Journal of Clinical Epidemiology 159 (2023) 10−30

Journal of Clinical Epidemiology

REVIEW ARTICLE

Systematic review highlights high risk of bias of clinical prediction models for blood transfusion in patients undergoing elective surgery

Paula Dhiman[a,b,*], Jie Ma[a], Victoria N. Gibbs[c], Alexandros Rampotas[d], Hassan Kamal[a,e], Sahar S. Arshad[a], Shona Kirtley[a], Carolyn Doree[d], Michael F. Murphy[b,d,f], Gary S. Collins[a,b], Antony J.R. Palmer[c,f,g]

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

[b]*NIHR Oxford Biomedical Research Centre, Oxford University Hospitals NHS Foundation Trust, Oxford, UK*

[c]*Nuffield Department of Orthopaedics, Rheumatology, and Musculoskeletal Sciences, University of Oxford, Oxford, UK*

[d]*Systematic Review Initiative, NHS Blood & Transplant, John Radcliffe Hospital, Oxford, UK*

[e]*School of Medicine, University of Dundee, Ninewells Hospital & Medical School, Dundee, Scotland DD1 9SY*

[f]*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*

[g]*Oxford University Hospitals, Nuffield Orthopaedic Centre, Windmill Road, Headington, Oxford OX3 7HE, UK*

Accepted 1 May 2023; Published online 6 May 2023

Abstract

Background: Blood transfusion can be a lifesaving intervention after perioperative blood loss. Many prediction models have been developed to identify patients most likely to require blood transfusion during elective surgery, but it is unclear whether any are suitable for clinical practice.

Study Design and Setting: We conducted a systematic review, searching MEDLINE, Embase, PubMed, The Cochrane Library, Transfusion Evidence Library, Scopus, and Web of Science databases for studies reporting the development or validation of a blood transfusion prediction model in elective surgery patients between January 1, 2000 and June 30, 2021. We extracted study characteristics, discrimination performance (c-statistics) of final models, and data, which we used to perform risk of bias assessment using the Prediction model risk of bias assessment tool (PROBAST).

Results: We reviewed 66 studies (72 developed and 48 externally validated models). Pooled c-statistics of externally validated models ranged from 0.67 to 0.78. Most developed and validated models were at high risk of bias due to handling of predictors, validation methods, and too small sample sizes.

Hassan Kamal and Sahar S. Arshad were visiting students at the Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford.

Registration: CRD42019132342.

Funding: This study did not receive any funding. PD and GSC are supported by the NIHR Biomedical Research Centre, Oxford. GSC, SK, and JM are supported by Cancer Research UK (programme grant: C49297/A27294). AR is funded by a NIHR Academic Clinical Fellowship (grant number: ACF-2018-13-011). MFM, VNG and CD are employed by NHS Blood & Transplant and the Oxford University Hospitals NHS Foundation Trust. AJRP is employed by the Oxford University Hospitals NHS Foundation Trust and supported by the NIHR Blood and Transplant Research Unit in Data Driven Transfusion Practice (NIHR203334). 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.

Ethical approval and consent to participate: Not applicable.

Data availability: No additional data available.

Competing interests: 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.

Author Contributions: PD, MM, AJRP, and GSC conceived and designed the study. PD, CD, and SK developed the search strings. PD, JM, VNG, AR, HK, and SA extracted the data. PD analysed the data and wrote the first draft of the manuscript. All authors revised the manuscript and approved the final version of the submitted manuscript.

* Corresponding author. Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford OX3 7LD, UK. Tel.: +01865-225667; fax: +01865-227966.

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

https://doi.org/10.1016/j.jclinepi.2023.05.002

Conclusion: Most blood transfusion prediction models are at high risk of bias and suffer from poor reporting and methodological quality, which must be addressed before they can be safely used in clinical practice.

Keywords: Blood transfusion; Prediction model; Risk of bias; Systematic review; Meta-analysis; Surgery

1. Introduction

Blood transfusion can be a lifesaving intervention when there is perioperative blood loss during elective/planned surgery [1]. In the United Kingdom, approximately a third of the 1.5 million units of red blood cells (RBCs) issued each year are used in surgery [2−4]. Preoperative blood typing and antibody screening (procedures conducted under a 'group and save' or 'type and screen,' is where a patient's blood group and presence of atypical red cell antibodies in their blood is determined). Cross-matching (a laboratory process to ensure compatibility of donated blood products for a specific patient) ensures appropriate blood products are available when needed.

An inability to accurately predict which patients will require blood transfusion causes inefficiencies and other problems. Assuming patients need blood transfusions who ultimately do not results in unnecessary preoperative preparations, increasing costs, potentially decreasing patient satisfaction as they attend additional hospital visits and phlebotomy, and wasting blood products [5−7], which is of increasing concern given the recent UK blood shortage [8,9]. Conversely, incorrectly assuming patients will not require a blood transfusion may result in same-day surgery cancellations, that potentially lifesaving blood products are not available, or patient morbidity or mortality through untreated anaemia, delayed administration of blood products, or administration of incompatible blood products [10].

Increasing the efficiency of surgical pathways is crucial for tackling the unprecedented surgical waiting lists following the COVID-19 pandemic [11]. Accurately predicting which patients will require blood transfusion and should receive group and save would improve pathway efficiency and lower costs. Prediction modelling and prediction model research is widely used in health care and has improved care delivery in many clinical areas through development, validation, and implementation of multivariable models for prediction of individual diagnosis and prognosis [12−14]. It is an attractive strategy for selecting patients for preoperative group and save or cross-matching. Predicting risk of blood transfusion is closely aligned with predicting blood loss volume and may also facilitate the use of optimal blood management strategies [15] and reduce variation in transfusion practice [16−18].

The value of identifying patients at greatest risk of blood transfusion has driven the development of many prediction models in diverse clinical settings [19−22]. However, their quality and performance has not been evaluated, and none have been incorporated into clinical guidelines. We aimed to describe and critically appraise studies developing and validating blood transfusion prediction models used perioperatively in patients undergoing elective surgery. We also aimed to identify and list common blood transfusion predictors and summarize the predictive performance of these models.

2. Methods

We conducted a systematic review of risk prediction models that predict blood transfusion indication, amount, or timing in patients undergoing elective surgery and covering pre, intra, and postoperative periods (PROSPERO registration CRD42019132342). We focussed our review on patients undergoing elective surgery as there is more time to preoperatively assess these patients and plan for surgery compared with patients admitted for emergency procedures. Elective surgery provides an opportunity in clinical care order blood product in timely manner and to also reduce blood wastage by preventing unnecessary ordering for patients with low risk of needing it.

Blood transfusion was defined as transfusion of donor blood, not autologous transfusion. We refer to 'donor blood transfusion' as 'blood transfusion.' We report this study following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement and PRISMA extension for reporting literature searches (PRISMA-S) [23,24].

2.1. Eligibility criteria

Eligible studies were primary research studies reporting development or validation of blood transfusion prediction models in elective surgery patients as their primary aim. A blood transfusion model was defined as a multivariable (two or more predictors) risk prediction model for transfusion of red cell concentrates, fresh frozen plasma, platelets, or cryoprecipitate.

We included models intended for use in secondary and tertiary care and where the outcome was indication of blood transfusion, number of units transfused, or timing of transfusion. To ensure prediction model studies more applicable to and reflective of current clinical practice were included, we limited the search to studies published in the English language after January 01, 2000, including epub ahead of print, in-process, and other nonindexed citations. We

What is new?

Key findings

- Common predictors of red blood cell (RBC) transfusion were preoperative haemoglobin, age, and sex, which were used across most clinical specialties.
- Included models for blood transfusion showed moderate to good discriminatory performance (pooled estimates ranged from 0.67 to 0.78).
- However, most blood transfusion prediction models are at high risk of bias and suffer from poor reporting and methodological quality.

What this adds to what is known?

- High risk of bias for most studies was driven by the methodological issues found in the analysis.
- Common issues were inadequate samples sizes for development and validation, excluding missing data at the study entry and analysis level, using univariable predictor selection and categorizing continuous predictors, leading to increased risk of overfitting.
- Poor reporting of key information often inhibited the risk of bias assessment and meta-analysis of performance estimates, where the recommended minimum set of reporting items (TRIPOD) were not [fully] reported.

What is the implication and what should change now?

- The models currently available to predict blood transfusion in elective surgery should not be implemented into routine clinical care without further validation and research that adheres to current reporting and methodological recommendations for prediction modelling.

excluded studies that were unavailable in the English language or whose data comprised of more than 20% urgent or emergency surgeries.

2.2. Information sources and search strategy

On 16 April 2019, we searched MEDLINE (OVID), Embase (OVID), PubMed (via www.pubmed.ncbi.nlm.nih.gov), Cochrane Library (via www.cochranelibrary.com), Transfusion Evidence Library (Evidentia Publishing), Scopus (Elsevier), and Web of Science (via Clarivate Analytics) databases for studies developing and validating prediction models published from January 01, 2000 to April 16, 2019.

On June 30, 2021, we ran an update search on the seven databases. We used the original search strategies, except for Embase, where free-text terms were searched for in the keyword heading field (.kw.) rather than the keyword heading word field (.kf.), which was not a search field that was available when the update search was run, and limited the search to articles published January 01, 2000 to June 30, 2021.

References identified by the original and update searches were combined. We searched for and removed duplicate references, including those already screened in the original search, generating the list of additional references identified by the update search.

The search strategy combined blood transfusion search terms (e.g., 'blood transfusion', 'red blood cell', 'blood loss'), general elective surgery search terms (e.g., 'elective surgical procedures'), specific elective surgery names (e.g., 'hysterotomy', 'arthroplasty'), and prediction modelling search terms (e.g., 'prediction model', 'risk score'), searched as controlled vocabulary headings (e.g., MeSH or EMTREE) or free-text terms. We also searched for named transfusion prediction models: transfusion risk and clinical knowledge (TRACK) score, transfusion risk score (TRS), transfusion risk understanding scoring tool (TRUST), and PORT score for perioperative risk of blood transfusion in cardiac surgery by ACTA (ACTA-PORT).

No other limits were applied to the searches. The full original and update search strategies for the seven databases are reported in Supplementary material–Appendix A. Information specialists CD and SK developed and ran the original and updated search strategy for all databases, respectively.

2.3. Study selection

Studies identified by the search strategy were imported into EndNote citation manager [25] and deduplicated, then imported into Covidence [26] for title, abstract, and full-text screening. Duplicate articles were identified by comparing author names, journal information, and article titles.

