Journal of Clinical Epidemiology 186 (2025) 111902

Journal of Clinical Epidemiology

ORIGINAL RESEARCH

External validation, impact assessment and clinical utilization of clinical prediction models: a prospective cohort study

Banafsheh Arshi[a,b], Laura Elizabeth Cowley[c], Eline Rijnhart[a], Kelly Reeve[d], Luc J. Smits[a], Laure Wynants[a,e,f,*]

[a]*Faculty of Health, Medicine and Life Sciences, Department of Epidemiology, CAPHRI School for Public Health and Primary Care, Maastricht University, Maastricht, The Netherlands*

[b]*Institute for Risk Assessment Sciences (IRAS), Utrecht University, Utrecht, The Netherlands*

[c]*Population Data Science, Swansea University Medical School, Swansea, Wales, United Kingdom*

[d]*Department of Epidemiology, Biostatistics and Prevention Institute, Department of Biostatistics, University of Zurich, Hirschengraben 84, Zurich CH-8001, Switzerland*

[e]*Department of Development and Regeneration, KU Leuven, Leuven, Belgium*

[f]*Leuven Unit for Health Technology Assessment Research (LUHTAR), KU Leuven, Leuven, Belgium*

Accepted 11 July 2025; Published online 16 July 2025

Abstract

Objectives: We aimed to assess paths taken by clinical prediction models (CPMs) after development by quantifying external validation, impact assessment, and utilization in clinical practice.

Study Design and Setting: We followed a random sample of 109 regression-based CPM development articles published between 1995 and 2020 by performing a forward citation search. We estimated 5- and 10-year probabilities of validation and impact assessment after development of CPMs using Kaplan-Meier analysis. In addition, we conducted a survey among the authors of the development articles to determine whether the CPMs had been used in clinical settings.

Results: Eighteen (17%) CPM development articles reported a CPM that was externally validated after development. Five- and 10-year probabilities of validation were 0.13 (0.06−0.19) and 0.16 (0.08−0.23), respectively. Only 1 article had a CPM with impact assessment during follow-up (10-year probability: 0.01 [0−0.04]). Among the 34 (31%) articles with a survey response, 17 (50%) had CPMs that had been used in clinical practice, in a median of five sites (interquartile range: 1−347). Of these models, only 4 (24%) were externally validated, and none had undergone impact assessment.

Conclusion: Despite evidence of utilization in clinical settings, few models are externally validated after development, and published impact assessment is scarce. To prevent compromising patient safety, it is crucial to intensify efforts to promote external validation and impact assessment of prediction models.

Keywords: Prediction research; Clinical prediction models; External validation; Impact assessment; Clinical utilization; Implementation

1. Introduction

Clinical prediction models (CPMs) can be useful aids in decision-making and patient management by offering the estimated probability of the presence or future occurrence of a particular outcome relevant to diagnosis or prognosis [1,2]. Ultimately, they can improve health-care quality or improve the cost-effectiveness of intervention strategies [3,4]. While prediction model development is increasing at a staggering pace [5−7], their adoption rate in clinical practice is unknown [4,8,9].

Methodologic guidelines recommend external validation and impact assessment of CPMs to judge their suitability for eventual utilization in clinical practice [3,9,10]. External validation of a CPM examines its reliability and transportability in populations outside the development population [11]. CPMs that show potential via adequate

Funding: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

* Corresponding author. Faculty of Health, Medicine and Life Sciences, Department of Epidemiology, CAPHRI School for Public Health and Primary Care, Maastricht University, Peter Debyeplein 1, 6229 HA Maastricht P.O. Box 616, 6200 MD, Maastricht, The Netherlands.

E-mail address: Laure.Wynants@maastrichtuniversity.nl (L. Wynants).

https://doi.org/10.1016/j.jclinepi.2025.111902

What is new?

Key findings

- Among a representative cohort of regression-based clinical prediction model (CPM) development publications, 17% reported a model that was externally validated after development. Ten-year probability of validation was 16% (8%−23%).
- Only 1 article had a CPM that had undergone impact assessment. Ten-year probability of impact assessment was 1% (0%−4%).
- Despite evidence of clinical use, the majority of used models were never externally validated and none had a published impact assessment.

What this adds to what was known?

- For the first time, we quantified external validation, impact assessment and clinical utilization of a representative cohort of regression-based CPMs after publication, across all medical fields.
- The practical pathway to clinical utilization does not align with the ideal evidence-based pathway which underlines external validation and impact assessment of models before clinical utilization.
- The generalizability, harms, and benefits of most models used in clinical practice are unknown or unpublished.

What is the implication, what should change now?

