Journal of Clinical Epidemiology 173 (2024) 111433

Journal of Clinical Epidemiology

ORIGINAL RESEARCH

Planned but ever published? A retrospective analysis of clinical prediction model studies registered on clinicaltrials.gov since 2000

Nicole White[a,*], Rex Parsons[a], David Borg[a,b], Gary Collins[c], Adrian Barnett[a]

[a]*Australian Centre for Health Services Innovation and Centre for Healthcare Transformation, School of Public Health and Social Work, Queensland University of Technology, Kelvin Grove, Queensland, Australia*

[b]*School of Exercise and Nutrition Sciences, Queensland University of Technology, Kelvin Grove, Queensland, Australia*

[c]*Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology & Musculoskeletal Sciences, Botnar Research Centre, University of Oxford, Oxford, United Kingdom*

Accepted 12 June 2024; Published online 17 June 2024

Abstract

Objectives: To describe the characteristics and publication outcomes of clinical prediction model studies registered on clinicaltrials.gov since 2000.

Study Design and Setting: Observational studies registered on clinicaltrials.gov between January 1, 2000, and March 2, 2022, describing the development of a new clinical prediction model or the validation of an existing model for predicting individual-level prognostic or diagnostic risk were analyzed. Eligible clinicaltrials.gov records were classified by modeling study type (development, validation) and the model outcome being predicted (prognostic, diagnostic). Recorded characteristics included study status, sample size information, Medical Subject Headings, and plans to share individual participant data. Publication outcomes were analyzed by linking National Clinical Trial numbers for eligible records with PubMed abstracts.

Results: Nine hundred twenty-eight records were analyzed from a possible 89,896 observational study records. Publications searches found 170 matching peer-reviewed publications for 137 clinicaltrials.gov records. The estimated proportion of records with 1 or more matching publications after accounting for time since study start was 2.8% at 2 years (95% CI: 1.7%, 3.9%), 12.3% at 5 years (9.8% to 14.9%) and 27% at 10 years (23% to 33%). Stratifying records by study start year indicated that publication proportions improved over time. Records tended to prioritize the development of new prediction models over the validation of existing models (76%; 704/928 vs. 24%; 182/928). At the time of download, 27% of records were marked as complete, 35% were still recruiting, and 14.7% had unknown status. Only 7.4% of records stated plans to share individual participant data.

Conclusion: Published clinical prediction model studies are only a fraction of overall research efforts, with many studies planned but not completed or published. Improving the uptake of study preregistration and follow-up will increase the visibility of planned research. Introducing additional registry features and guidance may improve the identification of clinical prediction model studies posted to clinical registries.

Keywords: Diagnosis; Prognosis; Development; Validation; Study registration; Publication bias

1. Introduction

Diagnostic and prognostic risk information contributes to decisions that affect patients' health. Clinical prediction models estimate patient risk by applying regression or machine learning models to data available during decision-making. Current interest in applying machine learning methods for clinical prediction [1,2] would suggest that clinical prediction model studies are a relatively new field. Yet prediction models have been developed for many years using regression methods, and there are several well-known examples being used in practice, for example, the Framingham risk score for coronary heart disease [3], and the Nottingham Prognostic Index for breast cancer [4]. These successes, combined with the promises made by big data and personalized medicine, have spurred an explosion of clinical prediction model publications [5], describing newly

* Corresponding author. Australian Centre for Health Services Innovation and Centre for Healthcare Transformation, 60 Musk Avenue, Kelvin Grove, 4059, Australia.

E-mail address: nm.white@qut.edu.au (N. White).

https://doi.org/10.1016/j.jclinepi.2024.111433

What is new?

Key findings

- Clinicaltrials.gov is being used more to register clinical prediction model studies.
- Planned studies prioritize developing new models over evaluating existing models.
- 27% of study registrations published their planned prediction model study within 10 years of registration.

What this adds to what was known?

- Nonpublication of clinical prediction model studies is a source of research waste.

What is the implication and what should change now?

- Better visibility of planned models is needed to reduce waste from duplication.
- Methods to follow-up study outcomes and model uptake can proactively reduce waste.

developed models or findings from evaluating existing models in new patient cohorts.

The premise that clinical prediction models will advance personalized medicine has regrettably been met with significant research waste. Systematic reviews have identified waste across several clinical areas [6−10], driven by poor study design, inappropriate statistical methods, and poor reporting. These sources of bias are partly driven by spin practices, where researchers have likely exaggerated prediction model performance to increase their chances of publication [11,12]. Expert-led initiatives to address these issues have sought to improve the quality of clinical prediction model studies and their critical appraisal during peer review and postpublication [13,14]. While these initiatives have led to modest improvements in conduct and reporting among published studies [15], it is also possible that their introduction has made it more difficult for both poorly designed and ‘‘negative’’ studies to get published.

Poorly designed and incompletely reported studies are 2 important sources of avoidable research waste [16]. Nonpublication is a third source of waste, where planned studies are never published, possibly because analysis was not feasible based on available data or analysis results were disappointing. Study registration and protocols aim to reduce publication bias by making study details available before research starts, which can then be linked to publications [17]. If publication bias exists in the clinical prediction model literature, then more models will have been planned but never completed or completed but never published. We tested this hypothesis by analyzing data on studies registered with clinicaltrials.gov that proposed developing or evaluating a clinical prediction model. Clinicaltrials.gov is an online study registry that was launched in 2000 to improve access to information on planned, ongoing and completed clinical studies [18]. Our analysis summarized the characteristics of planned prediction models and publication outcomes, as an indicator of research waste.

2. Materials and Methods

We analyzed clinical prediction model studies registered with clinicaltrials.gov from January 1, 2000, when clinical trials.gov was established, to March 3, 2022. Record-level data were downloaded from clinicaltrials.gov in XML format. Individual records were indexed by National Clinical Trial (NCT) number. All data processing and analysis were performed in R. Code developed for this research is available from https://github.com/nicolemwhite/prediction_clinicaltrials.

2.1. Record screening

Eligible records were found through keyword searches, followed by manual record screening. A record was returned if it contained one or more of the terms: ‘‘machine learning,’’ ‘‘artificial intelligence,’’ ‘‘deep learning,’’ ‘‘prediction model,’’ ‘‘predictive model,’’ ‘‘prediction score,’’ ‘‘predictive score,’’ ‘‘warning score,’’ ‘‘risk score,’’ ‘‘risk prediction,’’ ‘‘prognostic model,’’ and ‘‘diagnostic model.’’ Terms were combined with ‘‘prognos*’’ or ‘‘diagnos*’’ to identify diagnosis and prognosis applications. Search terms were applied to record titles, descriptions, outcome measures, and keywords. Searches were filtered to observational (noninterventional) study types only, therefore removing models being trialled in a prospective research study [19]. Identified records were manually reviewed by at least 2 authors to determine final eligibility. Review decisions were based on details stated in the clinicaltrials.gov record only. A formal sample size calculation was not used, as we aimed to include all eligible studies based on information available at the time of screening.