For the original search, two independent reviewers separately screened the title and abstract of all studies using the defined eligibility criteria (VNG and AR). Three independent reviewers then screened the full text of eligible studies for inclusion. PD screened the full text of all articles and VNG and AR each screened half. For the updated search, two independent reviewers separately screened the title and abstract of all studies using the defined eligibility criteria (PD and JM). Two independent reviewers screened the full text of additional eligible studies for inclusion (PD and VNG). Screening disagreements for both searches were discussed and adjudicated by an additional reviewer (PD or GSC), where necessary.

2.4. Data collection and data items

Data were extracted separately for every blood transfusion prediction model developed or validated in each study. A standardized data extraction form was developed using the critical appraisal and data extraction for systematic reviews of prediction modelling studies (CHARMS), Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD), and a tool to assess the risk of bias and applicability of prediction model studies (PROBAST) tools [27−29]. CHARMS guided the extraction of performance measures, TRIPOD essential reporting items, and PROBAST risk of bias items. The data extraction form is provided supplementary material−Appendix B.

The data extraction form was piloted on three articles by PD, VNG, AR, and JM and amended as needed. It was piloted on two articles by SA and HK to standardize data extraction from studies included in the update. Extracted data were collected using REDCap [30] and included items, such as the data source, target population, predicted outcome, prediction time span, candidate predictors, sample size, missing data, study type, and underlying model (e.g., logistic regression, random forests).

Four independent researchers extracted the data from the original included studies (PD, VNG, AR, and JM). Three independent reviewers extracted the data from the updated included studies (PD, HK, and SA). Each study was extracted twice, with PD extracting all articles. Disagreements were discussed and adjudicated by an additional reviewer (GSC), if necessary.

Article authors were contacted for further information and clarification, if needed. Study authors were not blinded to the article authors and their institutions.

2.5. Summary measures

We extracted numerical model performance measures and confidence intervals or standard errors for the developed or validated prediction models. Calibration measures agree between observed and predicted outcomes. We extracted how authors assessed calibration, including calibration plots with calibration-in-the-large or calibration slope measures, observed:expected event ratios, and Hosmer-Lemeshow tests. Discrimination measures how well a prediction model separates individuals with and without the outcome. We extracted the c-statistic (or c-index) and D-statistic discrimination measures. Data were also extracted on the reporting of additional performance measures (e.g., classification measures, decision curves), but numerical values were not extracted.

2.6. Data transformation and results synthesis

Data were summarized using descriptive statistics, visual plots, and narrative synthesis. Predictors used in the final models developed for each study were ranked and their frequency reported. Apparent and optimism-corrected performance measures before and after internal validation, if reported, were summarized. We performed a random-effects meta-analysis on all models with at least two external validations and pooled their discrimination performance using the reported c-statistic values. If 95% confidence intervals and standard error for the c-statistic point estimate were not reported, we calculated them using the reported number of events, number of participants, and formulae described by Debray et al. [31].

For the meta-analysis, we transformed the c-statistic and its 95% confidence interval to the logit scale, then pooled the estimates for each externally validated model [31]. We back-transformed the pooled c-statistic estimates with their 95% confidence intervals before plotting the results in a forest plot.

2.7. Risk of bias

Risk of bias was assessed using PROBAST for each prediction model development and validation analysis using study-level information [29,32]. Developed and validated prediction models were ranked as at low, high, or unclear risk of bias in the participant, predictor, outcome, and analysis domains with 20 signalling questions. Supplementary material−Appendix C provides more information on how risk of bias was assessed with PROBAST.

All analyses were carried out in Stata v15 and R statistical software [33,34]. We used the 'forestplot', 'tidyr', and 'dplyr' packages in R [35−37].

3. Results

The original search retrieved 6,645 published studies, and the update search retrieved 9,270 published studies. Forty-nine studies were included from the original search and another 17 studies from the update search. Studies were primarily excluded from either search due to publication type ($n = 50$, e.g., commentary, review) and study design ($n = 18$, e.g., epidemiology study). Seventeen studies were excluded for including more than 20% nonelective surgery in their patient samples. Figure 1 shows a PRISMA flow diagram of the original and updated searches.

We included 66 studies containing 120 model development and validation analyses: 37 development-only, 25 development-with-external validation, and four external-validation-only. Seventy-two models were developed from 62 development or development-with-external validation studies. Forty-eight models were validated from 29 validation or development with external validation studies, of which 26 were validations of existing models and 22 were validations of models created in that study. Supplementary material−Figure 1 shows the number of published studies by year of publication.

Studies developing and validating prediction models for blood transfusion were predominantly in cardiothoracic

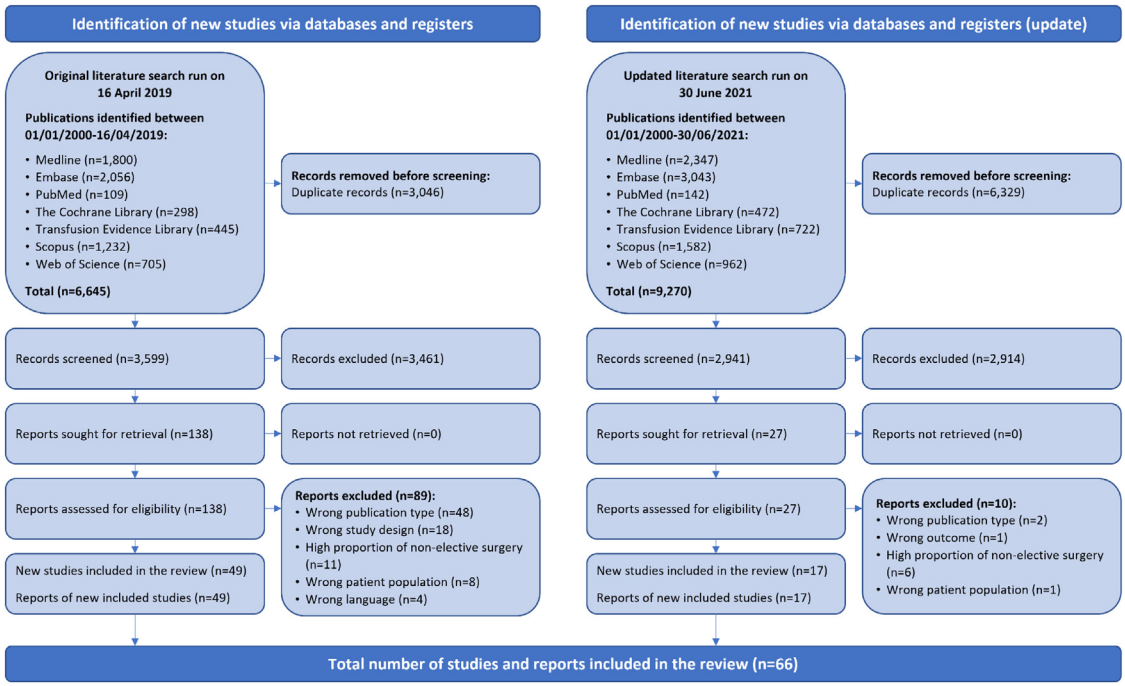

Fig. 1. PRISMA flow diagrams of the original search (left, run on April 16, 2019) and updated search (right, run on June 30, 2021).

(n = 19/66 studies, 29%) and general (n = 17/66 studies, 26%) surgery (Table 1). Models were most often developed and validated for binary blood transfusion outcomes (n = 113/120 analyses, 94.2%) and transfusion of RBCs (n = 110/120 analyses, 92%). Over half of the developed and validated models predicted the outcome during the 'intraoperative and postoperative period' (n = 66/120 analyses, 55%). Where the timing of the outcome included the postoperative period (n = 87/120 analyses, 73%), the prediction horizon was commonly reported as 'end of hospital stay' (n = 13/87 analyses, 15%) or '24 hours postsurgery' (n = 12/87 analyses, 14%). The prediction horizon was not reported for 44% of models (n = 38/87 analyses). Supplementary material–Table 1 provides details about the target population, outcomes, and timing for each development and validation analysis.

Seventy-six percent of developed models (n = 55/72 development analyses) and 96% of validated models (n = 46/48 validation analyses) used logistic regression. Thirteen studies included nonelective surgeries in their development data and four studies in their validation data. A median of 8% (IQR: 3.7 to 10.9%, range: 0.6 to 20%, n = 13 studies) nonelective surgeries were used for model development and 13% (IQR: 8 to 20%, range: 8 to 20%, n = 3 studies) for model validation.

3.1. Predictors for blood transfusion outcomes

A median of 20 candidate predictors were considered in each study developing a model (n = 62, IQR: 15 to 29, range: 3 to 79). Less than a third of studies provided rationale for their candidate predictors (n = 19/62, 31%). The number of candidate predictors was unclear in two studies [38,39].

A median of six predictors were included in the final model (IQR: 4 to 8, range: 2 to 17, n = 68 development analyses). The predictors included in the final model were unclear for four developed models. We found 135 unique predictors across all final models developed in the 68 development analyses that reported their final model predictors. The most frequently included predictors in the final models were haemoglobin (n = 43/68 development analyses, 63%), age (n = 38/68, 56%), and procedure detail (e.g., surgery route, surgery type, and transplantation procedure) (n = 32/68, 47%). Supplementary material–Figure 2 lists the top 10 predictors in the final models across all the clinical procedures. Supplementary material–Table 2 summarises the model type and final predictors for all developed models.

Twenty-four percent of studies (n = 15/62 development studies) examined interactions between predictors. Two studies included these interactions in their final model: between age and sex for orthopaedic surgery [40] and between haematocrit and diabetes and age and left ventricular ejection fraction for cardiothoracic surgery [41].

