Discrepancies in reported results between trial registries and journal articles for AI clinical research

Zixuan He,[a,b] Lan Yang,[a,b] Xiaofan Li,[c] and Jian Du[b,*]

[a]Institute of Medical Technology, Health Science Center of Peking University, Beijing, China
[b]National Institute of Health Data Science, Peking University, Beijing, China
[c]Department of Statistics & Data Science, National University of Singapore, Singapore

eClinicalMedicine
2025;80: 103066

Published Online xxx
https://doi.org/10.1016/j.eclinm.2024.103066

Summary

Background Complete and unbiased reporting of clinical trial results is essential for evaluating medical advances, yet publication bias and reporting discrepancies in research on the clinical application of artificial intelligence (AI) remain unknown.

Methods We conducted a comprehensive search of research publications and clinical trial registries focused on the application of AI in healthcare. Our search included publications in Dimensions.ai and pre-registered records from ClinicalTrials.gov and the EU Clinical Trials Registry before 31 December 2023. We linked registered trials to their corresponding publications, analysed the registration, reporting and different dissemination patterns of results, identified discrepancies between clinical trial registries and published literature, and assessed the use of these results in secondary research.

Findings We identified 28,248 publications related to the use of AI in clinical settings and found 1863 publications that included a clinical trial registration ID. The clinical trial registry search identified 3710 trials evaluating the use of AI in clinical settings, of which 1106 trials are completed, yet only 101 trials have published results. By linking the trials to their corresponding publications, we found that 26 trials had results available from both registries and publications. There were more results in trial registries than in articles, but researchers showed a clear preference for rapid dissemination of results through peer-reviewed articles (37.6% published within one year) over trial registries (15.8%). Discrepancies and omissions of results were common, and no complete agreement was observed between the two sources. Selective reporting of publications occurred in 53.6% of cases, and the underestimation of the incidence of adverse events is alarming.

Interpretation This research uncovers concerns with the registration and reporting of AI clinical trial results. While trial registries and publications serve distinct yet complementary roles in disseminating research findings, discrepancies between them may undermine the reliability of the evidence. We emphasise adherence to guidelines that promote transparency and standardisation of reporting, especially for investigator-initiated trials (IITs).

Funding The authors declare no source of funding.

Keywords: Clinical trial; Artificial intelligence; Discrepancy; Clinical evidence

Introduction

Clinical trials are pivotal in the advancement of medical science, serving as the primary method for evaluating the safety and efficacy of novel interventions. Accurate and complete reporting of their outcomes is critical for informing clinical decisions and shaping policies. Yet, a notable phenomenon of publication bias in scientific literature favours positive outcomes, with 12.5% of researchers publishing their negative findings.[1,2] Data repositories, such as Clinicaltrials.gov, offer routes for sharing negative results. The reliability of clinical evidence hinges on the accuracy and consistency of trial results, regardless of whether they are published in peer-reviewed journals, clinical trial registries, or preprints. However, previous research has highlighted significant inconsistencies between the results posted in clinical trial registries and those published in the literature, raising concerns about the reliability of these

*Corresponding author. National Institute of Health Data Science, Peking University, No.38 Xueyuan Road, Beijing 100191, China.
E-mail address: dujian@bjmu.edu.cn (J. Du).

Research in context

Evidence before this study

With the emerging application of artificial intelligence (AI) in clinical settings, accurate and complete reporting of clinical trial outcomes is essential for decision-making. The reliability of clinical evidence depends on the accuracy and consistency of trial results, however, publication bias and reporting discrepancies in the clinical application of AI remain unexplored.

Added value of this study

This study highlights concerns regarding the registration and reporting of AI clinical trial results. By examining trials that published results in both clinical trial registry and journal article, we found limited evidence available across these dissemination platforms, accompanied by prevalent discrepancies, omissions, publication bias, and selective reporting. While registries and publications serve distinct roles, they also act as complementary sources for disseminating research findings.

Implications of all the available evidence

The study emphasises the importance of utilising multiple data sources, adhering to established reporting guidelines, and enhancing the accuracy and transparency of AI clinical trial evidence.

sources.[3–7] Such discrepancies can result in confusion and misinterpretation, undermining the evidence's utility for clinical practice and research. Re-analyses that incorporate results data from clinical trial registries into meta-analyses indicate a reduction in treatment effect estimates compared to those derived from published sources.[8] Moreover, a comprehensive review of 698 Cochrane meta-analyses revealed that published trials may exaggerate pharmaceutical treatment effects, with unpublished trials typically showing less pronounced outcomes.[9] This suggests a need for an evidence synthesis that includes all available data to prevent the overstatement of treatment efficacy and to ensure an unbiased representation of medical interventions' true efficacy.

Similar concerns may extend to clinical evidence that assesses the application of artificial intelligence (AI) in clinical settings. There is a growing consensus that AI has the potential to transform various aspects of healthcare. Numerous studies and reports highlight its potential benefits in improving diagnostic accuracy, enhancing the efficiency of healthcare systems, and managing chronic diseases.[10–15] Despite the enthusiasm for AI's potential in healthcare, there is a shortage of solid clinical evidence underpinning AI-driven therapies. The exaggeration of AI performance becomes a major concern as it could potentially misguide clinical decisions with the overstated benefits. AI needs evidence-based validation, but currently, the quality of AI evidence is generally at a lower level, and there are concerns over 'exaggerated' claims of AI outperforming doctors.[16–19]

The need to establish a comprehensive and rigorous evidence base for AI in clinical practice is urgent. This paper aims to conduct a thorough analysis of clinical trials involving AI, concentrating on their registration and reporting compliance, the results toward AI interventions, and the transparency and potential biases of these results. Afterwards, this study examines discrepancies between clinical trial registry postings and publication narratives. Additionally, it assesses how evidence from registries and publications is integrated into secondary analyses, including systematic reviews and meta-analyses. This study contributes to understanding the current landscape of AI clinical evidence, the accessibility of their outcomes, the consistency of information across platforms, and the evidence's role in subsequent research, thereby offering critical insights into AI treatment efficacy, publication bias, and the dependability of clinical evidence.