2.2. Inclusion and exclusion criteria

Studies that planned the development and validation of a new prediction model or the validation of an existing prediction model were included. No restrictions were placed on modeling methods (eg, regression and machine learning methods). Studies that planned to extend (eg, add new predictors) or update (eg, recalibrate) an existing model were included; these instances were classified as model development or model development and validation within the same study based on text descriptions [14]. Records could describe clinical prediction modeling as the primary research objective or an objective within a larger study.

Records that described identifying or validating prognostic factors or "biomarkers" [20] without reference to the subsequent multivariable modeling for individualized prediction were excluded.

Model outcomes were classified broadly as diagnostic or prognostic, using the stated outcome of interest and time horizon for prediction [21]. Records were marked as unknown if there were insufficient details to determine either the study type or the model's predicted outcome.

Eligible records described models for predicting individual-level risk. Model predictors could include patient-level characteristics (eg, demographics) or processed features (eg, extracted from medical images), with or without group-level variables (eg, socioeconomic status). Models that contained only group-level variables were excluded.

2.3. Data extraction

Extracted data on eligible records included study start and end dates, overall study status, Medical Subject Headings (MeSH), sample size details, sponsorship/funding, and plans to share individual participant data. Additional data from record histories captured updates to sample sizes and study status after registration.

Manual review coded the planned outcomes to be predicted by the model (prognostic, diagnostic, or both), and study type (model development, validation, or both). Final codings were validated by 4 authors (NW, RP, DB, AB) by manually re-reviewing a 20% random sample. Finally, we checked for duplicate study registrations to account for the possibility of the same study being registered under multiple NCT numbers. Potential duplicates were identified by applying the Jaccard similarity index to record titles [22]. Record pairs with a Jaccard index greater than 0.5 were manually reviewed and excluded as appropriate.

2.4. Publication matching

Identifying publications involved reviewing titles and abstracts for all NCT-linked papers and papers submitted by study investigators to clinicaltrials.gov records with a known PubMed identifier (PMID). NCT-linked papers were found in 2 ways. First, we reviewed all papers that were automatically populated by clinicaltrials.gov. Second, we used the R package rentrez to search for instances of each NCT number cited in PubMed abstracts as metadata or in the abstract text.

Additional publications were identified by using a web-based tool that was developed for finding clinical trial results [23]. To improve the chances of finding a match and keep the number of candidate publications reviewed manageable, PMIDs found by the machine learning classifier were filtered to hits from a verified MEDLINE search strategy for finding clinical prediction model studies [24]. The PMID with the highest similarity score after filtering was reviewed as a potential matching publication.

We excluded PMIDs where the publication date was before the clinicaltrials.gov record registration date, and duplicates (eg, online version ahead of print).

For NCT numbers with one or more matching publications, time to publication was defined as the time between the study start date reported in clinicaltrials.gov and the first publication date on PubMed.

2.5. Statistical analysis

Time to publication was estimated using cumulative incidence functions. Our primary analysis estimated the proportion of all records with one or more matching publications up to 10 years since study start (as recorded in the clinicaltrials.gov record). Right censoring was applied to records at 10 years since study start or the last search date (November 17, 2023), whichever occurred first. Sensitivity analyses were conducted by fitting stratified cumulative incidence functions to examine publication outcomes based on the study start year. We further compared functions for NCT-linked abstracts with and without matching publications identified by the machine learning classifier.

Results were summarized by MeSH descriptors, to examine studies across different health conditions. Descriptors were used as listed, and aggregated to major disease concepts, based on the 2022 MeSH tree structure [25]. Record status histories were summarized as stacked probability plots, up to 10 years following record registration. Records are marked as unknown by clinicaltrials.gov if they had passed their study end date, for studies yet to be completed, or had not been updated in the last 2 years [26]. Records with missing data were excluded from the denominator of corresponding descriptive statistics.

3. Results

3.1. Search results

Targeted keyword searches returned 1465 records, from a possible 89,869 observational studies (Fig 1). Keywords were most found in record description fields, and matches to planned modeling approaches (eg, deep learning) were more frequent than generic model descriptors (eg, warning score) (Supplementary Figure 1). Manual review confirmed 940 records for inclusion. The final sample size was 928 records after removing duplicates.

3.2. Publication outcomes

We identified and extracted 697 NCT-linked or investigator-submitted publications, from 254 clinicaltrials.gov records. Review outcomes confirmed 170 matching publications attributed to 137 clinicaltrials.gov records. Of these