3.2. Reporting and methodology

Only three studies explained their sample size [42–44]. Model development (n = 72 models) used a median of 884

Table 1. Study characteristics of the 66 included studies and their 120 development ($n = 72$) and validation ($n = 48$) analyses

All study characteristics $n = 66$ development and validation studies, 120 analyses	***n***	**%**
Surgery type		
Cardiothoracic surgery	19	28.8
General surgery	17	25.8
Obstetrics and gynaecology surgery	3	4.6
Orthopaedic surgery	7	10.6
Orthopaedic surgery (tumour)	1	1.5
Plastic surgery	4	6.1
Spinal surgery	5	7.6
Transplant surgery	9	13.6
Vascular surgery	1	1.5
Continent		
Australia	2	3.0
Europe	14	21.2
North America	27	40.9
Asia	15	22.7
Unclear	8	12.1
Setting		
Secondary	13	19.7
Secondary/tertiary	1	1.5
Tertiary	48	72.7
Unclear	4	6.1
Centres		
Single centre	41	22.6
Multicentre	17	64.5
Not reported	8	12.9
Study design		
Development-only	37	56.1
Development-with-external validation	25	37.9
External validation-only	4	6.1
Any blood product[a]		
RBCs	110	91.7
Platelets	1	0.8
Fresh frozen plasma	1	0.8
Any (RBC/Platelet/Fresh frozen plasma/ Cryoprecipitate)	8	4.2
Outcome[a]		
Binary (transfusion, yes/no)	113	94.2
Continuous (amount transfused)	4	3.3
Count (amount transfused)	2	1.7
Ordinal	1	0.8
Timing of outcome[a]		
Intraoperative	27	22.5
Intraoperative and postoperative	66	55
Postoperative	18	15
Preoperative/intraoperative/ postoperative	3	2.5
Not reported	6	5
Prediction horizon[b]		
24 hours	12	13.7
48 hours	1	1.1

(*Continued*)

Table 1. Continued

All study characteristics $n = 66$ development and validation studies, 120 analyses	***n***	**%**
72 hours	2	2.2
5 days	2	2.2
7 days	7	8
14 days	4	4.6
30 days	7	8
End of hospital stay/before discharge	13	14.9
End of time spent in intensive care	1	1.1
Not specified	38	43.7
Development characteristics $n = 62$ development studies, 72 analyses	***n***	**%**
Data source		
Case-control/case-cohort study (using NSQIP registry)	1	1.6
Prospective cohort (hospital database)	15	24.2
Registry	5	8.1
Retrospective cohort (database)	39	62.9
Benchmark studies	2	3.2
Model type		
Classification and regression tree (CART)	2	2.8
Logistic regression	55	76.4
Random forest	4	5.6
Linear regression	2	2.8
CatBoost classifier	1	1.49
Generalized regression (exponential distribution and ridge estimation)	1	1.49
Gradient-boosting model	1	1.49
Light gradient boosting machine	1	1.49
Mixed logistic regression	1	1.49
Naïve Bayes classifier	1	1.49
Ordinal logistic regression	1	1.49
Truncated negative binomial	1	1.49
XGBOOST	1	1.49
Validation characteristics $n = 29$ validation studies, 48 analyses	***n***	**%**
Data source		
Prospective cohort (hospital database)	10	34.5
Registry	4	13.8
Retrospective cohort (hospital database)	10	34.5
Not reported	5	17.2
Model type		
Gradient-boosting model	1	2.1
Logistic regression	46	95.8
XGBOOST	1	2.1

Abbreviations: RBC, red blood cells; NSQIP, ACS National Surgical Quality Improvement Program.

[a] out of 120 development and external validation analyses.

[b] out of 87 (where timing of outcome includes postoperative timing).

patients (IQR: 530 to 3,034, range: 150 to 10,477, $n = 71$ models) and 227 events (IQR: 106 to 449, range: 25 to 8,635, $n = 57$ models). External validation ($n = 48$ models) used a median of 946 patients (IQR: 323 to 2,371, range: 52 to 8,982, $n = 48$ models) and 204 events (IQR: 97 to 539, range: 7 to 1,439, $n = 42$ models). Missing data were mentioned in 59% of studies ($n = 39/66$ studies). Most studies performed complete-case analysis ($n = 24/39$, 62%). Only four studies used multiple imputations.

Although all development studies used continuous predictors, their handling of these predictors was inadequate with 65% of studies categorizing their continuous data ($n = 40/62$ development studies). Of these studies, most provided no rationale for deciding their cut-points ($n = 29/40$, 73%) with only three using clinically informed cut-points ($n = 3/40$, 8%). Only one-fifth of studies examined nonlinear terms ($n = 13/62$ development studies, 21%). Three studies included nonlinear predictors in their final model, presented as log (base 10) transformation for platelets in general hepatobiliary surgery [20], categories informed from restricted cubic splines for body mass index and parity in obstetrics and gynaecology surgery [45], and from piecewise linear functions for platelets in transplant surgery [44].

Nearly 30% of development studies unnecessarily reduced their available sample size by using the split-sample approach for internal validation ($n = 18/62$ development studies, 29%). Fifteen of these studies used a random split with a cut-point of 70% to allocate for model development. A quarter of models were validated using more appropriate bootstrapping procedures ($n = 16/62$, 26%) and 10% using cross-validation ($n = 6/62$). Ten studies did not internally validate their models, of which nine used development data to assess their models. Internal validation was unclear in 12 studies ($n = 12/62$, 19%).

Of the 48 validation analyses, 21 (44%) were fully independent validations where the data and study author team differed from the original model development team. Twenty models (42%) were geographically and temporally validated. Six models (13%) were validated using independent data but included at least one member of the original study author team. Table 2 shows full internal and external validation results.

Although 89% of studies ($n = 59/66$) reported discrimination, only 56% ($n = 37/66$) reported calibration. Calibration was most often assessed using the Hosmer–Lemeshow test ($n = 26/37$, 70%), including one study that reported the calibration slope and intercept alongside it. Only 10 studies presented a calibration plot ($n = 10/37$, 27%). Thirty-nine studies reported other performance measures ($n = 39/66$, 59%), including classification measures such as sensitivity, specificity, and accuracy.

Supplementary material–Table 3 provides more information on the reporting and methodological characteristics of the included studies.

3.3. Validation of existing models–meta-analysis

Eight existing models predicting blood transfusion were externally validated more than once. These included four unnamed models, developed for hepatic resection by Pulitano et al. [62] (pooled c-statistic 0.73 [0.71 to 0.75]), hepatectomy by Sima et al. [20] (0.68 [0.66 to 0.7]), and liver resection by Cockbain et al. [59] (0.7 (0.67 to 0.72]) and Quan et al. [66] (0.78 [0.75 to 0.82]). Named models included transfusion risk and clinical knowledge (TRACK) for cardiac surgery [22] (c-statistic 0.74 [0.73 to 0.75]), the three-point transfusion risk score for hepatectomy [70] (0.67 [0.65 to 0.7]), predictive model of transfusion in spine surgery (PMTSS) [91] (0.73 [0.65 to 0.79]), and transfusion risk understanding scoring tool (TRUST) for cardiac surgery [21] (0.72 [0.7 to 0.73]). The development paper for TRUST was not included in this review as its development data included a high proportion (>40%) of nonelective surgeries. Figure 2 shows all pooled c-statistic results.

3.4. Risk groups and model presentation

Less than half of the studies reported their models in enough detail to allow for new risk calculation ($n = 29/62$ development studies, 47%). Most of the studies reported coefficients from regression-based models and cut-points in classification models (e.g., random forest, CART) ($n = 62/72$ development analyses, 86%). Although the intercept is needed to calculate predictions from regression-based models, it was often not reported ($n = 32/61$ developed regression-based models, 52%). Fifteen studies created risk categories to stratify patients ($n = 15/62$ development studies, 24%).

Two studies produced screening questionnaires based their models [46,77], and over half simplified their models for clinical use ($n = 33/62$, 53%). Reasons for simplifying prediction models included ease of use ($n = 17/33$, 52%), clinical utility ($n = 2/33$, 6%), and to aid doctor-patient communication and facilitation ($n = 2/33$, 6%). Models were also presented as nomograms ($n = 9/33$, 27%), creating an integer-based points system, counting present predictors, a sum score, or simplified risk score ($n = 24/33$, 73%). Six studies included a link or reference to a web calculator [42,45,66,82,95,97]. One study provided their model's code [83].

3.5. Risk of bias

The development and validation of most models was considered at high risk of bias ($n = 58/72$ development analyses, 81%; $n = 29/48$ validation analyses, 60%) (Fig. 3). The development of only four models was at low risk of bias: models predicting blood transfusion risk in patients undergoing total knee arthroplasty [81], elective hepatectomy [20,70], and cardiac surgery [46].

Validation of only five models from three studies was found to be at low risk of bias [20,66,70]: four models

Table 2. Characteristics and discrimination performance estimates from internal and external validation for studies and analyses (n = 120 analyses)

Clinical specialty/ surgery type	Study	Study design	Analysis	Surgical procedure	Outcome (product)	Timing	Model	Validation type	Events fraction (%)	Reported c-statistic
Cardiothoracic surgery	Al-Khabori 2014 [46]	D	d	Cardiac surgery	Binary: transfusion (RBCs)	Intra/post	Logistic regression	Internal: Bootstrapping	264/413 (63.9)	0.749 (0.701-0.797)
	Cevenini 2013 [47]	D	d	Cardiac surgery	Ordinal: transfusion (RBCs)	Post	Naïve Bayes Classifier	Internal: Cross validation	1,107/3,182 (34.8)	Not reported
	Cirasino 2000 [48]	D	d	Lung cancer	Binary: transfusion (RBCs)	Intra/post	Logistic regression	Internal: Development dataset	25/173 (14.5)	Not reported
	Covin 2003 [49]	D	d d d	CABG	Binary: transfusion (Platelets) Binary: transfusion (FFP) Binary: transfusion (RBCs)	Intra	Logistic regression	Internal: Bootstrapping	313/3,034 (10.3) 101/3,034 (3.3) 362/3,034 (11.9)	0.694 0.724 0.752
	Hayn 2016 [38]	D	d	Hip and knee and CABG	Continuous: transfusion volume (RBCs)	Not reported	Random forest	Internal: Cross validation	na/6,530	Not reported
	Hayn 2017 [38]	D	d	Hip and knee and CABG	Continuous: transfusion volume (RBCs)	Intra/post	Random forest	Internal: Cross validation	na/6,530	Not reported
	Doshi 2021 [39]	D	d	Cardiac surgery	Binary: transfusion (RBCs)	Intra/post	Logistic regression	Internal: Split sample	170/345 (49.3)	0.715
	Lu 2020 [50]	D	d d d	CABG	Binary: transfusion (RBCs)	Intra Intra/post Post	Logistic regression	Internal: Development dataset	87/1,253 (6.9) 364/1,253 (29.1) 334/1,253 (26.7)	0.869 (0.849-0.887) 0.830 (0.808-0.851) 0.821 (0.799-0.842)
	Paiva 2021 [51]	D	d	CABG	Binary: transfusion (RBCs)	Intra	Logistic regression	Internal: Bootstrapping	91/530 (17.2)	0.962 (0.945-0.980)
	Liu 2021 [52]	D	d	Cardiac surgery	Binary: transfusion (RBCs)	Intra	CatBoost classifier	Internal: Split sample	missing/473	0.888 (0.845-0.909)