- Researchers, funding agencies, and stakeholders are recommended to pursue evaluation of existing models and outline a dissemination plan before development of new ones.
- Formulation of an intervention based on the CPM should be considered at development and validation by developers, stakeholders, and funding agencies.
- Health-care professionals and other stakeholders are recommended to abstain from using models without supporting evidence, or to take caution.

external validation do not always impact the decision-making process or improve health outcomes. Therefore, impact assessment is a crucial step in evaluating the extent to which actual use of a CPM can effectively improve outcomes or enhance decision-making or cost-effectiveness compared to usual care [9,12,13].

However, external validations remain uncommon [4,7,14,15], and publications on impact assessments appear to be rare [8,12,16]. This implies a myriad of CPMs with unknown validity or effect, hindering clinical utilization of promising models. The upsurge in model development further raises concern about the potential waste of resources in prediction research [12]. To date, a systematic empirical investigation of the paths taken by CPMs after development is lacking. The existing reports are from specific clinical fields or models, or are restricted by their cross-sectional nature. What is more, model utilization in clinical practice for decision-making or risk counseling is unclear and has not been systematically investigated.

In the present study, we investigated the fate of a representative cohort of CPM development publications to quantify external validation and impact assessment after CPM development. In addition, we assessed utilization of the developed CPMs in clinical practice.

2. Methods

2.1. Study cohort

This study is a follow-up of a cohort of studies developing multivariable regression-based CPMs for a health outcome. The cohort was randomly selected from prediction model publications between 1 January 1995 and 31 December 2020 in PubMed and Embase, using a validated search strategy (sensitivity: 98.2% [91.5%−99.9%]) [17,18]. We set a target sample size of 100 model development articles balancing precision of statistical estimates of interest and screening workload [18]. To reach this target, 2860 records—110 per publications year—were randomly selected for screening based on published search string metrics. As the target of 100 CPM development articles, with at least 1 per year, was not met, we increased the yearly sample by 30 until the goal was reached. Articles not developing a model, single-predictor CPMs, diagnostic questionnaires, nondiagnostic or nonprognostic models, preclinical studies, and methodological studies were excluded. Other exclusions were articles that were not available in English, conference abstracts, and articles with no available full text [18].

Extracted data from these articles included the medical domain (based on outcome and/or target population); publishing journal; country; publication year; study design; study setting; population; data sources; sample size; outcomes; number of individuals with the outcome; number of models developed; type of model (diagnostic/prognostic); number and type of predictors; accompanying internal or external validation; validation type; performance measures (discrimination and calibration); and model presentation (model equation, nomogram, sum score, etc).

This study follows the Strengthening the Reporting of OBservational studies in Epidemiology guidance [19]. The forward citation search follows the Preferred Reporting Items for Systematic Reviews and Meta-Analyses [20]. The protocol for this study was published [18] and registered in

the Open Science Framework Registry (https://osf.io/nj8s9). Ethical clearance was obtained from the Faculty of Health, Medicine and Life Sciences Research Ethics Committee of Maastricht University (case number: FHML-REC/2023/066).

2.2. Outcomes after development publication

2.2.1. Validation, impact assessment, reviews, and guidelines

We followed all 109 CPM development articles by performing a forward citation search on the Scopus and Web of Science online platforms. Six reviewers, in pairs, screened abstracts and full texts independently with the primary goal of identifying any published external validation or impact assessment studies. Discrepancies were resolved by consensus and arbitration in team meetings. Eligible articles were those that reported external validation or impact assessment of the original CPMs. We defined external validation as studies that evaluate the model's predictive performance in a different population or in a different time period. Impact assessment studies could be either health-economic analyses, decision-analytic evaluations or empirical intervention studies. Empirical intervention studies could include experimental, quasiexperimental and pre-experimental studies (eg, prepost studies). Additional eligible articles were guidelines and consensus statements that mentioned the CPMs; (systematic) reviews and meta-analyses with a focus on prediction; studies describing decision aids based on the models; and studies using the CPM for other research purposes (eg, elevated predicted risk as inclusion criterion). The date of the forward citation search was 27-02-2022. We also checked citations of published study protocols picked up by our search to avoid missing any eligible articles.

The extracted information from the follow-up articles included the country; publication year; number of shared authors with the development article; publishing journal; setting; study design; sample size; population characteristics; primary outcome; and secondary outcomes (if applicable). From the external validation articles, we additionally extracted information on the number of outcomes, reported performance measures, and model presentation. From the impact assessment articles, we additionally extracted information on the target population; methods; evaluated risk threshold(s); risk-based management options; and recommendations to use in clinical practice. We also extracted information from articles developing a decision aid (description; target population; intended user and moment of application), (systematic) reviews/meta-analyses (performance measures; number of included models (in case of model comparisons)), and guidelines and consensus articles (reporting risk thresholds; recommendations regarding clinical use). