TrialScout links published results to trial registrations using a large language model

Authors: Love von Schreeb[1], Till Bruckner[1,2,3], Darya Ava Aspromonti[1], Laura Caquelin[1], Jamie Cummins[4,5], Nicholas J. DeVito[5], Cathrine Axfors[6,7,8], John P.A. Ioannidis[8,9,10,11], Gustav Nilsonne[1,8,12]

[1]: Karolinska Institutet, Department of Clinical Neuroscience, Stockholm, Sweden

[2]: UiT the Arctic University of Tromsø, Tromsø, Norway

[3]: TranspariMED, Stockholm, Sweden

[4]: Department of the Psychology of Digitalisation, University of Bern, Bern, Switzerland

[5]: Bennett Institute for Applied Data Science, Nuffield Department of Primary Care Health Sciences, University of Oxford, UK

[6]: Center for Pharmacoepidemiology, Division of Clinical Epidemiology, Department of Medicine Solna, Karolinska Institutet, Stockholm, Sweden

[7]: Department of Women's and Children's Health, Uppsala University, Uppsala, Sweden

[8]: Meta-Research Innovation Center at Stanford (METRICS), Stanford University School of Medicine, Stanford, California, USA

[9]: Stanford Prevention Research Center, Department of Medicine, Stanford University School of Medicine, Stanford, California, USA

[10]: Department of Epidemiology and Population Health, Stanford University School of Medicine, Stanford, California, USA

[11]: Department of Biomedical Data Science, Stanford University School of Medicine, Stanford, California, USA

[12]: International Globally Distributed Organization for Research and Education (IGDORE), Sweden

Abstract

Background and Objective: Multiple stakeholders need to locate results of registered clinical trials but frequently struggle to find them. Summary results of clinical trials are often not published in trial registries, and publications containing trial results are often not explicitly linked to their respective trial registrations. Finding these results is important to researchers, systematic reviewers, research funders, regulators, clinical practitioners, and patients. **Methods:** We developed TrialScout, a computer program that uses a large language model to match clinical trials registered on ClinicalTrials.gov with corresponding result publications indexed in PubMed. TrialScout's performance was evaluated through comparison to human-coded matches from previous studies of results reporting rates. Subsequently, TrialScout was applied to a random sample of 9,600 completed or terminated trials. **Results:** TrialScout had a sensitivity of 92.5% and a specificity of 81.2% compared to human coders. Manual review of 200 cases where TrialScout disagreed with human researchers showed that a majority (123/200, 61.5%, 95% CI, 54.4–68.3%) of disagreements were due to human errors. When used on 9,600 sampled trials in ClinicalTrials.gov, TrialScout found result publications for 6,110 (63.6%) of trials. **Conclusions:** TrialScout reliably located results of completed clinical trials. The tool offers benefits in terms of speed and efficiency. Estimating TrialScout's accuracy is limited by the lack of a true gold standard. TrialScout can accelerate the process of locating trial results in the scientific literature and can assist in monitoring trial reporting practices.

Keywords: Clinical Trials, Evidence Synthesis, Data Science, Publication Rates, Publication Bias, Research Transparency

What is new

- TrialScout matches published results to clinical trial registrations
- TrialScout uses a large language model to compare publications to registrations
- Accuracy rivals searching by human researchers
- Most disagreements between TrialScout and humans were due to human error
- TrialScout found published results for 63.6% of 9,600 trials from clinicaltrials.gov

1. Background

Clinical trials advance medical research by evaluating the efficacy of interventions. However, advancement of knowledge occurs only if results are reported (1). The Declaration of Helsinki

articulates an ethical obligation to publish clinical trial results in a timely manner (2). Non-reporting hinders medical research by biasing evidence reviews (3) and leading to redundant research (1,4), which can expose trial participants to avoidable risks and harms (5) without scientific justification. Non-reporting of trial results is thus unethical (6) and can be considered scientific misconduct (3,7).

Despite guidelines and legislation (8,9) promoting timely publication of clinical trial results, results are not always made public (6,10–13). A recent Cochrane systematic review covering 204 research reports on interventional studies registered in ClinicalTrials.gov estimated that only 53% of trials had results published in a journal (14). When completed trials lack results in clinical trial registries, it is often unclear whether results have been published elsewhere. ClinicalTrials.gov sometimes links to publications from trial registrations via researcher submissions and automatic detection of National Clinical Trial numbers (NCT-ID) found in PubMed abstracts and metadata. However, links remain incomplete. In a study of trials conducted by German university medical centres, 50% of registrations did not contain links to their result publications (15). Furthermore, many trial publications do not reference their trial registration numbers (15), despite established guidelines (16–18). Finding trial results often requires searching multiple databases, reading candidate publications, and cross-referencing registration data, which is arduous and time-consuming.

Automated methods to detect results have been developed. Powell-Smith and Goldacre developed TrialsTracker, which detects results published to ClinicalTrials.gov and PubMed publications mentioning the trial's NCT-ID in the abstract or metadata (19). However, TrialsTracker's reliance on explicit NCT-ID links limits its sensitivity. Smalheiser and Holt developed a logistic regression model that predicts linkage between trial registrations and PubMed articles using metadata similarity, reporting a sensitivity of 84.6% and a positive predictive value of 90.4% (20). Goodwin and colleagues developed a neural network approach for the same purpose (21). The Smalheiser-Holt model is reported to perform better, but takes several minutes to run per registration since it performs pairwise comparisons across many publications (20). As the PubMed database grows, this computational cost increases, limiting scalability. Appendix 2 gives an overview of currently available tools for automated results detection.

