BMC Medicine

RESEARCH

Open Access

Practice of data sharing plans in clinical trial registrations and concordance between registered and published data sharing plans: a cross-sectional study

Jingyi Zhang[1], Barbara E. Bierer[2,3], Harriette G. C. Van Spall[4,5,13], Yingxin Liu[1], Xuerui Bai[6], Lehana Thabane[5,7], Gregory Y. H. Lip[8,9], Xin Sun[10], David Moher[11,12] and Guowei Li[1,7*]

Abstract

Background The International Committee of Medical Journal Editors (ICMJE) recommends that trial authors must specify data sharing plans when trials are registered and published, yet this uptake remains unclear. We aimed to assess the practice of data sharing plans in trial registration platforms and the concordance between registered and published data sharing plans.

Methods We included clinical trials published between 2021 and 2023 in six high-profile journals (*The Lancet*, *The New England Journal of Medicine*, *JAMA*, *BMJ*, *JAMA Internal Medicine*, and *Annals of Internal Medicine*) that enrolled participants no earlier than 2019 and registered on clinical trial platforms. One study outcome was data sharing plans in the trial registration platform, where trials clearly responding a "yes" to "Plan to share" were considered as planning to share data (including study protocols, statistical analysis plans, analytic codes, and individual participant data). The concordance between registered and published plans to share data was also assessed, which included plans to either share data (Yes/Yes) and not to share data (No/No) in both registration and publications. Univariate analyses were used to assess associations between trial characteristics and registered plans to share data and between trial characteristics and concordance.

Results Of the 383 included registration IDs, only 44.6% (171/383) planned to share data in registration. Trials with drug versus non-drug interventions had increased odds of registering plans to share data (*OR* = 2.71, 95% *CI*: 1.63, 4.63). There were seven trial publications, each pooling two trials and having two registration IDs. We selected the registration IDs with a later start date, resulting in 376 trial publications for concordance assessment. Over half (216/376, 57.4%) had discordance between registration and publications. COVID-19-related trials were associated with decreased odds of data sharing concordance (*OR* = 0.59, 95% *CI*: 0.37, 0.91). Additionally, significant discordance was consistently found in statistical analysis plans or study protocols, analytic codes, and individual participant data.

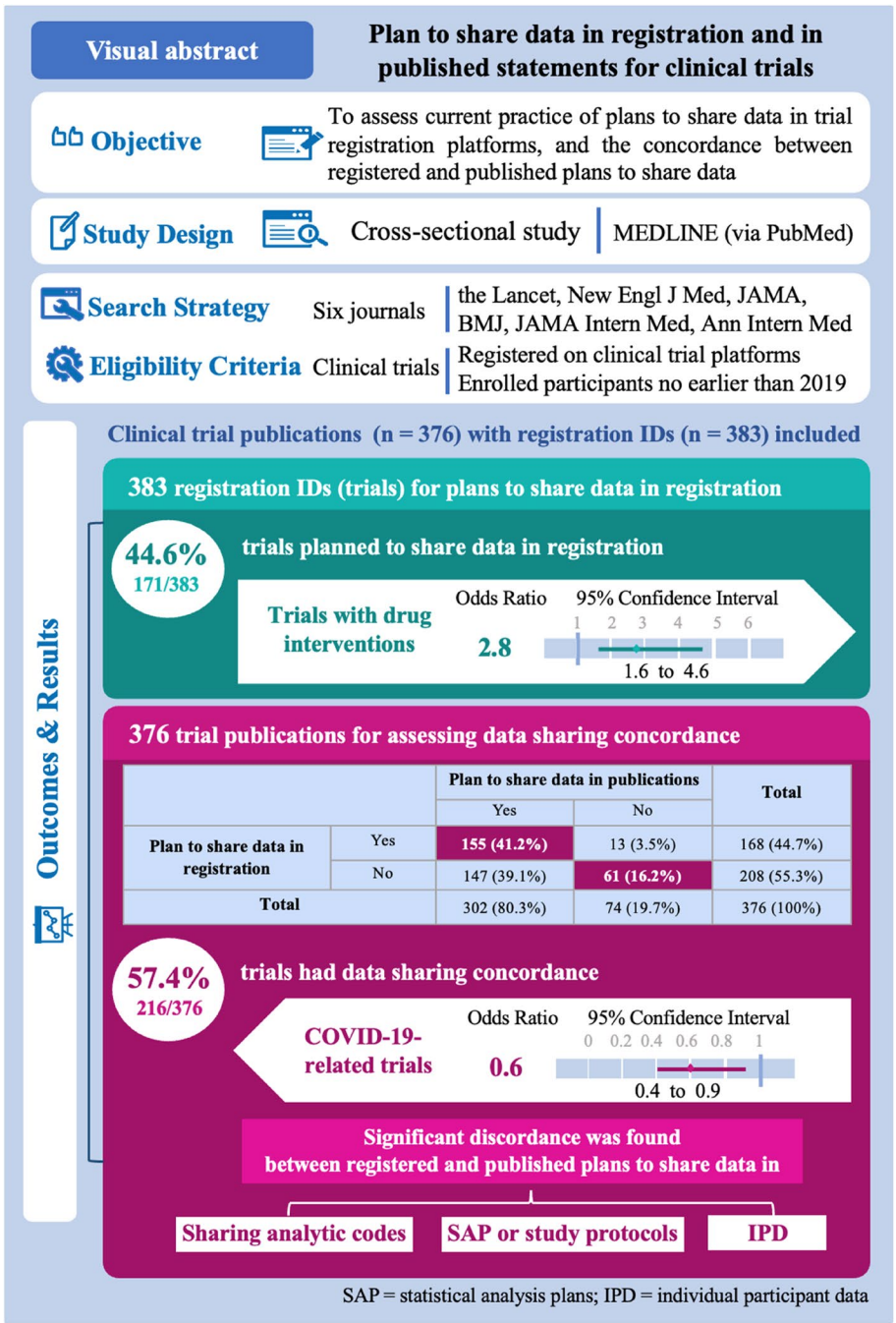
Conclusions Most registered trials do not specify plans to share data. More than half of published trials have data sharing discordance between registration and publication. Efforts are required to improve the reporting and reliability of plans to share clinical trial data.

*Correspondence:
Guowei Li
ligw@gd2h.org.cn
Full list of author information is available at the end of the article

Trial registration This study was registered on the Open Science Framework (https://osf.io/k6etb)

Keywords Plan to share data, Data sharing, Trial registration, Trial publication, Clinical trial

Graphical Abstract



Background

Sharing data (including individual participant data [IPD] and other supporting materials) in clinical trials as one component of open science practices can foster research integrity, promote transparency, and meet ethical obligations to participants, ultimately advancing scientific research and innovation [1]. The trial registration platform *ClinicalTrials.gov* has added a component of "Plan to share IPD" for trial authors when registering clinical trials since 2015 [2]. To further enhance data sharing, the International Committee of Medical Journal Editors (ICMJE) recommended that trial authors must specify a plan to share data for trials registered from January 2019 onwards [3]. Subsequently, a statement as to whether data are planned to be shared, and if so how, became a mandatory element on some registry platforms to align with the ICMJE recommendation [4].

A previous survey investigating clinical trials primarily sponsored by China's institutions found an increase in responding to the question about whether data would be shared in registration platforms after 2016, yet no significant change in the proportion of trials indicating "yes" to share data was observed from 2016 to 2018 [5]. Nevertheless, the current practice of plans to share data in clinical trial registration and the evolving impact of ICMJE recommendations on trial registration are largely unknown. Moreover, the concordance between registered and published plans to share data is unclear [6]. While trial registration can improve reporting transparency, there may be discordance between what is registered and subsequently published in peer-reviewed manuscripts [7–10]. For instance, some trials explicitly indicated "Yes" to the "Plan to share data" option in their registration platforms; however, their subsequent data sharing statements from publications indicated that their data would not be shared [11–13]. Exploring the discordance between registered and published plans to share data is therefore important to advance our understanding of data sharing reporting transparency and consistency.

