Journal of Clinical Epidemiology 169 (2024) 111287

ORIGINAL RESEARCH

Prediction model protocols indicate better adherence to recommended guidelines for study conduct and reporting

Paula Dhiman[a,*], Jie Ma[a], Shona Kirtley[a], Elizabeth Mouka[a,1], Caitlin M. Waldron[a,1], Rebecca Whittle[a,b], Gary S. Collins[a]

[a]*Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, Centre for Statistics in Medicine, University of Oxford, Oxford, OX3 7LD, UK*

[b]*NIHR Blood and Transplant Research Unit in Data Driven Transfusion Practice, Nuffield Division of Clinical Laboratory Sciences, Radcliffe Department of Medicine, University of Oxford, Oxford, UK*

Accepted 15 February 2024; Published online 20 February 2024

Abstract

Background and Objective: Protocols are invaluable documents for any research study, especially for prediction model studies. However, the mere existence of a protocol is insufficient if key details are omitted. We reviewed the reporting content and details of the proposed design and methods reported in published protocols for prediction model research.

Methods: We searched MEDLINE, Embase, and the Web of Science Core Collection for protocols for studies developing or validating a diagnostic or prognostic model using any modeling approach in any clinical area. We screened protocols published between Jan 1, 2022 and June 30, 2022. We used the abstract, introduction, methods, and discussion sections of The Transparent Reporting of a multivariable prediction model of Individual Prognosis Or Diagnosis (TRIPOD) statement to inform data extraction.

Results: We identified 30 protocols, of which 28 were describing plans for model development and six for model validation. All protocols were open access, including a preprint. 15 protocols reported prospectively collecting data. 21 protocols planned to use clustered data, of which one-third planned methods to account for it. A planned sample size was reported for 93% development and 67% validation analyses. 16 protocols reported details of study registration, but all protocols reported a statement on ethics approval. Plans for data sharing were reported in 13 protocols.

Conclusion: Protocols for prediction model studies are uncommon, and few are made publicly available. Those that are available were reasonably well-reported and often described their methods following current prediction model research recommendations, likely leading to better reporting and methods in the actual study.

Keywords: Prediction model; Protocol; Reporting; TRIPOD; Systematic Review; Methodology

1. Introduction

A study protocol is a 'roadmap' that provides a detailed plan for a research study before it is conducted. Protocols provide key details on the study rationale, objectives, design, data collection, methodology, statistical considerations, dissemination plans, and organization of a research study; they are a crucial component in mapping out any research project [1]. Developing a study protocol prompts

Funding statement: Gary Collins, Shona Kirtley, Jie Ma, and Rebecca Whittle are supported by Cancer Research UK (programme grant: C49297/A27294). Rebecca Whittle is also supported by the NIHR Blood and Transplant Research Unit in Data Driven Transfusion Practice (NIHR203334). Paula Dhiman is supported by Cancer Research UK (project grant: PRCPJT-Nov21\100021). This publication presents independent research funded by Cancer Research UK, and the National Institute for Health Research (NIHR). The views expressed are those of the author(s) and not necessarily those of the Cancer Research UK, the NHS, the NIHR or the Department of Health and Social Care.

Registration: osf.io/a8mcj/.

[1] Elizabeth Mouka and Caitlin Waldron were visiting students at the Center for Statistics in Medicine, Nuffield Department of Orthopedics, Rheumatology and Musculoskeletal Sciences, University of Oxford.

* Corresponding author. Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, Centre for Statistics in Medicine, University of Oxford, Oxford, OX3 7LD, UK.

E-mail address: paula.dhiman@csm.ox.ac.uk (P. Dhiman).

https://doi.org/10.1016/j.jclinepi.2024.111287
0895-4356/

What is new?

Key findings

- Protocols of prediction model studies showed good reporting and conduct, often reporting and describing their methods that follow current prediction model research recommendations.

What this adds to what was known?

- Compared to the systematic reviews of final results publications of prediction model studies that have highlighted poor reporting and methodological weaknesses in studies, we observed generally good reporting and the conduct of protocols describing plans for prediction model development or validation.

What is the implication and what should change now?

- We recommend that research teams develop and make available a protocol before embarking on their prediction model study to enhance research integrity and transparency and to improve methodological conduct and reporting quality in the final study reports.

the research team to think about and plan all steps in the study, minimizing the risk of encountering problems once the study has started. Along with study registration, developing a protocol enhances transparency, replication, research integrity, and promotes good research practice.

Protocols are mandatory for some study types, such as randomized controlled trials [2] (eg, for funding, ethical approval, and increasingly for manuscript submission), but they are an important document and should be mandatory for other types of research. Several studies have highlighted poor reporting and methodological weaknesses or flaws in publications describing the development and validation of a prediction model [3–7]. While having an available protocol is no guarantee of a well-designed and conducted study, it is worth noting that prediction model studies rarely cite the existence or availability of a protocol, suggesting these studies are rarely protocol driven.

However, the mere existence of a protocol is insufficient if key details are omitted. Only a complete and transparently reported protocol that provides a full, detailed plan addressing all aspects of the study design, conduct (including data collection), and proposed analysis has intrinsic value. A well-reported research protocol for a prediction model study will describe all key and study-specific design features, such as strategies to ensure the data are representative of the target population and approaches for dealing with missing data.

To achieve maximum impact, the protocol should ideally be freely available and accessible by others to view and review, either as a published peer-reviewed article (if study funds permit), on a registration platform, or on a preprint server. However, there is limited evidence about the availability of study protocols and their reporting completeness. The aim of this study is to review the content and reported details of the planned methodological conduct of published protocols or protocols available on preprint websites (for studies planning the development or validation of a clinical prediction model, regardless of whether regression or machine learning methods have been used).

2. Methods

2.1. Protocol registration and reporting standards

This study was registered on the Open Science Framework (osf.io/a8mcj) on 21 July 2022. We reported our study following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guideline and its extension for Reporting Literature Searches in Systematic Reviews (PRISMA-S) [8,9].

2.2. Information sources

We searched the MEDLINE (via OVID), Embase (via OVID), and Web of Science Core Collection (via Clarivate Analytics) medical literature databases for protocols for studies developing or validating a diagnostic or prognostic model using any modeling approach in any clinical area on June 30, 2022. The search strategy was limited to protocols that were published between January 1, 2020 and June 30, 2022 (the date that the search was conducted). The results set was then further restricted to protocols published between January 1, 2022 and June 30, 2022 (which were then entered into the study for screening).