Despite the growing body of evidence of non-reporting of trial results, the extent of the problem is known only for subsets of trials that have been studied, not necessarily representing the population at large. Furthermore, most previous studies have relied on labour-intensive manual searches to find published results, limiting sample sizes. In this study we therefore developed and evaluated a novel tool for the discovery of published clinical trial results. Subsequently, we applied this tool to a large random sample of completed or terminated trials registered in ClinicalTrials.gov, as a proof-of-principle and to provide initial evidence on overall reporting rates.

2. Methods

2.1 Definitions

We defined the following operational terms: *Summary results* are tabular, structured results of clinical trials submitted to the trial registry. *Published results* are results of a clinical trial contained in a peer-reviewed scientific publication, in a thesis, preprint, or published conference abstract. *Reported results* are results of a trial available as summary results and/or published results. This definition of what constitutes results publication is adopted from our previous study on result reporting across the Nordic countries (22).

2.2 Development of TrialScout

We developed an automated tool, TrialScout, using Node.js (version 22.7.0) (23). TrialScout retrieves trial registration metadata, searches for candidate publications using a predefined algorithm, and prompts an LLM with both the trial registration and publication abstracts to determine whether any candidate publications report trial results. This filters irrelevant candidate publications and distinguishes result publications from other publication types such as study protocols, systematic reviews, and meta-analyses. TrialScout was designed to find *published results* indexed in PubMed for clinical trials registered on ClinicalTrials.gov, EU Clinical Trials Register (EUCTR), and the German Clinical Trials Register (DRKS). A key constraint of TrialScout is that it can only find results in PubMed with an available abstract.

TrialScout was iteratively designed. Prototypes were run on a subset of the *validation dataset* (described below) constituting <10% of the total dataset. Search strategies and LLM prompts were refined based on manual review of misclassifications. Development concluded when preliminary tests showed high agreement with human data (approximately 85-90% sensitivity and specificity). Figure 1 illustrates TrialScout's technical implementation. A web version is

freely available at https://metaresearch.se/trialscout, and further technical details are provided in Appendix 1 and the source code (https://github.com/lahnstrom/trialscout). TrialScout used ChatGPT-5.1 (Version "gpt-5.1-2025-11-13") for all inferences, with temperature set to 1.0 and the reasoning effort parameter set to "medium". The average cost was 0.043 USD per trial. Runtime varied depending on external API rate limits, hardware, and model settings. TrialScout processed the validation dataset with 5,774 trials in approximately one day on an M2 MacBook Air with 8GB of RAM.

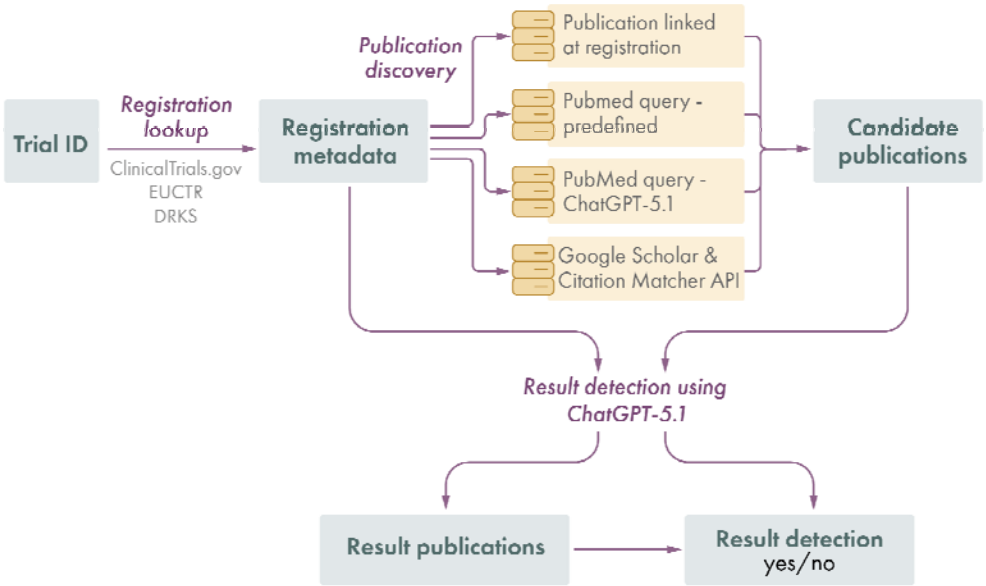

Figure 1. Overview of TrialScout's result detection algorithm. Dark blue rectangles represent data. Light blue ellipses represent processes. Light grey cylinders represent search strategies. Arrows represent the flow of data. Abbreviations: API: Application Programming Interface; GPT: Generative Pre-trained Transformer; EUCTR: EU Clinical Trials Register; DRKS: German Clinical Trials Register.

2.3 Validation of TrialScout Against Human Assessment

We constructed a *validation dataset* with human-curated data as a reference standard, by merging the IntoValue dataset of German clinical trials from 2009–2017 (N=3,790) (24,25) and the dataset from our previous study on Nordic trials completed between 2016–2019 (N=2,112) (22,26). Trials were registered on ClinicalTrials.gov (Nilsonne et al.: N=2,007, IntoValue: N=3,132), EU Clinical Trials Register (EUCTR) (Nilsonne et al.: N=431), and the German Clinical Trials Register (DRKS) (IntoValue: N=658). In the Nordic dataset, 326 trials were cross-registered in both EUCTR and ClinicalTrials.gov; each was classified under the registry of its primary identifier in the source dataset (174 under EUCTR, 152 under

ClinicalTrials.gov). After removing 128 duplicates, the validation dataset contained 5,774 registrations.