Methods

Identifying and screening clinical trials

Data sources and search strategy

For this research, we used a systematic approach to identify, screen, and extract data from clinical trials that evaluate the application of AI in clinical settings. We investigated different clinical trial registry platforms, including the WHO International Clinical Trials Registry Platform, and discovered that only ClinicalTrials.gov and the EU Clinical Trials Register (EU-CTR) had available posted results. Consequently, we conducted systematical searches of clinical trial data in the ClinicalTrials.gov registry and EU-CTR. Systematic multi-string search strategies were developed using a combination of text words and indexes related to different forms of AI with searching fields restricted to intervention/treatment (full searching strategy refers to Supplementary 1a). Trials registered before December 31, 2023 are included.

Inclusion and exclusion criteria

We include clinical trials that met the following criteria: 1) Participants: No restriction; 2) Intervention: Interventions containing AI algorithm; 3) Comparison: No restriction; 4) Study Type: Interventional or observational. Clinical trials not targeting humans or not registered in English were excluded. For analyses of discrepancies, further screening was limited to trials

that were completed and had accessible results posted on the registry.

Identifying and screening of publications

Data sources and search strategy

We searched publications on the application of AI in clinical settings in Dimensions.ai, a scholarly database containing publications from multiple platforms worldwide that were recorded before December 31, 2023. Systematic multi-string search strategies were developed to identify different forms of AI and clinical trials, with search fields restricted to titles (Supplementary 1a).

Inclusion and exclusion criteria

We include publications that report the results of clinical trials assessing the application of AI and excluded publications that assessed non-AI intervention, were not in a clinical setting or did not state a clinical trial registration ID. Publications that are study protocols, case medical research, secondary analysis, preprints or abstracts that have been published, and trial summary publications that contain multiple clinical trials ID numbers, not targeting humans or full text not in English are further excluded.

Linking clinical trials with publications

To associate clinical trials with their corresponding publications, we first extract the clinical trial registration ID in all included publications using Python 3.8.8 and pair them with the trial ID from registries. Then, we conducted an additional search using all clinical trial IDs identified from ClinicalTrials.gov and the EU-CTR platform, ensuring that the search for trials-related publications was all-encompassing and no publication was excluded in error during the extraction of trial ID.

Identifying systematic reviews and meta-analyses on AI in clinical applications

In addition to the analyses of discrepancies, we searched systematic reviews and meta-analyses publications in the Dimensions.ai database to analyse the types of evidence included in reviews assessing the effectiveness of AI in clinical applications. Specifically, we examined whether these reviews integrated results from trial registries and publications and whether they utilised evidence from multi-databases.

Data extraction

For all included clinical trials, we extracted trial ID, study type, status and type of sponsor to conduct overall analyses. For trials that are completed with available results, we further collected the title, disease or conditions, details of the intervention, enrolment number, registration date, result posting date, all outcomes number, description and their associating results, as well as the description of adverse events and their associating occurrence number. Correspondingly, for each associated publication, we extracted information including publication and trial registration ID, title and abstract, publication date, intervention, enrolment, specifics of primary and secondary outcomes, capturing its number, description, results, and the details of adverse events. For the systematic reviews and meta-analyses on AI in clinical applications, we specifically extracted the information on the databases searched during their identification process.

For all the above screening and extraction processes, two researchers independently screened clinical trials and publications eligible for inclusion and linked clinical trials with their corresponding publications. We calculated the Cohen's Kappa index to evaluate the level of agreement between the two researchers during the screening process. Three researchers extracted data to the standardised collection form in Excel, each extracting 2/3 of the included trials and publications, ensuring that at least two researchers cross-checked all extracted information. Uncertainties or disagreements were resolved through discussions among all authors until a consensus was reached.

Terminology and definition

Investigator-initiated trials (IITs)

To investigate who is leading the research in the clinical application of AI, we examined the funding status of the trials and identified whether they were investigator-initiated trials (IITs), based on the lead funder disclosed in the trial registry. Recognising the ambiguous definition of IITs in previous research, we classified trials as IITs if they were led by medical or academic institutions, received funding from non-commercial sources (including government agencies, non-profit organisations, or institutional research funds) and with no commercial funding involved. This approach allowed us to obtain a strict classification of IITs.[20,21]

Reporting discrepancy

In the context of this research, the identification and definition of discrepancies between clinical trial registry records and their corresponding publications are critical. Publications are deemed to be in 'complete agreement' if no discrepancies are observed. Any discrepancies between the registry and publication reports in arm groups, primary outcome, secondary outcome, and adverse events are categorised as 1) discrepancy in study enrolment or arm groups information (the study is considered as selective reporting if only report a subgroup result); 2) discrepancy in the number of primary outcomes; 3) discrepancy in primary outcome description; 4) discrepancy in primary outcome's effect size; 5) discrepancy in primary outcome's effect direction; 6) discrepancy in the number of secondary outcomes; 7) discrepancy in the proportion of resulting positive secondary outcomes; 8) discrepancy in

the number of adverse events; 9) discrepancy in adverse events' incidence rate (Supplementary Fig. S1).