The full search strategies for the three databases are provided in Supplementary Boxes 1–3. The search terms included relevant Mesh and EMTREE headings and free-text terms. Protocol-related search terms were based on a search string developed by Madden et al., which was modified to account for differing database-controlled vocabulary headings and to search in the title and/or keyword fields [10]. An experienced and qualified information specialist (SK) was involved in the development of the search strategy for all three databases.

2.3. Eligibility criteria

We included protocols for studies developing or validating a prediction model for individualized health outcomes (including response to treatment) using any regression or machine learning approach within any clinical field in 2022. We included protocols that involved developing or validating a prediction model containing two or more predictors.

We excluded protocols for imaging studies, or studies using imaging parameters as candidate predictors in the model; speech recognition/voice pattern studies, or studies using speech parameters as candidate predictors; genetic studies, or studies using genetic risk factors as candidate predictors; and molecular studies, or studies using molecular markers as candidate predictors. We also excluded protocols investigating risk or prognostic factors, secondary research (eg, reviews of prediction models), and conference abstracts. Protocols to be included were limited to English language studies only.

2.4. Study selection, data extraction, and data management

Publications identified from the search strategies were imported into Endnote reference software where they were deduplicated and then imported into Rayyan web application where they were screened [11,12].

Two independent researchers (PD and JM) screened the titles and abstracts of the identified records and then the full text of protocols potentially meeting the eligibility criteria. Two independent researchers, from a combination of four reviewers (PD, JM, EM, and CW), performed a duplicate data extraction of eligible publications after initially piloting the form on five protocols and subsequent amendments. Disagreements were discussed and adjudicated by a third reviewer (GSC), where necessary.

Details from the included articles were extracted using data extraction forms, comprising article publication information and reporting items from the abstract, introduction, methods, and discussion sections (not results) of The Transparent Reporting of a multivariable prediction model of Individual Prognosis Or Diagnosis (TRIPOD) statement [13], the recommended reporting guideline for prediction model studies. We supplemented TRIPOD reporting items with additional items relating to data structure (whether data was clustered and if and how this was handled) and reporting of study registration, ethics, patient and public involvement, data sharing, and funding statements.

We extracted details on the proposed methodology to address key study design features of prediction model studies, including the study type (eg, development and/or validation), clinical specialty, outcome to be predicted, time span of prediction, intended moment of using the prediction models, data source, number of models developed or validated, modeling approach (eg, logistic, survival, random forests, artificial neural networks), method of internal validation, methods to handle continuous predictors (where relevant), and methods to handle missing data.

The data extraction form was implemented using MS Excel.

2.5. Data items, summary measures, and synthesis of results

Findings were summarized using descriptive statistics and a narrative synthesis. We described levels of reporting using numbers and percentages of studies that fully report, partially report (used when an item included multiple details and some but not all details were reported), or do not report the specified items. Items were extracted separately for the development and, if planned, for the validation of the models. All analyses were carried out in Stata v17 [14].

3. Results

3.1. Study flow

The search in Embase, Medline, and Web of Science retrieved 17,577 protocols published between January 1, 2020 and June 30, 2022, and 2380 protocols published between January 1, 2022 and June 30, 2022 were screened for inclusion (after deduplication). Thirty protocols met the eligibility criteria and were included in our review after title and abstract and then full text screening. Twenty-four protocols were describing plans for developing a prediction model; four were describing plans for developing and also validating a prediction model; and two protocols were only planning to validate an existing prediction model. Figure 1 provides the study flow for our review. Citations for all included protocols are provided in Supplementary Table 1.

3.2. Study characteristics

Table 1 describes the publication and study characteristics of the included protocols. Prediction models were commonly being developed and validated for oncology (n = 6/30, 20%) [15−20] and cardiovascular (n = 5/30; 17%) [21−25] clinical specialties and were often predicting health outcomes (n = 24/30; 80%) [15−22,25−40] for patient target populations (n = 18/30; 60%) [16,19,21−25,27,29,31,34,35,37−42]. Development and validation of a prediction model were often the primary aims reported in the protocol (n = 22/30; 73%) [15−20,22,24,26,28−32,34−37,41−44]. The majority of models were prognostic (n = 18/30; 60%) [20−27,29,33−37,39,41,42,44], and a majority of models were predicting a binary outcome (n = 20/30; 63%) [16,19,21,23−26,28−34,36−38,40−42]. The approach to be used to develop the prediction models (eg, logistic regression, XGBoost) was reported in 27 out of the 28 protocols (96%) [15−18,21,23−40,42−44]. A total of 49 models were planned to being developed in 28 protocols, with a median of 1 model developed per study (n = 28; IQR: 1−2; range 1−6).

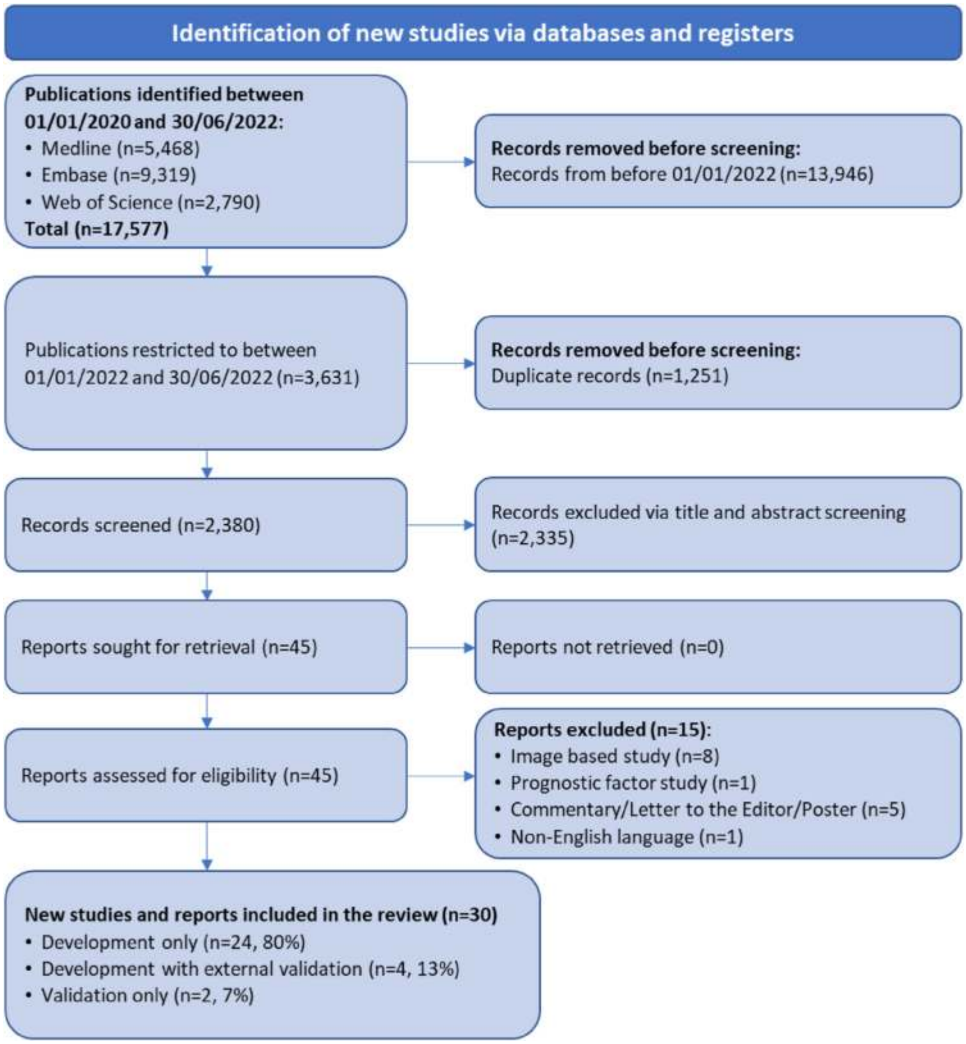

Figure 1. Flow diagram of included studies. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Over a third of the included protocols were published in BMJ Open (n = 11/30; 37%) [15,21,24,26,27,28,29,32,33,38,42], and 1 protocol was a preprint in medRxiv [17] (Supplementary Table 1).

3.3. *Reporting quality*

3.3.1. *Title, abstract, and introduction*

Half of the protocols clearly reported that they were developing or validating a prediction model in the title (n = 15/30; 50%) [15,17−19,24,26,28−30,32,34,37,39,42,43], while over two-thirds (n = 21/30, 70%) [17−19,21,24−26,28−30,32,33,35−39,41−44] and over three-quarters (n = 23/30; 77%) [15,17−19,22,24−29,31,33−41,43,44] clearly reported the target population they wanted to predict for and the outcome to be predicted, respectively. Only one-third fully reported all three pieces of information (n = 10/30; 33%) [17−19,24,26,28,29,37,39,43].

Seven protocols (n = 7/30; 23%) did not explicitly specify in the title that they were developing a prediction model or the outcome to be predicted—broadly reporting predicting a health outcome and not sufficiently describing the target population—and only reporting the clinical area for prediction [20,21,23,27,30,31,33].

Seven protocols (n = 7/30; 23%) fully or partly reported the study aims, study design, clinical setting, summary of participants (ie, target population), sample size, predictors, outcome, and summary of analysis methods for the development or validation of their model in the abstract [21,26,30,33,36,37,43]. Half of the protocols did not report any sample size information (n = 15/30; 50%) [15,17,18,19,20,23,24,27,29,31,32,34,38,40,41], and over 50% did not provide a summary of predictors (n = 16/30) [15,18−20,22,24,25,27−29,32,35,38−41]. Two-thirds of the protocols either did not report or only partly reported the analysis methods for model development or validation (n = 20/30; 67%) [15,19,21,23,25,26,27,28,31,33,34,35,35,

Table 1. Characteristics for all included protocols (n = 30)

Study characteristic	n	%
Clinical specialty		
Oncology	6	20
Cardiovascular	5	17
Maternal health/Obstetrics and gynecology	3	10
Musculoskeletal/Orthopedics	3	10
Endocrinology	2	7
Neurology/Neuropsychology	2	7
Psychiatry	2	7
Surgical	2	7
Infectious diseases	1	3
Health service	1	3
Respiratory	1	3
Stroke	1	3
Urology	1	3
Target population to be predicted for		
Patients	18	60
General population	4	13
Neonates/children	2	7
Pregnant women and young mothers	2	7
Healthy women	1	3
GP referrals	1	3
Stroke survivors	1	3
Cancer survivors	1	3
Primary outcome to be predicted		
Health outcome	24	80
Measurement/Score/Number	4	13
Treatment response	2	7
Was development or validation of the prediction model the primary or secondary aim in the protocol?		
Primary aim	22	73
Secondary aim	8	27
Type of prediction		
Prognostic	18	60
Diagnostic	8	27
Risk prediction (outcome in future in healthy people)	4	13
Type of outcome to be predicted		
Binary	20	67
Time to event	6	20
Continuous	2	7
Count	1	3
Unclear	1	3
Model to be developed[a]		
Logistic regression	12	24
Cox proportional hazard model	6	12
Deep learning model	5	10
Support vector machine	4	8
XGBoost	4	8
Convolutional neural network	4	8
Random forest[b]	3	6
Linear regression	3	6
Nearest neighbor algorithms	2	4
Decision tree	2	4
Competing risk model	1	2
Elastic net (underlying model not specified)	1	2
Negative binomial regression	1	2
Unclear[c]	1	2
Model to be validated[d]		
Logistic regression	2	18
XGBoost	2	18
Time to event with competing risks	2	18
Elastic net regression (underlying model not specified)	1	9
Cox proportional hazard model	1	9
Random survival forest	1	9
Deep learning' neural networks	1	9
Artificial neural network	1	9

(*Continued*)

[a] Denominator is the total 49 models that are reported to be developed in the 28 development or development with validation protocols.
[b] Includes one random survival forest.
[c] Protocol specifies 'such as SVM [support vector machine] or linear discriminant analysis'.
[d] Denominator is the total 11 models that are reported to be validated in the six development with validation or validation only study protocols.