Thus, we conducted a cross-sectional study to assess the current practice of plans to share data in trial registration platforms and explored the concordance between what was registered and what was published in trial publications. We also evaluated the associations between trial characteristics and plans to share data in trial registration and between trial characteristics and the concordance between registered and published plans to share data.

Methods

We reported this study according to the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guideline [14]. This study was registered on the Open Science Framework (https://osf.io/k6etb).

Search strategy and eligibility criteria

We first identified the top 10 medical general journals that frequently published clinical trials and were ranked by journal impact factor in the category of "Medicine, General & Internal" based on Journal Citation Reports (as of June 2023). After excluding those journals that primarily focused on basic science or published less than 10 clinical trials annually, a total of 6 journals were chosen, including *The Lancet*, *The New England Journal of Medicine (New Engl J Med)*, *Journal of the American Medical Association (JAMA)*, *British Medical Journal (BMJ)*, *JAMA Internal Medicine*, and *Annals of Internal Medicine (Ann Intern Med)*. Subsequently, we searched MEDLINE (via PubMed) to systematically retrieve clinical trials published in these journals (Additional file 1: Table S1 shows the search strategy used). Given that ICMJE required a data sharing plan in trial registration from Jan 2019 onwards, we only included trials that started participant enrollment on or after Jan 1, 2019, and trial publications published between Jan 1, 2021, and Dec 31, 2023, to allow sufficient time and samples of trials for evaluation. We included clinical trial publications with primary results; methods papers, publications of secondary results, relevant reviews, commentaries, perspectives, or editorials were excluded. The detailed selection process is presented in Additional file 1: Fig. S1.

If the same trial was registered on different platforms, we only extracted and analyzed information from *ClinicalTrials.gov*. Some publications may pool different trials for analysis, thereby having ≥ two registration identifiers (IDs). We treated such publications as different trials by their corresponding registration IDs for data extraction and analysis; i.e., each registration ID represented an individual trial. Some publications with updated data may share the same registration ID with prior publications; in this case, the most recent publication was kept for analysis to avoid double counting.

Study outcomes

The outcomes were the inclusion of a plan to share data in the trial registration and the concordance between registered and published plans to share data.

Trials that clearly responded with a "Yes" to "Plan to share" in registration were considered as planning to share data, while those reporting with a "No" were considered as not planning to share data. In this study, the data that were planned to share included study protocols, statistical analysis plans, analytic codes, and IPD, where trials reporting a plan to share any of these data were considered to "plan to share data" in registration. We searched trial registration platforms based on registration IDs to determine the stated plan to share

data. If a trial had multiple registration records, we used the latest registration record before the trial publication was published. On the registration platform, all information on the data sharing plan description was extracted, including plans to share IPD and supporting information (study protocols, statistical analysis plans, and analytic codes). If the question "Plan to share IPD" was left blank or answered "Undecided," responses were pooled as "undecided/missing".

We further explored the data sharing concordance between registered and published plans to share. From data sharing statements in trial publications, trials that clearly stated a willingness to share data were defined as published plans to share data. We also treated trials as having published plans to share data if a link to a data repository was provided, even if the shared data were accessible only after the user registered and signed a data use agreement. Trials that were unwilling to share data or did not report/obtain data sharing statements were considered not to have a published plan to share. Subsequently, data sharing concordance was assessed: (1) plan to share data in registration and publication (both Yes in registration and publication, i.e., "Yes/Yes") and (2) no plan to share data in registration and publication (both No in registration and publication, i.e., "No/No"). Discordance between registered and published plans to share data included (1) plan to share data in registration but no plan to share data in the publication (Yes in registration but No in publication, i.e., "Yes/No") and (2) no plan to share in registration but a plan to share in publication (No in registration but Yes in publication, i.e., "No/Yes"). Seven trial publications pooled two trials/registration IDs in which case the registration IDs with a later study start date were used to assess data sharing concordance.

We also assessed the details of the data sharing plans, which are elaborated in registration platforms and statements in trial publications. Therefore, the specific information extracted included the following: (1) data sharing content (analytic code, statistical analysis plan or study protocol, IPD), (2) data access time after publication or trial completion (< 12 months, ≥ 12 months, unclear), and (3) data access method (public, private, unclear). If trial authors clearly stated that the shared data would be publicly available, we considered the trials to have a public data access method. If the shared data were only available from trial authors, funders, or trial review committees after review, trials were grouped to have a private data access method. Trials were categorized as having an unclear data access method if no relevant details were provided regarding how the trial authors would share data.

Data extraction

Data extraction and coding were completed independently by four study authors in pairs (J. Z. and X. B., Y.L., and G. L.). Any disagreement was resolved by discussion between the study authors and, if no consensus could be reached, resolved through consultation with the senior author (D. M.).

Data on trial characteristics from registration platforms were extracted, including whether the trial was multicenter, country of origin, design information (with or without control group, parallel or crossover, with or without randomization), trial phase (1–2, or 3–4), planned sample size, intervention type (drug or other, where "drug" included both drug alone or drug in combination with non-drug), whether the trial was COVID-19-related, and funding source (industry or other, where "industry" included industry alone or the combination of industry and non-industry funder) [15]. For those that did not report a trial phase, they were classified as phase 3–4 if they planned to enroll ≥ 400 participants and grouped as phase 1–2 if the planned sample size was < 400 [10, 16].

We predefined data extraction from trial publications, where the extracted data included the year of publication, whether trial publication mentioned authors' conflict of interest (yes or no), and the risk of bias (ROB). If the trial publication mentioned authors' conflict of interest, we further categorized the conflict of interest as either financial, non-financial, or both [17]. We did not aim to assess the ROB for each outcome of the included trials; therefore, the ROB 1.0 tool was used to evaluate the overall ROB for individual trials [18]. A trial was grouped as having high ROB if at least one domain (random sequence generation, allocation concealment, blinding of participants and personnel, blinding of outcome assessment, incomplete outcome data, selective reporting, and others) was rated as high ROB. Trials were defined to have low ROB if all domains were rated as low ROB, while they were considered to have unclear ROB if there was ≥ one domain rating as unclear ROB [19].

Statistical analysis

We described continuous trial characteristics with medians and lower and upper quartiles (Q1, Q3) and categorical variables using counts and percentages. We used the McNemar's test to evaluate whether the concordance between registered and published plans to share data was significant [20]. We plotted the proportions of trials with plans to share data in registration and proportions of data sharing concordance from 2021 to 2023 by country of trial origin and journal.

We assessed the associations between trial characteristics and registered plans to share data and between trial characteristics and data sharing concordance. Trials with undecided/missing plans to share data were treated as "No" plan to share data in our association analysis. The univariate logistic regression analysis was used to explore trial characteristics in relation to registered plans to share data, taking those trials without registered plans to share data as reference. For the association between trial characteristics and data sharing concordance (Yes/Yes, and No/No), trials with the two types of discordance (Yes/No, and No/Yes) were combined as the reference group, given the concern over small sample sizes for each type of discordance. This approach to combine these two types of discordance was also used by some previous methodological studies [7, 21].

