrepancies resolved via discussion, or a third author (MJP) if necessary. Where domains were rated as unclear, clarification was sought from meta-research authors. Given the purpose of the tool was to differentiate between studies at a high risk of bias from those with a low risk, a study was only classified as low risk of bias if all criteria were assessed as low risk. We did not assess the likelihood of publication bias affecting the findings of the review (e.g., using a funnel plot), nor did we assess certainty in the body of evidence, as available methods are not well suited for methodology reviews such as ours.

Statistical analysis

A 'two-stage' approach to IPD meta-analysis was used, whereby summary statistics were computed from available IPD, abstracted from included study reports, or obtained directly from meta-research authors, then pooled using conventional meta-analysis techniques. We calculated proportions and 95% confidence intervals (CI) for all prevalence outcomes. Where possible, we calculated risk ratios with 95% confidence intervals for all association outcomes. For primary outcome measures, we considered the methodological characteristics of the included studies to determine which were appropriate for aggregation and decided that we would pool studies that met the following criteria: i) did not use non-random sampling methods, ii) did not restrict primary article evaluations to specific journals, preprint servers, funders, institutions, or data types, and iii) reported outcome data on primary articles published after 2016. These criteria were specifically chosen to minimise biasing of estimates (i.e., reduce upward or downward biasing of pooled estimates due to the overrepresentation of studies of journals with mandatory sharing policies, certain study designs, etc), and to provide a modern picture of data and code sharing (i.e., an estimate of sharing since the introduction of the FAIR principles [10]). The same criteria were applied to secondary outcome measures and subgroup analyses unless specified otherwise.

We pooled prevalence estimates by first stabilising the variances of the raw proportions using arcsine square root transformations, then applied random-effects models using the Hartung-Knapp-Sidik-Jonkman method which has shown to be preferable to the DerSimonian and Laird method when including a small number of studies, and when including studies with differing sample sizes [11]. 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 (2004) [12] was applied to studies reporting zero events in a single group (double zero-cell events were excluded from the main analysis). Statistical heterogeneity was assessed via visual inspection of forest plots, the size of the I^2 statistics and their 95% confidence intervals, and via 95% prediction intervals (PI) where more than four studies were included. Data deduplication, preparation, analysis and visualisation was performed in R (R Foundation for Statistical Computing, Vienna, Austria, v4.2.1) using the meta (v5.5) [13], metafor (v3.8) [14] and altmeta (v4.1) [15] packages. Risk of bias plots were created using robvis [16]. The Python (v3.10.7) client Dimcli (v0.9.9.1) was used to access Dimensions Analytic's API and retrieve required primary article meta-data (e.g., DOIs, PMIDs, ANZSRC FOR codes). All R and Python scripts are publicly available on the project's OSF page (https://osf.io/p5ehb/).

Subgroup and sensitivity analyses

We planned to conduct the following subgroup analyses to investigate whether prevalence estimates of public data sharing differed depending on i) the data type, or whether primary articles: ii) were subject to any mandatory sharing policies by the funders of the research or not, or iii) posted a preprint prior to

publication or not. Furthermore, we also investigated the influence of publication year on data and code sharing rates by fitting three-level mixed-effects meta-regression models on arcsine-transformed proportions. A multi-level model was used to account for dependencies between effect estimates due to some studies contributing multiple yearly estimates. Due to substantially differing levels of variation of point estimates prior to 2014 and after 2020, to preserve assumptions of homoscedasticity we only modelled changes in sharing rates between 2014 and 2022.

We also performed sensitivity analyses to assess changes in pooled estimates when excluding meta-research studies that i) were rated as high or unclear risk of bias, ii) did not provide IPD for the review, iii) were at high risk of overlap with other meta-research studies, iv) did not assess compliance with the FAIR principles, v) did not manually assess primary articles and vi) did not examine COVID-19-related research. Finally, we also examined differences in pooled proportions and risk ratios when using generalised linear mixed models (GLMMs) to aggregate findings, which have been specifically recommended in situations when the probability of the event of interest is rare [17,18]. Such methods also circumvent the need to add arbitrary continuity corrections to zero events, which can produce biased results when most cases are zero events, and group sample sizes are highly imbalanced [12]. For meta-analyses of risk ratios, we report the results of analyses both excluding and including studies with no events in both groups.

SUPPLEMENTARY RESULTS.

Study selection and IPD retrieval

The search of Ovid MEDLINE, Ovid Embase and the medRxiv, bioRxiv and MetaArXiv preprint servers, once deduplicated, identified 4,952 potentially eligible articles for the review, of which 4,736 were excluded following the screening of titles and abstracts. Of the remaining 216 articles, full-text articles were retrieved for all papers, and 70 were adjudicated as eligible for the review. Furthermore, the additional searches revealed another 44 eligible reports for inclusion, resulting in a total of 114 eligible meta-research studies examining a combined total of 2,254,031 primary articles for the review [6, 19-134].

Following confirmation of eligibility, we searched for publicly available IPD for the 114 meta-research studies. Of these studies, 70 had already made complete IPD publicly available (61%), 20 studies had posted partial IPD (18%), and 24 had not publicly shared any IPD (21%), with three of the latter articles declaring upfront that IPD could not be shared. Of the 70 complete datasets that were originally posted publicly, 60 (86%) were deposited into data repositories, 36 (51%) had a DOI, 26 (37%) provided a data dictionary, and 14 (20%) applied a license to the data. Most data were archived in Microsoft Excel (N=33, 47%) or CSV (N=25, 36%) formats, with a minority of meta-researchers storing their data in PDFs (N=5, 7%) and Microsoft Word documents (N=3, 4%).