38,39,40,41,42,43,44], with details about model building procedures, internal and external validation methods, or model performance measures often missing (n = 9/20, 45%) [17,21,26,28,35,39,40,43,44].

All protocols fully reported the objectives in the introduction, including whether the study described the development or validation of the model or both. Most protocols also fully reported the medical context of model development or validation (n = 27/30, 90%) [15−19,21−26,28−30,32−44]. All studies provided a rationale for model development and validation, of which 50% (n = 15/30) [17,18,20−22,24,26−28,34,36−39,42] included details of the target population and two-thirds (n = 20/30) [15,17,18,20−22,24,27,29,30,32−34,36−41,43] included details of the clinical decision the model was planned to support.

3.3.2. Methodology

In the 30 included protocols, 28 development [15−18,20,21,23−44] and six validation analyses [15,19,22,28,34,36] were described. Figure 2 describes the reporting completeness of the 28 development analyses. The completeness of reporting for each of these analyses is described in Supplementary Table 2.

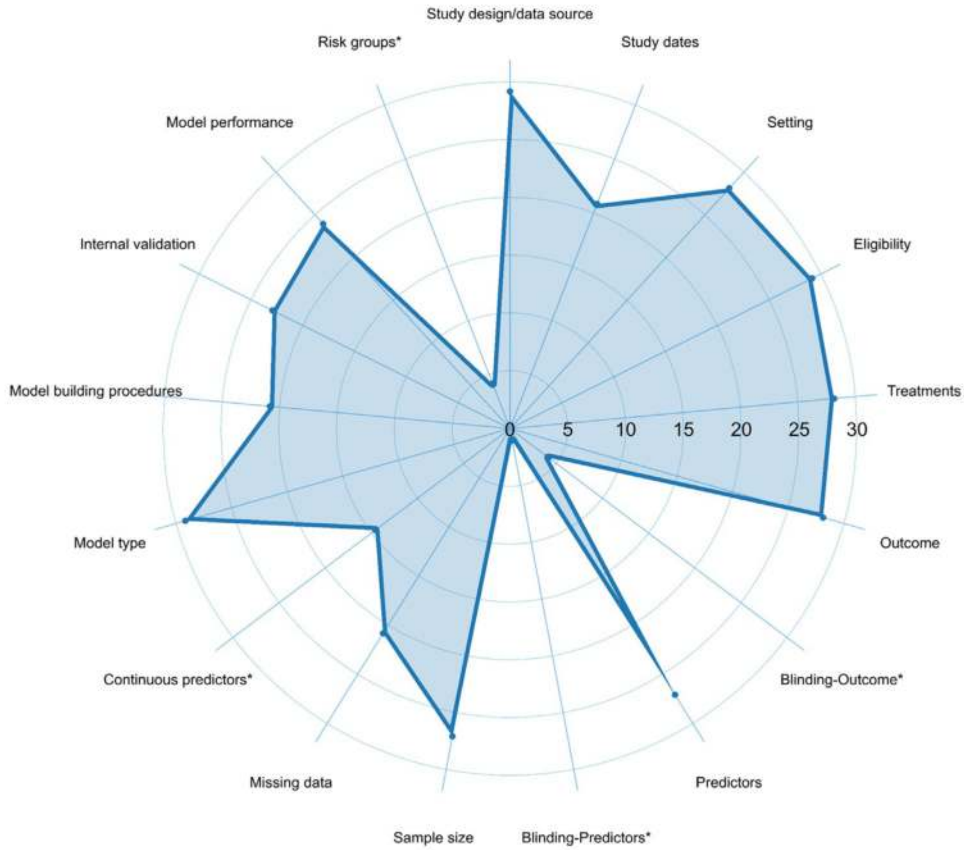

Figure 2. Methodology reporting completeness of the development analyses described in 28 protocols. *These items account for protocols that were not applicable: blinding of the outcome was not applicable for five analyses; blinding of predictors was not applicable for seven analyses; handling of continuous predictors was not applicable for one analysis; details on risk groups was not applicable for 21 analyses. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

3.3.2.1. Study design and data source. Forty-seven percent of protocols (n = 14/30) described using a cohort study design [16,20−23,25,29,31,33,36,37,39,42,44], including two protocols [20,23] that also proposed to use data from randomized controlled trials. A 10th of protocols (n = 3/30, 10%) proposed to use a cross-sectional study design [28,35,43], and for one protocol, the study design was unclear [41].

One-third of protocols proposed to use electronic health records (n = 10/30; 33%) [15,17−19,24,30,32,34,38,40], one proposed to use data from an individual patient data meta-analysis [27], and one proposed to use a registry [26]. Half of the protocols reported prospectively collecting data (n = 15/30; 50%) [20−22,25,26,31,33,35−37,39,41−44], and half described using an existing dataset (n = 15/30; 50%) [15−19,23,24,27−30,32,34,38,40] (including one development and validation study that would use an existing dataset for model development and a prospectively collected dataset for model validation [28]).

The study design or data source was justified in seven protocols (n = 7/30, 23%) [15,18,28−30,35,40]. One protocol cited the size of their data (including outcome number) and the availability of predictors in the data when justifying their development data [29]. Six protocols (n = 6/30; 20%) cited representativeness when justifying their development data [15,18,28,30,35,40], including two protocols that also cited the size of the dataset [15,18] and one protocol that cited size and data availability when justifying their validation dataset [15].

3.3.2.2. Handling clustered data. Over two-thirds of the protocols described using data with a clustered structure (n = 21/30, 70%) (20 multicentre datasets [15−20,22−26, 29,31,34,36,38−41,43] and one multistudy individual patient dataset [27]) to develop or validate their models, six protocols will use single center data (n = 6/30, 20%) [21,30,32,35,37,44], and for three protocols, data structure was unclear (n = 3/30; 10%) [28,33,42].

Of the 21 study protocols using a clustered dataset for development or validation, 14 mentioned clustering in their methodology (n = 14/21, 67%) [15,17,18,22−25,27,29, 34,36,39,41,43], but only seven reported methods to account for it (n = 7/21, 33%) [15,17,18,23,27,29,34]. Three protocols planned to use internal-external cross validation methods (n = 3/30; 10%) [15,18,27]. One protocol reported 'clustering techniques to identify patient subgroups' [23], one protocol reported that 'internal validation will be undertaken using bootstrapping of the entire development data set (accounting for the clustering of participants within studies) [29], one protocol reported that they would use a random effects model [34], and one protocol reported that they would use Wald's methods' agglomerative hierarchical clustering, [17].