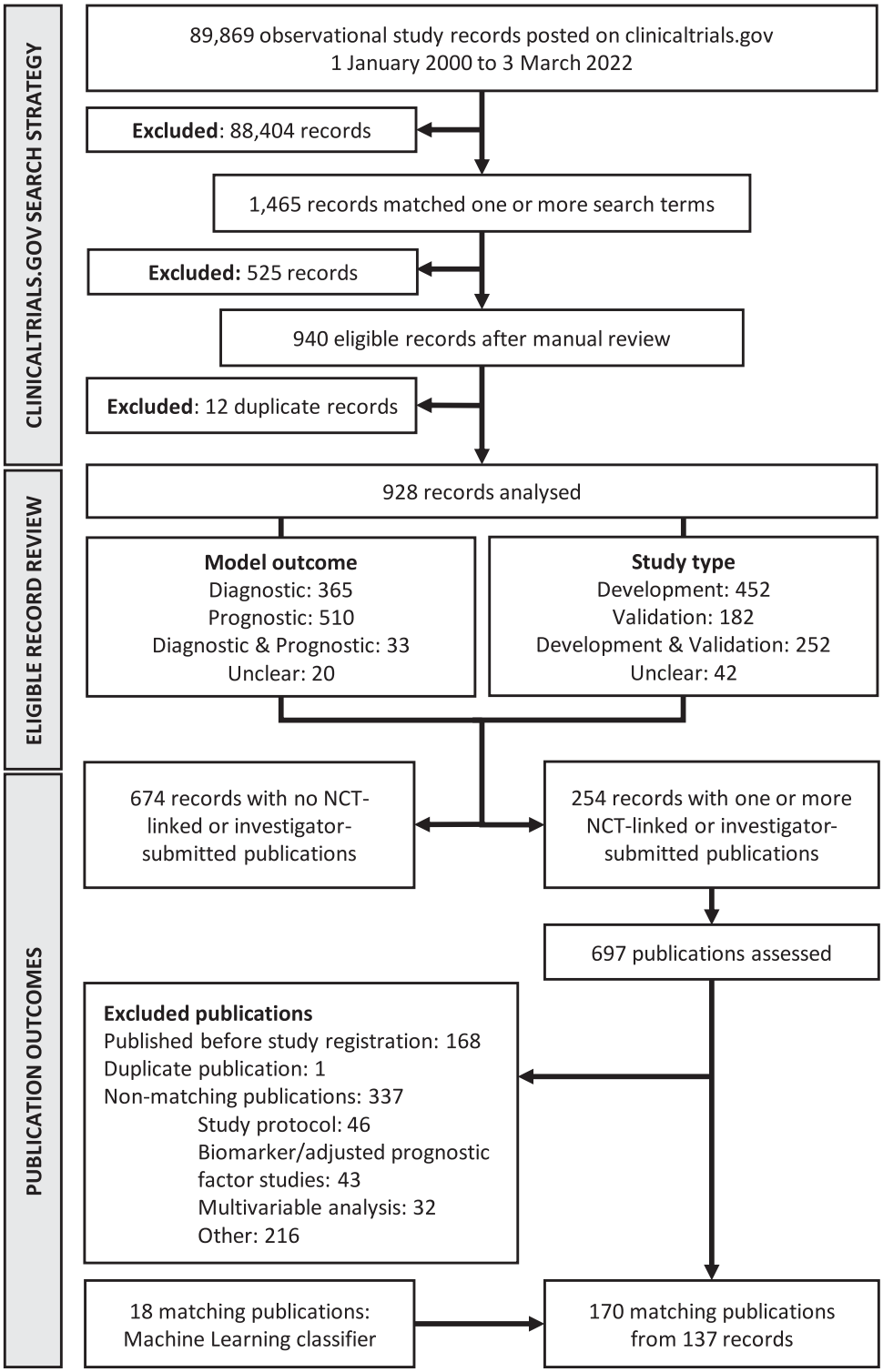

Figure 1. clinicaltrials.gov search strategy, record classification, and publication outcomes.

matching publications, 152 publications were NCT-linked or investigator-submitted, and 18 were identified by the machine learning classifier. Unmatched abstracts included study protocols ($n = 46$), biomarker/adjusted prognostic factor studies ($n = 43$) and multivariable analyses ($n = 32$). A random subset of matching publication titles is in Supplementary File 1.

These results led to an observed publication proportion of 14.8%. Publication proportions among MeSH

Table 1. Clinical prediction model study publication rates for common MeSH descriptors included in clinicaltrials.gov records

MeSH descriptor	clinicaltrials.gov records found	Matching publications	Publication rate (%)
Chest pain	11	5	46
Carcinoma, hepatocellular	8	3	38
Syndrome	19	7	37
Prostatic neoplasm	9	3	33
Heart failure	24	7	30
Coronary artery disease	29	6	21
Emergencies	21	4	19
Critical illness	11	2	18
Atrial fibrillation	20	3	15
Colorectal neoplasms	14	2	14
Sepsis	14	2	14
Breast neoplasms	30	4	13
Diabetic retinopathy	8	1	13
Parkinson disease	9	1	11
Stomach neoplasms	9	1	11
COVID-19	46	5	11
Stroke	22	2	9
Atherosclerosis	11	1	9
Lung neoplasms	19	0	0
Acute kidney injury	13	0	0

MeSH, Medical Subject Headings.

descriptors were highest for Chest Pain (n = 5/11; 46%) and Hepatocellular Carcinoma (n = 3/8; 38%) (Table 1). No matching publications were found for Lung Neoplasms and Acute Kidney Injury.

The median time to publication for matched records was 3.5 years (IQR: 2.2 to 6.1 years). Cumulative incidence functions for time to publication accounting for censoring showed steady publication increases over 10 years (Fig 2A). Estimated publication rates were 2.8% at 2 years (95% CI: 1.7%, 3.9%), 12.3% at 5 years (95% CI: 9.8%, 14.9%) and 27.8% at 10 years (95% CI: 22.6%, 33.0%) (Fig 2A). Publication rates at 2 years varied between 0% and 4.2% when stratified by study start year, and cumulative publications improved in later years (Fig 2B). Results stratified by publication matching strategy are presented in Supplementary Figure 2. Sensitivity analyses showed slightly higher publication rates for combined development and validation studies compared with development-only studies (Supplementary Figure 3). No discernible differences in publication outcomes were found between studies that predicted diagnostic vs prognostic outcomes (Supplementary Figure 4).

3.3. Record status over time

At the time of download, 326 records (35%) were marked as recruiting participants, and 252 (27%) records were marked as complete (Table 2). Seventy-four records (7.9%) were stated as complete at the time of registration (Fig 3). Approximately half of the records were registered between 2020 and 2022 (Fig 4A).

3.4. Modeling study type

Study registration emphasized developing new models over validating existing models (Fig 4B). One in 3 model development studies mentioned validation as part of the same study (252/704 records); however, less than half stated specific validation methods to be used. Nonstandard terminology to describe model validation was the primary source of reviewer disagreement, based on our random sample (11/180 records reassessed; 6.1%). The remaining disagreements were due to poor reporting; all were originally marked as unclear study types (n = 4). Validation studies were the least common study design across major disease groups (Supplementary Figure 5).

Targeted keyword searches largely found records based on a single keyword, or as a combination of "Artificial Intelligence", "Machine Learning" and/or "Deep Learning" (Fig 5). Other common matches to a single keyword were "Prediction model" (n = 86), "Risk score" (n = 70) and "Predictive model" (n = 66).

3.5. Model outcomes

Fifty-eight percent of records (n = 543) described prediction modeling for prognostic outcomes. Prognostic modeling studies were consistently more common than diagnostic outcomes over major disease concepts and over time (Supplementary Figures 5 and 6). MeSH descriptor summaries for prognostic outcomes found 33 records on COVID-19, 21 on coronary artery disease, and 19 records each on heart failure and stroke. MeSH descriptors for diagnostic outcomes included 18 for COVID-19, 17 for breast neoplasms, 14 for diabetes, and 13 records each for retinal diseases, polyps, and atrial fibrillation (Supplementary File 2).