Of the 44 meta-research studies that had not posted complete IPD, three groups stated in the study report that data could not be released, and for one we were unable to source contact information. Consequently, we contacted 40 authorship teams and asked them to share article-level IPD for the review. We received 32 responses to our 40 requests (80%), of which 20 meta-researchers (50%) shared the required IPD by the census date. The median time taken to receive IPD was 7 days (range: 0-216 days). For the 20 articles where complete IPD was not assembled, 10 studies had useable IPD and/or summary data. The nine studies that were eligible for the review but could not be included in the quantitative analysis are outlined in Supplementary Table 3. They are also included in relevant forest plots, without providing usable data for the meta-analysis. Ultimately, 108 reports of 105 meta-research studies collecting information from a total of 2,121,580 primary articles were included in the quantitative analysis [6, 19-125], with complete IPD available for 90 studies, a combination of partial IPD and summary data for 10 studies, and only summary data available for 5 studies. Refer to Figure 1 for the full PRISMA flow diagram.

IPD integrity checks

In total, 100 meta-researchers' datasets (90 complete and 10 partial) were obtained for the review. For the 90 complete datasets, sample sizes, as well as data and/or code sharing rates reported in study reports, were reproduced in all but five cases (94%), with the reasons for irreproducibility being due to simple typographical errors in the report (N=2), unclear data filtering steps (N=2) and an error in the meta-researchers' code (N=1). For the ten partial datasets, we were able to independently verify sample sizes and sharing estimates for all but one case due to the receipt of an incorrect version of the data.

Of the 105 included meta-research studies examining 2,121,580 primary articles, we were able to retrieve identifying details (i.e., DOIs, PMIDs) for 2,121,197 primary articles (99.98%) from 100 studies (95%). After the removal of non-medical articles and duplicate articles observed within each of the 100 datasets, we were left with 1,849,828 primary articles with which to explore the extent of overlap between eligible studies. Of these 1,849,828 primary articles, we observed that 704,310 (38%) were flagged as having been sampled by more than one included meta-research study (some articles being repeatedly sampled by up to five studies). Notably, articles examined by the three largest studies

by Serghiou et al [25], Colavizza et al [35] and Federer et al [42] were implicated in 681,595 of the 704,310 flagged cases (96.77%). Further, for some studies, all sampled primary articles had been completely assessed by other included studies (e.g., Sumner et al [113], Strcic et al [114]), whereas others demonstrated zero overlap (e.g. Rufiange et al [20]) (see Supplementary Figure 3 for further details).

For the five meta-research studies where identifying details for the primary articles were unavailable, only a single study was deemed to be at high risk of overlap [65]. Furthermore, for the nine meta-research studies excluded from the quantitative analysis, 127,985 of the 132,451 observations (97%) would have come from two meta-research studies of articles published in PLOS One, which would have had a high risk of overlap with the included studies by Serghiou et al [25], Colavizza et al [35] and Federer et al [42]. Given the likelihood of high overlap, our inability to include these nine meta-research studies in the quantitative analyses is unlikely to have influenced our results.

Study characteristics

Summary information on the 105 meta-research studies that are included in the quantitative analysis of this review is outlined in Table 1. Eligible meta-research studies examined a median of 195 primary articles (IQR: 113-475; sample size range: 10-1,475,401), with a median publication year of 2015 (IQR: 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%), addiction medicine, clinical psychology, and oncology (each N=5, 5%). Eleven studies (10%) examined COVID-19-related articles.

Most meta-research studies did not set any restrictions concerning data types (N=63) or journals of interest (N=56). However, when data restrictions were imposed, they were most often limited to trial data (N=16), sequence data (N=6), gene expression data and review data (each N=5). When journal restrictions were incorporated, the scope was most often limited to papers published in 'high impact' journals (variably defined by authors) (N=18), one or two journals of interest (N=10 and 5 respectively), or multiple journals subjectively deemed relevant to a field (N=7). Of the 105 meta-research studies, 31 and 4 also evaluated compliance with journal data and code sharing policies, respectively. However, 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 how compliant publicly shared data was with the FAIR principles. In contrast, 10, 4 and 2 studies, respectively, assessed whether study data, code, or both data and code could be retrieved in response to a private request (i.e., actual private availability). Of these 16 studies, the stated reasons underpinning requests were: to perform a re-analysis (N=6), for a meta-research study (N=5), to populate a registry (N=1), to validate their findings (N=1) and for interest and coursework (N=1), with the remaining two not reporting what reason they gave. Of the 14 meta-research articles that shared the request templates they used, 12 meta-researchers provided primary article authors with an honest account of why they wished to source data and/or code, whereas two used deception.

Risk of bias assessment

The overall and individual results of the risk of bias assessments are reported in Supplementary Figures 1 & 2. 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, a minority of meta-

research studies were judged to be at low risk of selective reporting bias (N=45, 42%) and article selection bias (N=24, 23%) (i.e., shared study protocols and information on which primary articles were excluded and why). Similarly, only half of meta-research studies (N=54, 51%) were judged to have used a primary article coding strategy considered to be at low risk of errors. Ultimately, only eight studies (8%) were classified as low risk of bias for all four domains.

Public data and code sharing estimates

Combination of eligible studies in a random-effects meta-analysis suggests that 8% of medical articles published since 2016 declare data to be publicly available (95% CI: 5-11%, 95% PI: 0-30%, k = 27 studies, o = 700,054 primary articles, I^2 = 96%; Figure 2) and 2% actually share data publicly (95% CI: 1-3%, 95% PI: 0-11%, k = 25, o = 11,873, I^2 = 90%; Figure 3). Despite the included meta-research studies following similar methodologies, we do note high I^2 values for both analyses, with influence analyses showing that the greatest contributors to between-study heterogeneity for declared data sharing were the high precision findings of Uribe et al [117] and Serghiou et al [25], who used automated coding strategies. For actual data sharing, the high estimate by Hamilton et al [6], who assessed partial sharing of data rather than complete, was the largest contributor to between-study heterogeneity. However, when leaving these studies out of the meta-analyses I^2 values did not drop meaningfully. Ultimately, given the consistency of the estimates, and the narrow width of the prediction intervals, and the low impact leave-one-analyses had on the reported I^2 values, we do not believe this indicates concerning levels of variability.