3.3.2.3. Sample size calculation or justification. Sample size was reported for 26 (n = 26/28; 93%) development [15−18,20,23−39,41−44] and four (n = 4/6; 67%) validation analyses [15,19,23,34]. Nine (n = 9/28; 32%) of the model development analyses based their development sample size calculations on formulae by Riley et al. [15,17,18,24,25,29,30,35,44], six (n = 6/28; 21%) used events per predictor [16,27,32,34,39,42], three (n = 3/28; 11%) used power calculations [21,33,41], three (n = 3/28; 11%) based it on outcome prevalence [36,38,43], one used a learning curve [31], one used Kendall m criterion [37], and one protocol used 'piecewise regression' [23]. The final two protocols reported that a formal or statistical sample size calculation was not possible because of the complexity of methods [machine learning] [26] and study design [proof of concept study] [20] of which one protocol then used previous literature to inform their sample size [26]. For the validation analyses, one protocol reported using events per variable [34], one study used precision [22], one study used reaching an area under the curve of at least 0.7 [19], and one protocol used outcome prevalence [28]. A summary of the sample size used to develop and validate the models is provided in Table 2.

3.3.2.4. Handling missing data and continuous predictors. Twenty-one protocols described how the researchers planned to handle missing data (n = 21/30; 70%) [15,17−20,24,25,27,29−32,34−38,40−42,44], of which 10 protocols describe multiple imputation methods [15,17,18,24,29,32,35,37,42] and an additional four protocols describe multiple imputation with methods to exclude missing data or conduct a complete case analysis depending on the missingness mechanism and amount of missing data [27,30,36,38]. One protocol reported they will use a single imputation analysis [44], one protocol mentioned the missing indicator method [40], one protocol reported they will carry out a complete case analysis approach [34], one protocol reported manual searching to resolve missing data [25], one protocol reported an artificial neural network [19], and for two protocols, the approach to handling missing data was unclear [20,31].

Twenty-two protocols planned to develop at least one model using regression-based methods (n = 22/28; 79%) [15−18,21,24−29,32−35,37−39,41−44]. Of these protocols, 64% reported planned methods to handle continuous predictors (n = 14/22; 64%) [15,17,18,21,23,27−29,32,34,35,38,39,42]. Four protocols reported restricted cubic splines [15,27,34,38], three protocols reported fractional polynomials [17,18,35] and 1 protocol reported plans to use both of these methods to handle continuous predictors [29]. One protocol planned transformations [28], two reported plans to categorise their continuous predictors or to treat them as linear [32,39] and one protocol reported they would use autoencoder neural networks but did not specify any method for their planned logistic regression model [23]. For two protocols, detailed methods planned to handle continuous predictors was unclear [21,42].

3.3.2.5. Internal and external validation methods and model performance measures. Of the 28 protocols that included a model development, plans and details to carry out an internal validation were reported in 21 (n = 21/28; 75%) [15−18,20,24−30,34−37,39,40,42−44]. Seven of these protocols planned to use random split sample methods [17,25,26,28,39,43,44], including 1 protocol that also planned bootstrapping methods [39], and two protocols will use a temporal split sample approach [34,40].

Six protocols will use cross-validation [16,20,24,30,37,42] (including one protocol that will also use bootstrapping [16]), three protocols will use an internal-external cross-validation [15,18,27], and three protocols will use bootstrapping [29,35,36]. Four protocols explicitly reported their plan to produce optimism-adjusted performance measures [16,27,29,35].

A summary of the intended performance measures is provided in Table 3.

3.3.3. Discussion and supporting information

Study limitations (such as sample size limitations, a nonrepresentative study sample) were fully reported in 77% (n = 23/30; 77%) [15,17,18,21,23,24,26−35,37−39,41−44] of protocols and potential clinical use and implications from model development and validation was discussed in 83% of protocols (n = 25/30; 83%) [16−23,26,28−35,37−44].

Table 2. Summary of the minimum required sample size for model development or validation

Model development (n = 28)	Protocols reported (%)	Median	p25-p75	Range
Minimum total required sample size	26 (93%)	838	429−4000	114−450,000
Number of events	4 (14%)	608	227−1057	154−1199
Model validation (n = 6)				
Minimum total required sample size	3 (50%)	1014	300−1350	300−1350
Number of events	0	-	-	-

Table 3. Summary of intended model performance measures

Performance measures	n (%)
Discrimination	26 (87%)
Area under the receiver operating characteristic curve/C-statistic	26 (87%)
Royston's D-statistic	3 (10%)
Area under the precision-recall curve	2 (7%)
Calibration	19 (63%)
Plot	14 (47%)
Hosmer-Lemeshow test	6 (20%)
Calibration (slope or intercept)	4 (13%)
Brier score	2 (7%)
No metric specified	2 (7%)
Decision curve analysis	5 (17%)
Classification measures	13(43%)
Sensitivity/specificity	11 (37%)
PPV/NPV	7 (23%)
Accuracy	4 (13%)
Classification table/confusion matrix	2 (7%)
F1 score	2 (7%)
Likelihood ratios	1 (3%)
Other	4 (13%)
R-squared	2 (7%)
Akaike's information criterion	1 (3%)
Matthew correlation coefficient	1 (3%)
Net reclassification index	1 (3%)
Youden index	1 (3%)

Values are all out of $n = 30$ protocols.
NPV, negative predicted value, PPV, positive predicted value.

Only half of the protocols reported details and a link to study registration ($n = 16/30$; 53%) [16,19,20,22,24,25,27,28,30,31,35,37,38,41,43,44]. All protocols reported a statement on ethics approval, although six protocols ($n = 6/30$; 20%) did not provide a supporting ethics approval number [16,27,29,32,36,41]. The sources of funding and the role of the funders were reported in all protocols: 27 protocols ($n = 27/30$; 90%) reported funding from government or academic institutions [15–34,36–41,44], two reported funding from industry [35,43] and 1 protocol reported receiving no funding [42]. A data sharing statement was reported in 13 protocols [16,19,22,24,27,28,30,36,38–41,44], of which five protocols specified that their data will be available on request [19,24,28,36,38], 1 protocol reported making metadata and code available [30], four reported making anonymized data (freely or publicly) available [16,22,27,41], and three protocols specified data sharing was not applicable [39,40,44].