Among them, the effect direction was defined as 'positive' if the treatment effect significantly surpassed the comparison group, defined as 'negative' and vice versa, and defined as 'no difference' if the effect was not statistically significant. The proportion of positive direction is calculated accordingly. For discrepancies in the number of outcomes, we further distinguished those publications that reported fewer outcomes than the registry platform as selective reporting and those that reported more outcomes than registry as complementary reporting. Two researchers independently assessed for discrepancies and calculated the positive rate and adverse event incidence rate. Any uncertainties were resolved through discussions among all authors.

We then navigate the discrepancy from two aspects. First, the distributional differences of basic characteristics between trials that posted results on registries and published results in journals, which include the time frame from the start of the trial to the result posted or published, the distribution of disease, as well as the type of lead sponsorship. Second, the reporting discrepancy in enrolment, outcomes, and adverse events. In addition, we investigate whether both clinical trials registry posted results and peer-reviewed publications serve as evidence sources for AI-related systematic reviews and meta-analyses. We employed a descriptive approach to provide an overview of clinical trials on the application of AI in clinical settings, and analyse discrepancies by comparing the number and proportions of trials reporting results in registry posts and publications.

Role of the funding source

The authors declare no source of funding.

Results

We identified a total of 3710 registered clinical trials in ClinicalTrials.gov and the EU Clinical Trials Registry that were registered in English, targeted human populations, and involved AI algorithm-based interventions. Among these, 1106 trials were completed, and 101 trials had posted results on the registries. An associated publication search for these 3710 clinical trial IDs within Dimensions.ai yielded 1558 potential matches. After removing duplicates and excluding irrelevant studies based on title and abstract screening, the dataset was refined to 488 publications, corresponding to 380 trials. To assess consistency, we matched the publications to the trials with posted results, resulting in 26 trials corresponding to 28 publications. Data extraction from these matched sources yielded 419 outcomes, of which 110 were from publications and 309 from registries. The Cohen's Kappa index between researchers is 0.7, as detailed in Supplementary 3.

The identification and screening process illustrates the pre-registration rate of AI clinical research and the low reporting rate of trial results. From the publication side, 6.60% (1863/28,248) of AI clinical research publications reported pre-registration ID, among which 1558 (83.63%) can be related to trial registered in ClinicalTrials.gov or EU-CTR. With 3710 unique pre-registered clinical trial IDs identified in registries, 1106 trials are completed. 9.14% (101/1106) had results posted on the clinical trial registry, while 34.37% (380/1106) had results published as articles. However, 2.35% (26/1106) of trials had results in both the registration platforms and published literature (Fig. 1).

Starting from 50 trials ten years ago to the peak of 704 trials in 2022, the rapid increase indicates the expanding activity of AI clinical research (Fig. 2). Based on the latest update, a total of 1106 trials have been completed. 1215 trials are still ongoing, while 79 trials are suspended or terminated. For 3688 trials with lead sponsor and funder declaration, we see IITs are leading the AI clinical research, taking up 57.7%–66.1% of all registered trials in the past five years. Interventional and observational trials account for about an equal share of registered clinical trials, with observational studies slightly outnumbering interventional studies since 2020.

Preferences in reporting results

The analysis of result reporting across dissemination sources reveals a notable disparity, with a higher number of results reported in publications (380) compared to those posted on trial registries (101). Although IITs represent 61.1% of the total registered AI clinical trials, the majority of trials (354 out of 380) did not have corresponding registry results posted. With 455 trials that have either registry-posted or journal-published results, IITs and non-IITs contribute almost equally to the current evidence base regardless of dissemination sources (Fig. 3). Interventional trials, on the other hand, contribute to 62.9% of all results (57.9% of publications and 91.9% registry post results), with a 47.5% registration share, indicating higher completion in reporting results.

Fig. 4 illustrates the time differences in the reporting of results between publication and registry posts. 16 trials (15.84%) reported results on registry platforms within one year since the start of trial. The percentage of trials posting results on registries increases over time, with 39.60% (40 trials) reporting results within two years and 57.43% (58 trials) within three years. By the four-year mark, 68 (67.33%) registries result have been posted. In contrast, publication percentages show a steeper initial increase, with 37.63% (181 publications) publishing results within one year, rising to 56.34%