Validation data had been collected through rigorous processes detailed in the source studies (22,24), where at least two independent researchers conducted manual searches of PubMed and Google to find trial results, with discrepancies resolved by consensus (22,24). In the dataset from Nilsonne and colleagues, reviewers fully agreed on 75% of trials, while in 11% of cases one reviewer found a publication that the other missed (22). Researchers were instructed to locate only the earliest available result publication, meaning that there is no full publication record for each trial. The validation dataset also contained only pooled consensus judgments rather than per-reviewer data. The definition of a published result varies slightly between the two datasets: the IntoValue dataset included peer-reviewed journal articles and dissertations (24), while the Nilsonne et al. dataset additionally included preprints, congress abstracts, and letters to the editor (22). However, peer-reviewed journal articles make up the vast majority of identified publications in both datasets. In total, 74.4% (n=4,295) of trials had reported results in any format, and 72.5% (n=4,186) had an identified publication.

TrialScout was applied to all 5,774 trials in the validation dataset, and its findings were compared to the human-collected data. Performance characteristics were calculated, including sensitivity, specificity, positive and negative predictive values, and F-score. A true positive was defined as the identification of *any* valid result publication for a trial, rather than the exact one found by humans, since many trials had multiple publications whereas the validation dataset contained only the earliest. Publications identified by TrialScout published after the manual search date cutoffs (November 11, 2020, for the IntoValue dataset; February 15, 2023, for Nilsonne et al.) were excluded to avoid penalizing TrialScout for correctly finding publications that appeared after data collection was completed.

To further validate TrialScout's accuracy, we manually reviewed a random sample of 200 cases where its assessment disagreed with human reviewers. This sample included 100 false positives (publications found by TrialScout but not humans) and 100 false negatives (publications missed by TrialScout). Based on the search manual by Nilsonne et al. (22), two reviewers independently judged whether the publications matched the trial registrations. Any conflicts were resolved by consensus. Each confirmed TrialScout error was then investigated and categorized.

2.4 Performance of TrialScout on ClinicalTrials.gov data

We applied TrialScout to a random sample of trials from ClinicalTrials.gov. A total of 571,121 trial registrations were retrieved and downloaded on February 12, 2026 (Figure 2). Only completed or terminated interventional trials were included, as non-interventional studies generally do not fall under the same legal and ethical requirements as interventional trials. Trials completed before September 30, 2022 were eligible, allowing approximately 3.5 years to report results. Completion was defined by the primary completion date, or the study completion date if the primary completion date was not available. Trials lacking a completion or termination date were excluded. From eligible trial registrations (N=220,167), 9,600 were randomly sampled using the "sample_n" function from the R (version 4.2.2) (27) package "dplyr" (version 1.1.4) (28). The sample size was chosen based on a precision-based sample size calculation with a desired 95% confidence interval of ±1% and an estimated proportion of results reporting of 50%, using the "prec_prop" function from the R package "presize" (version 0.3.7) (29).

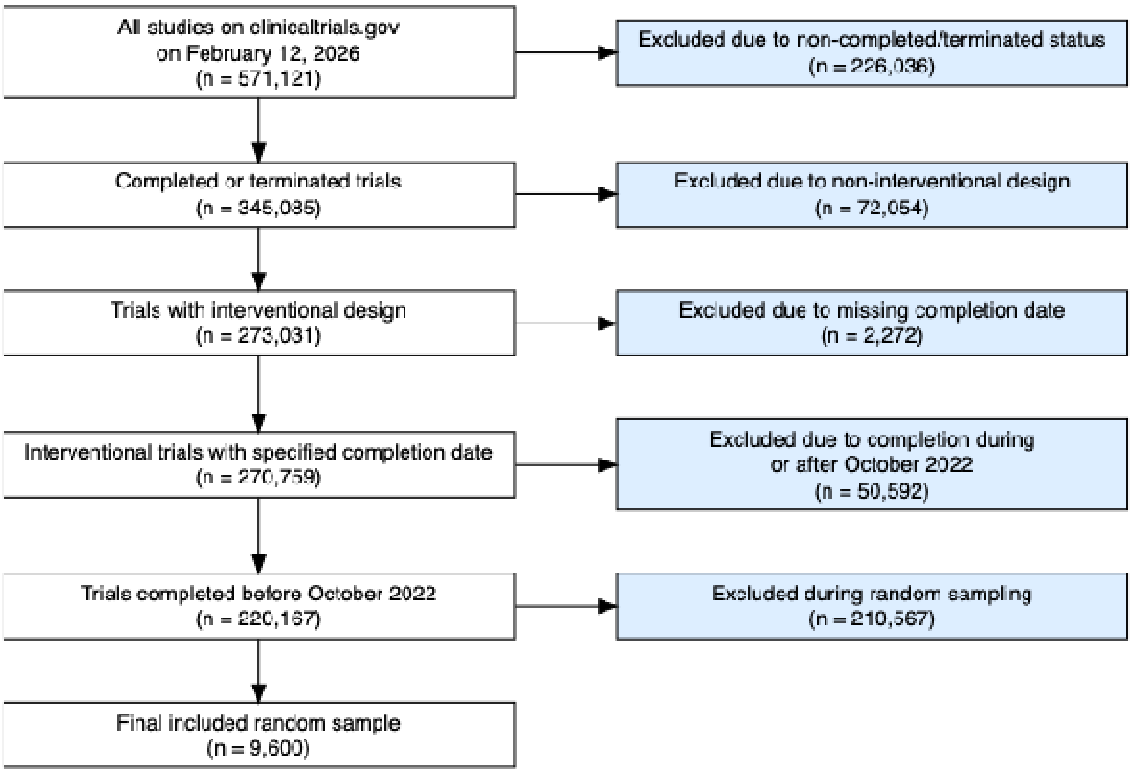

Figure 2. Inclusion flowchart. This figure shows the inclusion and exclusion process for clinical trials retrieved from ClinicalTrials.gov. Only interventional trials with valid date information completed before October 2022 were eligible for random sampling, of which 9,600 were selected.