We performed univariable logistic regression analysis for each trial characteristic in relation to registered plans to share data, including the year of publication, whether being multicenter, funding source, planned sample size, whether being a COVID-19 trial, intervention type, country of origin, phase of clinical trial, and whether a parallel design. Similarly, we conducted univariable analysis to investigate whether the trial characteristics (including the year of publication, whether being multicenter, planned sample size, whether being a COVID-19 trial, intervention type of drug, country of origin, trial phase, whether a parallel design, funding source, authors' conflict of interest, and ROB) were associated with data sharing concordance between registered and published plans to share data. Odds ratios (ORs) with 95% confidence intervals (CIs) were used for the relationship between trial characteristics and registered plans to share data and between trial characteristics and data sharing concordance. An $OR > 1.0$ presented that the trial characteristic was associated with increased odds of registered plans to share data in registration or data sharing concordance.

Regarding the associations between trial characteristics and registered plans to share data, we performed a prespecified sensitivity analysis by removing trials with undecided/missing plans to share data from the association analysis. We performed another post hoc sensitivity analysis by excluding non-randomized trials from the association analysis.

We redescribed the counts and percentages of data sharing concordance between registered and published plans by removing trials with undecided/missing plans to share data and by treating trials with undecided/missing plans as having registered plans to share data. Moreover, for trial characteristics in relation to data sharing concordance, we conducted two post hoc sensitivity analyses by replacing the seven registration IDs that had a later study start date with those having an earlier study start date and by excluding non-randomized trials. We performed a third post hoc sensitivity analysis by using the two types of discordant groups as a separate control group for the association analysis (i.e., Yes/Yes and No/No *vs* Yes/No, Yes/Yes and No/No *vs* No/Yes).

Furthermore, we evaluated the differences in data sharing content, data access time after publication or trial completion, and data access method among those trials having Yes/Yes in registration and publications. The McNemar's test was used to assess whether a significant discordance existed.

All statistical tests were two-sided with a significance level of 0.05. Analyses were conducted in R software version 4.4.1.

Results

Registered plans to share data in registration

We included a total of 376 clinical trial publications and 383 registration IDs for analyses. The registration IDs representing 383 individual trials were used to assess the current practice of plans to share data in registration. As shown in Table 1, 42.6% of trials were published in the *New Engl J Med*, and 29.5% were COVID-19-related. The vast majority of trials were randomized controlled trials, registered in *ClinicalTrials.gov*, and had a parallel design. The median planned sample size of trials was 572. There were 63.2% of trials rated as low ROB.

A total of 171 (44.6%) trials reported having a plan to share data in registration, 133 (34.7%) trials did not plan to share data, and 79 (20.6%) trials had undecided/missing plans to share data (Table 1). A slightly increased proportion of trials planning to share data in registration was observed in 2023 when compared to 2021. Less than half of trials registered on *ClinicalTrials.gov* had a plan to share data in registration. Among the 171 trials planning to share data, the majority of trials had a drug intervention (86.0%) and were non-COVID-19-related (74.3%). Figure 1a and b depicts the proportions of trials planning to share data in registration by country of trial origin and journal temporally. A steady growth in planning to share data was found in the UK. China seemed to face a decline in planning to share data, yet the observed decline was based on very few observations. The proportion of trials planning to share data in registration consecutively descended in *Ann Intern Med* and *JAMA*.

Figure 2 shows the trial characteristics in relation to registered plans to share data. Trials with an intervention type of drug were significantly associated with elevated odds of registered plans to share data compared to non-drug intervention ($OR = 2.71$, 95% *CI*: 1.63, 4.63). Compared to industry-funded trials, trials supported by other funding sources were significantly associated with decreased odds of registered plans to share data

Table 1 Description of included trials' characteristics and comparisons between trials with and without plans to share data in registration[a]

Trial characteristics	Overall ($n=383$)	Whether trials plan to share data in registration		
		No ($n=133$)	Undecided /missing ($n=79$)	Yes ($n=171$)
Year of publication				
2021	86 (22.5)	28 (21.1)	19 (24.1)	39 (22.8)
2022	118 (30.8)	39 (29.3)	34 (43.0)	45 (26.3)
2023	179 (46.7)	66 (49.6)	26 (32.9)	87 (50.9)
Journal				
BMJ	25 (6.5)	14 (10.5)	4 (5.1)	7 (4.1)
JAMA	59 (15.4)	21 (15.8)	16 (20.3)	22 (12.9)
LANCET	94 (24.5)	25 (18.8)	18 (22.8)	51 (29.8)
New Engl J Med	163 (42.6)	51 (38.3)	32 (40.5)	80 (46.8)
JAMA Internal Medicine	22 (5.7)	11 (8.3)	5 (6.3)	6 (3.5)
Annals of Internal Medicine	20 (5.2)	11 (8.3)	4 (5.1)	5 (2.9)
Registration platform name				
Clinicaltrials.gov	346 (90.3)	123 (92.5)	68 (86.1)	155 (90.6)
Others	37 (9.7)	10 (7.5)	11 (13.9)	16 (9.4)
Country of origin[b]				
USA	80 (20.9)	39 (29.3)	11 (13.9)	30 (17.5)
UK	19 (5.0)	7 (5.3)	3 (3.8)	9 (5.3)
China	12 (3.1)	8 (6.0)	1 (1.3)	3 (1.8)
Others	272 (71.0)	79 (59.4)	64 (81.0)	129 (75.4)
Planned sample size: median (Q1, Q3)	572 (276, 1,600)	492 (215, 1,770)	700 (292, 2,280)	600 (321, 1,440)
Funding source[c]				
Others	147 (38.7)	55 (42.0)	36 (46.2)	56 (32.7)
Industry	233 (61.3)	76 (58.0)	42 (53.8)	115 (67.3)
With control group				
No	7 (1.8)	3 (2.3)	0	4 (2.3)
Yes	376 (98.2)	130 (97.7)	79 (100.0)	167 (97.7)
Randomization				
No	17 (4.4)	7 (5.3)	4 (5.1)	6 (3.5)
Yes	366 (95.6)	126 (94.7)	75 (94.9)	165 (96.5)
Parallel design				
No	10 (2.6)	3 (2.3)	1 (1.3)	6 (3.5)
Yes	373 (97.4)	130 (97.7)	78 (98.7)	165 (96.5)
Phase of clinical trial				
1 ~ 2	110 (28.7)	38 (28.6)	28 (35.4)	44 (25.7)
3 ~ 4	273 (71.3)	95 (71.4)	51 (64.6)	127 (74.3)
Multicenter				
No	41 (10.7)	19 (14.3)	8 (10.1)	14 (8.2)
Yes	342 (89.3)	114 (85.7)	71 (89.9)	157 (91.8)
Intervention type				
Non-drug	89 (23.2)	42 (31.6)	23 (29.1)	24 (14.0)
Drug	294 (76.8)	91 (68.4)	56 (70.9)	147 (86.0)
COVID-19-related				
No	270 (70.5)	100 (75.2)	43 (54.4)	127 (74.3)
Yes	113 (29.5)	33 (24.8)	36 (45.6)	44 (25.7)
Risk of bias[d]				
High	64 (16.7)	25 (18.8)	17 (21.5)	22 (12.9)
Some concerns	77 (20.1)	29 (21.8)	12 (15.2)	36 (21.1)
Low	242 (63.2)	79 (59.4)	50 (63.3)	113 (66.1)

New Engl J Med (*The New England Journal of Medicine*), *JAMA* (*Journal of the American Medical Association*), and *BMJ* (*British Medical Journal*)

[a] Results shown as count (%) unless otherwise specified

[b] Results only shown for countries having > 10 clinical trials

[c] Results shown for trials that provided information on funding ($n = 380$)

[d] The ROB 1.0 tool was used to assess the risk of bias in trials

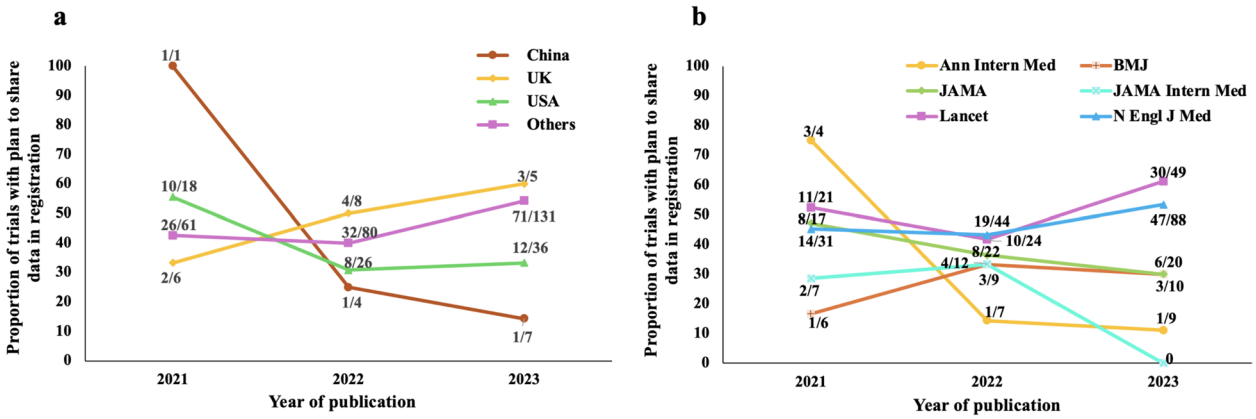


Fig. 1 Proportions of trials with plan to share data in registration from 2021 to 2023 by country of trial origin and journal (**a** registered plan to share data by country of origin, **b** registered plan to share data by journal). Note: The numbers shown within the lines represented the counts of trials with plans to share in registration over all the included trials by year of publication, country of trial origin, or journal

(*OR* = 0.63, 95% *CI*: 0.41, 0.96). Results from the sensitivity analysis after removing trials with undecided and missing plans to share data (Additional file 1: Table S2) and by excluding non-randomized trials (Additional file 1: Table S3) were in general consistent with the main findings.

Concordance between registered and published plans to share data

A total of 376 trials were included for assessing the concordance, among which 216 (57.4%) had concordance between registered and published plans to share data ($n = 155$ and 61 for Yes/Yes and No/No, respectively; Table 2). For the discordance, there were 13 (3.5%) and 147 (39.1%) trials as Yes/No and No/Yes, respectively. A significant discordance was observed between registered and published plans to share data ($P < 0.001$).

Among the 216 trials having data sharing concordance (Table 3), the majority were non-COVID-19-related trials (75.0%), multicenter trials (92.1%), and mentioned conflict of interest (93.9%). An increased proportion of trials with data sharing concordance from 2021 to 2023 was observed in the UK and *New Engl J Med* (Fig. 3a and b). COVID-19-related trials were found to significantly relate to decreased odds of the concordance (*OR* = 0.59, 95% *CI*: 0.37, 0.91). Multicenter trials (*OR* = 2.07, 95% *CI*: 1.08, 4.05) and authors' conflict of interest (*OR* = 2.35, 95% CI: 1.15, 4.97) were significantly associated with increased odds of concordance (Additional file 1: Fig. S2).

Similar results were found from sensitivity analyses by using the seven registration IDs that had an earlier study start date (Additional file 1: Table S4) and by excluding non-randomized trials (Additional file 1: Table S5). Results from using the two types of discordance as separate control groups were also consistent with the main findings in general (Additional file 1: Table S6). For the sensitivity analyses by removing 78 trials with undecided/missing plans and treating them as having registered plans to share data, the percentages of the concordance became 64.7% (Yes/Yes: 52.0%, No/No: 12.7%) and 65.9% (Yes/Yes: 55.8%, No/No: 10.1%), respectively (Additional file 1: Table S7).

Additional file 1: Fig. S3 displays the counts of trials that registered and published plans to share data by data sharing content. Among the 155 trials having Yes/Yes in registration and publications, significant discordance was found in analytic code, SAP/study protocol, and IPD after comparing registered with published plans to share data (all *P*-values < 0.05; Fig. 4). The number of trials planning to share IPD increased by nearly 30% (from 96 in registration to 123 in publications), while the number of trials planning to share SAP/study protocol dropped from 91 in registration to 43 in publications. A significant discordance in the data access method was also observed comparing registered with published plans to share data (Fig. 5).

Discussion

In this cross-sectional study, we systematically explored plans to share data in clinical trial registration and the concordance between registered and published plans to share data. It was found that more than half of trials published in high-profile journals did not plan to share data in registration, and over 40% were discordant between registration and publication plans. Trials with an intervention of drug were associated with increased odds of registered plans to share data, while COVID-19-related trials were related to lower odds of concordance. Additionally, significant discordance was consistently found for specific data sharing contents, including statistical

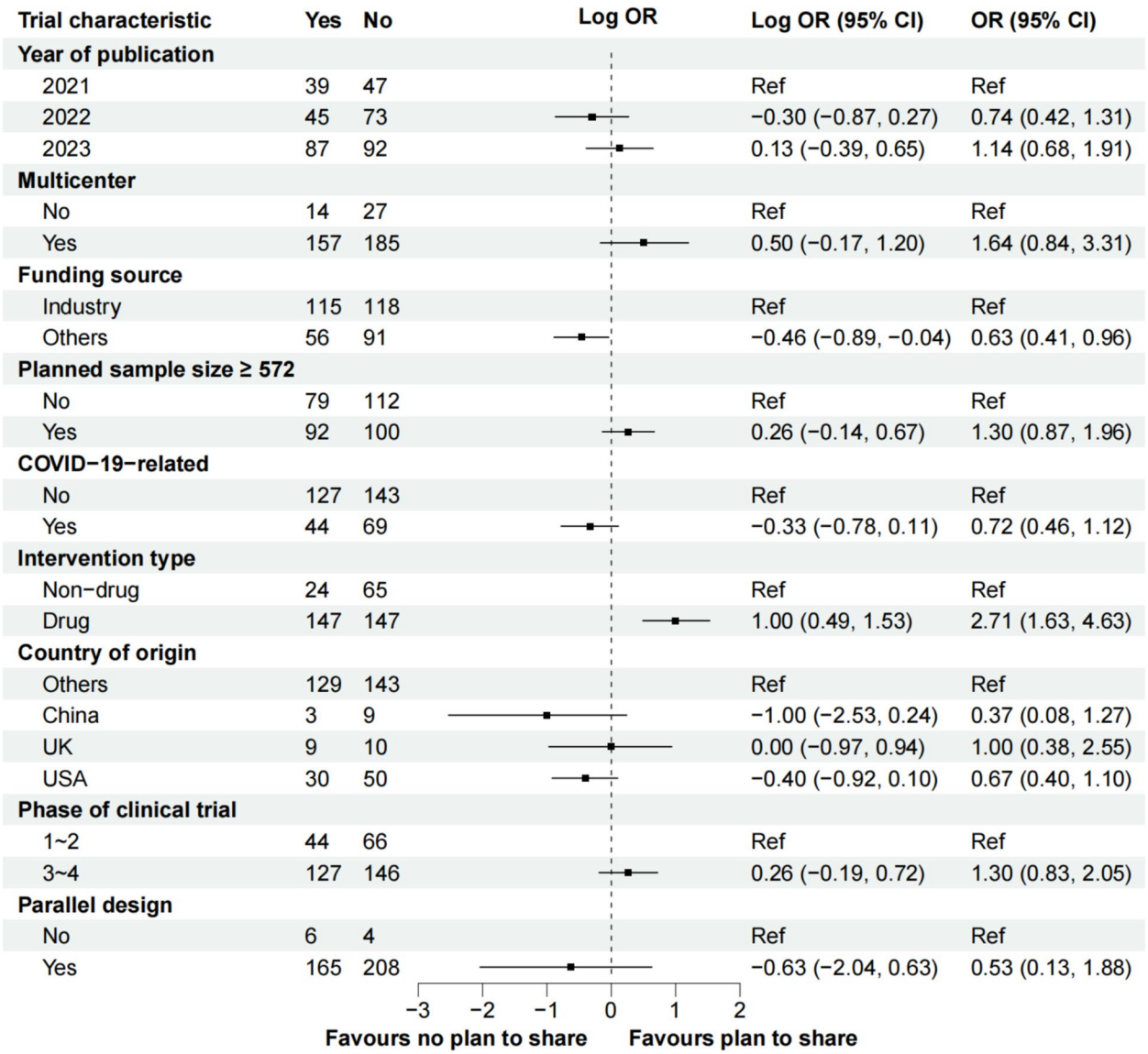


Fig. 2 Univariate logistic analysis results for the relationship between trial characteristics and registered plans to share data. Note: The median planned sample size among all the trials was 572. OR, odds ratio; CI, confidence interval

Table 2 Counts and percentages of comparing registered with published plans to share data in the included trials[a]

		Plan to share data in publication		Total
		Yes	No	
Plan to share data in registration	Yes	155 (41.2%)	13 (3.5%)	168 (44.7%)
	No	147 (39.1%)	61 (16.2%)	208 (55.3%)
Total		302 (80.3%)	74 (19.7%)	376 (100%)

[a] Results shown as count (%) unless otherwise specified

analysis plans, study protocols, analytic codes, and individual participant data.

There were over 20% of trials with undecided/missing plans to share data in registration. This highlights the need for awareness and education of trial authors about the value of data sharing and its consideration during trial planning. A previous study that included all trials registered in the 6 months preceding the ICMJE registration requirement (January 2019) found that only 5.5% reported data sharing plans in *ClinicalTrials.gov* [22]. Our study observed that 45% of trials registered to share

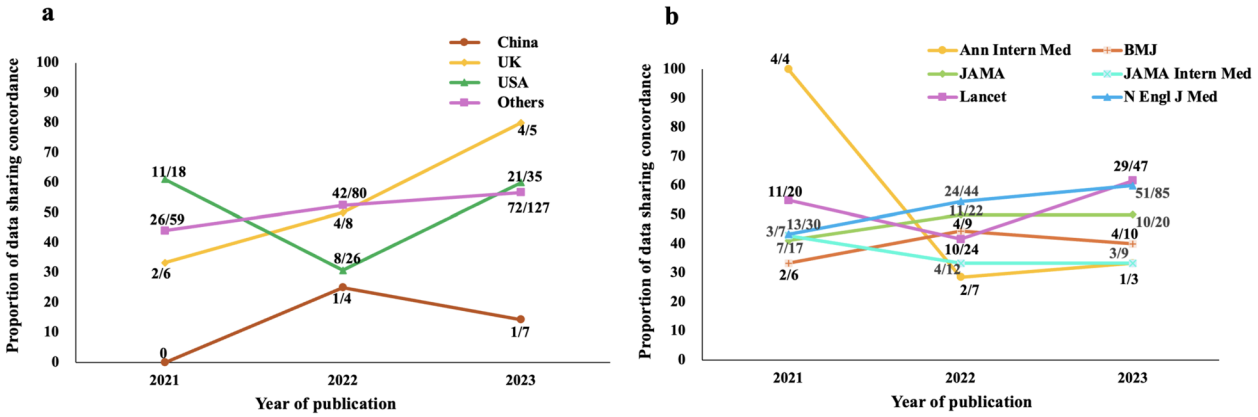


Fig. 3 Proportions of data sharing concordance from 2021 to 2023 by country of trial origin and journal (**a** data sharing concordance by country of origin, **b** data sharing concordance by journal). Note: The numbers shown within the lines represent the counts of trials with data sharing concordance over all the included trials by year of publication, country of trial origin, or journal

data, possibly reflecting temporal improvements after the ICMJE recommendation was issued. Nevertheless, we only included six high-profile journals for assessment and may overestimate the current practice of data sharing in registration.

Trials with drug interventions received close attention regarding data sharing and reporting transparency, partly explaining a significant association between trials with drug interventions and increased odds of the plan to share data in registration [23]. Of note, sharing IPD from phase 1 trials is not required or anticipated [24], consistent with our finding of a lower proportion of plans to share data in phase 1 trials when compared with phase 2–4 trials, as demonstrated in our post hoc analysis: 38.9% (7/18) vs 44.9% (164/365), respectively.

Concerns about the challenges of data sharing, including participant privacy, necessary additional financial resources, funders' expectations, and personal interest, may be significant barriers to adhering to data sharing plans [25]. In addition, trial authors may change their data sharing statements in publications from plans aiming to cater to journals' implicit expectation or to preserve their advantage of future research output, among others. Notably, the ICMJE recommended that any change in planning to share data should be updated in the registration record and reflected in and consistent with statements in trial publications [3]. Despite the ICMJE requirement, we found discordant data sharing plans in over 40% of trials, a finding that requires attention to increase consistency and reliability. Compliance is difficult, however, in that updating one element requires updating the entire registration record, at least on *ClinicalTrials.gov*. Given that discordant data sharing plans in this study result, we recommend ICMJE reassess that expectation (i.e., very late changes to the registration record). While available, accessing prior registration records and identifying relevant changes are burdensome.

The National Institutes of Health (NIH) has suggested that SAPs and trial protocols should be publicly available at the time of publishing any summary trial results [26], especially given that analytic codes, SAPs, and protocols could help identify and prevent inappropriate analysis and selective or biased reporting [27]. However, an earlier survey reported a low percentage of cancer trial publication statements willing to share analytic code (12/274, 4.4%) [28]. This low percentage is consistent with our findings of the plan to share analytic codes both in registration (20/383, 5.2%) and in trial publications (25/376, 6.6%; Additional file 1: Fig. S3). More importantly, a substantial decline in planning to share SAP/study protocol from registration to publication was observed. While previous studies focused on the declaration to share IPD or the comparison between declared and actual data sharing, we recognize a wide gap between registered and published plans to share analytic codes and SAPs/protocols. Efforts are thus required to enhance the concordance of relevant sharing contents (including analytic code, SAP/study protocol, and IPD) for reporting transparency and consistency.

Promptly sharing data had been strongly advocated during public health emergencies and pandemics [29–32]. However, we found no difference between COVID-19 and non-COVID-19 trials in planning to share data in registration, similar to previous studies [33, 34].

Table 3 Comparisons between trials with and without data sharing concordance[a]

Trial characteristics	Overall ($n = 376$)	Concordance between registered and published plans to share data	
		($n = 160$)	Yes($n = 216$)
Year of publication			
2021	84 (22.3)	39 (24.4)	45 (20.8)
2022	118 (31.4)	51 (31.9)	67 (31.0)
2023	174 (46.3)	70 (43.8)	104 (48.1)
Journal			
BMJ	25 (6.6)	15 (9.4)	10 (4.6)
JAMA	59 (15.7)	24 (15.0)	35 (16.2)
LANCET	91 (24.2)	39 (24.4)	52 (24.1)
New Engl J Med	159 (42.3)	57 (35.6)	102 (47.2)
JAMA Internal Medicine	22 (5.9)	14 (8.8)	8 (3.7)
Annals of Internal Medicine	20 (5.3)	11 (6.9)	9 (4.2)
Registration platform name			
ClinicalTrials.gov	339 (90.2)	143 (89.4)	196 (90.7)
Others	37 (9.8)	17 (10.6)	20 (9.3)
Country of origin[b]			
USA	79 (21.0)	33 (20.6)	46 (21.3)
UK	19 (5.1)	9 (5.6)	10 (4.6)
China	12 (3.2)	8 (5.0)	4 (1.9)
Others	266 (70.7)	110 (68.8)	156 (72.2)
Planned sample size: Median (Q1, Q3)	583 (276, 1,650)	570 (273, 2,290)	583 (280, 1,420)
With control group			
No	7 (1.9)	1 (0.6)	6 (2.8)
Yes	369 (98.1)	159 (99.4)	210 (97.2)
Randomization			
No	17 (4.5)	7 (4.4)	10 (4.6)
Yes	359 (95.5)	153 (95.6)	206 (95.4)
Parallel design			
No	10 (2.7)	3 (1.9)	7 (3.2)
Yes	366 (97.3)	157 (98.1)	209 (96.8)
Phase of clinical trial			
1 ~ 2	107 (28.5)	46 (28.8)	61 (28.2)
3 ~ 4	269 (71.5)	114 (71.3)	155 (71.8)
Multicenter			
No	41 (10.9)	24 (15.0)	17 (7.9)
Yes	335 (89.1)	136 (85.0)	199 (92.1)
Intervention type			
Non-drug	287 (76.3)	115 (71.9)	172 (79.6)
Drug	89 (23.7)	45 (28.1)	44 (20.4)
COVID-19 related			
No	264 (70.2)	102 (63.8)	162 (75.0)
Yes	112 (29.8)	58 (36.3)	54 (25.0)
Funding source[c]			
Industry	226 (60.6)	88 (55.3)	138 (64.5)
Others	147 (39.4)	71 (44.7)	76 (35.5)
Whether trials mentioned conflict of interest[d]			
No	34 (9.1)	21 (13.2)	13 (6.1)
Yes	339 (90.9)	138 (86.8)	201 (93.9)

Table 3 (continued)

Trial characteristics	Overall ($n = 376$)	Concordance between registered and published plans to share data	
		($n = 160$)	Yes($n = 216$)
Type of conflict interest			
Financial	48 (14.2)	24 (17.4)	24 (11.9)
Non-financial	12 (3.5)	8 (5.8)	4 (2.0)
Both	279 (82.3)	106 (76.8)	173 (86.1)
Risk of bias[e]			
High	64 (17.0)	31 (19.4)	33 (15.3)
Some concerns	77 (20.5)	34 (21.3)	43 (19.9)
Low	235 (62.5)	95 (59.4)	140 (64.8)

New Engl J Med (*The New England Journal of Medicine*), *JAMA* (*Journal of the American Medical Association*), and *BMJ* (*British Medical Journal*)

[a] Results shown as count (%) unless otherwise specified

[b] Results only shown for countries having > 10 clinical trials

[c] Results shown for trials that provided information on funding ($n = 373$)

[d] Results shown for trials that provided information on conflict of interest ($n = 373$)

[e] The RoB 1.0 tool was used to assess the risk of bias in trials

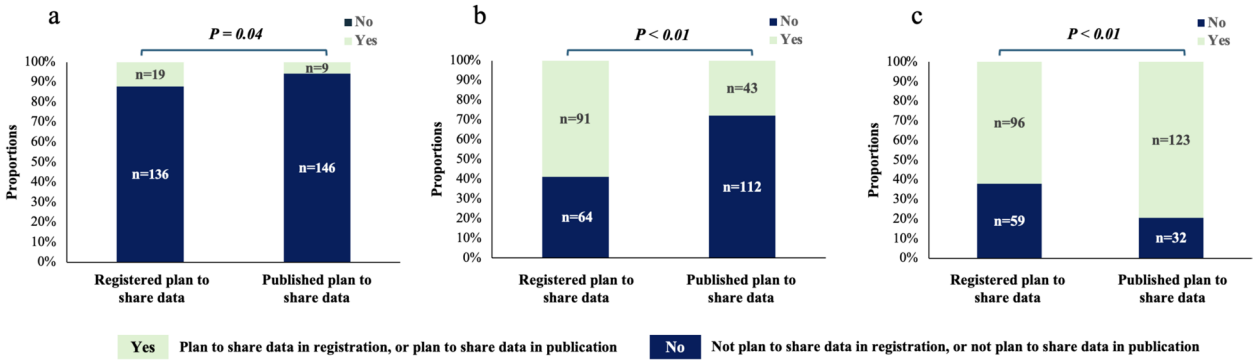


Fig. 4 Data sharing concordance between registered and published plans by data sharing content (**a** for analytic code, **b** for SAP/study protocol, **c** for IPD). Note: SAP, statistical analysis plan; IPD, individual participant data

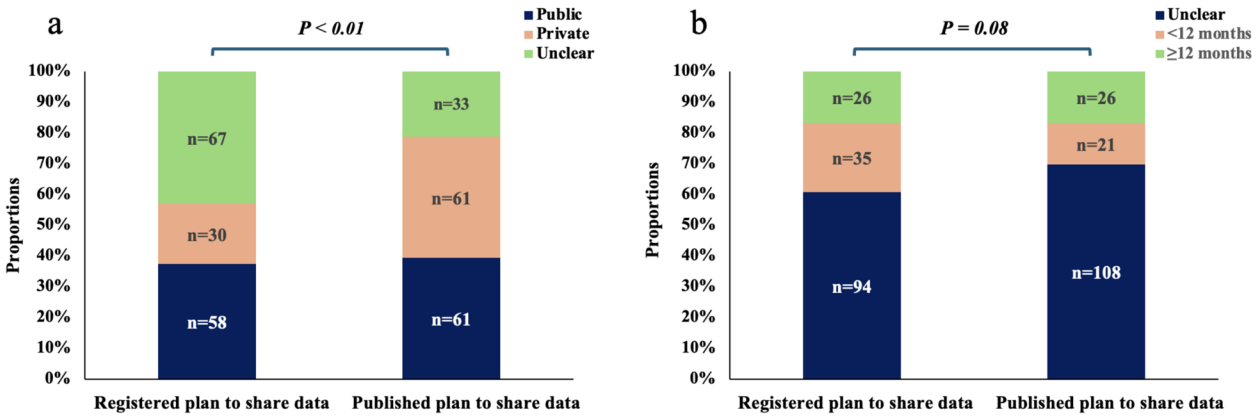


Fig. 5 Data sharing concordance comparing registered with published plans by data access method and time (**a** for data access method, **b** for data access time after publication or trial completion)

Furthermore, COVID-19-related trials were associated with decreased odds of data sharing concordance, possibly due to researchers' increased concerns about losing research output and patients' sensitization to data privacy as the COVID-19 pandemic rapidly progressed [35]. While one survey in 2021 reported that only 2.7% of COVID-19 trials had discordance between trial registration and published data sharing statements [36], we observed a substantially higher proportion of trials with data sharing discordance. The small discordance reported from this previous survey may be because it only included trials published at the early stage of the COVID-19 pandemic (up to Jun 2020) and did not include data from the evolving pandemic nor the impact of ICMJE recommendation.