For public code sharing, declared and actual code sharing prevalence estimates since 2016 are estimated to be 0.3% (95% CI: 0-1%, 95% PI: 0-8%, k = 26, o = 707,943, I^2 = 89%; Figure 4) and 0.1% (95% CI: 0-0.3%, 95% PI: 0-1%, k = 21, o = 3,843, I^2 = 52%; Figure 5), respectively. Like declared data sharing estimates, despite similar methodologies, declared code sharing estimates were also associated with high I^2 values. Again, influence analyses revealed high precision estimates from Uribe et al [117] and Serghiou et al [25], in addition to the high estimate by Kobres et al [70], who evaluated the sharing of model code from Zika virus forecasting and prediction research, were the biggest contributors to between-study heterogeneity. However, for the same reasons reported above we do not believe this indicates concerning levels of variability.

Private data and code sharing estimates

In contrast to declarations of public availability, 'available upon request' declarations were not commonly observed in primary articles published since 2016 for data (2%, 95% CI: 1-4%, 95% PI: 0-10%, k = 23, o = 3,058, I^2 = 80%) or code (0%, 95% CI: 0-0.1%, 95% PI: 0-0.5%, k = 22, o = 2,825, I^2 = 0%) (refer to Supplementary Figures 4 & 5 for forest plots). For actual private data and code availability prevalence estimates, we could not combine the findings of eligible meta-research studies due to methodological differences, particularly in journal restrictions (i.e., policy differences), as well as the type of data being requested, both of which are explored via subgroup analyses below.

Overall, we observed that the prevalence of success in privately obtaining data and code from authors of published medical research ranged between 0-37% (k = 12, I^2 = 88%) and 0-23% (k = 5, I^2 = 94%) respectively (Figure 6). However, we note that when authors who declared data and code to be 'available on request' were asked for these products by meta-researchers, the upper limits of success increased to 100% (k = 7, I^2 = 83%) and 43% (k = 4, I^2 = 86%) respectively. In comparison, when requests for data and code were made to authors who did not include a statement concerning availability, the prevalence of success dropped to between 0-30% (k=7, I^2 = 65%) and 0-12% (k=3, I^2 = 89%) respectively. Lastly, and unsurprisingly, we also note that attempts to obtain data from authors explicitly declaring it to be unavailable were associated with a 0% sharing prevalence estimate (k = 2, I^2 = 0%).

See Supplementary Figure 6 for the full results. Interestingly, we also noted during the IPD deduplication process that two of four primary article authors who were asked to share data by two independent meta-research teams on two separate occasions responded differently, providing some anecdotal evidence that requestor and requestee characteristics likely also play a role in success.

Secondary outcomes

Insufficient data were available to evaluate the first three secondary outcome measures (i.e., outcomes concerning data and code availability statements), due to only a single study recording information about both the prevalence of statements and journal policies across a random sample of articles [6]. Similarly, very few meta-research studies recorded information on compliance with multiple data sharing policies across random samples of primary articles. This review was therefore also unable to evaluate the fourth and fifth secondary outcomes measures (i.e., direct comparison of mandatory and 'share on request' policies with non-mandatory data sharing policies).

However, for journals implementing mandatory data sharing policies, we estimate that 65% of primary articles (95% CI: 36-88%, 95% PI: 2-100%, k = 5, o = 28,499, I^2 = 99%) declared data to be publicly available and 33% actually shared data (95% CI: 5-69%, k = 3, o = 429, I^2 = 93%). In contrast, we estimate the prevalence of success in retrieving data from authors subject to 'share on request' policies to be 21% (95% CI: 4-47%, k = 3, o = 166, I^2 = 30%). For comparison, the prevalence of declared and actual data sharing under 'encourage' systems are estimated to be 17% (95% CI: 0-62%, k = 6, o = 1,010, I^2 = 98%) and 8% (95% CI: 0-48%, k = 3, o = 284, I^2 = 90%) respectively. Similarly, the prevalence of declared and actual sharing for articles published in journals with no sharing policy are estimated to be 17% (95% CI: 0-59%, k = 4, o = 686, I^2 = 95%) and 4% (95% CI: 0-95%, k = 2, o = 198, I^2 = 83%) respectively. Refer to Supplementary Figure 7 for the prevalence of declared and actual public code sharing according to journal policies.

We were able to assess the last three secondary outcomes. Our data suggest that triallists are 31% less likely to declare data are publicly available in comparison to non-triallists (RR: 0.69, 95% CI: 0.45-1.07, 95% PI: 0.12-4.13, k = 23, I^2 = 0%). However, when examining actual data sharing, neither group appears more or less likely to share their data than the other (RR: 0.96, 95% CI: 0.53-1.72, 95% PI: 0.15-5.95, k = 19, I^2 = 0%) (see Figure 7). We also estimate that researchers using data derived from human participants are also 35% less likely to declare data to be publicly available than researchers working with non-human participants (RR: 0.65, 95% CI: 0.42-0.99, 95% PI: 0.12-3.61, k = 19, I^2 = 57%). However, this decreased likelihood became more pronounced when examining the prevalence of actual data sharing (RR: 0.44, 95% CI: 0.24-0.81, 95% PI: 0.05-3.57, k = 16, I^2 = 28%) (see Figure 8). Lastly, we estimate that researchers who declare that their data are publicly available are eight times more likely to declare code to be available also (RR: 8.03, 95% CI: 2.86-22.53, 95% PI: 0.33-194.43, k = 12, I^2 = 32%). Additionally, researchers who are verified to have made data available are estimated to be 42 times more likely than researchers who withheld data to share code as well (RR: 42.05, 95% CI: 12.15-145.52, 95% PI: 0.94-1879.62, k = 7, I^2 = 0%) (Supplementary Figure 8).

Subgroup analyses

Insufficient data were available to evaluate whether prevalence estimates of public data sharing differed depending on whether primary articles were subject to any mandatory sharing policies by the funders of the research or posted as a preprint prior to publication. However, we did observe that the prevalence of both declared and actual public data sharing significantly differed according to the data type, with the highest prevalence of actual data sharing occurring among authors working with sequence data (57%, 95% CI: 12-96%, k = 3, o = 444, I^2 = 86%), review data (6%, 95% CI: 0-77%, k = 2, o = 372, I^2 = 75%) then trial data (1%, 95% CI: 0-6%, k = 3, o = 235, I^2 = 6%) (Supplementary Figures 9 & 10).