(*Continued*)

Table 2. Continued

Clinical specialty/ surgery type	Study	Study design	Analysis	Surgical procedure	Outcome (product)	Timing	Model	Validation type	Events fraction (%)	Reported c-statistic
	Simeone 2011 [53]	D	d	Heart	Count: transfusion units (RBCs)	Post	Linear regression	Internal: Split sample	567/1,204 (47.1)	Not reported
	Welsby 2010 [41]	D	d	CABG	Binary: transfusion (RBCs)	Intra/post	Ordinal logistic regression	Internal: Split sample	missing/3,876	0.78
	Arora 2004 [54]	D + EV	d ev	CABG	Binary: transfusion (Any blood product)	Intra/post	Logistic regression	Internal: Bootstrapping External: Temporal	704/3,046 (23.1) 499/2,117 (23.6)	0.84 (0.82-0.86) 0.79 (0.77-0.81)
	Karkouti 2006 [55]	D + EV	d v	Cardiac surgery	Binary: transfusion (RBCs)	Post	Logistic regression	Internal: Development dataset External: Temporal	476/6,651 (7.2) 499/4,016 (11)	0.88 Not reported
	Karkouti 2001 [56]	D + EV	d ev	CABG	Binary: transfusion (RBCs)	Intra/post	Logistic regression	Internal: Unclear External: Temporal	190/717 (26.5) 106/290 (36.6)	0.86 Not reported
	Klein 2017 [19]	D + EV	d ev ev	Cardiac surgery	Binary: transfusion (RBCs)	Intra/post	Logistic regression	Internal: Development dataset External: Geographical External: Fully independent	8,635/20,036 (43.1) 97/1,047 (9.3) 97/1,047 (9.3)	0.762 0.835 (0.810-0.859) 0.781
	Madhu Krishna 2019 [57]	D + EV	D ev ev	Cardiac surgery	Binary: transfusion (RBCs)	Intra/post	Logistic regression	Internal: unclear External: Fully independent	773/1,014 (76.2) 773/1,014 (76.2) 773/1,014 (76.2)	0.749 (0.722-0.776) 0.756 (0.729-0.782) 0.72 (0.692-0.748)
	Ranucci 2009 [22]	D + EV	d ev ev ev ev	Cardiac with cardiopulmonary	Binary: transfusion (RBCs)	Intra/post	Logistic regression	Internal: Development dataset External: Geographical External: Fully independent	4,978/8,989 (55.4) 1,399/2371 (59) 1,399/2371 (59) 1,399/2371 (59) 1,399/2371	0.73 (0.718-0.743) 0.71 (0.681-0.724) 0.713 (0.692-0.734) 0.686 (0.665-0.708) 0.679 (0.657-0.701)

(*Continued*)

Table 2. Continued

Clinical specialty/ surgery type	Study	Study design	Analysis	Surgical procedure	Outcome (product)	Timing	Model	Validation type	Events fraction (%)	Reported c-statistic
									(59)	
	Leff 2019 [58]	EV	ev ev ev ev ev	Cardiac surgery	Binary: transfusion (Any blood product) Binary: transfusion (RBCs)	Intra/post Intra Post Intra Post	Logistic regression	External: Fully independent	1,439/2,776 (51.8) 829/2,776 (29.9) 1,161/2,776 (41.8) missing/2776 missing/2776	0.768 (0.750-0.785) 0.775 (0.775-0.794) 0.713 (0.694-0.732) 0.817 (0.800-0.835) 0.722 (0.703-0.741)
General surgery	Cockbain 2010 [59]	D	d	Liver resection	Binary: transfusion (RBCs)	Intra/post	Logistic regression	Internal: Unclear	100/589 (17)	0.777
	Huang 2021 [60]	D	d	Gastrectomy	Binary: transfusion (RBCs)	Intra	Logistic regression	Internal: Split sample	163/666 (24.5%)	0.859
	Kim 2016 [61]	D	d	Hepatopancreatic-biliary and colorectal	Binary: transfusion (RBCs)	Intra/post	Logistic regression	Internal: Bootstrapping	1,549/4,961 (31.2)	0.75
	Pulitano 2007 [62]	D	d	Hepatectomy	Binary: transfusion (RBCs)	Intra/post	Logistic regression	Internal: Split sample	108/320 (33.8)	0.89, se 0.02
	Wang 2015 [63]	D	d	Liver resection	Binary: transfusion (RBCs)	Intra/post	Logistic regression	Internal: Split sample	missing/1,080	0.709
	Feng 2021 [64]	D	d	Nonspecific	Binary: transfusion (RBCs)	Intra/post	Light gradient boosting machine	Internal: Split sample	missing/104777	0.908 (0.907-0.913)
	Yamamoto 2014 [65]	D	d	Hepatectomy	Binary: transfusion (RBCs)	Intra	Logistic regression	Internal: Unclear	38/168 (22.6)	0.76
	Quan 2020 [66]	D + EV	d ev ev ev ev ev ev	Liver resection	Binary: transfusion (RBCs)	Intra/post	Logistic regression	Internal: Development dataset External: Fully independent External: Geographical	204/878 (23.2) 204/878 (23.2) 204/878 (23.2) 204/878 (23.2) 58/306 (15.9) 148/691 (21.4)	0.833 (0.801-0.864) 0.66 0.71 0.77 0.89 0.777 (0.712-0.841) 0.786 (0.743-0.829)
	Bagante 2016 [67]	D + EV	d ev	Hepatectomy	Binary: transfusion (RBCs)	Intra/post	Logistic regression	Internal: Unclear External: Independent	449/1,345 (33.4) 1,439/1,345	0.73 0.69

(*Continued*)

Table 2. Continued

Clinical specialty/ surgery type	Study	Study design	Analysis	Surgical procedure	Outcome (product)	Timing	Model	Validation type	Events fraction (%)	Reported c-statistic
								data	(107)	
	Bansal 2018 [68]	D + EV	d d ev	Pancreaticoduodenectomy	Binary: transfusion (RBCs)	Post	Logistic regression	Internal: Unclear Internal: Unclear External: Fully independent	missing/555 missing/620 missing/555	0.735 (0.678-0.792) 0.651 (0.595-0.706) 0.700 (0.640-0.760)
	Cucchetti 2014 [69]	D + EV	d ev ev ev	Hepatectomy	Binary: transfusion (RBCs)	Intra/post	Logistic regression	Internal: Cross validation External: Fully independent	106/323 (32.8) 106/323 (32.8) 106/323 (32.8) 106/323 (32.8)	0.79 (0.74-0.83) 0.64 (0.58-0.71) 0.69 (0.63-0.75) 0.67 (0.61-0.74)
	Lemke 2017 [70]	D + EV	d ev ev ev	Hepatectomy	Binary: transfusion (RBCs)	Intra/post	Logistic regression	Internal: Unclear External: Fully independent	341/1,287 (26.5) 341/1,287 (26.5) 341/1,287 (26.5) 341/1,287 (26.5)	0.66 (0.63-0.69) 0.66 (0.63-0.69) 0.66 (0.63-0.70) 0.68 (0.64-0.71)
	Sima 2009 [20]	D + EV	d ev	Liver resection	Binary: transfusion (RBCs)	Intra/post	Logistic regression	Internal: Not done External: Temporal	567/1,204 (47.1) 234/555 (42.2)	- 0.71
	van Klei 2002 [71]	D + EV	d ev	Surgery under general or regional anaesthesia	Binary: transfusion (RBCs)	Intra/post	Logistic regression	Internal: Split sample External: Geographical	651/9,033 (7.2) 539/8,982 (6)	0.89 (0.87-0.90) 0.76 (0.73-0.78)
	Guo 2021 [72]	D + EV	d ev	Pheochromocytoma	Binary: transfusion (RBCs)	Intra	Logistic regression	Internal: Split sample External: Geographical	61/189 (32.3) 7/56 (12.5)	0.831 (0.750-0.822) 0.924 (0.766-1,000)
	Janny 2015 [73]	EV	ev	Liver resection	Binary: transfusion (RBCs)	Not reported	Logistic regression	External: Fully independent	48/205 (23.4)	0.68 (0.59-0.77)
	Lemke 2018 [74]	EV	ev	Hepatectomy	Binary: transfusion (RBCs)	Intra/post	Logistic regression	External: Independent data	525/2,854 (18.4)	0.68 (0.66-0.70)
Obstetrics and gynaecology	Klebanoff 2021 [75]	D	d	Hysterectomy/Myomectomy	Binary: transfusion (RBCs)	Intra/post	Logistic regression	Internal: Bootstrapping	623/6,387 (9.8)	0.792 (0.790-0.794)
	Ackroyd 2018 [76]	D + EV	d ev	Hysterectomy/Myomectomy	Binary: transfusion	Intra/post	Logistic regression	Internal: Development	missing/2004 missing/1,466	0.8 (0.78-0.83) 0.69 (0.66-0.72)

(*Continued*)