In February 2026, TrialScout was applied to all clinical trials in the random sample (n=9,600). The primary outcomes were the rates of *published results*, defined as findings published in a peer-reviewed journal found by TrialScout; and *reported results*, a broader measure indicating that summary results were submitted directly to the trial registry or that published results were found by TrialScout. Data on summary results availability was obtained directly from the ClinicalTrials.gov Application Programming Interface (API). Results reporting rates for trial subgroups based on phase, completion status, participant sex, and funder type were analysed using descriptive statistics and Pearson's chi-squared test. Omnibus tests on the full contingency tables were followed by post-hoc chi-squared tests for hypotheses of lower reporting among industry-funded, male-only, and early-phase (phase 1 and early phase 1) trials. Separately, logistic regression models examined associations between trial enrolment and two outcomes: result reporting and result publication. R (version 4.2.2) was used for all statistical analyses. The analysis code is available on GitHub (https://github.com/lahnstrom/trialscout).

3. Results

3.1 Validation of TrialScout Against Human Assessment

In the validation dataset, TrialScout performed well against the human reference standard (N=5,774 trials) (Table 1). The F-score was 92.7%. The tool detected published results for 72.3% (n=4,173) of trials in the validation dataset, which aligns well with the proportion of trials for which humans found results, 72.5% (n=4,186). Among trials with an associated PubMed-indexed result publication found during manual searches (n=3,914), TrialScout found that publication in 95.4% of cases (n=3,734) and identified that publication as containing results in 91.6% of cases (n=3,585). Performance varied by how the publication was originally found: for trials with results found via links from the registration (n=2,140), TrialScout found results for 98.2% (n=2,101), compared to 86.2% (n=1,668) for trials found through systematic Google search (n=1,934). TrialScout's agreement with humans varied between registries (Table 1).

Table 1. TrialScout validation performance by trial registry.

	ClinicalTrials.gov (n=4,837)	EUCTR (n=279)	DRKS (n=658)	All (N=5,774)

	ClinicalTrials.gov (n=4,837)	EUCTR (n=279)	DRKS (n=658)	All (N=5,774)
True positive	3,221	209	444	3,874
False positive	251	9	39	299
False negative	235	13	64	312
True negative	1,130	48	111	1,289
Sensitivity	93.2%	94.1%	87.4%	92.5%
Specificity	81.8%	84.2%	74%	81.2%
Positive predictive value	92.8%	95.9%	91.9%	92.8%
Negative predictive value	82.8%	78.7%	63.4%	80.5%
F-score	93.0%	95.0%	89.6%	92.7%

Performance characteristics of TrialScout when compared to human searches, by trial registry. Abbreviations: EUCTR: EU Clinical Trials Register; DRKS: Deutsches Register Klinischer Studien (German Clinical Trials Register).

To investigate discrepancies between TrialScout and human assessment, a random sample of 200 discrepant cases was manually reviewed by two independent reviewers. Of 100 false positive cases, 79 (79.0%, 95% CI, 69.7–86.5%) were in fact true positives missed by the original manual search (and published before the respective search cutoff dates). For the remaining 21 confirmed false positives, errors were categorized as publications describing a similar but non-matching trial (n=19), or sub-analyses of trial outcomes not listed in the registration (n=2). Similarly, of 100 false negative cases, 44 (44.0%, 95% CI, 34.1–54.3%) were found to be true negatives upon review. For the 56 confirmed false negatives, the primary reason for failure was TrialScout's search algorithm not finding the correct publication during the search phase (n=51). Of these 51 publications, 51% (n=26) were indexed by PubMed whereas the remaining 49% (n=25) lacked a PubMed entry. The remaining 5 publications were found by TrialScout but classified as not containing results. Overall, across all 200 cases with discrepancies, human error accounted for a larger proportion (123/200, 61.5%, 95% CI, 54.4–68.3%) than TrialScout error.

3.2 Random Sample of Completed or Terminated Clinical Trials: Characteristics

TrialScout was applied to 9,600 ClinicalTrials.gov registrations; their characteristics are summarised in Table 2. Enrolment numbers showed a nearly log-normal distribution with a median of 56 participants and an interquartile range of 25-141 participants. The median completion year was 2015. Notably, only 28.6% (n=2,743) of trials had summary results in ClinicalTrials.gov.

Table 2. Characteristics of included trials.

Characteristic	Number of trials (total N=9,600)
Study status	
Completed	8,561 (89.2%)
Terminated	1,039 (10.8%)
Phase	
Early Phase 1	104 (1.1%)
Phase 1	1,301 (13.6%)
Phase 1/Phase 2	331 (3.4%)
Phase 2	1,505 (15.7%)
Phase 2/Phase 3	167 (1.7%)
Phase 3	1,095 (11.4%)
Phase 4	915 (9.5%)
Missing/Not applicable	4,182 (43.6%)
Completion Year	
<2005	352 (3.7%)
2005–2009	1,479 (15.4%)
2010–2014	2,573 (26.8%)
2015–2019	3,278 (34.1%)
2020–2024	1,918 (20.0%)
Enrolled participants	

Characteristic	Number of trials (total N=9,600)
1-99	6,294 (65.6%)
100-499	2,524 (26.3%)
500+	678 (7.1%)
Missing	104 (1.1%)
Summary results	
No	6,857 (71.4%)
Yes	2,743 (28.6%)
Participant Sex	
All	8,119 (84.6%)
Female only	914 (9.5%)
Male only	561 (5.8%)
Missing	6 (0.1%)
Lead Sponsor Type	
Other (Universities, Organizations, Networks, Non-U.S. Governmental Agencies, Individuals, Unknown, Ambiguous)[1]	6,116 (63.7%)
Industry	3,120 (32.5%)
U.S. Federal Agency/NIH[2]	364 (3.8%)

Characteristics of the 9,600 randomly sampled clinical trial registrations included in the study. Abbreviations: NIH: National Institutes of Health.