Information about the availability of supplementary resources (such as statistical analysis plans, web calculator, and datasets) was reported in 14 protocols ($n = 14/30$; 47%) [19,20,22,24–28,30,35,36,38,43,44]. In these 14 protocols, supplementary material included patient information sheets [20,22,26,35], patient consent forms [20,22,26,28,35], data monitoring information [19], study questionnaires [38,43], ethics approval form and letters [20,22,30], predictor and outcome information and definitions [24,35,44], sample size calculations [35], completed reporting checklists [22,36,44], data management and linkage plans [30], and statistical analysis plans [20,22,27].

Plans to disseminate the models were fully or partly reported in two-thirds of protocols ($n = 19/30$; 63%) [15–18,21,26–30,32–34,36,38,40–43], of which 16 specified disseminating via peer-reviewed publications [15–17,21,26–30,32,33,36,38,40–42], 11 via national and international conferences [16–18,21,26,29,30,32,33,36,42], two via an international meeting [27,38], one via health clinics and services [33], three via social media engagement [30,40,42], and four via study websites [17,18,29,43], app development, or electronic health systems [18,34].

4. Discussion

4.1. Summary of findings

We reviewed 30 study protocols describing plans to develop or validate a prediction model. Overall, we found a reasonable level of detail in the minimum reporting items listed in the TRIPOD reporting guideline.

The included protocols also used methods that were generally consistent with research recommendations for prediction model research [13,45–48]. Multiple imputations were most commonly planned to handle missing data; sample size formulae described by Riley et al. were most often used; fractional polynomials, restricted cubic splines, and transformations were often planned to handle continuous predictors; and calibration was commonly planned as a measure of model performance in addition to discrimination. Though some of the planned studies may still be at high or unclear risk of bias, it is encouraging that appropriate methodology is being considered before the study is conducted.

4.2. Current literature

There is a lack of research on the prevalence, availability, and reporting completeness of protocols for prediction model studies. Compared to the numerous systematic reviews of the final publications of prediction model studies [3–7], we observed better reporting of the minimum reporting criteria than would be expected when publishing the development and validation of the prediction model. In a review of prediction models developed using supervised machine learning, sample size and calibration model performance measures were reported in only 18% and 13% of the included studies, respectively [6]. In our review, however, we found that a sample size calculation or justification was reported in over 90% of the protocols, and calibration was planned to be assessed in almost two-thirds of the protocols. In both of these cases, we might expect the quality of reporting and conduct to carry over to the eventual final study publications.

It is likely the case that researchers developing and making a protocol available are atypical and that those who write protocols are more likely to be more cognizant and up-to-date in prediction model research methods. Though it is difficult to assess methodological conduct and the risk of bias without a formal assessment of the completed studies, based on our findings, we can extrapolate and imply that studies with an available protocol at the outset of the study may lead to more methodologically robust and better-reported final study publications. As key study design and analysis considerations are being made before the study commences, there is less (known) risk of deviation from the protocol (without documentation), leading to a potentially better-reported and conducted study. Therefore, increased awareness and the mandatory practice of making protocols available are recommended.

4.3. Strength and limitations

Most protocols were published open-access in peer-reviewed journals, with one made available on a preprint server. Though the protocol preprint was not peer reviewed at the time of this review and may be subject to change after peer review, this would not change our findings [49]. It is important to include protocols made available on preprint servers, as this is often an option to quickly make protocols freely available to the scientific community.

We did not search the Open Science Framework or other similar registers for protocols that may have been submitted when registering a study. We appreciate that these protocols are better than not having any, and by focusing our review of published protocols, we risk skewing our results based on the quality of reporting and methodology of the available protocols for prediction model research. However, given research that has found little improvement in reporting from preprint to peer-reviewed publication [49], we do not anticipate the inclusion of these protocols to change our findings.

One protocol identified by our search was coauthored by a coauthor of this review (GSC). This coauthor was not involved in the data extraction of the protocols in this study, and we declare this in the conflict of interest statement.

4.4. Future research and recommendations

We recommend that all research teams develop a protocol for their study at the outset to enhance research integrity and transparency and to improve methodological conduct and reporting quality in the final study reports.

Reporting guidelines are useful of final study publications to ensure the key details for the study are well reported. In our review, we amended the TRIPOD reporting guideline to describe the reporting quality of protocols. There is currently no guidance to support the reporting of protocols at the design stage, though this guidance is currently being developed [50]. This reporting guideline will be aimed at studies using any type of modeling method (eg, statistics or machine learning) to have the maximum impact on the prediction modeling field.

We also recommend that journals, preprint servers, and online platforms mandate both study and protocol registration to encourage research teams to not only consider and develop protocols for their research but to make them freely available and accessible. We recommend that protocols (particularly those not published in a journal) be included, or at least referenced, in the submission of the final results paper.

5. Conclusion

Protocols are invaluable documents for any research study, especially for prediction model studies. Though protocols for prediction model studies are rarely developed and made publicly available, we found that protocols for prediction model studies showed good reporting quality and often detailed their methods in line with current prediction model research recommendations, which is likely to lead to better reporting in the final study report and methods in the actual study.

Ethics approval and consent to participate

Not applicable.

CRediT authorship contribution statement

Paula Dhiman: Conceptualization, Data curation, Formal analysis, Methodology, Project administration, Writing – original draft, Writing – review & editing. **Jie Ma:** Data curation, Writing – review & editing. **Shona Kirtley:** Methodology, Writing – review & editing. **Elizabeth Mouka:** Data curation, Writing – review & editing. **Caitlin M. Waldron:** Data curation, Writing – review & editing. **Rebecca Whittle:** Writing – review & editing. **Gary S. Collins:** Conceptualization, Methodology, Writing – review & editing.

Data availability

Data that has informed the analysis can be found on the Open Science Framework (https://osf.io/a8mcj/).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

None.

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jclinepi.2024.111287.