Many organizations, academic institutions, and pharmaceutical companies have developed internal data sharing policies to make their research data publicly or privately available within a reasonable time frame [37]. Nevertheless, our study found that a sizable proportion of trials did not specify data access methods or timing in registration or in publications. Significant discordance was also detected in data access methods, comparing registered with published plans to share data. This again reflects an important gap in the reporting transparency of data sharing in clinical trials.

Research implications

The ICMJE recommendation was an important step in promoting data sharing in clinical trials [38]; we demonstrate, however, that further attention to whether and how data sharing plans are described in registration and in publications is needed, even in trials published in high-profile journals. Trial registration platforms may consider a more structured and standardized approach to guide the reporting of data sharing plans. Endeavors to promote awareness of journals and trial authors are also needed for completing and/or updating the plan to share data at trial registration, including data sharing content, data access time, and method. Moreover, incorporating recommendations for data sharing plans into future SPIRIT (Standard Protocol Items: Recommendations for Interventional Trials) iterations may help improve reporting transparency, consistency, and reliability of data sharing plans [39].

Strengths and limitations of the study

Our study systematically surveyed the current practice of data sharing plans in registration of clinical trials and in publications in order to identify areas of needed improvement and enhance data sharing reporting transparency and reliability and, ultimately, data sharing itself. Study results were strengthened by rigorous methodology and comprehensive analyses.

Several limitations should be noted. First, this cross-sectional study did not assess temporal changes in trial registration, because we only analyzed the last registration record prior to the trial publication date. Our analysis only included trials published in high-profile journals, which may underestimate the discordance between registered and published data sharing plans and thus weaken the generalizability of our findings. We classified trials without a reported trial phase using a cut-off sample size of 400 participants, potentially introducing misclassification in the subgroup analysis. Moreover, while this study aimed to evaluate the reporting transparency of plans to share data in registration and in trial publications, we did not contact trial authors to assess whether the data had been shared. Further research is needed to explore the causes and mechanisms of data sharing discordance between registered and published plans to share data. In addition, due to the small sample size of the included trials, we did not perform statistical tests to explore whether there were significant temporal trends in proportions of trials with plans to share data in registration and data sharing concordance by country of trial origin and journal. The association analysis results from this study were hypothesis-generating of an exploratory nature, given the potential for residual or unmeasured confounding and biases.

Conclusions

Most registered clinical trials published in high-profile journals do not specify a plan to share data. More than half of published trials have discordance between data sharing plans that they submit in trial registration versus publications. Joint efforts are required to improve the reporting and reliability of plans to share clinical trial data.

Abbreviations

IPD	Individual participant data
ICMJE	International Committee of Medical Journal Editors
STROBE	Strengthening the Reporting of Observational Studies in Epidemiology
New Engl J Med	*The New England Journal of Medicine*
JAMA	*Journal of the American Medical Association*
BMJ	*British Medical Journal*
Ann Intern Med	*Annals of Internal Medicine*
ROB	Risk of bias
SAP	Statistical analysis plan
OR	Odds ratio
Cis	Confidence intervals
NIH	National Institutes of Health

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12916-025-04328-z.

Additional file 1: Figure S1. Flow diagram for this study. Figure S2. Univariate logistic analysis results for the relationship between trial characteristics and data sharing concordance comparing registered with published plans to share data. Figure S3. Numbers of trials registered and published plans to share data by data sharing content among the 383 registration IDs and 376 clinical trial publications. Table S1. The search strategy used for this study. Table S2. Results for the relationship between trial characteristics and registered plans to share data after removing trials with undecided and missing plans to share data. Table S3. Sensitivity analysis results of the relationship between trial characteristics and registered plans to share data by excluding non-randomized trials. Table S4. Sensitivity analysis results of the relationship between trial characteristics and data sharing concordance by using the 7 registration IDs with earlier study start dates. Table S5. Sensitivity analysis results of the relationship between trial characteristics and data sharing concordance by excluding non-randomized trials. Table S6. Sensitivity analysis results of the relationship between trial characteristics and data sharing concordance by using the two discordant groups as separate control group. Table S7. Counts and percentages of data sharing concordance comparing registered with published plans to share data, after removing 78 trials with undecided/missing plans to share dataand treating trials with undecided/missing plans as having registered plans to share data.

Acknowledgements
We would like to thank Prof An-Wen Chan for his helpful comments on the study design and data interpretation. We also acknowledged Yaoyao Wang and Likang Li for their help with data collection and checking for data accuracy.

Authors' contribution
JZ, BB, LT, YL, DM and GL conceived the idea of the study and designed the study. JZ, YL, XB, and GL contributed to records screening and data collection. JZ, YL and GL performed statistical analyses and interpretation, and drafted the manuscript. BB, HGCVS, GYHL, XS, MX, LT and DM provided professional support and made several critical revisions to the manuscript. All authors read and approved the final manuscript. GL acts as the guarantor of this work.

Funding
This study was funded by the National Natural Science Foundation of China (82473612), the Natural Science Foundation of Guangdong Province of China (2025A1515010779), the Science Foundation of Guangdong Second Provincial General Hospital (YY2018-002), the Young Top Talent Project in the Special Support Plan for Training High-level Talents in Guangdong (0720240244), and the Science Foundation of Guangdong Second Provincial General Hospital (TJGC-2025006). The funders had no roles in study design, data collection and analysis, publication decisions, or manuscript preparation.

Data availability
Data were collected from the literature and registration platforms. All the relevant data and coding were available on the OSF (https://osf.io/vwb7r/files/osfstorage).

Declarations

Ethics approval and consent to participate
Not required because this study used publicly available materials and did not involve humans. Neither patients nor public representatives were involved in this study.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Center for Clinical Epidemiology and Methodology (CCEM), The Affiliated Guangdong Second Provincial General Hospital of Jinan University, Guangzhou, Guangdong 510317, China. [2]Department of Medicine, Harvard Medical School, Boston, MA, USA. [3]The Multi-Regional Clinical Trials Center of Brigham and Women's Hospital and Harvard, Boston, MA, USA. [4]Department of Medicine, Faculty of Health Sciences, McMaster University, Hamilton, ON, Canada. [5]Department of Health Research Methods, Evidence, and Impact (HEI), McMaster University, Hamilton, ON, Canada. [6]Department of Epidemiology, School of Medicine, Jinan University, Guangzhou, China. [7]Father Sean O'Sullivan Research Centre, St Joseph's Healthcare Hamilton, Hamilton, ON, Canada. [8]Liverpool Centre for Cardiovascular Sciences at University of Liverpool, Liverpool Heart & Chest Hospital, Liverpool John Moores University, Liverpool, UK. [9]Department of Clinical Medicine, Aalborg University, Aalborg, Denmark. [10]Chinese Evidence-Based Medicine Center and Cochrane China Center, West China Hospital, Sichuan University, Chengdu, China. [11]School of Epidemiology and Public Health, University of Ottawa, Ottawa, ON, Canada. [12]Methodological and Implementation Research Program, Centre for Journalology, Ottawa Hospital Research Institute, Ottawa, ON, Canada. [13]Population Health Research Institute, Hamilton, ON, Canada.