[1] Other n=5,823, Non-U.S. Governmental Agencies n=191, Network n=82, Individuals n=17, Unknown n=2, Ambiguous n=1.

[2] NIH n=253, Other U.S. Federal Agency n=111

3.3 Identification of Results from Randomly Sampled Trials

TrialScout detected results published in a scientific journal for 63.6% of trials (n=6,110). The proportion of trials with reported results, i.e., with publications or summary results, was 72.9% (n=6,998) (Figure 3). TrialScout found 125,528 candidate publications, of which it classified 9% (n=11,256) as results publications. For trials with identified result publications, the mean number per trial was 1.84, with most having a single publication (n=4,163), but

some having 10 or more (n=81). Detailed descriptions of each search strategy are available in Appendix 1.

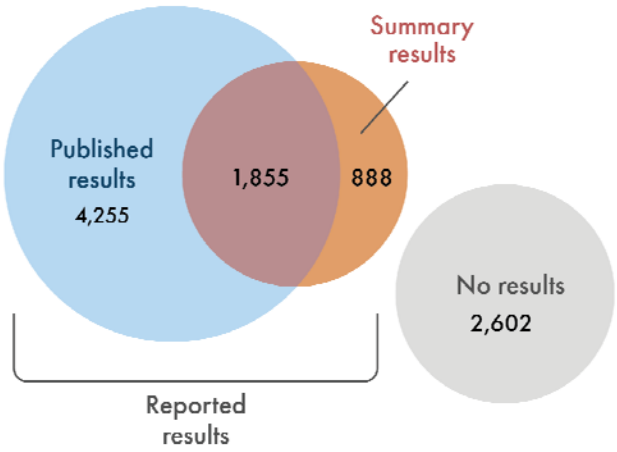

Figure 3. Overlap between published and summary results among 9,600 clinical trials. Published results refer to trial results identified in peer-reviewed publications by TrialScout. Summary results refer to tabular results posted directly on ClinicalTrials.gov. Of the 6,998 trials (72.9%) with any reported results, 4,255 had only published results, 888 had only summary results, and 1,855 had both.

Higher participant enrolment (log-transformed) was strongly associated with having published (β=0.385, p<0.001) and reported (β=0.284, p<0.001) results (Figure 4). The relationship between results reporting and completion year was not monotonic: results availability increased from the early 2000s, plateaued around 2010-2015, before declining in recently completed trials, presumably because of more limited time to publish results (Figure 5). Pearson's chi-squared test showed differences in reporting rates across study status, trial phase, and funder type (Table 3, omnibus tests). Subsequent post-hoc tests showed that reporting was significantly lower for industry-funded trials (70.5% vs. 74.1%, p<0.001), trials on male-only subjects (68.6% vs. 73.2%, p=0.022), and for early-phase (phase 1 and early phase 1) versus later-phase trials (60.4% vs. 79.6%, p<0.001).

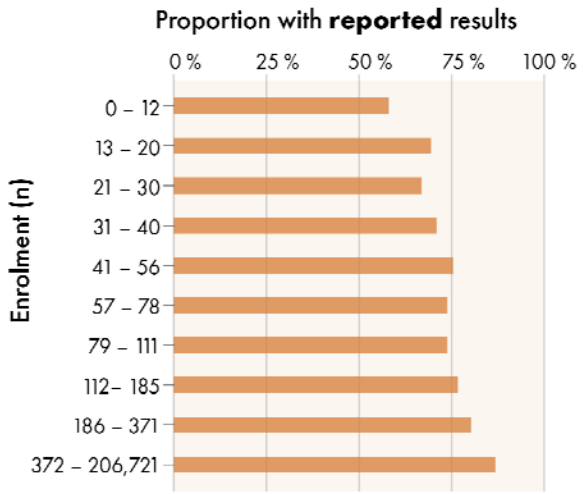

Figure 4. Proportion of clinical trials with reported results by participant enrolment decile. Proportion of clinical trials for which results were reported (either as summary results on ClinicalTrials.gov or in a journal article found by TrialScout) based on the number of participants enrolled in each trial. Sample size is grouped in deciles.

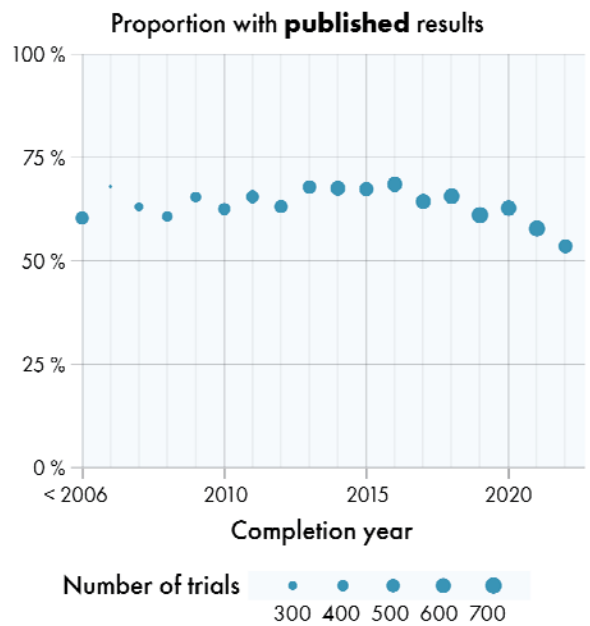

Figure 5. **Proportion of clinical trials with results availability by completion year.** Circles represent years and are sized according to the number of trials completed during that year. Trials from 2005 and earlier have been merged into a single data point as they were comparatively few.