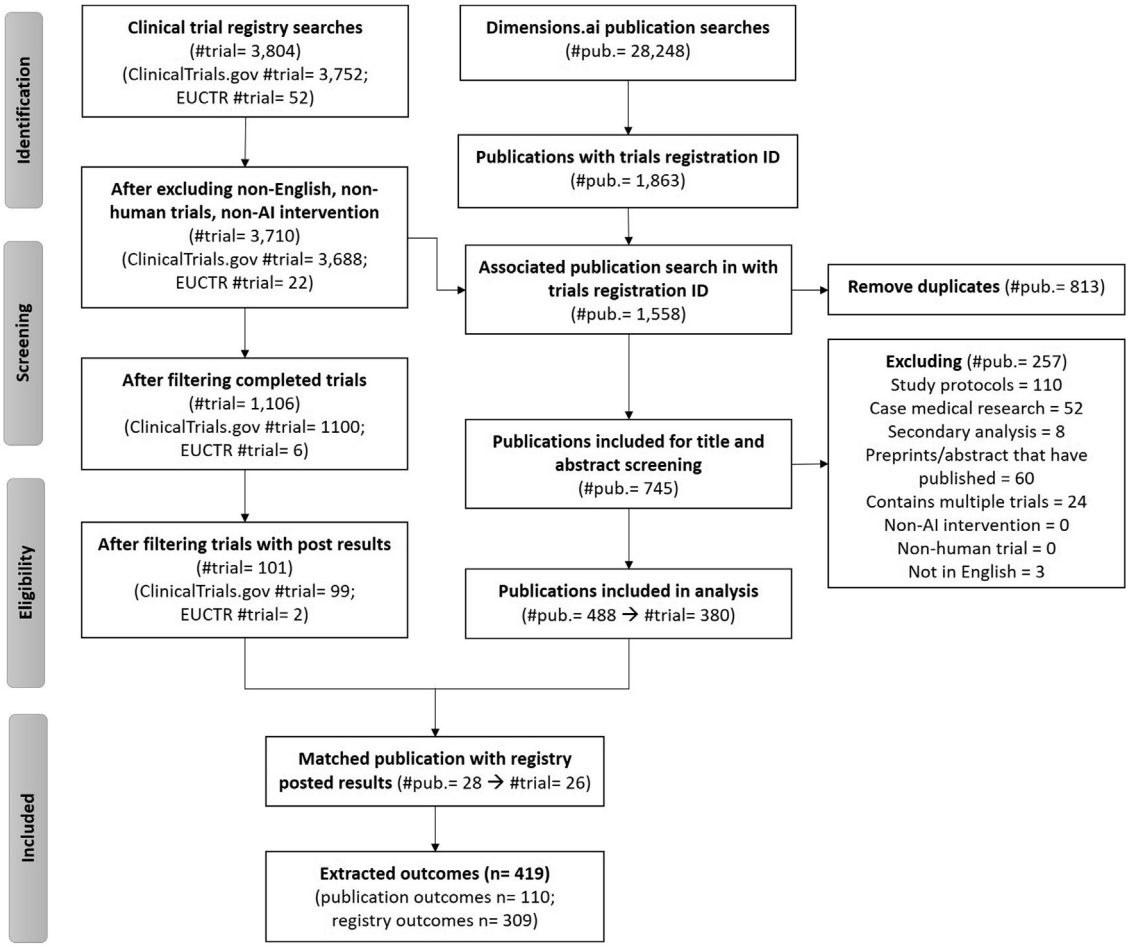

***Fig. 1:* Data identification and screening flowchart.** Note: #pub. stands for number of publications; #trial stands for number of registered clinical trials.

within two years (additional 90 publications). By the time trials reach three years, 70.48% (additional 68 publications) had published in journals, and this proportion and further rise to 78.59% (additional 39 publications) within four years. This indicates a more rapid dissemination of results through publications, with a steady increase in both registry posting and publication over time.

Disease types between dissemination sources

The comparative analysis of disease types in 101 pre-registered AI clinical trials with available posted results and 488 publications with available pre-register IDs reveal a distinct variation of disease types. Registry posts are heavily concentrated in nervous system diseases, with 22 trials taking up 21.8%, followed by mental disorders, with 11 trials taking up 10.9%. Peer-reviewed publications, on the other hand, exhibit a wider distribution across disease types but a notable emphasis on neoplasms (12.9%) and cardiovascular diseases (10.8%) (Supplementary Fig. S2).

Discrepancies and selective reporting

Table 1 classified discrepancies observed between clinical trial registry posts and their corresponding peer-reviewed publications. Notably, there were no cases of complete agreement, indicating a discrepancy in each instance examined. The most frequent discrepancies were related to the reporting of secondary outcomes and adverse events, with 67.9% (19/28) and 57.1% (16/28) of cases showing a discrepancy, respectively. Discrepancies in enrolment information were found in 32.1% of cases, while differences in the primary outcomes scored 21.4%. In addition, the omission of data is prevalent, leaving most of the secondary and adverse event outcomes size incomparable.

Out of the 28 discrepancies identified, selective reporting in publications was a prominent issue,

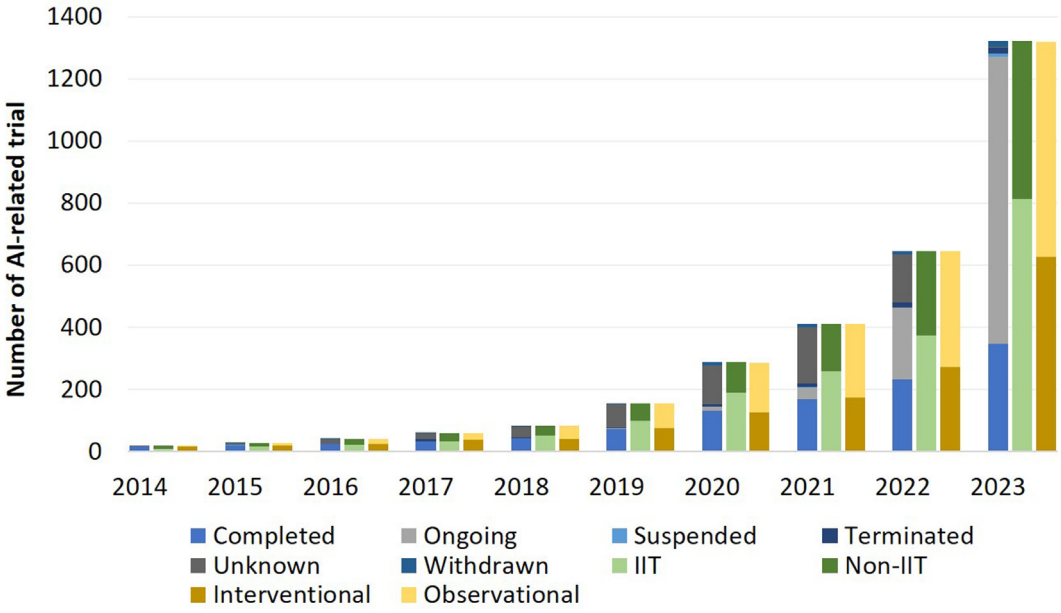

***Fig. 2:* Distribution of AI clinical trial registrations (2014–2023).** The combined bar chart showed analyses of clinical trial registrations of AI in clinical settings, categorised by trial status, sponsor type, and research design characteristics. Over the past decades, AI clinical trials increased significantly. As of the latest update, 1106 trials have been completed and 1215 trials are still ongoing. Investigator-initiated trials (IITs) have been leading the research registrations for the past five years, and observational trials have outnumbered interventional trials since 2020.