Received: 12 November 2024 Accepted: 6 August 2025
Published online: 01 September 2025

References

1. Bauchner H, McDermott MM, Butte AJ. Data sharing enters a new era. Ann Intern Med. 2023;176(3):400–1.
2. Bergeris A, Tse T, Zarin DA. Trialists' intent to share individual participant data as disclosed at clinicaltrials.gov. JAMA. 2018;319(4):406–8.
3. Taichman DB, Sahni P, Pinborg A, Peiperl L, Laine C, James A, et al. Data sharing statements for clinical trials: a requirement of the International Committee of Medical Journal Editors. Ethiop J Health Sci. 2017;27(4):315–8.
4. Merson L, Ndwandwe D, Malinga T, Paparella G, Oneil K, Karam G, et al. Promotion of data sharing needs more than an emergency: an analysis of trends across clinical trials registered on the International Clinical Trials Registry Platform. Wellcome Open Res. 2022;7:101.
5. Xu Y, Dong M, Liu X. Characteristics and trends of clinical studies primarily sponsored by China in WHO primary registries between 2009 and 2018: a cross-sectional survey. BMJ Open. 2020;10(11): e037262.
6. 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(1):e2033972.
7. TARG Meta-Research Group & Collaborators. Estimating the prevalence of discrepancies between study registrations and publications: a systematic review and meta-analyses. BMJ Open. 2023;13(10): e076264.
8. Hartung DM, Zarin DA, Guise JM, McDonagh M, Paynter R, Helfand M. Reporting discrepancies between the ClinicalTrials.gov results database and peer-reviewed publications. Ann Intern Med. 2014;160(7):477–83.
9. Riveros C, Dechartres A, Perrodeau E, Haneef R, Boutron I, Ravaud P. Timing and completeness of trial results posted at ClinicalTrials.gov and published in journals. PLoS Med. 2013;10(12): e1001566.
10. Chan AW, Pello A, Kitchen J, Axentiev A, Virtanen JI, Liu A, et al. Association of trial registration with reporting of primary outcomes in protocols and publications. JAMA. 2017;318(17):1709–11.
11. Iyasere CA, Wing J, Martel JN, Healy MG, Park YS, Finn KM. Effect of increased interprofessional familiarity on team performance, communication, and psychological safety on inpatient medical teams: a randomized clinical trial. JAMA Intern Med. 2022;182(11):1190–8.
12. Mailankody S, Devlin SM, Landa J, Nath K, Diamonte C, Carstens EJ, et al. GPRC5D-targeted CAR T cells for myeloma. N Engl J Med. 2022;387(13):1196–206.
13. Sharma A, Boelens JJ, Cancio M, Hankins JS, Bhad P, Azizy M, et al. *CRISPR-Cas9* editing of the HBG1 and HBG2 promoters to treat sickle cell disease. N Engl J Med. 2023;389(9):820–32.
14. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The strengthening the reporting of observational studies in

epidemiology (STROBE) statement: guidelines for reporting observational studies. J Clin Epidemiol. 2008;61(4):344–9.
15. Fundytus A, Wells JC, Sharma S, Hopman WM, Del Paggio JC, Gyawali B, et al. Industry funding of oncology randomised controlled trials: implications for design, results and interpretation. Clin Oncol (R Coll Radiol). 2022;34(1):28–35.
16. Zhang J, Van Spall HGC, Wang Y, Thabane L, Wang R, Li G. Enrollment of Black, Indigenous, and other people of color in multicountry randomized controlled trials of diabetes conducted in North America and Europe. Diabetes Care. 2022;45(7):e116–7.
17. Perlis RH, Perlis CS, Wu Y, Hwang C, Joseph M, Nierenberg AA. Industry sponsorship and financial conflict of interest in the reporting of clinical trials in psychiatry. Am J Psychiatry. 2005;162(10):1957–60.
18. Higgins JP, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ. 2011;343: d5928.
19. Babic A, Vuka I, Saric F, Proloscic I, Slapnicar E, Cavar J, et al. Overall bias methods and their use in sensitivity analysis of Cochrane reviews were not consistent. J Clin Epidemiol. 2020;119:57–64.
20. Fagerland MW, Lydersen S, Laake P. The McNemar test for binary matched-pairs data: mid-p and asymptotic are better than exact conditional. BMC Med Res Methodol. 2013;13:91.
21. Shi X, Ross JS, Amancharla N, Niforatos JD, Krumholz HM, Wallach JD. Assessment of concordance and discordance among clinical studies posted as preprints and subsequently published in high-impact journals. JAMA Netw Open. 2021;4(3):e212110.
22. Statham EE, White SA, Sonwane B, Bierer BE. Primed to comply: individual participant data sharing statements on ClinicalTrials.gov. PLoS One. 2020;15(2): e0226143.
23. Hudson KL, Collins FS. Sharing and reporting the results of clinical trials. JAMA. 2015;313(4):355–6.
24. Alliance CRDS. A review of biopharma sponsor data sharing policies and protection methodologies. https://crdsalliance.org/crdsa_resources/a-review-of-biopharma-sponsor-data-sharing-policies-and-protection-methodologies/. Accessed 4 Aug 2025.
25. 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.
26. Department of Health and Human Services. U.S. Department of Health and Human Services. Clinical trials registration and results information submission - a rule by the Health and Human Services Department. https://www.federalregister.gov/documents/2016/09/21/2016-22129/clinical-trials-registration-and-results-information-submission. Accessed 4 Aug 2025.
27. Sharma NK, Ayyala R, Deshpande D, Patel Y, Munteanu V, Ciorba D, et al. Analytical code sharing practices in biomedical research. PeerJ Comput Sci. 2024;10: e2066.
28. 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(1):438.
29. Cosgriff CV, Ebner DK, Celi LA. Data sharing in the era of COVID-19. Lancet Digit Health. 2020;2(5):e224. https://doi.org/10.1016/S2589-7500(20)30082-0.
30. Modjarrad K, Moorthy VS, Millett P, Gsell PS, Roth C, Kieny MP. Developing global norms for sharing data and results during public health emergencies. PLoS Med. 2016;13(1): e1001935.
31. Moorthy V, Henao Restrepo AM, Preziosi MP, Swaminathan S. Data sharing for novel coronavirus (COVID-19). Bull World Health Organ. 2020;98(3):150.
32. Rios RS, Zheng KI, Zheng MH. Data sharing during COVID-19 pandemic: what to take away. Expert Rev Gastroenterol Hepatol. 2020;14(12):1125–30.
33. Dron L, Dillman A, Zoratti MJ, Haggstrom J, Mills EJ, Park JJH. Clinical trial data sharing for COVID-19-related research. J Med Internet Res. 2021;23(3): e26718.
34. Ramdjee B, Husson M, Hajage D, Tubach F, Estellat C, Dechartres A. COVID-19 trials were not more likely to report intent to share individual data than non-COVID-19 trials in clinicaltrials.gov. J Clin Epidemiol. 2023;158:10–7.
35. Tan AC, Askie LM, Hunter KE, Barba A, Simes RJ, Seidler AL. Data sharing-trialists' plans at registration, attitudes, barriers and facilitators: a cohort study and cross-sectional survey. Res Synth Methods. 2021;12(5):641–57.
36. Li R, von Isenburg M, Levenstein M, Neumann S, Wood J, Sim I. COVID-19 trials: declarations of data sharing intentions at trial registration and at publication. Trials. 2021;22(1):153.
37. Modi ND, Kichenadasse G, Hoffmann TC, Haseloff M, Logan JM, Veroniki AA, et al. A 10-year update to the principles for clinical trial data sharing by pharmaceutical companies: perspectives based on a decade of literature and policies. BMC Med. 2023;21(1):400.
38. Kuntz RE, Antman EM, Califf RM, Ingelfinger JR, Krumholz HM, Ommaya A, et al. Individual patient-level data sharing for continuous learning: a strategy for trial data sharing. NAM Perspect. 2019. https://doi.org/10.31478/201906b.
39. Chan AW, Tetzlaff JM, Altman DG, Laupacis A, Gøtzsche PC, Krleža-Jerić K, et al. SPIRIT 2013 statement: defining standard protocol items for clinical trials. Ann Intern Med. 2013;158(3):200–7.

Publisher's Note