3.6. Other record information

Most records included sample size information, albeit only an anticipated sample size (n = 557, 60%; Table 2). The Bland-Altman plot of planned and actual sample sizes showed average sample sizes were concentrated between

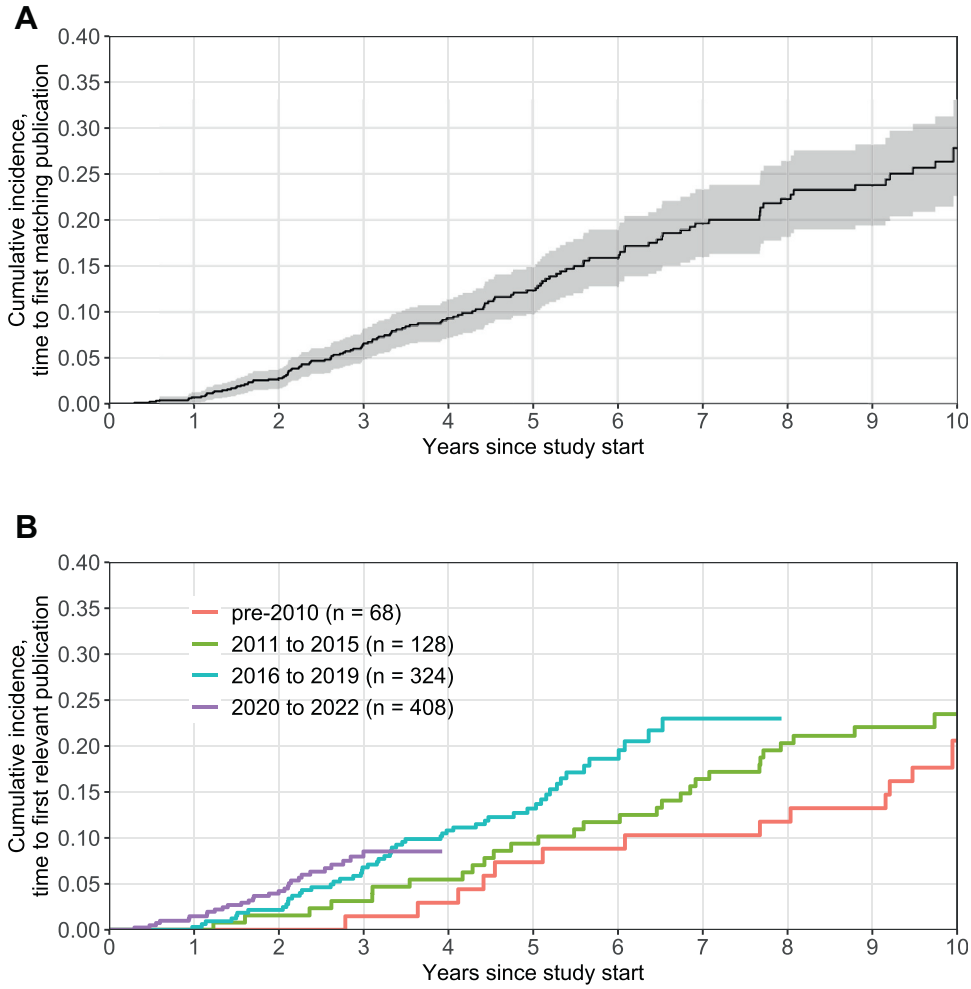

Figure 2. Cumulative incidence functions for time to first matching publication since reported study start. A: All records combined; B: Records by study start year. The gray shaded area in the top figure is the 95% confidence interval. A similar figure showing cumulative censoring over time is presented in Supplementary Figure 7. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

115 (10th percentile) and 6200 (90th percentile) (Supplementary Figure 8). The mean actual to planned sample size ratio was 0.92, indicating that final sample sizes were slightly lower than planned. There was little intention to share individual participant data (n = 66; 7.1%; Supplementary File 3).

Approximately 1 in 20 records were industry-sponsored, and 1 in 100 were sponsored by US federal agencies (Table 2). Common sponsor types for remaining records (n = 874) were universities (239; 27%) and hospitals (450; 52%). Grant funding was reported as metadata in 43 records (4.5%).

4. Discussion

Clinical prediction model studies should be scrutinized due to their potential to influence patient-level decision-making. Systematic and scoping reviews have evaluated factors that affect published model quality and revealed flaws in study design, analysis, and reporting. Our analysis expands this evidence base to evaluate study registration and the follow-up of planned clinical prediction model studies. The finding that results for only around 1 in 3 registered studies were published is disappointing but not unexpected compared to publication rates for other registered study designs [27–33].

The large nonpublication uncovered here challenges the hype surrounding clinical prediction research. Successfully predicting a patient's diagnosis or prognosis can help inform decision-making, but obtaining accurate predictions are difficult to do well. The QRISK3 tool for 10-year cardiovascular risk [34] and the National Early Warning Score [35] for sepsis risk stratification are rare examples of high-quality and widely validated prediction models that are recommended in clinical guidelines. These models are implemented in practice as simple scoring systems that use demographic and clinical variables available at the time

Table 2. Record characteristics extracted from clinicaltrials.gov

Characteristic	Total records, *n* (%)
Overall status	
Recruiting	326 (35%)
Completed	252 (27%)
Unknown status	136 (15%)
Not yet recruiting	109 (12%)
Active, not recruiting	72 (7.8%)
Enrolling by invitation	24 (2.5%)
Suspended/terminated/withdrawn	9 (1.0%)
Sample size information available	
Anticipated only	557 (60%)
Actual only	97 (10%)
Anticipated and actual	267 (29%)
Not mentioned	7 (0.8%)
Plan to share individual participant data[a]	
Yes	58 (7.4%)
No	338 (43%)
Undecided	181 (23%)
Not mentioned	209 (27%)
Grant funding reported	42 (4.5%)
Lead sponsor type	
Industry	45 (4.8%)
United States government[b]	96 (1.0%)
Other	874 (94%)
Other sponsor types[c]	
University	239 (27%)
Hospital	450 (52%)
Academic centers/Research institutes	53 (6.1%)
Societies/foundations	11 (1.3%)
Other	121 (14%)

[a] Only for studies posted since January 1, 2016 (n = 786).
[b] Combines registry categories National Institutes of Health (n = 6) and U.S. Federal government (n = 3).
[c] n = 874; Categories determined from word and bigram analysis of lead sponsor information reported in clinicaltrials.gov record. Hospital includes university teaching hospitals and academic medical centers.

of decision-making. Systematic reviews comparing regression-based predictions to predictions made by more complex models have shown that the latter often provide similar or worse performance [36,37].

Study registration and protocols are uncommon for clinical prediction model studies [38] despite being recommended as a strategy to improve research transparency [39,40]. Awareness of planned studies can reduce the chances of duplication and, therefore, potentially wasted research efforts [41,42]. That at least some models were registered is a promising sign that researchers are making a concerted effort to improve the visibility of planned clinical prediction model studies. However, our finding that developing new models was far more popular than evaluating existing models indicates that researchers may not be doing their due diligence by searching for similar models already developed [43].