Additionally, we also observed substantial differences in compliance with journal policies depending on the data type (Table 2). For example, estimates from a single study by Page et al [89] showed that the prevalence of actual data sharing among systematic review authors decreased from 28% for mandatory sharing policies, to 1% and 0% for encourage and no policy systems, respectively. Whereas in the context of sequence and gene expression data, decreases in the prevalence of actual sharing between mandatory policies (67% and 43%), encourage policies (57% and 43%) and no policy (46% and NA) were much less apparent.

Finally, changes in the prevalence of public data and code sharing over time were investigated by fitting three-level mixed-effect meta-regression models to arcsine-transformed data (refer to Supplementary Table 4 for the full results). Publication year was found to be a significant moderator of declared data sharing prevalence estimates (β=0.017, 95% CI: 0.008-0.025, p=0.0001, between-study I^2 = 91%, within-study I^2 = 9%) but not actual data sharing prevalence estimates (β = 0.004, 95% CI: -0.005-0.013, p = 0.3589, between-study I^2 = 75%, within-study I^2 = 3%). Specifically, we note an estimated rise in the prevalence of declared data sharing from 4% in 2014 (95% CI: 2-6%, 95% PI: 0-18%) to 9% in 2020 (95% CI: 6-12%, 95% PI: 0-26%). Refer to Figure 9 for a bubble plot comparing declared and actual data sharing prevalence estimates over time. Comparatively, both declared and actual code sharing prevalence estimates did not appear to have meaningfully increased over time.

Sensitivity analyses

The results of the sensitivity analyses of the primary outcomes are reported in Table 3. For public data and code sharing outcomes, meta-analysis of prevalence estimates using GLMMs did not result in any substantial changes to combined estimates in comparison to the standard inverse-variance aggregation methods. Similarly, limiting analyses to meta-research studies in which authors manually coded articles (i.e., removal of meta-research studies that used automated or unclear coding methods) did not result in any meaningful changes. When limiting analyses to meta-research studies where summary data were only derived from available IPD, no changes were observed to the declared data availability analysis. Insufficient data were available to evaluate whether findings from meta-research studies that assessed compliance with FAIR or were classified as low risk of bias resulted in meaningful changes to pooled estimates. Similarly, with respect to the impact of overlapping primary articles, removing the only meta-research study that was deemed to be at risk of overlapping with other included meta-research studies had no impact on any of the analyses. Lastly, we estimate declared and actual public data sharing prevalence estimates for studies investigating COVID-19 (including both preprints or peer-reviewed publications) to be 9% (95% CI: 0-57%, k=3, o = 7,804, I^2 = 95%) and 11% (95% CI: 0-76%, k=3, o = 934, I^2 = 84%) respectively. Both of which compare favourably to our best estimates for declared (8%) and actual data sharing (2%) since 2016. The findings of the sensitivity analyses of secondary outcomes and subgroup analyses are reported in Supplementary Table 5. Most notably, we observed stronger associations between data and code sharing when including studies with no events in both groups.

SUPPLEMENTARY REFERENCES.

1. 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, May 28, 2021. DOI: 10.17605/OSF.IO/7SX8U.
2. Hamilton DG, Fraser H, Fidler F, McDonald S, Rowhani-Farid A, Hong K & Page MJ. 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]. F1000Research 2021, 10: 491. DOI: 10.12688/f1000research.53874.2.
3. Page MJ, McKenzie JE, Bossuyt PM, et al.: The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. Syst Rev. 2021; 10(1): 89.
4. Stewart LA, Clarke M, Rovers M, et al.: Preferred Reporting Items for Systematic Review and Meta-Analyses of individual participant data: the PRISMA-IPD Statement. JAMA. 2015; 313(16): 1657–1665.
5. 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.
6. 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 Medicine. 2022; 20. DOI: 10.1186/s12916-022-02644-2.
7. Porter SJ & Hook DW. Recategorising research: Mapping from FoR 2008 to FoR 2020 in Dimensions. arXiv. 2022: 2209.00104v1. DOI: 10.48550/arXiv.2209.00104.
8. 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; 10.
9. Page MJ, McKenzie JE, Kirkham J, Dwan K, Kramer S, Green S, Forbes A. 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(10): MR000035.
10. Wilkinson M, Dumontier M, Aalbersberg I, et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data. 2016; 3: 160018.
11. 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(1): 25. DOI: 10.1186/1471-2288-14-25.
12. Sweeting MJ, Sutton A, Lambert PC. What to add to nothing? Use and avoidance of continuity corrections in meta-analysis of sparse data. Statistics in Medicine. 2004; 23(9): 1351-75.
13. Balduzzi S, Rücker G, Schwarzer G. How to perform a meta-analysis with R: a practical tutorial, Evidence-Based Mental Health 2019; 22: 153-160. DOI: 10.1136/ebmental-2019-300117.
14. Viechtbauer, W. Conducting meta-analyses in R with the metafor package. Journal of Statistical Software, 2010; 36(3): 1-48. DOI: 10.18637/jss.v036.i03.
15. Lin L, Chu H (2022). altmeta: Alternative Meta-Analysis Methods. R package version 4.1. https://CRAN.R-project.org/package=altmeta.
16. McGuinness, LA, Higgins, JPT. Risk-of-bias VISualization (robvis): An R package and Shiny web app for visualizing risk-of-bias assessments. Res Syn Meth. 2020; 1- 7. https://doi.org/10.1002/jrsm.1411
17. Stijnen T, Hamza TH, Özdemir P. Random effects meta-analysis of event outcome in the framework of the generalized linear mixed model with applications in sparse data. Statist Med. 2010 Dec 20;29(29):3046–67.
18. Chu H, Nie L, Chen Y, Huang Y, Sun W. Bivariate random effects models for meta-analysis of comparative studies with binary outcomes: methods for the absolute risk difference and relative