affecting 53.6% of cases for both secondary outcomes and adverse events. Additionally, 14.3% of publications provided complementary results for both primary and secondary outcomes. In other words, there were more reported outcomes in trial registries than in peer-reviewed literature. By comparing the discrepancies in primary outcome effect direction, effect size, and overall adverse events incidence rate of each trial, we also noticed that the understatement of adverse event incidence rate is emerging and worth noting (Fig. 5).

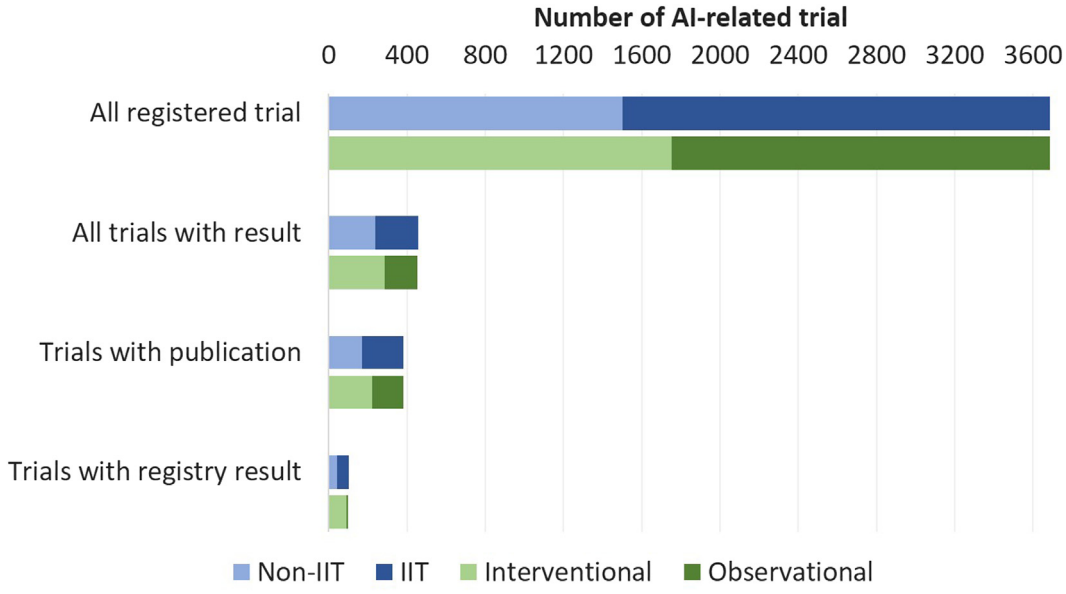

***Fig. 3:* Comparison of reporting preferences in publications and trial registries.** The bar chart shows that reporting through publications is favoured over registry posts. Investigator-initiated trials (IITs) and non-IITs contribute equally to the evidence base. Interventional trials account for 62.9% of the evidence base.

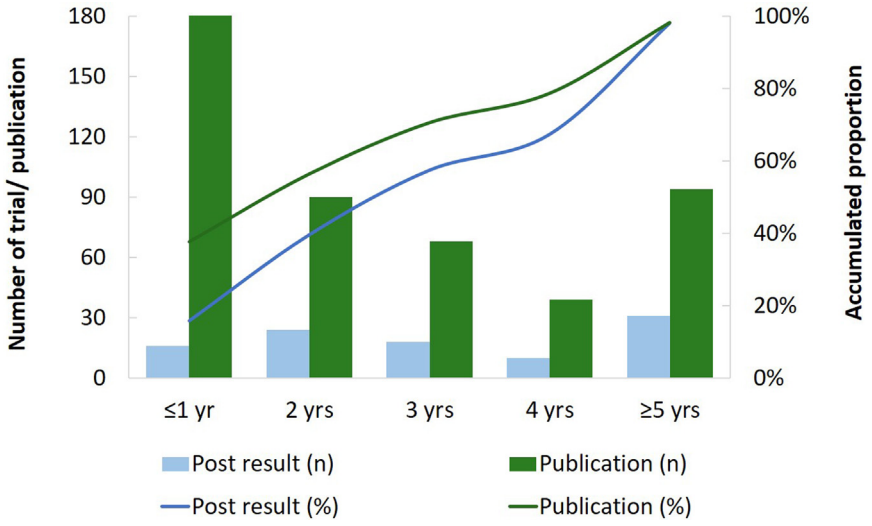

Fig. 4: **Time differences of reporting results in publication and registry posts.** This combo chart shows a significant increase in result accessibility over time. 37.6% of publications occur within a year of the trial start, while registry postings steadily increase from 15.8% within one year to 67.3% by the four-year mark.

Use of trial registries in AI evidence synthesis

The examination of data source usage in systematic reviews and meta-analyses pertaining to AI clinical research reveals a notable absence of evidence posted in clinical trial databases, with 96.23% of systematic reviews and meta-analyses extracted from a variety of publication databases, including Web of Science, PubMed, Embase, etc. Studies that used trial registry databases or both registry and publication databases are collectively represented by less than 4% combined (Supplementary Fig. S3).