There are likely several reasons why a clinical prediction model study is planned but not published. Developing or validating a clinical prediction model depends on collecting sufficient data to produce accurate predictions [44]. Sample sizes reported in this analysis varied, and few provided a sample size justification. Several studies in this analysis proposed the establishment of a patient data registry as an objective before planned prediction modeling. Prospective data collection by patient registries can support targeted research in well-defined patient populations, however, subsequent data collection burden and resources needed to maintain data quality can impact recruitment targets [45]. Another reason for nonpublication is that funding for prediction models is relatively uncommon. A recent study that included all interventional studies registered on clinicaltrials.gov reported that 27% had industry as a lead sponsor, which is much higher than the 4.8% in our sample [46]. Trials and other interventional studies may generally have more funding than prediction models, and this could partly explain the higher nonpublication rate shown here compared with registered trials [47].

Nonpublication may itself be due to journal-initiated publication bias, where studies may be less likely to be accepted for publication if a model's predictive performance falls short of arbitrary thresholds or perceived good performance [48,49]. Our previous work has provided evidence of ''AUC-hacking'', where excessive numbers of studies report an area under the receiver operating characteristic curve statistic just above such arbitrary thresholds of 0.7, 0.8 and 0.9 [12]. Nonpublication based on final model performance likely affects studies irrespective of quality and may bias perceptions about clinical prediction modeling more broadly. When a study produces disappointing results compared with already published models, researchers may experience barriers to publishing or may not submit their results for publication at all. Performance reported by published models may therefore inflate expectations for how well models perform. These interrelated behaviors negatively impact the current evidence base; however, their extent is unknown.

The scope of our analysis was clinical prediction model studies registered onclinicaltrials.gov, where final eligibility was determined by a systematic search strategy. Decisions were based on information reported in study records at the time of screening, which might differ from details reported in subsequent publications. While clinicaltrials.gov is a widely recognized registry, the sample size analyzed does not represent the full breadth of planned clinical prediction model research. Prospective registration of observational study designs is currently not required and is therefore less

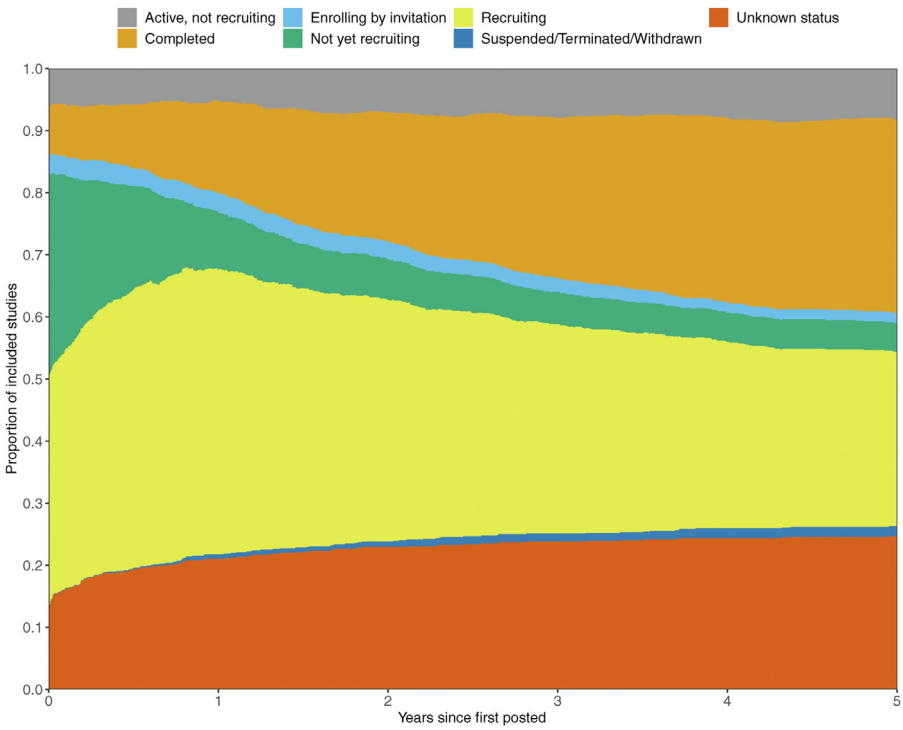

Figure 3. Stacked probability plot of record status history, up to 5 years since registration. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

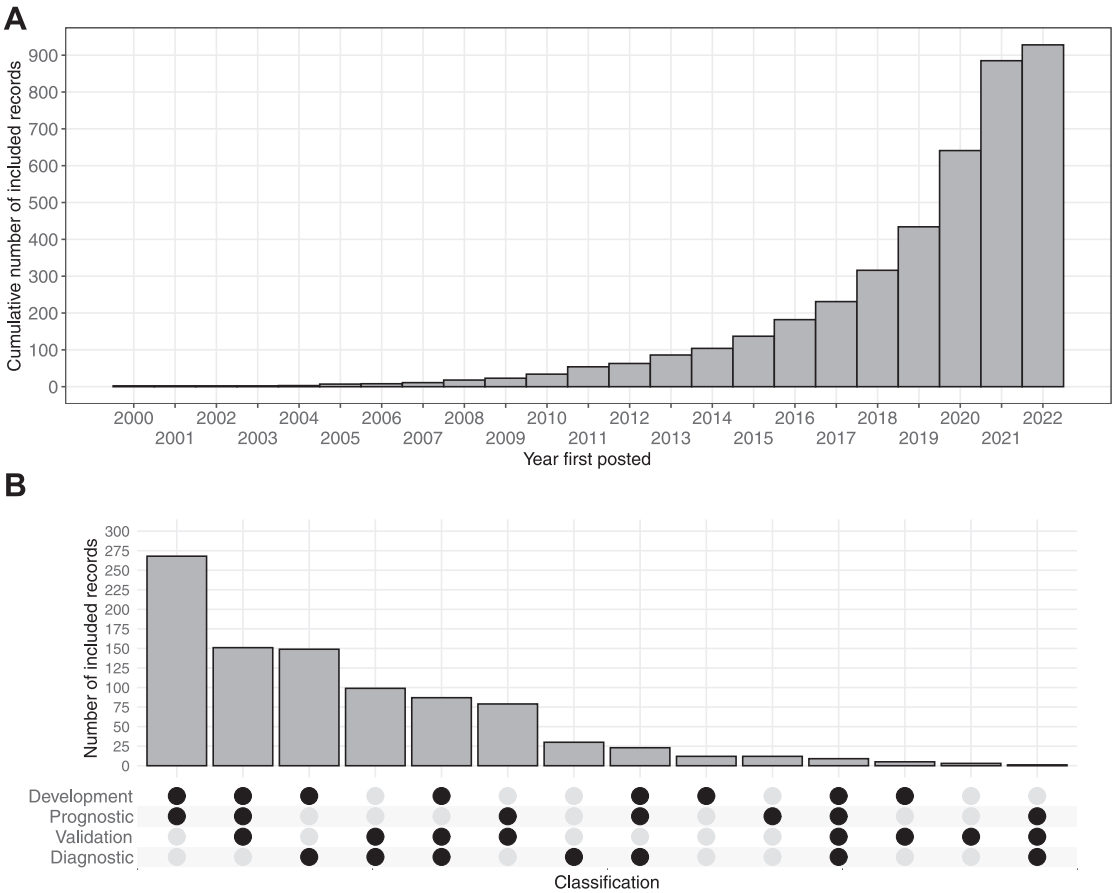

Figure 4. Included study characteristics; A: Total records posted by year. B: Record frequency by study type and model outcome(s) reported.