risk. Statistical Methods in Medical Research 2012; 21(6): 621-633. DOI: 10.1177/0962280210393712.
19. Milia N, Congiu A, Anagnostou P, Montinaro F, Capocasa M, Sanna E, Bisol GD. 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. Veterinary Record Open. 2019; 6. DOI: 10.1136/vetreco-2018-000322.
21. Witwer KW. Data Submission and Quality in Microarray-Based MicroRNA Profiling. Clinical Chemistry. 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. 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. Journal of Clinical Epidemiology. 2019; 109: 111-116. DOI: 10.1016/j.jclinepi.2019.01.012.
24. Nguyen PY, Kanukula R, McKenzie JE, Alqaidoom Z, Brennan SE, Haddaway NR, 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 Nov 22; 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 Biology. 2021; 19: e3001107. DOI: 10.1371/journal.pbio.3001107.
26. Adewumi MT, Vo N, Tritz D, Beaman J, Vassar M. An evaluation of the practice of transparency and reproducibility in addiction medicine literature. Addictive Behaviors. 2021; 112: 106560. DOI: 10.1016/j.addbeh.2020.106560.
27. 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.
28. 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.
29. Anderson JM, Wright B, Rauh S, Tritz D, Horn J, Parker I, Bergeron D, Cook S, Vassar M. Evaluation of indicators supporting reproducibility and transparency within cardiology literature. Heart. 2020; 107: 120-126. DOI: 10.1136/heartjnl-2020-316519.
30. Ascha M, Katabi L, Stevens E, Gatherwright J, Vassar M. Reproducible Research Practices in the Plastic Surgery Literature. Plastic & Reconstructive Surgery. 2022; 149: 810e-823e. DOI: 10.1097/prs.0000000000008956.
31. 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 Network Open. 2022; 5: e2215209. DOI: 10.1001/jamanetworkopen.2022.15209.
32. 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. Research in Social and Administrative Pharmacy. 2022; 18: 2259-2268. DOI: 10.1016/j.sapharm.2021.06.002.
33. 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. 2022; DOI: 10.1101/2022.08.10.22278643.

34. Cenci MS, Franco MC, Raggio DP, Moher D, Pereira-Cenci T. Transparency in clinical trials: Adding value to paediatric dental research. International Journal of Paediatric Dentistry. 2020; 31: 4-13. DOI: 10.1111/ipd.12769.
35. 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.
36. Collins A, Alexander R. Reproducibility of COVID-19 preprints. Scientometrics. 2022; 127: 4655-4673. DOI: 10.1007/s11192-022-04418-2.
37. 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 Network Open. 2021; 4: e2033972. DOI: 10.1001/jamanetworkopen.2020.33972.
38. DeBlanc J, Kay B, Lehrich J, Kamdar N, Valley TS, Ayanian JZ, Nallamothu BK. Availability of Statistical Code From Studies Using Medicare Data in General Medical Journals. JAMA Internal Medicine. 2020; 180: 905. DOI: 10.1001/jamainternmed.2020.0671.
39. Duan Y, Luo J, Zhao L, Zhang X, Miao J, Moher D, Bian Z. 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.
40. Errington TM, Denis A, Perfito N, Iorns E, Nosek BA. Challenges for assessing replicability in preclinical cancer biology. eLife. 2021; 10. DOI: 10.7554/elife.67995.
41. 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. 2019; DOI: 10.1101/773473.
42. Federer LM, Belter CW, Joubert DJ, Livinski A, Lu Y-L, Snyders LN, Thompson H. Data sharing in PLOS ONE: An analysis of Data Availability Statements. PLOS ONE. 2018; 13: e0194768. DOI: 10.1371/journal.pone.0194768.
43. Fladie IA, Adewumi TM, Vo NH, Tritz DJ, Vassar MB. An Evaluation of Nephrology Literature for Transparency and Reproducibility Indicators: Cross-Sectional Review. Kidney International Reports. 2020; 5: 173-181. DOI: 10.1016/j.ekir.2019.11.001.
44. Fladie IA, Evans S, Checketts J, Tritz D, Norris B, Vassar M. Can Orthopaedics become the Gold Standard for Reproducibility? A Roadmap to Success. 2019; DOI: 10.1101/715144.
45. Gabelica M, Bojcic R, Puljak L. Many researchers were not compliant with their published data sharing statement: a mixed-methods study. Journal of Clinical Epidemiology. 2022; 150: 33-41. DOI: 10.1016/j.jclinepi.2022.05.019.
46. Gkiouras K, Nigdelis MP, Grammatikopoulou MG, Goulis DG. Tracing open data in emergencies: The case of the COVID-19 pandemic. European Journal of Clinical Investigation. 2020; 50. DOI: 10.1111/eci.13323.
47. Gorman DM. Availability of Research Data in High-Impact Addiction Journals with Data Sharing Policies. Science and Engineering Ethics. 2020; 26: 1625-1632. DOI: 10.1007/s11948-020-00203-7.
48. Grant R, Hrynaszkiewicz I. The impact on authors and editors of introducing Data Availability Statements at Nature journals. International Journal of Digital Curation. 2018; 13: 195-203. DOI: 10.2218/ijdc.v13i1.614.
49. Grayling MJ, Wheeler GM. A review of available software for adaptive clinical trial design. Clinical Trials. 2020; 17: 323-331. DOI: 10.1177/1740774520906398.
50. Hanson KA, Almeida N, Traylor JI, Rajagopalan D, Johnson J. Profile of Data Sharing in the Clinical Neurosciences. Cureus. 2020; DOI: 10.7759/cureus.9927.