Discussion

The rapid advancement of AI presents unprecedented opportunities to enhance patient outcomes. This study highlights key issues and trends in the dissemination of AI research, based on clinical trials extracted from registries such as ClinicalTrials.gov and publications identified via Dimensions.ai. A significant number of trial results are reported solely in trial registries or publications, underscoring the complementary roles these sources play in disseminating research findings. However, the discrepancies between clinical trial registry posts and corresponding publications are concerning. Selective reporting, particularly of secondary outcomes and adverse events, was prevalent, with over half of the cases exhibiting discrepancies, and the underreporting of adverse events was common. These findings raise concerns about the reliability of the current evidence base on the application of AI in clinical settings. Our results align with previous studies on reporting discrepancies across sources and contribute further insight into this important issue.[3–6] The prominence of selective reporting in publications suggests that not all relevant outcomes may be reported, potentially leading to a biased representation of trial results. Jer'ome Adda (2020) argued that selectiveness could explain the excess of significant results in phase III for large industry sponsors, hinting at potential selective reporting and data manipulation.[22] Ensuring transparency in the reporting of AI methodologies and results is essential for the credibility and trustworthiness of AI applications in healthcare.

Classification of discrepancy (n = 28 pairwise trial-publications)	Discrepancy No. (%)	Omission No. (%)
All	28 (100.0%)	0 (0.0%)
Enrolment	9 (32.1%)	0 (0.0%)
Number of primary outcomes	6 (21.4%)	0 (0.0%)
Selective reporting of publication	2 (7.1%)	/
Complementary reporting of publication	4 (14.3%)	/
Primary outcome description	4 (14.3%)	0 (0.0%)
Primary outcome's effect size	6 (21.4%)	9 (32.1%)
Primary outcome's effect direction	2 (7.1%)	9 (32.1%)
Number of secondary outcomes	19 (67.9%)	0 (0.0%)
Selective reporting of publication	15 (53.6%)	/
Complementary reporting of publication	4 (14.3%)	/
Secondary outcomes' effect proportional direction	0 (0.0%)	26 (92.9%)
Number of adverse events	16 (57.1%)	0 (0.0%)
Selective reporting of publication	15 (53.6%)	/
Complementary reporting of publication	1 (3.6%)	/
Adverse events' incidence rate	5 (17.9%)	21 (75.0%)

Table 1: **Descriptive statistics of discrepancies between results posted on the clinical trial registry and their corresponding publications by classification.**