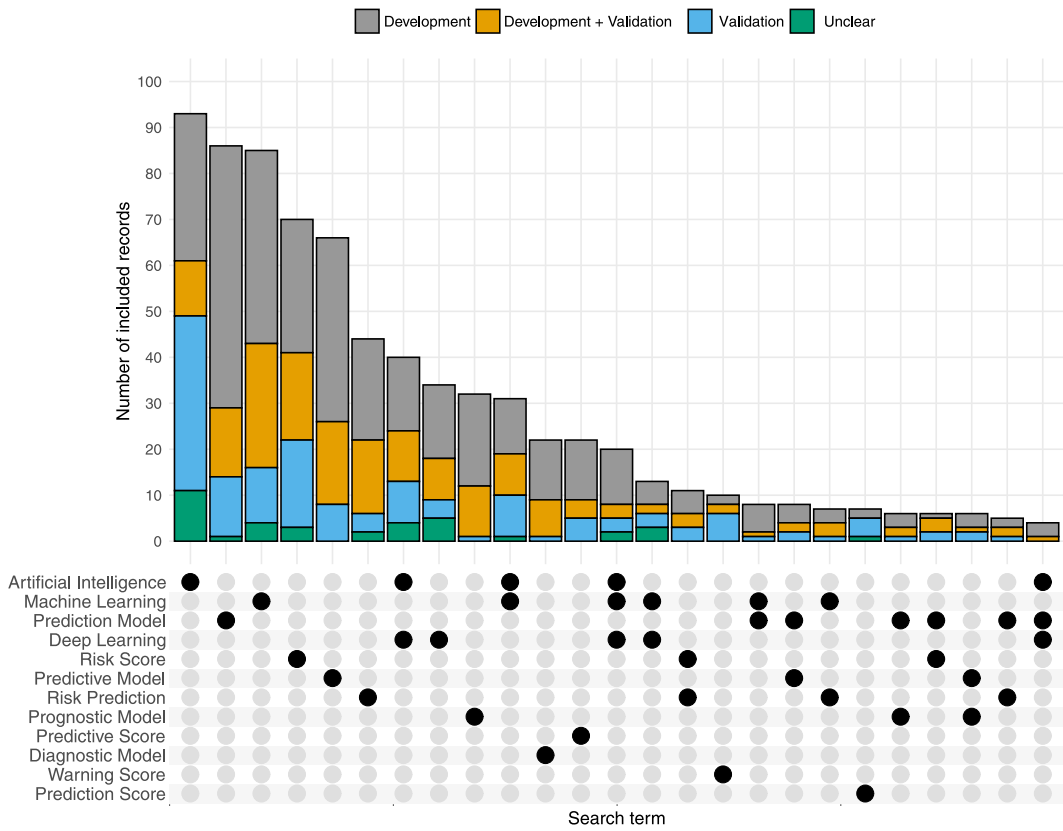

Figure 5. Targeted keyword search frequencies for all eligible records (n = 928), stratified by modeling study type. The top 25 search term combinations are shown.(For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

common than clinical trials. Nonetheless, making the details of planned research available can help address publication bias [50]. Our study represents the first attempt to quantify nonpublication bias affecting clinical prediction model studies, based on publicly available information on planned studies. Given that most studies in our sample were prospectively registered, it is possible that publication outcomes from our analysis are biased upwards and that nonpublication may be higher among registered and unregistered studies combined. Without access to planned studies that have never been registered, the true extent of nonpublication remains speculative. Study protocols are another way to make the key details of planned research publicly available. Study protocol guidelines for prediction model studies to complement the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis statement are currently being developed [51]. Future work should consider to what extent our findings generalize to studies that have not been preregistered.

Publications would have been missed if an NCT number was incorrectly reported or there were systematic differences in study registration descriptions and final abstracts. Publication searches were limited to articles indexed by PubMed, which covers more than 80% of the peer-reviewed biomedical literature [52]. Preprint servers allow researchers to share study results before peer review, however do not guarantee publication in a peer-reviewed journal [53,54]. While peer-reviewed publications should not be interpreted in isolation as a marker for study quality, preprints were excluded from this study as they are generally not peer reviewed.

Study registries are uniquely placed to improve research visibility and should not be limited to clinical trials [42,50]. Introducing new fields and filters into registry search functions may help researchers find clinical prediction model studies more easily and eliminate the need for a customized search strategy. Our search strategy involved a range of common terms related to clinical prediction model studies, yet manual review was still required to determine final record eligibility. Introducing new fields would require researchers to self-identify their work as a clinical prediction model, as is currently required for interventional and noninterventional study designs. Providing guidance on definitions relevant to prediction model studies may help to improve reporting completeness and transparency, as has been recommended for diagnostic accuracy studies [55]. Greater oversight of current and ongoing research may also improve awareness among clinicians and their patients as intended end-users.

Automated indexing by clinicaltrials.gov allowed us to link to publications based on a unique identifier. Available functionality can track publication outcomes for planned research,

but not the subsequent appraisal and uptake of promising models by others. Despite exponential increases in newly developed models, independent validations of promising models are rare, and even less is known about which models are successfully implemented into clinical practice [56]. Automation tools may increase the impact of high-quality clinical prediction models by synthesizing evidence about their reproducibility and clinical utility. Expert appraisal through open science platforms can provide additional validation for evidence found by automated tools. Expert crowdsourcing has successfully been used to synthesize published evidence on genetic variants to support decisions about personalized cancer therapy [57,58].

Research funding agencies are instrumental in maintaining standards for research conduct and dissemination. A survey of Australian health and medical research funders discovered varied incentives to encourage responsible research practices [59]. Greater scrutiny of funding applications that propose clinical prediction model studies is another actionable step toward improving the research quality and visibility. We further recommend that funding agencies strongly encourage researchers to demonstrate meaningful collaboration with statisticians at all project stages, given the technical expertise required to conduct high-quality clinical prediction model studies.