Table 3. Subgroup analyses for reported, published, and summary results.

Subgroup	Trials with reported results	Trials with published results	Trials with summary results	p

Subgroup	Trials with reported results	Trials with published results	Trials with summary results	p
Study status				
Completed	6407/8561 (74.8%)	5816/8561 (67.9%)	2307/8561 (26.9%)	<0.001
Terminated	591/1039 (56.9%)	294/1039 (28.3%)	436/1039 (42.0%)	
Phase				
Early Phase 1	64/104 (61.5%)	59/104 (56.7%)	13/104 (12.5%)	<0.001
Phase 1	784/1301 (60.3%)	691/1301 (53.1%)	223/1301 (17.1%)	
Phase 1/Phase 2	264/331 (79.8%)	204/331 (61.6%)	161/331 (48.6%)	
Phase 2	1176/1505 (78.1%)	933/1505 (62.0%)	707/1505 (47.0%)	
Phase 2/Phase 3	119/167 (71.3%)	108/167 (64.7%)	47/167 (28.1%)	
Phase 3	921/1095 (84.1%)	812/1095 (74.2%)	527/1095 (48.1%)	
Phase 4	714/915 (78.0%)	598/915 (65.4%)	315/915 (34.4%)	
Missing/Not applicable	2956/4182 (70.7%)	2705/4182 (64.7%)	750/4182 (17.9%)	
Lead Sponsor Type				
Other (Universities, Organizations, Networks, Non-U.S. Governmental Agencies, Individuals, Unknown, Ambiguous)	4504/6116 (73.6%)	4129/6116 (67.5%)	1269/6116 (20.7%)	<0.001
Industry	2199/3120 (70.5%)	1725/3120 (55.3%)	1320/3120 (42.3%)	
U.S. Federal Agency/NIH	295/364 (81.0%)	256/364 (70.3%)	154/364 (42.3%)	
Participant Sex				
All	5951/8119 (73.3%)	5189/8119 (63.9%)	2407/8119 (29.6%)	0.097
Female	658/914 (72.0%)	567/914 (62.0%)	230/914 (25.2%)	
Male	385/561 (68.6%)	350/561 (62.4%)	105/561 (18.7%)	
Missing	4/6 (66.7%)	4/6 (66.7%)	1/6 (16.7%)	

Subgroup analyses of 9,600 randomly sampled trials were conducted to examine the effect of certain trial characteristics on result detection. Reported results include those with summary results available on ClinicalTrials.gov or with published results found by TrialScout. Significance levels were determined using Pearson's chi-squared test on *reported results* between subgroups. Abbreviations: NIH: National Institutes of Health.

4. Discussion

TrialScout, our novel tool for linking published results to clinical trial registrations, performs well against a human reference standard. A manual review of disagreement between TrialScout and humans revealed that many of the false positive cases were in fact true positives, and similarly that nearly half of the false negative cases were in fact true negatives. TrialScout's true performance is therefore likely higher than the performance metrics suggest. When applied to a random sample of 9,600 completed or terminated clinical trials, TrialScout located published results for 63.6% of trials (n=6,110), which can be compared to a recent meta-analysis on clinical trial reporting practices that estimated 53% (14). We found TrialScout to be useful for accurate and scalable linking of clinical trials to published results.

Where existing automated methods rely on NCT-ID searches or metadata matching between trial registrations and publications (19–21), TrialScout uses an LLM to classify publications based on the text of the abstracts. This allows it to use information not present in metadata alone, distinguishing true result publications from protocols, secondary analyses, and reviews. Although an LLM is more computationally intensive per registration-publication comparison, TrialScout achieves faster runtime than the Smalheiser-Holt model (seconds vs. minutes per trial) (20) by combining aggressive filtering of the PubMed corpus (its search strategies) with the cloud infrastructure accessible through OpenAI's API. Due to the different way these previous models have been evaluated, data do not permit a head-to-head comparison to TrialScout.

A core methodological limitation of this study is our reliance on human researchers' output when validating TrialScout. We relied on curated datasets, with data collected through rigorous methods, using multiple independent reviewers (22,24). However, manual review of 200 randomly sampled cases where TrialScout and human reviewers disagreed showed that these datasets contain errors. In four fifths of the cases initially categorized as false positives, TrialScout had in fact found a publication that humans missed, making accurate estimation of the tool's performance difficult. Furthermore, the validation dataset only represents a subset

of all trials, not necessarily representative of the entire trial population. The validation dataset is thus best seen as a silver standard for evaluating TrialScout's performance.