References

[1] Peat G, Riley RD, Croft P, Morley KI, Kyzas PA, Moons KGM, et al. Improving the transparency of prognosis research: the role of reporting, data sharing, registration, and protocols. PLoS Med 2014;11(7):e1001671.
[2] Chan AW, Tetzlaff JM, Altman DG, Laupacis A, Gøtzsche PC, Krleža-Jerić K, et al. SPIRIT 2013 statement: defining standard protocol items for clinical trials. Ann Intern Med 2013;158:200−7.
[3] Dhiman P, Ma J, Andaur Navarro CL, Speich B, Bullock G, Damen JAA, et al. Methodological conduct of prognostic prediction models developed using machine learning in oncology: a systematic review. BMC Med Res Methodol 2022;22:101.
[4] 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.
[5] Andaur Navarro CL, Damen JAA, van Smeden M, Takada T, Nijman SWJ, Dhiman P, et al. Systematic review identifies the design and methodological conduct of studies on machine learning-based prediction models. J Clin Epidemiol 2023;154:8−22.
[6] Andaur Navarro CL, Damen JAA, Takada T, Nijman SWJ, Dhiman P, Ma J, et al. Completeness of reporting of clinical prediction models developed using supervised machine learning: a systematic review. BMC Med Res Methodol 2022;22:12.
[7] Wynants L, Calster BV, Collins GS, Riley RD, Heinze G, Schuit E, et al. Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal. BMJ 2020;369:m1328.
[8] Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71.
[9] Rethlefsen ML, Kirtley S, Waffenschmidt S, Ayala AP, Moher D, Page MJ, et al. PRISMA-S: an extension to the PRISMA statement for reporting literature Searches in systematic reviews. Syst Rev 2021;10(1):39.
[10] Madden K, Arseneau E, Evaniew N, Smith CS, Thabane L. Reporting of planned statistical methods in published surgical randomised trial protocols: a protocol for a methodological systematic review. BMJ Open 2016;6(6):e011188.
[11] The Endnote Team. Endnote. Philadelphia, PA: Clarivate Analytics; 2013.
[12] Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan — a web and mobile app for systematic reviews. Syst Rev 2016;5:210.
[13] Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. Ann Intern Med 2015;162:55−63.
[14] StataCorp. Stata Statistical Software: Release 17. College Station, TX: StataCorp LLC; 2021.
[15] Clift AK, Hippisley-Cox J, Dodwell D, Lord S, Brady M, Petrou S, et al. Development and validation of clinical prediction models for breast cancer incidence and mortality: a protocol for a dual cohort study. BMJ Open 2022;12(3):e050828.
[16] Hampton JS, Koo S, Dobson C, Stewart CJ, Neilson LJ, Montague K, et al. The COLO-COHORT (Colorectal Cancer Cohort) study: protocol for a multi-centre, observational research study and development of a consent-for-contact research platform. Colorectal Dis 2022;24(10):1216−26.
[17] Liao W, Burchardt J, Coupland C, Gleeson F, Hippisley-Cox J, DART initiative (WP6). Development, validation, and evaluation of prediction models to identify individuals at high risk of lung cancer for screening in the English primary care population using the QResearch® database: research protocol and statistical analysis plan [Internet]. medRxiv 2022. https://doi.org/10.1101/2022.01.07.22268789.
[18] Liao W, Jepsen P, Coupland C, Innes H, Matthews PC, Campbell C, et al. Development and validation of personalised risk prediction models for early detection and diagnosis of primary liver cancer among the English primary care population using the QResearch® database: research protocol and statistical analysis plan. Diagn Progn Res 2022;6(1):21.
[19] Skarping I, Dihge L, Bendahl PO, Huss L, Ellbrant J, Ohlsson M, et al. The NILS study protocol: a retrospective validation study of an artificial neural network based preoperative decision-making tool for noninvasive lymph node staging in women with primary breast cancer (ISRCTN14341750). Diagnostics 2022;12(3):582.
[20] Tzelves L, Manolitsis I, Varkarakis I, Ivanovic M, Kokkonidis M, Useros CS, et al. Artificial intelligence supporting cancer patients across Europe-The ASCAPE project. PLoS One 2022;17:e0265127.
[21] Carmona-Segovia ADM, Doncel-Abad MV, Becerra-Muñoz VM, Rodríguez-Capitán J, Sabatel-Pérez F, Flores-López M, et al. Prognostic stratification of older patients with multivessel coronary artery disease treated with percutaneous transluminal coronary angioplasty based on clinical and biochemical measures: protocol for a prospective cohort study. BMJ Open 2022;12(2):e058042.
[22] Choffat D, Farhoumand PD, Jaccard E, de la Harpe R, Kraege V, Benmachiche M, et al. Risk stratification for hospital-acquired venous thromboembolism in medical patients (RISE): protocol for a prospective cohort study. PLoS One 2022;17:e0268833.
[23] Elmer J, He Z, May T, Osborn E, Moberg R, Kemp S, et al. Precision care in cardiac arrest: ICECAP (PRECICECAP) study protocol and informatics approach. Neurocrit Care 2022;37(Suppl 2):237−47.
[24] Tefre S, Lilja-Cyron A, Arvidsson L, Bartek J, Corell A, Forsse A, et al. Endoscopic third ventriculostomy for adults with hydrocephalus: creating a prognostic model for success: protocol for a retrospective multicentre study (Nordic ETV). BMJ Open 2022;12(1):e055570.
[25] Harrison SL, Lane DA, Buckley BJR, Chatterjee K, Alobaida M, Shipley E, et al. The liverpool heart and bRain project (L-HARP): protocol for an observational cohort study of cardiovascular risk and outcomes following stroke. Vasc Health Risk Manag 2022;18:313−8.
[26] Bilal AM, Fransson E, Bränn E, Eriksson A, Zhong M, Gidén K, et al. Predicting perinatal health outcomes using smartphone-based digital phenotyping and machine learning in a prospective Swedish cohort (Mom2B): study protocol. BMJ Open 2022;12(4):e059033.
[27] Couillard S, Steyerberg E, Beasley R, Pavord I. Blood eosinophils, fractional exhaled nitric oxide and the risk of asthma attacks in randomised controlled trials: protocol for a systemic review and control arm patient-level meta-analysis for clinical prediction modelling. BMJ Open 2022;12(4):e058215.
[28] Dong W, Cheng WHG, Tse ETY, Mi Y, Wong CKH, Tang EHM, et al. Development and validation of a diabetes mellitus and prediabetes risk prediction function for case finding in primary care in Hong Kong: a cross-sectional study and a prospective study protocol paper. BMJ Open 2022;12(5):e059430.
[29] Farooq S, Hattle M, Dazzan P, Kingstone T, Ajnakina O, Shiers D, et al. Study protocol for the development and internal validation of Schizophrenia Prediction of Resistance to Treatment (SPIRIT): a clinical tool for predicting risk of treatment resistance to