Table 2. Continued

Clinical specialty/ surgery type	Study	Study design	Analysis	Surgical procedure	Outcome (product)	Timing	Model	Validation type	Events fraction (%)	Reported c-statistic
					(RBCs)			dataset External: Temporal		
	Stanhiser 2017 [45]	D + EV	d ev	Gynaecological	Binary: transfusion (Any blood product)	Intra/post	Logistic regression	Internal: Bootstrapping External: Temporal	239/12,219 (2) missing/6,100	0.906 (0.890-0.928) 0.915 (0.872-0.954)
Orthopaedic surgery	Ahmed 2012 [77]	D	d	Hip and knee	Binary: transfusion (RBCs)	Post	Logistic regression	Internal: Unclear	227/2281 (10)	0.74 (0.70-0.775)
	Huang 2018 [78]	D D	d d	Hip and knee Hip and knee	Binary: transfusion (RBCs) Binary: transfusion (RBCs)	Intra/post Intra/post	Logistic regression Random forest	Internal: Cross validation Internal: Cross validation	2,867/15,187 (18.9) 2,867/15,187 (18.9)	0.84 (0.81-0.87) 0.77(0.74-0.79)
	Noticewala 2012 [79]	D	d	Hip and knee	Binary: transfusion (RBCs)	Post	Logistic regression	Internal: Split sample	71/644 (11)	Not reported
	Rashiq 2004 [80]	D	d	Hip and knee	Binary: transfusion (RBCs)	Intra/post	Logistic regression	Internal: Split sample	239/884 (27)	0.76
	Hu 2020 [81]	D + EV	d ev	Hip and knee	Binary: transfusion (RBCs)	Post	Logistic regression	Internal: Development dataset External: Temporal	391/5,402 (7.2) 148/1,116 (13.3)	0.884 (0.865-0.903) 0.839 (0.773-0.905)
	Jo 2020 [82]	D + EV	d ev	Hip and knee	Binary: transfusion (RBCs)	Post	Gradient boosting model	Internal: Cross validation External: Independent data	108/1,686 (6.4) 7/400 (1.8)	0.842 (0.820-0.856) 0.880 (0.844-0.910)
	To 2017 [40]	D + EV	d ev	Hip and knee	Binary: transfusion (RBCs)	Pre/intra/post	Logistic regression	Internal: Unclear External: Independent data	184/737 (25) 25/653 (3.8)	0.80 (0.77 to 0.84) 0.84 (0.79 to 0.88)
Orthopaedic surgery (tumour)	Thompson 2014 [83]	D	d d	Musculoskeletal tumour	Count: transfusion units (RBCs) Binary:	Not reported	Truncated negative binomial Logistic regression	Internal: Unclear Internal: Unclear	na/1,319 417/1,319 (31.6)	Not reported Not reported

(*Continued*)

Table 2. Continued

Clinical specialty/ surgery type	Study	Study design	Analysis	Surgical procedure	Outcome (product)	Timing	Model	Validation type	Events fraction (%)	Reported c-statistic
					transfusion (RBCs)					
Plastic Surgery	Su 2021 [84]	D	d	Head and neck with free flap	Binary: transfusion (RBCs)	Pre/intra/post	Logistic regression	Internal: Development dataset	101/385 (26.2)	0.826 (0.781-0.871)
	Kolbenschlag 2016 [85]	D	d	Free tissue transfer	Binary: transfusion (RBCs)	Intra	Logistic regression	Internal: Unclear	166/398 (41.7)	0.86
	Shah 2010 [86]	D	d	Head and neck with free flap	Binary: transfusion (RBCs)	Intra/post	Logistic regression	Internal: Bootstrapping	144/585 (24.6)	0.754
	Krupp 2003 [87]	EV	ev	Head and neck cancer	Binary: transfusion (RBCs)	Intra/post	Logistic regression	External: Fully independent	122/539 (22.6)	0.72 (0.62-0.81)
Spinal surgery	Durand 2018 [88]	D	d	Spinal deformity	Binary: transfusion (RBCs)	Intra/post	Random forest	Internal: Split sample	missing/824	0.85 (0.80-0.90)
	Wang 2021 [89]	D	d	Spine fusion	Binary: transfusion (RBCs)	Post	Logistic regression	Internal: Bootstrapping	289/885 (32.7)	0.895
	Pennington 2021 [42]	D	d	Spinal tumour surgery	Binary: transfusion (RBCs)	Intra	Logistic regression	Internal: Bootstrapping	188/350 (53.7)	0.819
	Carabini 2014 [90]	D + EV	d ev ev	Spine fusion	Binary: transfusion (RBCs)	Intra	Logistic regression	Internal: Bootstrapping External: Temporal External: Fully independent	326/548 (59.5) 61/95 (64.2) 61/95 (64.2)	0.88 0.89 (0.80-0.90) 0.61 (0.48-0.71)
	Lenoir 2009 [91]	D + EV	d ev	Thoracolumbar spine	Binary: transfusion (RBCs)	Intra/post	Logistic regression	Internal: Unclear External: Unclear	74/230 (32.2) 46/125 (36.8)	0.86 (0.81-0.92) 0.83 (0.75-0.91)
Transplant surgery	Cywinski 2014 [92]	D	d d d d	Liver transplant	Continuous: number of units transfused (RBCs) Binary: transfusion (RBCs)	Intra	Linear regression CART Logistic regression CART	Internal: Split sample	na/804 156/804 (19.4) 75/804 (9.3) 75/804 (9.3)	Not reported 0.67 (0.60-0.73)

(*Continued*)

Table 2. Continued

Clinical specialty/ surgery type	Study	Study design	Analysis	Surgical procedure	Outcome (product)	Timing	Model	Validation type	Events fraction (%)	Reported c-statistic
	Liu 2015 [43]	D	d	Liver transplant	Binary: transfusion (Any blood product)	Intra/post	Logistic regression	Internal: Split sample	422/482 (87.6)	0.871
	Massicotte 2018 [93]	D	d	Liver transplant	Binary: transfusion (RBCs)	Intra	Mixed logistic regression	Internal: Bootstrapping	61/800 (7.6)	Not reported
	McCluskey 2006 [94]	D	d	Liver transplant	Binary: transfusion (RBCs)	Post	Logistic regression	Internal: Bootstrapping	193/460 (42)	0.82
	Metcalf 2018 [95]	D	d	Liver transplant	Continuous: transfusion volume (RBCs)	Intra	Generalized regression	Internal: Split sample	93 (44.9, ≤ 1250 ml), 59 (29.5, 1251-2000 ml), 55 (26.6, >2000 ml)/207	Not reported
	Pustavoitau 2020 [96]	D + EV	d ev	Liver transplant	Binary: transfusion (RBCs)	Intra	Logistic regression	Internal: Bootstrapping External: Independent data	60/203 (29.6) 111/403 (27.5)	0.77 (0.70-0.84) 0.69 (0.62-0.76)
	Liu 2021 [97]	D + EV	d ev	Liver transplant	Binary: transfusion (RBCs)	Intra/post	XGBOOST	Internal: Split Sample External: Independent data	602/835 (72.1) 31/52 (59.6)	0.813 Not reported
	Massicotte 2009 [98]	D + EV	d ev	Liver transplant	Binary: transfusion (RBCs)	Not reported	Logistic regression	Internal: Bootstrapping External: Temporal	179/406 (44.1) 22/109 (20.2%)	0.898 0.898
	Pustavoitau 2017 [44]	D + EV	d ev	Liver transplant	Binary: transfusion (RBCs)	Intra	Logistic regression	Internal: Bootstrapping External: Temporal	missing/150 missing/53	0.835 (0.781-0.888) 0.895 (0.809-0.982)
Vascular surgery	Stangenberg 2016 [99]	D	d	Carotid endarterectomy	Binary: transfusion (RBCs)	Intra/post	Logistic regression	Internal: Split sample	missing/8,039	0.81

Abbreviations: CABG, coronary artery bypass graft; pre, preoperative; intra, intraoperative; post, postoperative; D, development; D + EV, development with external validation; EV, external validation; na, not applicable.

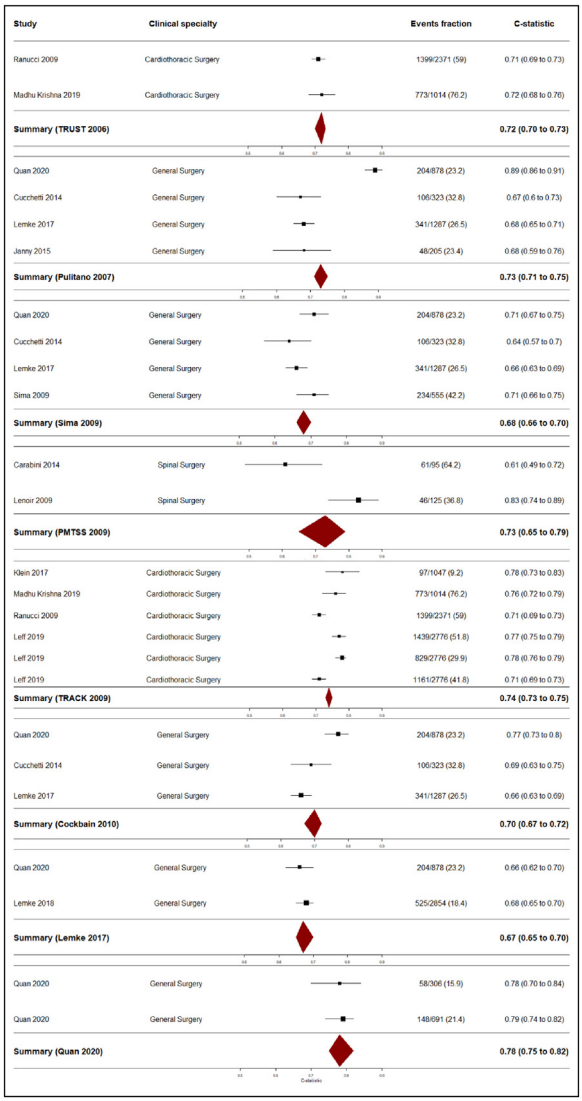

Fig. 2. Forest plot of the eight existing models that were externally validated more than once in the included studies. Two external validations from Leff et al. [58] were excluded as the number of events (RBC transfusion) and c-statistic confidence intervals were not rsported.

predicting blood transfusion risk for patients undergoing elective hepatectomy [20,70] and one model for liver resection [66]. Supplementary material—Figure 3 and Supplementary material—Table 4 present risk of bias assessments for each model development and validation analysis.

4. Discussion

4.1. Summary of findings

We reviewed 66 studies and 120 prediction models developed or validated for predicting blood transfusion in elective surgery (including predictions for pre, intra, and postoperative periods). Common predictors of RBC transfusion used across most clinical specialties were preoperative haemoglobin, age, and sex. We meta-analysed eight externally validated clinical prediction models in general, cardiothoracic, and spinal surgeries. The models showed moderate to good discriminatory performance, and pooled estimates ranged from 0.67 to 0.78. However, half of the external validations informing these meta-analyses were at high risk of bias and another third had unclear risk of bias. These models' performance should be interpreted with caution.

Almost all developed and validated blood transfusion models for elective surgery were at high risk of bias and demonstrated poor reporting and methodological quality. Only a total of four models, one developed for cardiac surgery by Al-Khabori et al. [46], two models developed for general surgery by Lemke et al. [70] and Sima et al. [20], and one model developed for orthopaedic surgery by Hu et al. [81], were at low risk of bias and would warrant further validation. The model developed by Sima et al. was also externally validated in its development study by the same study team and was independently validated by Lemke et al. [70], both of which were judged to be low risk bias validations and indicate consideration for further validation and possible implementation analyses for blood transfusion prediction in elective general surgery. Lemke et al. [70] also externally validated other models for general surgery that were judged to be at low risk of bias; these were models developed by Cockbain et al. [59] and Pulitano et al. [62].