51. 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.
52. 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). Perspectives on Psychological Science. 2021; 17: 239-251. DOI: 10.1177/1745691620979806.
53. Heckerman G, Tzng E, Campos-Melendez A, Ekwueme C, Mueller AL. Accessibility and Reproducible Research Practices in Cardiovascular Literature. 2022; DOI: 10.1101/2022.07.06.498942.
54. Heller N, Rickman J, Weight CJ, Papanikolopoulos N. The Role of Publicly Available Data in MICCAI Papers from 2014 to 2018. arXiv 2019; 1908.06830. DOI: 10.48550/arXiv.1908.06830.
55. Helliwell JA, Shelton B, Mahmood H, Blanco-Colino R, Fitzgerald JE, Harrison EM, Bhangu A, Chapman SJ. Transparency in surgical randomized clinical trials: cross-sectional observational study. BJS Open. 2020; 4: 977-984. DOI: 10.1002/bjs5.50333.
56. Huang Y-N, Patel NA, Mehta JH, Ginjala S, Brodin P, Gray CM, Patel YM, Cowell LG, Burkhardt AM, Mangul S. Data Availability of Open T-Cell Receptor Repertoire Data, a Systematic Assessment. Frontiers in Systems Biology. 2022; 2. DOI: 10.3389/fsysb.2022.918792.
57. Hughes T, Niemann A, Tritz D, Boyer K, Robbins H, Vassar M. Transparent and Reproducible Research Practices in the Surgical Literature. bioRxiv 2019; 779702. DOI: 10.1101/779702v1
58. Ioannidis JPA, Allison DB, Ball CA, Coulibaly I, Cui X, Culhane AC, Falchi M, Furlanello C, Game L, Jurman G, et al. Repeatability of published microarray gene expression analyses. Nature Genetics. 2009; 41: 149-155. DOI: 10.1038/ng.295.
59. Ioannidis JPA, Polyzos NP, Trikalinos TA. Selective discussion and transparency in microarray research findings for cancer outcomes. European Journal of Cancer. 2007; 43: 1999-2010. DOI: 10.1016/j.ejca.2007.05.019.
60. Iqbal SA, Wallach JD, Khoury MJ, Schully SD, Ioannidis JPA. Reproducible Research Practices and Transparency across the Biomedical Literature. PLOS Biology. 2016; 14: e1002333. DOI: 10.1371/journal.pbio.1002333.
61. Jalali MS, DiGennaro C, Sridhar D. Transparency Assessment of COVID-19 Models. The Lancet Global Health, 2020; 8(12): e1459-e1460. DOI: 10.1101/2020.07.18.20156851.
62. Janssen MA, Pritchard C, Lee A. On code sharing and model documentation of published individual and agent-based models. Environmental Modelling & Software. 2020; 134: 104873. DOI: 10.1016/j.envsoft.2020.104873.
63. Johnson AL, Torgerson T, Skinner M, Hamilton T, Tritz D, Vassar M. An assessment of transparency and reproducibility-related research practices in otolaryngology. The Laryngoscope. 2019; 130: 1894-1901. DOI: 10.1002/lary.28322.
64. Johnson B, Rauh S, Tritz D, Schiesel M, Vassar M. Evaluating Reproducibility and Transparency in Emergency Medicine Publications. Western Journal of Emergency Medicine. 2021; 22: 963-971. DOI: 10.5811/westjem.2021.3.50078.
65. Johnson JN, Hanson KA, Jones CA, Grandhi R, Guerrero J, Rodriguez J. Data Sharing in Neurosurgery and Neurology Journals. Cureus. 2018; DOI: 10.7759/cureus.2680.
66. Jurburg SD, Konzack M, Eisenhauer N, Heintz-Buschart A. The archives are half-empty: an assessment of the availability of microbial community sequencing data. Communications Biology. 2020; 3. DOI: 10.1038/s42003-020-01204-9.
67. Kaufmann I, Gattrell WT. Assessment of journal compliance with data sharing guidelines from the International Committee of Medical Journal Editors (ICMJE). Current Medical Research and Opinion. 2019 Apr 11; 35(sup2): 31-41. DOI: 10.1080/03007995.2019.1583496

68. Kemper JM, Rolnik DL, Mol BWJ, Ioannidis JPA. Reproducible research practices and transparency in reproductive endocrinology and infertility articles. Fertility and Sterility. 2020; 114: 1322-1329. DOI: 10.1016/j.fertnstert.2020.05.020.
69. Kirouac DC, Cicali B, Schmidt S. Reproducibility of Quantitative Systems Pharmacology Models: Current Challenges and Future Opportunities. CPT: Pharmacometrics & Systems Pharmacology. 2019; 8: 205-210. DOI: 10.1002/psp4.12390.
70. Kobres P-Y, Chretien J-P, Johansson MA, Morgan JJ, Whung P-Y, Mukundan H, Del Valle SY, Forshey BM, Quandelacy TM, Biggerstaff M, et al. A systematic review and evaluation of Zika virus forecasting and prediction research during a public health emergency of international concern. PLOS Neglected Tropical Diseases. 2019; 13: e0007451. DOI: 10.1371/journal.pntd.0007451.
71. Littmann M, Selig K, Cohen-Lavi L, Frank Y, Honigschmid P, Kataka E, Mosch A, Qian K, Ron A, Schmid S, et al. Validity of machine learning in biology and medicine increased through collaborations across fields of expertise. Nature Machine Intelligence. 2020; 2: 18-24. DOI: 10.1038/s42256-019-0139-8.
72. Lopez-Nicolas R, Lopez-Lopez JA, Rubio-Aparicio M, Sanchez-Meca J. A meta-review of transparency and reproducibility-related reporting practices in published meta-analyses on clinical psychological interventions (2000-2020). Behavior Research Methods. 2021; 54: 334-349. DOI: 10.3758/s13428-021-01644-z.
73. Louderback ER, Gainsbury SM, Heirene RM, Amichia K, Grossman A, Bernhard BJ, LaPlante DA. Open Science Practices in Gambling Research Publications (2016-2019): A Scoping Review. Journal of Gambling Studies. 2022; DOI: 10.1007/s10899-022-10120-y.
74. 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.
75. 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 Veterinary Research. 2021; 17. DOI: 10.1186/s12917-021-02971-1.
76. Miyakawa T. No raw data, no science: another possible source of the reproducibility crisis. Molecular Brain. 2020; 13. DOI: 10.1186/s13041-020-0552-2.
77. 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. Journal of Clinical Epidemiology. 2020; 119: 75-84. DOI: 10.1016/j.jclinepi.2019.12.002.
78. 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. Veterinary Anaesthesia and Analgesia. 2022; 49: 18-25. DOI: 10.1016/j.vaa.2021.09.004.
79. Naudet F, Sakarovitch C, Janiaud P, Cristea I, Fanelli D, Moher D, Ioannidis JPA. 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; k400. DOI: 10.1136/bmj.k400.
80. Noor MAF, Zimmerman KJ, Teeter KC. Data Sharing: How Much Doesn't Get Submitted to GenBank? PLoS Biology. 2006; 4: e228. DOI: 10.1371/journal.pbio.0040228.
81. Norris E, He Y, Loh R, West R, Michie S. Assessing Markers of Reproducibility and Transparency in Smoking Behaviour Change Intervention Evaluations. Journal of Smoking Cessation. 2021; 2021: 1-12. DOI: 10.1155/2021/6694386.
82. Norris E, Sulevani I, Finnerty AN, Castro O. Assessing Open Science practices in physical activity behaviour change intervention evaluations. BMJ Open Sport & Exercise Medicine. 2022; 8: e001282. DOI: 10.1136/bmjsem-2021-001282.