TrialScout's high performance allowed us to analyse a large sample of randomly selected clinical trials, making this one of the largest studies on result reporting (14). TrialScout's analysis gives an overall publication rate of 63.6%, which is higher than the figure of 53% in Showell and colleagues' meta-analysis. This discrepancy may reflect differences in the studied trial cohorts, methodologies, or limitations of TrialScout's accuracy. Furthermore, our sample is limited to trials registered on ClinicalTrials.gov, which likely disproportionately represents trials from the Americas rather than the global trials landscape. The higher rate that we observed may also reflect the fact that older trials (when reported results were less frequent) were more prominently represented in Showell and colleagues (14). This analysis demonstrates that TrialScout can be applied at scale, providing a starting point for investigations of reporting practices with humans in the loop. Given its high positive predictive value, TrialScout is perhaps best used as a first-pass screening tool, allowing researchers to focus manual searches on trials where TrialScout failed to find results. Furthermore, TrialScout can complement existing monitoring tools such as the FDAAA TrialsTracker (30) or ideally be embedded in the trial registries themselves.

Our subgroup analyses mostly confirmed established findings, with higher publication rates for trials with high enrolment and higher phases (14,22,24), and lower rates for industry-sponsored and terminated trials (31). We found a lower publishing rate for trials with exclusively male subjects, which to our knowledge is a finding not previously reported. However, our subgroup analyses were univariable and did not adjust for potential confounders such as study size. Additionally, we saw no clear increase in reporting rates since 2010. We did note a relative decline in the most recent years, but this is almost certainly because many recently completed trials have yet to report their results.

5. Conclusion

We show that TrialScout, a novel LLM-based tool, can be used to link published results to clinical trial registrations. The use of AI tools to link research objects may enable a scaling up of efforts to match clinical trial results to registrations, with potential value for metascientists as well as for all stakeholders engaged in clinical trials transparency and evidence synthesis.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Data availability

The datasets generated and analysed are available in our GitHub repository (https://github.com/lahnstrom/trialscout). TrialScout can be accessed at https://metaresearch.se/trialscout.

Competing interests

The authors declare that they have no competing interests.

Funding

None.

CRediT authorship contribution statement

Love von Schreeb: Conceptualization, Methodology, Software, Validation, Formal Analysis, Investigation, Data Curation, Writing - Original Draft, Visualization, Project Administration. **Till Bruckner**: Writing - Review & Editing. **Darya Ava Aspromonti**: Validation, Writing - Review & Editing. **Laura Caquelin**: Validation, Writing - Review & Editing. **Jamie Cummins**: Writing - Review & Editing. **Nicholas J. DeVito**: Writing - Review & Editing. **Cathrine Axfors**: Writing - Review & Editing. **John P.A. Ioannidis**: Supervision, Writing - Review & Editing. **Gustav Nilsonne**: Conceptualization, Methodology, Supervision, Writing - Review & Editing.

Acknowledgements

Meike Latz refined and improved the visual presentation of figures 1 and 3–5.

This paper originates from Love von Schreeb's Master Thesis in the medical program at Karolinska Institutet.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

The large language model “Claude” and its associated tool “Claude Code”, provided by Anthropic, were used to assist with development of TrialScout. This included the generation of code snippets, scripts, and documentation. All such generated content was reviewed, tested, and verified by the authors.

References

1. Zarin DA, Goodman SN, Kimmelman J. Harms From Uninformative Clinical Trials. JAMA. 2019 Sep 3;322(9):813. doi:10.1001/jama.2019.9892

2. Ethical Principles for Medical Research Involving Human Subjects [Internet]. WMA - The World Medical Association; [cited 2024 Sep 12]. Available from: https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/

3. Wallach JD, Krumholz HM. Not Reporting Results of a Clinical Trial Is Academic Misconduct. Ann Intern Med. 2019 Aug 20;171(4):293–4. doi:10.7326/M19-1273

4. Wallach JD, Gonsalves GS, Ross JS. Research, regulatory, and clinical decision-making: the importance of scientific integrity. J Clin Epidemiol. 2018 Jan 1;93:88–93. doi:10.1016/j.jclinepi.2017.08.021

5. Council for International Organizations of Medical Sciences (CIOMS). International Ethical Guidelines for Health-related Research involving Humans [Internet]. Council for International Organizations of Medical Sciences (CIOMS); 2016 [cited 2024 Sep 13]. Report No. Available from: https://cioms.ch/publications/product/international-ethical-guidelines-for-health-related-research-involving-humans/ doi:10.56759/rgxl7405

6. Strzebonska K, Wasylewski MT, Zaborowska L, Riedel N, Wieschowski S, Strech D, et al. Results dissemination of registered clinical trials across Polish academic institutions: a cross-sectional analysis. BMJ Open. 2020 Jan 22;10(1):e034666. doi:10.1136/bmjopen-2019-034666 PubMed PMID: 31974090; PubMed Central PMCID: PMC7044990.

7. Chalmers I. Underreporting Research Is Scientific Misconduct. JAMA. 1990 Mar 9;263(10):1405–8. doi:10.1001/jama.1990.03440100121018

8. Public Law 110 - 85 - Food and Drug Administration Amendments Act of 2007. H.R. 3580. 2007 Sep 27.

9. Regulation (EU) No 536/2014 of the European Parliament and of the Council of 16 April 2014 on clinical trials on medicinal products for human use, and repealing Directive 2001/20/EC Text with EEA relevance. OJ L [Internet]. 2014 Apr 16. Available from: http://data.europa.eu/eli/reg/2014/536/oj/eng

10. Jansen MS, Dekkers OM, Groenwold RHH, Siegerink B. Publication rates in small German trials remained low five years after trial completion. Contemp Clin Trials. 2022 Oct;121:106899. doi:10.1016/j.cct.2022.106899 PubMed PMID: 36038002.