The participant domain, which considers appropriate participant selection and data sources, was most often judged to be at low risk of bias for developed models. The predictor domain, which considers appropriate predictor definitions, blinding, and availability, was most often judged to be at low risk of bias for validated models.

Methodological issues in the analysis domain drove most studies' at high risk of bias. Common issues were inadequate samples sizes for development and validation, worsened by data-splitting in internal validation, excluding missing data at study entry or analysis, univariable predictor selection, and categorizing continuous predictors. These issues increase the risk of overfitting developed models and biasing performance estimates. Poor reporting of key information often inhibited risk of bias assessment, as the minimum information recommended by TRIPOD was not fully reported or was not reported at all.

4.2. Current literature

Although a plethora of blood transfusion prediction models have been developed and validated for elective surgery, to our knowledge, no systematic review has critically appraised them. Evidence is therefore limited on their combined performance and quality. However, reviews of models in other clinical specialties have found similar risk of bias profiles, reporting inadequacies, and methodological

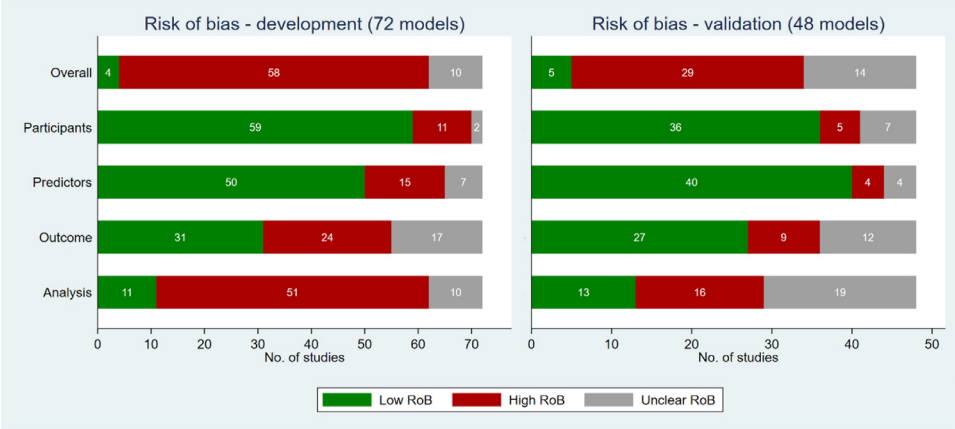

Fig. 3. Bar chart summarizing the risk of bias of all models developed and validated in the included studies.

flaws that we have found in our review [100–106]. For example, a review of prediction models for gestational hypertension and preeclampsia found poor compliance with TRIPOD [107]. A living systematic review of COVID-19 risk prediction models found that most were at high risk of bias due to inappropriate data sources, too-small sample sizes, inefficient internal validation methods, and absence of calibration assessment [103]. Although we found that data sources and participant selection were often appropriately addressed in blood transfusion risk prediction studies, methodological issues relating to reporting and analysis were prevalent.

4.3. Strengths and weaknesses

This is the first formal review and risk of bias assessment of blood transfusion prediction models developed for elective surgery. It provides a contemporary reflection of current practice in prediction modelling, generally and in each of the included clinical domains.

We described the quality of reporting using the formal recommendations made by TRIPOD and provided a PROBAST risk of bias assessment for each developed or validated model. We focused on models for elective surgeries. To ensure our results were specific to elective care, we excluded studies reporting >20% emergency surgeries in their study sample. We updated the study to include all models developed for blood transfusion from 2000 to 2021.

4.4. Clinical and research implications

Prediction models could help ensure the availability of compatible blood products when required, while improving clinical pathway efficiencies, reducing costs, and enhancing patient experiences. They could also guide the implementation of blood conservation strategies, such as antifibrinolytics [15] and cell salvage [108].

Future research should focus on updating and validating existing models and ensuring sufficient data are used to reliably estimate model performance during development and external validation [109–113]. The predictors and model performance estimates summarized and meta-analysed here can inform sample size calculations for future development and external validation studies. As well as ensuring sufficient data are used to develop and validate models, emphasis should also be placed on objective outcome definitions and data that reflect need of blood transfusion (ideally based on national guidelines) rather than data that reflects patients receiving blood transfusion. Using objective outcome measures will help avoid unnecessary blood transfusions and any inequality of blood product administration with respect to ethnicity and deprivation.

All studies developing, updating, or validating prediction models should adhere to TRIPOD and follow research recommendations [27,114]. Studies should use appropriate methods to handle missing data (e.g., multiple imputations) and avoid omitting missing data and conducting complete-case analyses [115,116]. Studies should test model assumptions, use appropriate methods to handle nonlinear continuous predictors (e.g., restricted cubic splines or fractional polynomials), and avoid categorizing continuous predictors [117–119]. Studies should use appropriate methods to internally validate developed models (e.g., bootstrapping and cross-validation) and avoid reducing data for model development by randomly splitting available data [120,121]. Studies should also assess and report both discrimination and calibration model performance measures [122,123].

5. Conclusion

Most blood transfusion models for elective surgery are at high risk of bias and demonstrate poor reporting and methodological quality. The models currently available to

predict blood transfusion should not be implemented into routine clinical care without further validation and research that adheres to reporting and methodological recommendations for prediction modelling.

Acknowledgments

We acknowledge English language editing by Dr Jennifer A de Beyer, Centre for Statistics in Medicine, University of Oxford.

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jclinepi.2023.05.002.

References

[1] Norfolk D. Handbook of Transfusion Medicine, United Kingdom Blood Services, 5th edition. 2013. Available at https://www.transfusionguidelines.org/transfusion-handbook. Accessed February 23, 2022.
[2] How blood is used. Available at https://www.blood.co.uk/why-give-blood/how-blood-is-used/. Accessed March 9, 2022.
[3] Annual reports. Available at https://www.bloodstocks.co.uk/resources/annual-reports/. Accessed February 6, 2022.
[4] NHS Blood and Transplant annual report and accounts: 2020 to 2021. Available at https://www.gov.uk/government/publications/nhs-blood-and-transplant-annual-report-and-accounts-2020-to-2021. Accessed March 9, 2022.
[5] Goodnough LT, Murphy MF. Do liberal blood transfusions cause more harm than good? BMJ 2014;349:g6897.
[6] Murphy MF. The Choosing Wisely campaign to reduce harmful medical overuse: its close association with Patient Blood Management initiatives. Transfus Med 2015;25(5):287−92.
[7] Murphy MF. Transfusing Wisely. BMJ 2015;350:h2308.
[8] Amber alert issued on blood stocks − existing O Neg and O Pos donors asked to make appointments. Available at https://www.blood.co.uk/news-and-campaigns/news-and-statements/amber-alert-issued-on-blood-stocks-existing-o-neg-and-o-pos-donors-asked-to-make-appointments/. Accessed October 31, 2022.
[9] Public responds to NHS amber alert over donor blood stocks. 2022. Available at https://www.bbc.com/news/health-63239924. Accessed October 31, 2022.
[10] Green L, Allard S, Cardigan R. Modern banking, collection, compatibility testing and storage of blood and blood components. Anaesthesia 2015;70 Suppl 1:3−9. e2.
[11] Gregory AJ, Grant MC, Boyle E, Arora RC, Williams JB, Salenger R, et al. Cardiac surgery-enhanced recovery programs modified for COVID-19: key steps to preserve resources, manage caseload backlog, and improve patient outcomes. J Cardiothorac Vasc Anesth 2020;34(12):3218−24.
[12] Hippisley-Cox J, Coupland C, Brindle P. Development and validation of QRISK3 risk prediction algorithms to estimate future risk of cardiovascular disease: prospective cohort study. BMJ 2017;357:j2099.
[13] National Institute for Health and Care Excellence. ColonFlag for identifying people at risk of colorectal cancer (Medtech innovation briefing MIB142). 2018. Available at https://www.nice.org.uk/advice/mib142. Accessed June 8, 2020.
[14] Green AR, Soria D, Stephen J, Powe DG, Nolan CC, Kunkler I, et al. Nottingham Prognostic Index Plus: validation of a clinical decision making tool in breast cancer in an independent series. J Pathol Clin Res 2016;2(1):32−40.
[15] Shah A, Palmer AJR, Klein AA. Strategies to minimize intraoperative blood loss during major surgery. Br J Surg 2020;107(2):e26−38.
[16] O'Malley SM, Sanders JO, Nelson SE, Rubery PT, O'Malley NT, Aquina CT. Significant variation in blood transfusion practice persists following adolescent idiopathic scoliosis surgery. Spine (Phila Pa 1976) 2021;46(22):1588−97.
[17] Joshi RV, Wilkey AL, Blackwell JM, Kwak J, Raphael J, Shore-Lesserson L, et al. Blood conservation and hemostasis in cardiac surgery: a survey of practice variation and adoption of evidence-based guidelines. Anesth Analg 2021;133:104−14.
[18] Lloyd TD, Neal-Smith G, Fennelly J, Claireaux H, Bretherton C, Carr AJ, et al. Peri-operative administration of tranexamic acid in lower limb arthroplasty: a multicentre, prospective cohort study. Anaesthesia 2020;75(8):1050−8.
[19] Klein AA, Collier T, Yeates J, Miles LF, Fletcher SN, Evans C, et al. The ACTA PORT-score for predicting perioperative risk of blood transfusion for adult cardiac surgery. Br J Anaesth 2017;119:394−401.
[20] Sima CS, Jarnagin WR, Fong Y, Elkin E, Fischer M, Wuest D, et al. Predicting the risk of perioperative transfusion for patients undergoing elective hepatectomy. Ann Surg 2009;250(6):914−21.
[21] Alghamdi AA, Davis A, Brister S, Corey P, Logan A. Development and validation of Transfusion Risk Understanding Scoring Tool (TRUST) to stratify cardiac surgery patients according to their blood transfusion needs. Transfusion 2006;46(7):1120−9.
[22] Ranucci M, Castelvecchio S, Frigiola A, Scolletta S, Giomarelli P, Biagioli B. Predicting transfusions in cardiac surgery: the easier, the better: the Transfusion Risk and Clinical Knowledge score. Vox Sang 2009;96(4):324−32.
[23] 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.
[24] 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.
[25] The Endnote Team. Endnote. Philadelphia, PA: Clarivate Analytics; 2013.
[26] Covidence systematic review software. Melbourne, Australia: Veritas Health Innovation; 2018.
[27] 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.
[28] Moons KGM, Groot JAH, Bouwmeester W, Vergouwe Y, Mallett S, Altman DG, et al. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. PLoS Med 2014;11(10):e1001744.
[29] Moons KGM, Wolff RF, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: a tool to assess risk of bias and applicability of prediction model studies: explanation and elaboration. Ann Intern Med 2019;170:W1−33.
[30] Harris P, Taylor R, Thielke R, Payne J, Gonzalez N, Conde J. Research electronic data capture (REDCap)-metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009;42(2):377e81.
[31] Debray TP, Damen JA, Riley RD, Snell K, Reitsma JB, Hooft L, et al. A framework for meta-analysis of prediction model studies with binary and time-to-event outcomes. Stat Methods Med Res 2019;28(9):2768−86.
[32] Wolff RF, Moons KGM, 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.