83. Ntzani EE, Ioannidis JP. Predictive ability of DNA microarrays for cancer outcomes and correlates: an empirical assessment. The Lancet. 2003 Nov; 362(9394): 1439-44. DOI: 10.1016/S0140-6736(03)14686-7.
84. 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: Psychology. 2017; 3. DOI: 10.1525/collabra.102.
85. Nutu D, Gentili C, Naudet F, Cristea IA. Open science practices in clinical psychology journals: An audit study. Journal of Abnormal Psychology. 2019; 128: 510-516. DOI: 10.1037/abn0000414.
86. Ochsner SA, Steffen DL, Stoeckert CJ, McKenna NJ. Much room for improvement in deposition rates of expression microarray datasets. Nat Methods. 2008 Dec; 5(12): 991-991.
87. 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; 729129. DOI: 10.1101/729129v1
88. Page MJ, Altman DG, Shamseer L, McKenzie JE, Ahmadzai N, Wolfe D, Yazdi F, Catal√°-L√≥pez F, Tricco AC, Moher D. Reproducible research practices are underused in systematic reviews of biomedical interventions. Journal of Clinical Epidemiology. 2018; 94: 8-18. DOI: 10.1016/j.jclinepi.2017.10.017.
89. Page MJ, Nguyen P-Y, Hamilton DG, Haddaway NR, Kanukula R, Moher D, McKenzie JE. Data and code availability statements in systematic reviews of interventions were often missing or inaccurate: a content analysis. Journal of Clinical Epidemiology. 2022; 147: 1-10. DOI: 10.1016/j.jclinepi.2022.03.003.
90. Papageorgiou SN, Antonoglou GN, Martin C, Eliades T. Methods, transparency and reporting of clinical trials in orthodontics and periodontics. Journal of Orthodontics. 2019; 46: 101-109. DOI: 10.1177/1465312519842315.
91. 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 Journal of Radiology. 2022; 23: 355. DOI: 10.3348/kjr.2021.0808.
92. Pellen C, Caquelin L, Jouvance-Le Bail A, Gaba J, Verin M, Moher D, Ioannidis JPA, Naudet F. Intent to share Annals of Internal Medicine's trial data was not associated with data re-use. Journal of Clinical Epidemiology. 2021; 137: 241-249. DOI: 10.1016/j.jclinepi.2021.04.011.
93. Piwowar HA, Chapman WW. Public sharing of research datasets: A pilot study of associations. Journal of Informetrics. 2010; 4: 148-156. DOI: 10.1016/j.joi.2009.11.010.
94. 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.
95. Rauh S, Bowers A, Rorah D, Tritz D, Pate H, Frye L, Vassar M. Evaluating the reproducibility of research in obstetrics and gynecology. European Journal of Obstetrics & Gynecology and Reproductive Biology. 2022; 269: 24-29. DOI: 10.1016/j.ejogrb.2021.12.021.
96. 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 Urology. 2022; 22. DOI: 10.1186/s12894-022-01059-8.
97. Rauh S, Torgerson T, Johnson AL, Pollard J, Tritz D, Vassar M. Reproducible and transparent research practices in published neurology research. Research Integrity and Peer Review. 2020; 5. DOI: 10.1186/s41073-020-0091-5.
98. 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-E987. DOI: 10.9778/cmajo.20200303.