11. Anderson ML, Chiswell K, Peterson ED, Tasneem A, Topping J, Califf RM. Compliance with Results Reporting at ClinicalTrials.gov. N Engl J Med. 2015 Mar 12;372(11):1031–9. doi:10.1056/NEJMsa1409364

12. DeVito NJ, Bacon S, Goldacre B. Compliance with legal requirement to report clinical trial results on ClinicalTrials.gov: a cohort study. The Lancet. 2020 Feb 1;395(10221):361–9. doi:10.1016/S0140-6736(19)33220-9 PubMed PMID: 31958402.

13. Psotka MA, Latta F, Cani D, Fiuzat M, Sbolli M, Barnett S, et al. Publication Rates of Heart Failure Clinical Trials Remain Low. J Am Coll Cardiol. 2020 Jun 30;75(25):3151–61. doi:10.1016/j.jacc.2020.04.068

14. Showell MG, Cole S, Clarke MJ, DeVito NJ, Farquhar C, Jordan V. Time to publication for results of clinical trials. Cochrane Database Syst Rev. 2024 Nov 27;11(11):MR000011. doi:10.1002/14651858.MR000011.pub3 PubMed PMID: 39601300; PubMed Central PMCID: PMC11600493.

15. Salholz-Hillel M, Strech D, Carlisle BG. Results publications are inadequately linked to trial registrations: An automated pipeline and evaluation of German university medical centers. Clin Trials. 2022 Jun 1;19(3):337–46. doi:10.1177/17407745221087456

16. Lamberink HJ, Vinkers CH, Lancee M, Damen JAA, Bouter LM, Otte WM, et al. Clinical Trial Registration Patterns and Changes in Primary Outcomes of Randomized Clinical Trials From 2002 to 2017. JAMA Intern Med. 2022 Jul;182(7):779–82. doi:10.1001/jamainternmed.2022.1551 PubMed PMID: 35575802; PubMed Central PMCID: PMC9112139.

17. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials | The BMJ [Internet]. [cited 2024 Sep 10]. Available from: https://www.bmj.com/content/340/bmj.c332

18. Joint statement on public disclosure of results from clinical trials [Internet]. [cited 2025 May 31]. Available from: https://www.who.int/news/item/18-05-2017-joint-statement-on-registration

19. Powell-Smith A, Goldacre B. The TrialsTracker: Automated ongoing monitoring of failure to share clinical trial results by all major companies and research institutions [Internet]. F1000Research; 2016 [cited 2024 Oct 23]. Available from: https://f1000research.com/articles/5-2629 doi:10.12688/f1000research.10010.1

20. Smalheiser NR, Holt AW. A web-based tool for automatically linking clinical trials to their publications. J Am Med Inform Assoc. 2022 May 1;29(5):822–30. doi:10.1093/jamia/ocab290

21. Goodwin TR, Skinner MA, Harabagiu SM. Automatically Linking Registered Clinical Trials to their Published Results with Deep Highway Networks. AMIA Summits Transl Sci Proc. 2018 May 18;2018:54–63. PubMed PMID: 29888040; PubMed Central PMCID: PMC5961767.

22. Nilsonne G, Wieschowski S, DeVito NJ, Salholz-Hillel M, Ahnström L, Bruckner T, et al. Results reporting for clinical trials led by medical universities and university hospitals in the nordic countries was often missing or delayed. J Clin Epidemiol. 2025 May 1;181:111710. doi:10.1016/j.jclinepi.2025.111710

23. Dahl R. Node.js [Internet]. [cited 2024 Sep 17]. Available from: https://nodejs.org/en

24. Riedel N, Wieschowski S, Bruckner T, Holst MR, Kahrass H, Nury E, et al. Results dissemination from completed clinical trials conducted at German university medical centers remained delayed and incomplete. The 2014 –2017 cohort. J Clin Epidemiol. 2022 Apr 1;144:1–7. doi:10.1016/j.jclinepi.2021.12.012

25. Wieschowski S, Riedel N, Wollmann K, Kahrass H, Müller-Ohlraun S, Schürmann C, et al. Result dissemination from clinical trials conducted at German university medical centers was delayed and incomplete. J Clin Epidemiol. 2019 Nov 1;115:37–45. doi:10.1016/j.jclinepi.2019.06.002

26. Axfors C, Nilsonne G. Nordic trial reporting project: Raw data from EU Clinical Trials Registry (EUCTR) and ClinicalTrials.gov [Internet]. Zenodo; 2023 [cited 2024 Sep 10]. Available from: https://zenodo.org/records/10091147 doi:10.5281/zenodo.10091147

27. R Core Team. R: A Language and Environment for Statistical Computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2022 [cited 2024 Sep 16]. Available from: https://www.r-project.org/

28. Wickham H, François R, Henry L, Müller K, Vaughan D. dplyr: A Grammar of Data Manipulation [Internet]. 2023. Available from: https://CRAN.R-project.org/package=dplyr

29. Haynes AG, Lenz A, Stalder O, Limacher A. `presize`: An R-package for precision-based sample size calculation in clinical research. J Open Source Softw. 2021;6(60):3118. doi:10.21105/joss.03118

30. DeVito NJ, Bacon S, Goldacre B. FDAAA TrialsTracker: A live informatics tool to monitor compliance with FDA requirements to report clinical trial results [Internet]. 2018 [cited 2024 Oct 30]. Available from: http://biorxiv.org/lookup/doi/10.1101/266452 doi:10.1101/266452

31. Nelson JT, Tse T, Puplampu-Dove Y, Golfinopoulos E, Zarin DA. Comparison of Availability of Trial Results in ClinicalTrials.gov and PubMed by Data Source and Funder Type. JAMA. 2023 Apr 25;329(16):1404–6. doi:10.1001/jama.2023.2351