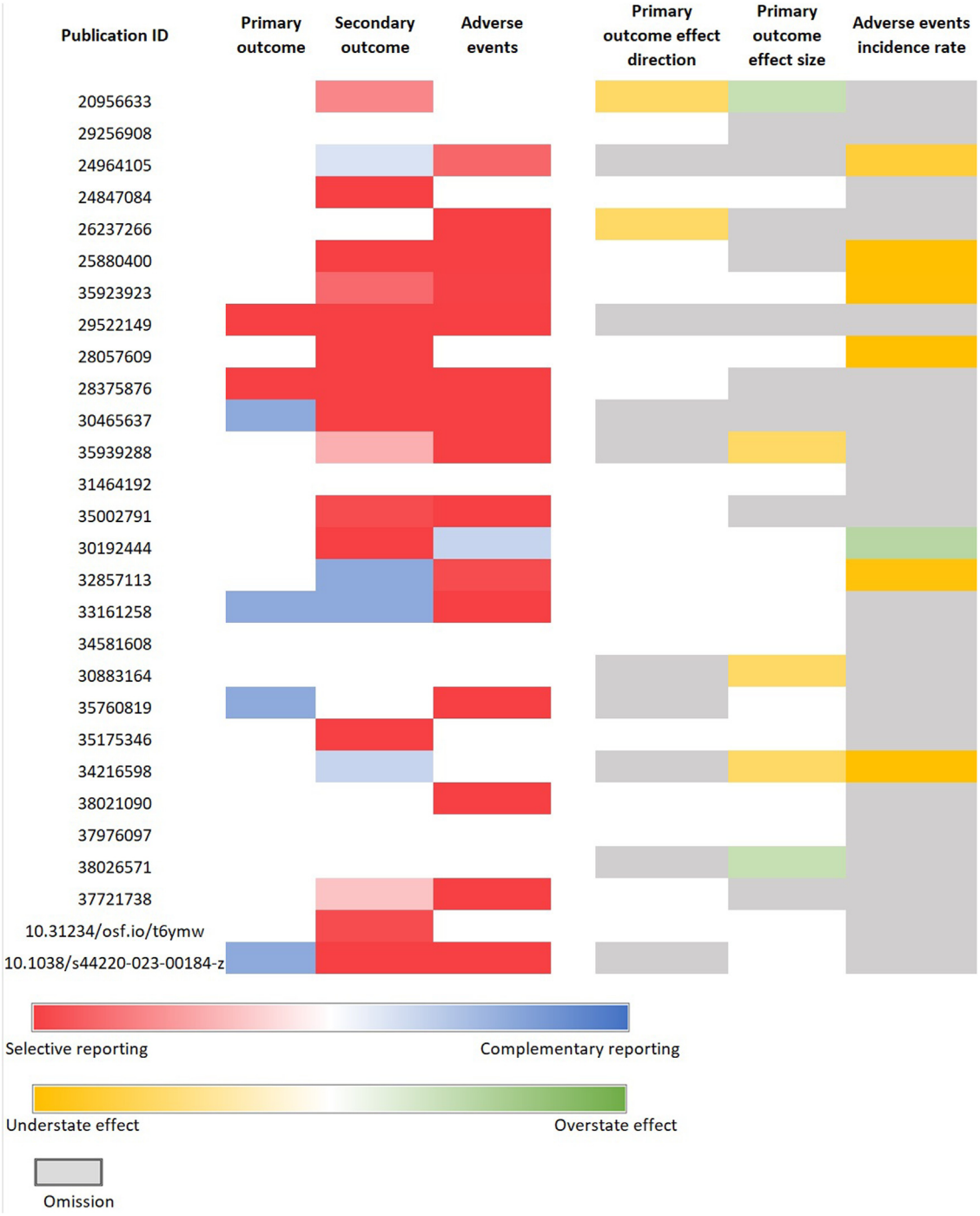

Fig. 5: **Heatmap of reporting characteristics in publications.** The heatmap illustrates reporting integrity and accuracy across the publications. Of the 28 discrepancies identified, 53.6% involved selective reporting, particularly concerning secondary outcomes and adverse events, while 14.3% of publications provided complementary results. The analysis also highlights an emerging trend of underreporting adverse event incidence rates in the publications. Note: The heatmap uses a colour gradient to represent discrepancies: the darkest colours at each end of the scale correspond to cases where the publication reports results 80% less or more than the registry. The white colour denotes no discrepancy, while grey indicates the omission of data.

On the aggregated level, the temporal trend in result accessibility reveals a clear distinction in the dissemination of results, with a preference for rapid sharing through publications over registry posts, despite that registry posts often require less effort and it is not conflict with seeking publication. This observation highlights the higher value placed on peer-reviewed results within the academic community. However, the swift appearance of results as journal articles suggests the possibility of post-registration rather than pre-registration, as we observed 181 publications that were published within one year of the trial start date, potentially due to the submission requirements of journals. We, therefore, emphasise that timely result posting is essential for maintaining scientific integrity and ensuring that all findings, positive or negative, contribute to the collective knowledge. Additionally, the heavy reliance on publication databases for secondary research in AI highlights an underutilisation of clinical trial data, similar to trends in pharmaceutical research. This could restrict the depth and accuracy of systematic reviews and meta-analyses, potentially leading to incomplete or biased evidence syntheses.

More importantly, a significant portion of trial research fails to comply with registration requirements, and results are not always publicly accessible. In September 2004, the International Committee of Medical Journal Editors (ICMJE) mandated that trials intended for publication in affiliated journals must be registered in a publicly accessible database before enrolling participants. However, there is ongoing confusion in AI research regarding which studies require registration. This distinction is critical, as only clinical trials are required to be registered. To address this uncertainty, some journals have implemented policies that require registration for any study involving the prospective collection of human data to assess intervention efficacy, regardless of how the study is categorised by ethical review boards.[23,24]

Further, when conducting the review, we found studies with significant omission in adverse event outcomes reached positive conclusions on AI intervention. Researchers have warned about the poor quality of AI applications in medical imaging publications and possible exaggeration of AI performance compared to clinicians.[17] The quality of evidence in many AI clinical research raises concerns about result interpretation, underscoring the need for standardised and interpretable guidelines for AI clinical trials. Fortunately, the growing field of AI in healthcare has led to the development of reporting standards to enhance research quality and transparency.[25] For instance, the SPIRIT-AI and the CONSORT-AI emphasis on transparent reporting of clinical trials assessing the performance of AI in clinical settings. Both sets of guidelines stress the importance of clearly detailing AI interventions, comparisons with other treatments, and data processing methods to ensure scientifically rigorous and clinically meaningful results.[26,27]

Building on the above, our study offers recommendations to enhance the transparency and accuracy of evidence reporting in AI clinical research. For researchers, we emphasise the importance of utilising multiple data sources, including publications and trial registries, to provide a more comprehensive evidence base. Given the scarcity of robust clinical evidence for AI applications and increasing demands from regulatory bodies such as the FDA for stronger evidence, researchers should aim to maximise the use of existing data while ensuring methodological rigor. For regulators, we recommend extending the principles of the 2007 FDA Amendments Act (FDAAA), which mandates trial registration and results reporting for most Phase II-IV FDA-regulated trials, to AI-based clinical trials. This would require pre-registration and mandatory reporting of participant demographics, outcomes, and adverse events, aligning AI trials with established practices for drug and device trials.

Additionally, systematic approaches should be adopted to ensure consistent reporting throughout the trial lifecycle, which would involve verifying trial features, tracking discrepancies, and providing explanations for any differences between registry and publication results. Finally, journal editors could require authors to submit a link to the trial's registration entry and an itemised comparison of key trial features during the manuscript review process. These measures would enhance transparency and accountability, ultimately strengthening the evidence base for AI in healthcare.

This study has several limitations. First, trial status is based on the most recent updates in the registries, but we cannot verify whether these updates accurately reflect the current trial status. Second, our search was limited to English-language publications, potentially excluding relevant non-English studies and affecting the comprehensiveness of our findings. Third, this study does not include a risk of bias analysis, as its primary objective is to examine discrepancies between sources. Our results could also be limited by the coverage of Dimensions.ai, we have supplemented our search with PubMed, Embase, and Web of Science, though incomplete database coverage may still affect the results. Besides, our analysis is limited by the observation period, with any unreported results considered selectively or complementarily reported. Similarly, post-publication revisions or updates may introduce discrepancies between registry data and published results.

In conclusion, while the application of AI in clinical research is expanding, significant issues persist regarding the registration and reporting of trial results. Although trial registries and publications serve complementary roles in research dissemination,

discrepancies between these sources raise concerns about the reliability and comprehensiveness of the current evidence base. Addressing these discrepancies and ensuring the timely and accurate reporting of all relevant outcomes is essential. To this end, we emphasise the importance of pre-registration, adherence to established reporting guidelines, and the implementation of systematic procedures to enhance transparency and accountability. By upholding rigorous methodological standards and ethical practices, the integration of AI in clinical settings can be advanced, with a more robust and reliable evidence base to support its application in healthcare.