99. Reidpath DD, Allotey PA. Data Sharing in Medical Research: An Empirical Investigation. Bioethics. 2001; 15: 125-134. DOI: 10.1111/1467-8519.00220.
100. 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. The Lancet Microbe. 2022; 3: e392-e398. DOI: 10.1016/s2666-5247(21)00250-0.
101. Riedel N, Kip M, Bobrov E. ODDPub - a Text-Mining Algorithm to Detect Data Sharing in Biomedical Publications. Data Science Journal. 2020; 19: 42. DOI: 10.5334/dsj-2020-042.
102. Rousi AM. Using current research information systems to investigate data acquisition and data sharing practices of computer scientists. Journal of Librarianship and Information Science. 2022; 096100062210930. DOI: 10.1177/09610006221093049.
103. Rowhani-Farid A, Barnett AG. Badges for sharing data and code at Biostatistics: an observational study. F1000Research. 2018; 7: 90. DOI: 10.12688/f1000research.13477.2.
104. 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.
105. Sarker A, Ginn R, Nikfarjam A, O'Connor K, Smith K, Jayaraman S, Upadhaya T, Gonzalez G. Utilizing social media data for pharmacovigilance: A review. Journal of Biomedical Informatics. 2015; 54: 202-212. DOI: 10.1016/j.jbi.2015.02.004.
106. 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.
107. 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.
108. Seibold H, Czerny S, Decke S, Dieterle R, Eder T, Fohr S, Hahn N, Hartmann R, Heindl C, Kopper P, 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.
109. Sherry CE, Pollard JZ, Tritz D, Carr BK, Pierce A, Vassar M. Assessment of transparent and reproducible research practices in the psychiatry literature. General Psychiatry. 2020; 33: e100149. DOI: 10.1136/gpsych-2019-100149.
110. Siebert M, Gaba JF, Caquelin L, Gouraud H, Dupuy A, Moher D, Naudet F. 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.
111. Smith CA, Nolan J, Tritz DJ, Heavener TE, Pelton J, Cook K, Vassar M. Evaluation of reproducible and transparent research practices in pulmonology. Pulmonology. 2021; 27: 134-143. DOI: 10.1016/j.pulmoe.2020.07.001.
112. Sofi-Mahmudi A, Raittio E. Transparency of COVID-19-Related Research in Dental Journals. Frontiers in Oral Health. 2022; 3. DOI: 10.3389/froh.2022.871033.
113. Sumner J, Haynes L, Nathan S, Hudson-Vitale C, McIntosh LD. Reproducibility and reporting practices in COVID-19 preprint manuscripts. 2020; DOI: 10.1101/2020.03.24.20042796.
114. Strcic J, Civljak A, Glozinic T, et al. Open data and data sharing in articles about COVID-19 published in preprint servers medRxiv and bioRxiv. Scientometrics. 2022; 127: 2791-2802. DOI: 10.1007/s11192-022-04346-1.
115. Tedersoo L, Kungas R, Oras E, Koster K, Eenmaa H, Leijen A, Pedaste M, Raju M, Astapova A, Lukner H, et al. Data sharing practices and data availability upon request differ across scientific disciplines. Scientific Data. 2021; 8. DOI: 10.1038/s41597-021-00981-0.
116. Thelwall M, Munafo M, Mas-Bleda A, Stuart E, Makita M, Weigert V, Keene C, Khan N, Drax K, Kousha K. 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.

117. Uribe SE, Sofi-Mahmudi A, Raittio E, Maldupa I, Vilne B. Dental Research Data Availability and Quality According to the FAIR Principles. Journal of Dental Research. 2022; 101: 1307-1313. DOI: 10.1177/00220345221101321.
118. 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.
119. Vassar M, Jellison S, Wendelbo H, Wayant C. Data sharing practices in randomized trials of addiction interventions. Addictive Behaviors. 2020; 102: 106193. DOI: 10.1016/j.addbeh.2019.106193.
120. Wallach JD, Boyack KW, Ioannidis JPA. Reproducible research practices, transparency, and open access data in the biomedical literature, 2015-2017. PLOS Biology. 2018; 16: e2006930. DOI: 10.1371/journal.pbio.2006930.
121. Walters C, Harter ZJ, Wayant C, Vo N, Warren M, Chronister J, Tritz D, Vassar M. 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.
122. Weissgerber T, Riedel N, Kilicoglu H, Labbe C, Eckmann P, ter Riet G, Byrne J, Cabanac G, Capes-Davis A, Favier B, et al. Automated screening of COVID-19 preprints: can we help authors to improve transparency and reproducibility? Nature Medicine. 2021; 27: 6-7. DOI: 10.1038/s41591-020-01203-7.
123. Wicherts JM, Crompvoets EAV. The poor availability of syntaxes of structural equation modeling. Accountability in Research. 2017; 24: 458-468. DOI: 10.1080/08989621.2017.1396214.
124. Womack RP. Research Data in Core Journals in Biology, Chemistry, Mathematics, and Physics. PLOS ONE. 2015; 10: e0143460. DOI: 10.1371/journal.pone.0143460.
125. Wright BD, Vo N, Nolan J, Johnson AL, Braaten T, Tritz D, Vassar M. An analysis of key indicators of reproducibility in radiology. Insights into Imaging. 2020; 11. DOI: 10.1186/s13244-020-00870-x.
126. Helliwell JA, Bolton WS, Burke JR, Tiernan JP, Jayne DG, Chapman SJ. Global academic response to COVID-19: Cross-sectional study. Learned Publishing. 2020; 33: 385-393. DOI: 10.1002/leap.1317.
127. Hemkens LG, Benchimol EI, Langan SM, Briel M, Kasenda B, Januel J-M, Herrett E, von Elm E. The reporting of studies using routinely collected health data was often insufficient. Journal of Clinical Epidemiology. 2016; 79: 104-111. DOI: 10.1016/j.jclinepi.2016.06.005.
128. Jiao C, Li K, Fang Z. Data sharing practices across knowledge domains: A dynamic examination of data availability statements in PLOS ONE publications. Journal of Information Science. 2022; 016555152211018. DOI: 10.1177/01655515221101830.
129. 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. F1000Research. 2017; 6: 1708. DOI: 10.12688/f1000research.12673.2.
130. Ramke J, Kuper H, Limburg H, Kinloch J, Zhu W, Lansingh VC, Congdon N, Foster A, Gilbert CE. Avoidable Waste in Ophthalmic Epidemiology: A Review of Blindness Prevalence Surveys in Low and Middle Income Countries 2000-2014. Ophthalmic Epidemiology. 2017; 25: 13-20. DOI: 10.1080/09286586.2017.1328067.
131. Rustici G, Williams E, Barzine M, Brazma A, Bumgarner R, Chierici M, Furlanello C, Greger L, Jurman G, Miller M, et al. Transcriptomics data availability and reusability in the transition from microarray to next-generation sequencing. 2021; DOI: 10.1101/2020.12.31.425022.
132. Stodden V, Seiler J, Ma Z. An empirical analysis of journal policy effectiveness for computational reproducibility. Proceedings of the National Academy of Sciences. 2018; 115: 2584-2589. DOI: 10.1073/pnas.1708290115.

133. Towse JN, Ellis DA, Towse AS. Opening Pandora's Box: Peeking inside Psychology's data sharing practices, and seven recommendations for change. Behavior Research Methods. 2020; 53: 1455-1468. DOI: 10.3758/s13428-020-01486-1.
134. Zhao M, Yan E, Li K. Data set mentions and citations: A content analysis of full-text publications. Journal of the Association for Information Science and Technology. 2017; 69: 32-46. DOI: 10.1002/asi.23919.