antipsychotics in first-episode schizophrenia. BMJ Open 2022;12(4):e056420.
[30] Farrow L, Ashcroft GP, Zhong M, Anderson L. Using artificial intelligence to revolutionise the patient care pathway in hip and knee arthroplasty (ARCHERY): protocol for the development of a clinical prediction model. JMIR Res Protoc 2022;11(5):e37092.
[31] Kishimoto T, Kinoshita S, Kikuchi T, Bun S, Kitazawa M, Horigome T, et al. Development of medical device software for the screening and assessment of depression severity using data collected from a wristband-type wearable device: SWIFT study protocol. Front Psychiatry 2022;13:1025517.
[32] Lin Y, Xu Z, Ding X, Chen L, Dai K. Development and validation of a clinical diagnostic model for pregnant women with renal colic in the emergency department in China: a protocol for a retrospective cohort study. BMJ Open 2022;12(5):e056510.
[33] Oyungu E, El Kebbi O, Vreeman R, Nyandiko W, Monahan PO, Tu W, et al. Predicting neurodevelopmental risk in children born to mothers living with HIV in Kenya: protocol for a prospective cohort study (Tabiri Study). BMJ Open 2022;12(4):e061051.
[34] Passos SC, Stahlschmidt A, Blanco J, Spader ML, Brandão RB, Castro SMJ, et al. Derivation and validation of a national multicenter mortality risk stratification model - the ExCare model: a study protocol. Braz J Anesthesiol 2022;72(3):316−21.
[35] Ratzki-Leewing A, Ryan BL, Zou G, Webster-Bogaert S, Black JE, Stirling K, et al. Predicting real-world hypoglycemia risk in American adults with type 1 or 2 diabetes mellitus prescribed insulin and/or secretagogues: protocol for a prospective, 12-wave internet-based panel survey with email support (the iNPHORM [investigating novel predictions of hypoglycemia occurrence using real-world models] study). JMIR Res Protoc 2022;11(2):e33726.
[36] Rees CA, Kisenge R, Ideh RC, Kamara J, Samma A, Godfrey E, et al. A Prospective, observational cohort study to identify neonates and children at risk of postdischarge mortality in Dar es Salaam, Tanzania and Monrovia, Liberia: the PPDM study protocol. BMJ Paediatr Open 2022;6(1):e001379.
[37] Shen Y, Li X, Yao J. Develop a clinical prediction model for postoperative cognitive dysfunction after major noncardiac surgery in elderly patients: a protocol for a prospective observational study. Gerontology 2022;68(5):538−45.
[38] Spek M, Venekamp R, De Groot E, Geersing GJ, Erkelens DC, van Smeden M, et al. Optimising telephone triage of patients calling for acute shortness of breath during out-of-hours primary care: protocol of a multiple methods study (Opticall). BMJ Open 2022;12(4):e059549.
[39] Zhang D, Gao L, Jia Y, Wang S, Wang H, Sun X, et al. Construction of progress prediction model of urinary incontinence in elderly women: protocol for a multi-center, prospective cohort study. Int J Environ Res Public Health 2022;19(2):734.
[40] Zhang J, Olatosi B, Yang X, Weissman S, Li Z, Hu J, et al. Studying patterns and predictors of HIV viral suppression using A Big Data approach: a research protocol. BMC Infect Dis 2022;22(1):122.
[41] Dolezalova I, Koritakova E, Souckova L, Chrastina J, Chladek J, Stepanova R, et al. Prediction of vagal nerve stimulation efficacy in drug-resistant epilepsy (PRECISE): prospective study for pre-implantation prediction/study design. Front Neurol 2022;13:839163.
[42] Kontakiotis N, Rushton AB, Billis E, Papathanasiou G, Gioftsos G. Development of a clinical prediction model to inform clinical decision making for classification of patients with sciatica, based on their clinical characteristics, in the Greek health system: protocol for a prospective predictive exploratory study. BMJ Open 2022;12(1):e052119.
[43] Kuwabara A, Nakatani E, Tsugawa N, Nakajima H, Sasaki S, Kohno K, et al. Development of a predictive model for vitamin D deficiency based on the vitamin D status in young Japanese women: a study protocol. PLoS One 2022;17:e0264943.
[44] Sardesai S, Solomon M J, Arumugam A, Guddattu V, Gorthi SP, Pai A, et al. Predicting post-stroke motor recovery of upper extremity using clinical variables and performance assays: a prospective cohort study protocol. Physiother Res Int 2022;27(2):e1937.
[45] Riley RD, Ensor J, Snell KIE, Harrell FE, Martin GP, Reitsma JB, et al. Calculating the sample size required for developing a clinical prediction model. BMJ 2020;368:m441.
[46] Moons KGM, Altman DG, Reitsma JB, Ioannidis JPA, Macaskill P, Steyerberg EW, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. Ann Intern Med 2015;162:W1−73.
[47] Steyerberg EW, Bleeker SE, Moll HA, Grobbee DE, Moons KGM. Internal and external validation of predictive models: a simulation study of bias and precision in small samples. J Clin Epidemiol 2003;56:441−7.
[48] Debray TPA, Collins GS, Riley RD, Snell KIE, Van Calster B, Reitsma JB, et al. Transparent reporting of multivariable prediction models developed or validated using clustered data: TRIPOD-Cluster checklist. BMJ 2023;380:e071018.
[49] Hudda MT, Archer L, van Smeden M, Moons KGM, Collins GS, Steyerberg EW, et al. Minimal reporting improvement after peer review in reports of COVID-19 prediction models: systematic review. J Clin Epidemiol 2023;154:75−84.
[50] 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.