[33] StataCorp. Stata Statistical Software: Release 15. College Station, TX: StataCorp LLC. College Station, TX: StataCorp LLC; 2017.
[34] R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2021.
[35] Gordon M, Lumley T. forestplot: Advanced Forest Plot Using 'grid' Graphics. 2021. Available at https://CRAN.R-project.org/package=forestplot. Accessed February 7, 2022.
[36] Wikham H, Girlich M. tidyr: Tidy Messy Data. 2022. Available at https://CRAN.R-project.org/package=tidyr. Accessed February 7, 2022.
[37] Wikham H, Francois R, Henry L, Muller K. dplyr: A Grammar of Data Manipulation. 2021. Available at https://CRAN.R-project.org/package=dplyr. Accessed February 7, 2022.
[38] Hayn D, Kreiner K, Kastner P, Breznik N, Hofmann A, Gombotz H, et al. Data driven methods for predicting blood transfusion needs in elective surgery. Stud Health Technol Inform 2016;223:9−16.
[39] Doshi KA, Shastry S, Pai VB. Transfusion requirement prediction score for patients undergoing cardiac surgery: an experience from a tertiary care set-up from South India. Transfus Med 2021;31(4):243−9.
[40] To J, Sinha R, Kim SW, Robinson K, Kearney B, Howie D, et al. Predicting perioperative transfusion in elective hip and knee arthroplasty: a validated predictive model. Anesthesiology 2017;127:317−25.
[41] Welsby I, Crow J, Bandarenko N, Lappas G, Phillips-Bute B, Stafford-Smith M. A clinical prediction tool to estimate the number of units of red blood cells needed in primary elective coronary artery bypass surgery. Transfusion 2010;50(11):2337−43.
[42] Pennington Z, Ehresman J, Feghali J, Schilling A, Hersh A, Hung B, et al. A clinical calculator for predicting intraoperative blood loss and transfusion risk in spine tumor patients. Spine J 2021;21(2):302−11.
[43] Liu C, Vachharajani N, Song S, Cooke R, Kangrga I, Chapman WC, et al. A quantitative model to predict blood use in adult orthotopic liver transplantation. Transfus Apher Sci 2015;53(3):386−92.
[44] Pustavoitau A, Lesley M, Ariyo P, Latif A, Villamayor AJ, Frank SM, et al. Predictive modeling of massive transfusion requirements during liver transplantation and its potential to reduce utilization of blood bank resources. Anesth Analg 2017;124:1644−52.
[45] Stanhiser J, Chagin K, Jelovsek JE. A model to predict risk of blood transfusion after gynecologic surgery. Am J Obstet Gynecol 2017;216:506.e1−506.e14.
[46] Al-Khabori M, Al-Riyami AZ, Mukaddirov M, Al-Sabti H. Transfusion indication predictive score: a proposed risk stratification score for perioperative red blood cell transfusion in cardiac surgery. Vox Sang 2014;107(3):269−75.
[47] Cevenini G, Barbini E, Massai MR, Barbini P. A naïve Bayes classifier for planning transfusion requirements in heart surgery. J Eval Clin Pract 2013;19:25−9.
[48] Cirasino L, Barosi G, Torre M, Crespi S, Colombo P, Belloni PA. Preoperative predictors of the need for allogeneic blood transfusion in lung cancer surgery. Transfusion 2000;40(10):1228−34.
[49] Covin R, O'Brien M, Grunwald G, Brimhall B, Sethi G, Walczak S, et al. Factors affecting transfusion of fresh frozen plasma, platelets, and red blood cells during elective coronary artery bypass graft surgery. Arch Pathol Lab Med 2003;127(4):415−23.
[50] Lu L, Che J, Xie W, Cheng W, Yang Z, Dong R, et al. Adjusted preoperative variables to predict perioperative red blood cell transfusion in coronary artery bypass grafting. Gen Thorac Cardiovasc Surg 2020;68(12):1377−87.
[51] Paiva PP, Leite FM, Antunes PE, Antunes MJ. Risk-prediction model for transfusion of erythrocyte concentrate during extracorporeal circulation in coronary surgery. Braz J Cardiovasc Surg 2021;36(3):323−30.
[52] Liu S, Zhou R, Xia XQ, Ren H, Wang LY, Sang RR, et al. Machine learning models to predict red blood cell transfusion in patients undergoing mitral valve surgery. Ann Transl Med 2021;9(7):530.
[53] Simeone F, Franchi F, Cevenini G, Marullo A, Fossombroni V, Scolletta S, et al. A simple clinical model for planning transfusion quantities in heart surgery. BMC Med Inform Decis Mak 2011;11:44.
[54] Arora RC, Légaré JF, Buth KJ, Sullivan JA, Hirsch GM. Identifying patients at risk of intraoperative and postoperative transfusion in isolated CABG: toward selective conservation strategies. Ann Thorac Surg 2004;78(5):1547−54.
[55] Karkouti K, O'Farrell R, Yau TM, Beattie WS. Reducing bleeding in cardiac surgery research group. Prediction of massive blood transfusion in cardiac surgery. Can J Anaesth 2006;53(8):781−94.
[56] Karkouti K, Cohen MM, McCluskey SA, Sher GD. A multivariable model for predicting the need for blood transfusion in patients undergoing first-time elective coronary bypass graft surgery. Transfusion 2001;41(10):1193−203.
[57] Madhu Krishna NR, Nagaraja PS, Singh NG, Nanjappa SN, Kumar KN, Prabhakar V, et al. Evaluation of risk scores in predicting perioperative blood transfusions in adult cardiac surgery. Ann Card Anaesth 2019;22(1):73−8.
[58] Leff J, Romano CA, Gilbert S, Nair S. Validation study of the transfusion risk and clinical knowledge (TRACK) tool in cardiac surgery patients: a retrospective analysis. J Cardiothorac Vasc Anesth 2019;33(10):2669−75.
[59] Cockbain AJ, Masudi T, Lodge JPA, Toogood GJ, Prasad KR. Predictors of blood transfusion requirement in elective liver resection. HPB (Oxford) 2010;12(1):50−5.
[60] Huang H, Cao M. Development and validation of a nomogram to predict intraoperative blood transfusion for gastric cancer surgery. Transfus Med 2021;31(4):250−61.
[61] Kim Y, Bagante F, Gani F, Ejaz A, Xu L, Wasey JO, et al. Nomogram to predict perioperative blood transfusion for hepatopancreaticobiliary and colorectal surgery. Br J Surg 2016;103(9):1173−83.
[62] Pulitanò C, Arru M, Bellio L, Rossini S, Ferla G, Aldrighetti L. A risk score for predicting perioperative blood transfusion in liver surgery. Br J Surg 2007;94(7):860−5.
[63] Wang HQ, Yang J, Yang JY, Wang WT, Yan LN. Development and validation of a predictive score for perioperative transfusion in patients with hepatocellular carcinoma undergoing liver resection. Hepatobiliary Pancreat Dis Int 2015;14(4):394−400.
[64] Feng Y, Xu Z, Sun X, Wang D, Yu Y. Machine learning for predicting preoperative red blood cell demand. Transfus Med 2021;31(4):262−70.
[65] Yamamoto Y, Ikoma H, Morimura R, Konishi H, Murayama Y, Komatsu S, et al. Predictive scoring system assessing the need for intraoperative blood transfusions during hepatectomy for hepatocellular carcinoma. Anticancer Res 2014;34(1):313−8.
[66] Quan B, Zhang WG, Serenari M, Liang L, Xing H, Li C, et al. A novel online calculator to predict perioperative blood transfusion in patients undergoing liver resection for hepatocellular carcinoma: an international multicenter study. HPB (Oxford) 2020;22(12):1711−21.
[67] Bagante F, Spolverato G, Ruzzenente A, Wilson A, Gani F, Conci S, et al. Validation of a nomogram to predict the risk of perioperative blood transfusion for liver resection. World J Surg 2016;40:2481−9.
[68] Bansal SS, Hodson J, Khalil K, Dasari B, Marudanayagam R, Sutcliffe RP, et al. Distinct risk factors for early and late blood transfusion following pancreaticoduodenectomy. Hepatobiliary Pancreat Dis Int 2018;17(4):349−57.
[69] Cucchetti A, Siniscalchi A, Cescon M, Mazzotti F, Ercolani G, Ravaioli M, et al. Assessment of perioperative transfusion requirement for cirrhotic patients undergoing elective hepatectomy. Minerva Anestesiol 2014;80(6):645−54.
[70] Lemke M, Law CHL, Li J, Dixon E, Tun Abraham M, Hernandez Alejandro R, et al. Three-point transfusion risk score in hepatectomy. Br J Surg 2017;104(4):434−42.
[71] van Klei WA, Rheineck Leyssius AT, Grobbee DE, Moons KGM. Identifying patients for blood conservation strategies. Br J Surg 2002;89(9):1176−82.