5. Conclusion

Clinicaltrials.gov is being used to register clinical prediction model studies. Our follow-up analysis of registered studies implies that few planned prediction models are published, likely creating a biased evidence base. Greater promotion of study registries for sharing details of planned research, additional registry features to identify clinical prediction studies, and tools to reliably follow-up studies from registration to publication and uptake by others will improve oversight of planned studies to proactively reduce research waste.

Ethics statement

All the data used in this study are publicly available and do not involve human participants. Ethics approval was therefore not required.

CRediT authorship contribution statement

Nicole White: Writing − review & editing, Writing − original draft, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Rex Parsons:** Writing − review & editing, Validation, Software, Formal analysis, Data curation, Conceptualization. **David Borg:** Writing − review & editing, Validation, Data curation, Conceptualization. **Gary Collins:** Writing − review & editing, Supervision, Conceptualization. **Adrian Barnett:** Writing − review & editing, Writing − original draft, Supervision, Software, Data curation, Conceptualization.

Data availability

Data and code used on included records are available at https://github.com/nicolemwhite/prediction_ clinicaltrials.

Declaration of competing interest

G.C. was supported by Cancer Research UK (programme grant: C49297/A27294). There are no competing interests for any other author.

Acknowledgments

The authors of this article acknowledge Neil Smalheiser and Arthur Holt, the authors of the machine learning−based algorithm used to match publications in this study. We acknowledge the computational resources and services used in this work provided by the eResearch Office, Queensland University of Technology, Brisbane, Australia.

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jclinepi.2024.111433.

References

[1] Obermeyer Z, Weinstein JN. Adoption of artificial intelligence and machine learning is increasing, but irrational exuberance remains. NEJM Catal Innov Care Deliv 2019;1(1). https://doi.org/10.1056/CAT.19.1090.
[2] Kolachalama VB, Garg PS. Machine learning and medical education. NPJ digital medicine 2018;1(1):54.
[3] D'Agostino RB, Vasan RS, Pencina MJ, Wolf PA, Cobain M, Massaro JM, et al. General cardiovascular risk profile for use in primary care. Circulation 2008;117:743−53.
[4] Haybittle J, Blamey R, Elston C, Johnson J, Doyle P, Campbell F, et al. A prognostic index in primary breast cancer. Br J Cancer 1982;45:361−6.
[5] Arshi B, Smits LJ, Wynants L, Cowley LE, Reeve K, Rijnhart E. Number of publications on new clinical prediction models: a systematic literature search. OSF Preprints 2023. https://doi.org/10.31219/osf.io/4txc6.
[6] Damen JAAG, Hooft L, Schuit E, Debray TPA, Collins GS, Tzoulaki I, et al. Prediction models for cardiovascular disease risk in the general population: systematic review. BMJ 2016;353:i2416.
[7] Dhiman P, Ma J, Navarro CA, Speich B, Bullock G, Damen JA, et al. Reporting of prognostic clinical prediction models based on machine learning methods in oncology needs to be improved. J Clin Epidemiol 2021;138:60−72.
[8] Dhiman P, Ma J, Andaur Navarro CL, Speich B, Bullock G, Damen JA, et al. Methodological conduct of prognostic prediction models developed using machine learning in oncology: a systematic review. BMC Med Res Methodol 2022;22:1−16.