Contributors

Z.H. and J.D. conceptualised and designed the study. Z.H., L.Y and X.L is responsible for data collection, data cleaning, data extraction and verification. Z.H conducted the analyses, and J.D. contributed to the interpretation of results. The initial manuscript was drafted by Z.H., amendments are made by Z.H. and J.D. All authors had access to the data and all authors have verified the published version of the manuscript.

Data sharing statement

All data utilised in this study were obtained from open-access sources or through institutional access. No data were collected directly from participants, and no additional documents were generated.

Declaration of interests

J.D. received funding from National Key R&D Program for Young Scientists (Project number 2022YFF0712000) and the National Natural Science Foundation of China (Project number 72074006). All other authors declare no competing interests.

Acknowledgements

None.

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.eclinm.2024.103066.

References

1 Herbet M-E, Leonard J, Santangelo MG, Albaret L. Dissimulate or disseminate? A survey on the fate of negative results. *Learn Publ*. 2022;35(1):16–29.
2 Han R, Acosta JN, Shakeri Z, Ioannidis JPA, Topol EJ, Rajpurkar P. Randomised controlled trials evaluating artificial intelligence in clinical practice: a scoping review. *Lancet Digit Health*. 2024;6(5):e367–e373.
3 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):477–483.
4 Choudhury MC, Chakraborty I, Saberwal G. Discrepancies between FDA documents and ClinicalTrials.gov for Orphan Drug-related clinical trial data. *PLoS Glob Public Health*. 2022;2(4):e0000261.
5 Becker JE, Krumholz HM, Ben-Josef G, Ross JS. Reporting of results in ClinicalTrials.gov and high-impact journals. *JAMA*. 2014;311(10):1063–1065.
6 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. discussion e.
7 Zarin DA, Fain KM, Dobbins HD, Tse T, Williams RJ. 10-Year update on study results submitted to ClinicalTrials.gov. *N Engl J Med*. 2019;381(20):1966–1974.
8 Baudard M, Yavchitz A, Ravaud P, Perrodeau E, Boutron I. Impact of searching clinical trial registries in systematic reviews of pharmaceutical treatments: methodological systematic review and reanalysis of meta-analyses. *BMJ*. 2017;356:j448.
9 Dechartres A, Atal I, Riveros C, Meerpohl J, Ravaud P. Association between publication characteristics and treatment effect estimates: a meta-epidemiologic study. *Ann Intern Med*. 2018;169(6):385–393.
10 LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521(7553):436–444.
11 Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*. 2017;542(7639):115–118.
12 Guan Z, Li H, Liu R, et al. Artificial intelligence in diabetes management: advancements, opportunities, and challenges. *Cell Rep Med*. 2023;4(10):101213.
13 Khosravi M, Zare Z, Mojtabaeian SM, Izadi R. Artificial intelligence and decision-making in healthcare: a thematic analysis of a systematic review of reviews. *Health Serv Res Manag Epidemiol*. 2024;11:23333928241234863.
14 Secinaro S, Calandra D, Secinaro A, Muthurangu V, Biancone P. The role of artificial intelligence in healthcare: a structured literature review. *BMC Med Inf Decis Making*. 2021;21(1):125.
15 World Health Organization. *Classification of digital interventions, services and applications in health: a shared language to describe the uses of digital technology for health, 2nd ed*. 2023. Geneva.
16 Denniston AK, Liu X. Responsible and evidence-based AI: 5 years on. *Lancet Digit Health*. 2024;6(5):e305–e307.
17 Nagendran M, Chen Y, Lovejoy CA, et al. Artificial intelligence versus clinicians: systematic review of design, reporting standards, and claims of deep learning studies. *BMJ*. 2020;368:m689.
18 Mello MM, Guha N. ChatGPT and physicians' malpractice risk. *JAMA Health Forum*. 2023;4(5):e231938.
19 Choudhury A, Shamszare H. Investigating the impact of user trust on the adoption and use of ChatGPT: survey analysis. *J Med Internet Res*. 2023;25:e47184.
20 Ito T. Differences in investigator-initiated trials between Japan and other countries: analyses of clinical trials sponsored by academia and government in the ClinicalTrials.gov registry and in the three Japanese registries. *PLoS One*. 2016;11(2):e0148455.
21 Paleoudis EG, Pinto C. Investigator-initiated vs. Investigator-sponsored research: definitions matter. *J Clin Med Res*. 2024;16(1):31–32.
22 Adda J, Decker C, Ottaviani M. P-hacking in clinical trials and how incentives shape the distribution of results across phases. *Proc Natl Acad Sci USA*. 2020;117(24):13386–13392.
23 Haug CJ, Drazen JM. Artificial intelligence and machine learning in clinical medicine, 2023. *N Engl J Med*. 2023;388(13):1201–1208.
24 Drazen JM, Haug CJ. Trials of AI interventions must Be preregistered. *NEJM AI*. 2024;1(4):AIe2400146.
25 Kolbinger FR, Veldhuizen GP, Zhu J, Truhn D, Kather JN. Reporting guidelines in medical artificial intelligence: a systematic review and meta-analysis. *Commun Med*. 2024;4(1):71.
26 Rivera SC, Liu X, Chan AW, Denniston AK, Calvert MJ. Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI Extension. *BMJ*. 2020;370:m3210.
27 Liu X, Rivera SC, Moher D, Calvert MJ, Denniston AK. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI Extension. *BMJ*. 2020;370:m3164.