[72] Guo Y, You L, Hu H, Tong A, Zhang X, Yan L, et al. A predictive nomogram for red blood cell transfusion in pheochromocytoma surgery: a study on improving the preoperative management of pheochromocytoma. Front Endocrinol (Lausanne) 2021;12:647610.
[73] Janny S, Eurin M, Dokmak S, Toussaint A, Farges O, Paugam-Burtz C. Assessment of the external validity of a predictive score for blood transfusion in liver surgery. HPB (Oxford) 2015;17(4):357−61.
[74] Lemke M, Mahar A, Karanicolas PJ, Coburn NG, Law CHL, Hallet J. Three point transfusion risk score in hepatectomy: an external validation using the American College of Surgeons − national Surgical Quality Improvement Program (ACS-NSQIP). HPB (Oxford) 2018;20(7):669−75.
[75] Klebanoff JS, Marfori CQ, Sparks AD, Barnes WA, Ingraham CF, Moawad GN. A clinically applicable prediction model for the risk of transfusion in women undergoing myomectomy. J Minim Invasive Gynecol 2021;28(10):1765−1773.e1.
[76] Ackroyd SA, Brown J, Houck K, Chu C, Mantia-Smaldone G, Rubin S, et al. A preoperative risk score to predict red blood cell transfusion in patients undergoing hysterectomy for ovarian cancer. Am J Obstet Gynecol 2018;219:598.e1−598.e10.
[77] Ahmed I, Chan JKK, Jenkins P, Brenkel I, Walmsley P. Estimating the transfusion risk following total knee arthroplasty. Orthopedics 2012;35(10):e1465−71.
[78] Huang Z, Huang C, Xie J, Ma J, Cao G, Huang Q, et al. Analysis of a large data set to identify predictors of blood transfusion in primary total hip and knee arthroplasty. Transfusion 2018;58(8):1855−62.
[79] Noticewala MS, Nyce JD, Wang W, Geller JA, Macaulay W. Predicting need for allogeneic transfusion after total knee arthroplasty. J Arthroplasty 2012;27(6):961−7.
[80] Rashiq S, Shah M, Chow AK, O'Connor PJ, Finegan BA. Predicting allogeneic blood transfusion use in total joint arthroplasty. Anesth Analg 2004;99:1239−44.
[81] Hu C, Wang YH, Shen R, Liu C, Sun K, Ye L, et al. Development and validation of a nomogram to predict perioperative blood transfusion in patients undergoing total knee arthroplasty. BMC Musculoskelet Disord 2020;21(1):315.
[82] Jo C, Ko S, Shin WC, Han HS, Lee MC, Ko T, et al. Transfusion after total knee arthroplasty can be predicted using the machine learning algorithm. Knee Surg Sports Traumatol Arthrosc 2020;28(6):1757−64.
[83] Thompson PA, May D, Choong PF, Tacey M, Liew D, Cole-Sinclair MF. Predicting blood loss and transfusion requirement in patients undergoing surgery for musculoskeletal tumors. Transfusion 2014;54(6):1469−77.
[84] Su JQ, Xie S, Cai ZG, Wang XY. Developing a predictive risk score for perioperative blood transfusion: a retrospective study in patients with oral and oropharyngeal squamous cell carcinoma undergoing free flap reconstruction surgery. Ann Transl Med 2021;9(10):854.
[85] Kolbenschlag J, Schneider J, Harati K, Goertz O, Hernekamp JF, Kempf R, et al. Predictors of intraoperative blood transfusion in free tissue transfer. J Reconstr Microsurg 2016;32:706−11.
[86] Shah MD, Goldstein DP, McCluskey SA, Miles BA, Hofer SO, Hofer S, et al. Blood transfusion prediction in patients undergoing major head and neck surgery with free-flap reconstruction. Arch Otolaryngol Head Neck Surg 2010;136(12):1199−204.
[87] Krupp NL, Weinstein G, Chalian A, Berlin JA, Wolf P, Weber RS. Validation of a transfusion prediction model in head and neck cancer surgery. Arch Otolaryngol Head Neck Surg 2003;129(12):1297−302.
[88] Durand WM, DePasse JM, Daniels AH. Predictive modeling for blood transfusion after adult spinal deformity surgery: a tree-based machine learning approach. Spine (Phila Pa 1976) 2018;43(15):1058−66.
[89] Wang H, Wang K, Lv B, Xu H, Jiang W, Zhao J, et al. Establishment and assessment of a nomogram for predicting blood transfusion risk in posterior lumbar spinal fusion. J Orthop Surg Res 2021;16(1):39.
[90] Carabini LM, Zeeni C, Moreland NC, Gould RW, Avram MJ, Hemmer LB, et al. Development and validation of a generalizable model for predicting major transfusion during spine fusion surgery. J Neurosurg Anesthesiol 2014;26(3):205−15.
[91] Lenoir B, Merckx P, Paugam-Burtz C, Dauzac C, Agostini MM, Guigui P, et al. Individual probability of allogeneic erythrocyte transfusion in elective spine surgery: the predictive model of transfusion in spine surgery. Anesthesiology 2009;110:1050−60.
[92] Cywinski JB, Alster JM, Miller C, Vogt DP, Parker BM. Prediction of intraoperative transfusion requirements during orthotopic liver transplantation and the influence on postoperative patient survival. Anesth Analg 2014;118:428−37.
[93] Massicotte L, Carrier FM, Denault AY, Karakiewicz P, Hevesi Z, McCormack M, et al. Development of a predictive model for blood transfusions and bleeding during liver transplantation: an observational cohort study. J Cardiothorac Vasc Anesth 2018;32(4):1722−30.
[94] McCluskey SA, Karkouti K, Wijeysundera DN, Kakizawa K, Ghannam M, Hamdy A, et al. Derivation of a risk index for the prediction of massive blood transfusion in liver transplantation. Liver Transpl 2006;12(11):1584−93.
[95] Metcalf RA, Pagano MB, Hess JR, Reyes J, Perkins JD, Montenovo MI. A data-driven patient blood management strategy in liver transplantation. Vox Sang 2018;113:421−9.
[96] Pustavoitau A, Rizkalla NA, Perlstein B, Ariyo P, Latif A, Villamayor AJ, et al. Validation of predictive models identifying patients at risk for massive transfusion during liver transplantation and their potential impact on blood bank resource utilization. Transfusion 2020;60(11):2565−80.
[97] Liu LP, Zhao QY, Wu J, Luo YW, Dong H, Chen ZW, et al. Machine learning for the prediction of red blood cell transfusion in patients during or after liver transplantation surgery. Front Med (Lausanne) 2021;8:632210.
[98] Massicotte L, Capitanio U, Beaulieu D, Roy JD, Roy A, Karakiewicz P. Independent validation of a model predicting the need for packed red blood cell transfusion at liver transplantation. Transplantation 2009;88(3):386−91.
[99] Stangenberg L, Curran T, Shuja F, Rosenberg R, Mahmood F, Schermerhorn ML. Development of a risk prediction model for transfusion in carotid endarterectomy and demonstration of cost-saving potential by avoidance of 'type and screen'. J Vasc Surg 2016;64(6):1711−8.
[100] 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.
[101] 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.
[102] Gerry S, Bonnici T, Birks J, Kirtley S, Virdee PS, Watkinson PJ, et al. Early warning scores for detecting deterioration in adult hospital patients: systematic review and critical appraisal of methodology. BMJ 2020;369:m1501.
[103] 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.
[104] Bridge J, Blakey JD, Bonnett LJ. A systematic review of methodology used in the development of prediction models for future asthma exacerbation. BMC Med Res Methodol 2020;20:22.
[105] Bradley A, Meer RVD, McKay CJ. A systematic review of methodological quality of model development studies predicting prognostic outcome for resectable pancreatic cancer. BMJ Open 2019;9(8):e027192.
[106] Collins GS, Omar O, Shanyinde M, Yu LM. A systematic review finds prediction models for chronic kidney disease were poorly reported and often developed using inappropriate methods. J Clin Epidemiol 2013;66:268−77.

[107] Antwi E, Amoakoh-Coleman M, Vieira DL, Madhavaram S, Koram KA, Grobbee DE, et al. Systematic review of prediction models for gestational hypertension and preeclampsia. PLoS One 2020;15:e0230955.
[108] Klein AA, Evans E. Association of Anaesthetists guidelines on cell salvage: a reply. Anaesthesia 2018;73(12):1577.
[109] Riley RD, Snell KIE, Ensor J, Burke DL, Harrell FE, Moons KGM, et al. Minimum sample size for developing a multivariable prediction model: Part I − continuous outcomes. Stat Med 2019;38:1262−75.
[110] Riley RD, Snell KI, Ensor J, Burke DL, Harrell FE, Moons KG, et al. Minimum sample size for developing a multivariable prediction model: PART II - binary and time-to-event outcomes. Stat Med 2019;38:1276−96.
[111] Snell KIE, Archer L, Ensor J, Bonnett LJ, Debray TPA, Phillips B, et al. External validation of clinical prediction models: simulation-based sample size calculations were more reliable than rules-of-thumb. J Clin Epidemiol 2021;135:79−89.
[112] Archer L, Snell KIE, Ensor J, Hudda MT, Collins GS, Riley RD. Minimum sample size for external validation of a clinical prediction model with a continuous outcome. Stat Med 2021;40:133−46.
[113] Riley RD, Debray TPA, Collins GS, Archer L, Ensor J, van Smeden M, et al. Minimum sample size for external validation of a clinical prediction model with a binary outcome. Stat Med 2021;40:4230−51.
[114] 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.
[115] Ambler G, Omar RZ, Royston P. A comparison of imputation techniques for handling missing predictor values in a risk model with a binary outcome. Stat Methods Med Res 2007;16(3):277−98.
[116] Janssen KJM, Donders ART, Harrell FE, Vergouwe Y, Chen Q, Grobbee DE, et al. Missing covariate data in medical research: to impute is better than to ignore. J Clin Epidemiol 2010;63:721−7.
[117] Gauthier J, Wu QV, Gooley TA. Cubic splines to model relationships between continuous variables and outcomes: a guide for clinicians. Bone Marrow Transplant 2020;55(4):675−80.
[118] Shepherd BE, Rebeiro PF. Assessing and interpreting the association between continuous covariates and outcomes in observational studies of HIV using splines. J Acquir Immune Defic Syndr 2017;74:e60−3.
[119] Sauerbrei W, Meier-Hirmer C, Benner A, Royston P. Multivariable regression model building by using fractional polynomials: description of SAS, STATA and R programs. Comput Stat Data Anal 2006;50(12):3464−85.
[120] Steyerberg EW. Validation in prediction research: the waste by data splitting. J Clin Epidemiol 2018;103:131−3.
[121] 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.
[122] Van Calster B, McLernon DJ, van Smeden M, Wynants L, Steyerberg EW, Bossuyt P, et al. Calibration: the Achilles heel of predictive analytics. BMC Med 2019;17(1):230.
[123] Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. Epidemiology 2010;21:128−38.