[9] Simon-Pimmel J, Foucher Y, L'eger M, Feuillet F, Bodet-Contentin L, Cinotti R, et al. Methodological quality of multivariate prognostic models for intracranial haemorrhages in intensive care units: a systematic review. BMJ Open 2021;11(9):e047279.
[10] Gao S, Albu E, Tuand K, Cossey V, Rademakers F, Wynants L, et al. Systematic review finds risk of bias and applicability concerns for models predicting central line-associated bloodstream infection (CLA-BSI). J Clin Epidemiol 2023;161:127−36.
[11] Navarro CLA, Damen JA, Takada T, Nijman SW, Dhiman P, Ma J, et al. Systematic review finds "Spin" practices and poor reporting standards in studies on machine learning-based prediction models. J Clin Epidemiol 2023;158:99−110.
[12] White N, Parsons R, Collins G, Barnett A. Evidence of questionable research practices in clinical prediction models. BMC Med 2023; 21(1):1−10.
[13] Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. Ann Intern Med 2015;162:55−63.
[14] Wolff RF, Moons KG, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. Ann Intern Med 2019; 170:51−8.
[15] Najafabadi AHZ, Ramspek CL, Dekker FW, Heus P, Hooft L, Moons KG, et al. TRIPOD statement: a preliminary pre-post analysis of reporting and methods of prediction models. BMJ Open 2020; 10(9):e041537.
[16] Chalmers I, Glasziou P. Avoidable waste in the production and reporting of research evidence. Lancet 2009;374(9683):86−9.
[17] Schwab S, Janiaud P, Dayan M, Amrhein V, Panczak R, Palagi PM, et al. Ten simple rules for good research practice. PLoS Comput Biol 2022;18(6):e1010139.
[18] Tse T, Fain KM, Zarin DA. How to avoid common problems when using ClinicalTrials. gov in research: 10 issues to consider. BMJ 2018;361:k1452.
[19] Kappen TH, van Klei WA, van Wolfswinkel L, Kalkman CJ, Vergouwe Y, Moons KG. Evaluating the impact of prediction models: lessons learned, challenges, and recommendations. Diagn Progn Res 2018;2(1):1−11.
[20] Riley RD, Hayden JA, Steyerberg EW, Moons KG, Abrams K, Kyzas PA, et al. Prognosis Research Strategy (PROGRESS) 2: prognostic factor research. PLoS Med 2013;10(2):e1001380.
[21] van Smeden M, Reitsma JB, Riley RD, Collins GS, Moons KG. Clinical prediction models: diagnosis versus prognosis. J Clin Epidemiol 2021;132:142−5.
[22] Mullen L. Textreuse: detect text reuse and document similarity. R package version 0.1.5 10.32614/CRAN.package.textreuse.
[23] Smalheiser NR, Holt AW. A web-based tool for automatically linking clinical trials to their publications. J Am Med Inform Assoc 2022 01; 29:82230.
[24] Geersing GJ, Bouwmeester W, Zuithoff P, Spijker R, Leeflang M, Moons K. Search filters for finding prognostic and diagnostic prediction studies in Medline to enhance systematic reviews. PLoS One 2012;7:e32844.
[25] MeSH Data 2022 Release. 2022. Available at: https://www.nlm.nih.gov/databases/download/mesh.html. Accessed November 2, 2022.
[26] Glossary of Common Site Terms. 2023. Available at: https://classic.clinicaltrials.gov/ct2/about-studies/glossary. Accessed September 26, 2023.
[27] Hooft L, Bossuyt PM. Prospective registration of marker evaluation studies: time to act. Clin Chem 2011;57:1684−6.
[28] Ross JS, Tse T, Zarin DA, Xu H, Zhou L, Krumholz HM. Publication of NIH funded trials registered in ClinicalTrials. gov: cross sectional analysis. BMJ 2012;344:d7292.
[29] Korevaar DA, Ochodo EA, Bossuyt PM, Hooft L. Publication and reporting of test accuracy studies registered in ClinicalTrials. gov. Clin Chem 2014;60:651−9.
[30] Rifai N, Bossuyt PM, Ioannidis JP, Bray KR, McShane LM, Golub RM, et al. Registering diagnostic and prognostic trials of tests: is it the right thing to do? Clin Chem 2014;60:1146−52.
[31] Baudart M, Ravaud P, Baron G, Dechartres A, Haneef R, Boutron I. Public availability of results of observational studies evaluating an intervention registered at ClinicalTrials. gov. BMC Med 2016; 14(1):1−11.
[32] Chen R, Desai NR, Ross JS, Zhang W, Chau KH, Wayda B, et al. Publication and reporting of clinical trial results: cross sectional analysis across academic medical centers. BMJ 2016;352:i637.
[33] Zwierzyna M, Davies M, Hingorani AD, Hunter J. Clinical trial design and dissemination: comprehensive analysis of Clinicaltrials. gov and PubMed data since 2005. BMJ 2018;361:k2130.
[34] National Institute for Health and Care Excellence. Cardiovascular disease: risk assessment and reduction, including lipid modification [NICE Guideline CG181]. 2014. Available at: https://www.nice.org.uk/guidance/cg181. Accessed May 24, 2023.
[35] National Institute for Health and Care Excellence. Sepsis: recognition, diagnosis and early management and diagnosis [NICE Guideline NG51]. 2016. Available at: https://www.nice.org.uk/guidance/cg181. Accessed September 13, 2017.
[36] Christodoulou E, Ma J, Collins GS, Steyerberg EW, Verbakel JY, Van Calster B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. J Clin Epidemiol 2019;110:12−22.
[37] Nusinovici S, Tham YC, Yan MYC, Ting DSW, Li J, Sabanayagam C, et al. Logistic regression was as good as machine learning for predicting major chronic diseases. J Clin Epidemiol 2020;122:56−69.
[38] Collins GS, Whittle R, Bullock GS, Logullo P, Dhiman P, de Beyer JA, et al. Open Science Practices need substantial improvement in prognostic model studies in Oncology using Machine Learning. J Clin Epidemiol 2023;165:111199. [Online ahead of print].
[39] Peat G, Riley RD, Croft P, Morley KI, Kyzas PA, Moons KG, et al. Improving the transparency of prognosis research: the role of reporting, data sharing, registration, and protocols. PLoS Med 2014;11(7): e1001671.
[40] Steyerberg EW, Moons KG, van der Windt DA, Hayden JA, Perel P, Schroter S, et al. Prognosis Research Strategy (PROGRESS) 3: prognostic model research. PLoS Med 2013;10(2):e1001381.
[41] Stewart L, Moher D, Shekelle P. Why prospective registration of systematic reviews makes sense. Syst Rev 2012;1(1):1−4.
[42] Serghiou S, Axfors C, Ioannidis JP. Lessons learnt from registration of biomedical research. Nat Human Behav 2023;7(1):9−12.
[43] van Royen FS, Moons KG, Geersing GJ, van Smeden M. Developing, validating, updating and judging the impact of prognostic models for respiratory diseases. Eur Respir J 2022;60(3):2200250.
[44] Riley RD, Ensor J, Snell KI, Harrell FE, Martin GP, Reitsma JB, et al. Calculating the sample size required for developing a clinical prediction model. BMJ 2020;368:m441.
[45] Gliklich RE, Leavy MB, Dreyer NA. Registry design. In: Registries for Evaluating Patient Outcomes: A User's Guide. 4th edition. Rockville (MD): Agency for Healthcare Research and Quality (US); 2020.
[46] Barnett AG, Glasziou P. Target and actual sample sizes for studies from two trial registries from 1999 to 2020: an observational study. BMJ Open 2021;11:e053377.
[47] Barnett AG, Wren JD. Examination of CIs in health and medical journals from 1976 to 2019: an observational study. BMJ Open 2019;9(11):e032506.
[48] Van Calster B, Steyerberg EW, Wynants L, van Smeden M. There is no such thing as a validated prediction model. BMC Med 2023; 21(1):70.
[49] de Hond AA, Steyerberg EW, van Calster B. Interpreting area under the receiver operating characteristic curve. Lancet Digit Health 2022; 4(12):e853−5.
[50] Williams RJ, Tse T, Harlan WR, Zarin DA. Registration of observational studies: is it time? Can Med Assoc J 2010;182(15):1638.

[51] Dhiman P, Whittle R, Van Calster B, Ghassemi M, Liu X, McCradden MD, et al. The TRIPOD-P reporting guideline for improving the integrity and transparency of predictive analytics in healthcare through study protocols. Nat Mach Intell 2023;5(8):816−7.
[52] Frandsen TF, Eriksen MB, Hammer DMG, Christensen JB. PubMed coverage varied across specialties and over time: a large-scale study of included studies in Cochrane reviews. J Clin Epidemiol 2019;112:59−66.
[53] Anderson KR. bioRxiv: trends and analysis of five years of preprints. Learn Publ 2020;33(2):104−9.
[54] Eckmann P, Bandrowski A. PreprintMatch: a tool for preprint to publication detection shows global inequities in scientific publication. PLoS One 2023;18:e0281659.
[55] Korevaar DA, Hooft L, Askie LM, Barbour V, Faure H, Gatsonis CA, et al. Facilitating prospective registration of diagnostic accuracy studies: a STARD initiative. Clin Chem 2017;63:1331−41.
[56] Arshi B, Wynants L, Rijnhart E, Reeve K, Cowley LE, Smits LJ. Protocol: what proportion of clinical prediction models make it to clinical practice? Protocol for a two-track follow-up study of prediction model development publications. BMJ Open 2023;13(5): e073174.
[57] Good BM, Ainscough BJ, McMichael JF, Su AI, Griffith OL. Organizing knowledge to enable personalization of medicine in cancer. Genome Biol 2014;15(8):1−9.
[58] Landrum MJ, Lee JM, Benson M, Brown G, Chao C, Chitipiralla S, et al. ClinVar: public archive of interpretations of clinically relevant variants. Nucleic Acids Res 2016;44:D862−8.
[59] Diong J, Kroeger CM, Reynolds KJ, Barnett A, Bero LA. Strengthening the incentives for responsible research practices in Australian health and medical research funding. Res Integr Peer Rev 2021; 6(1):11.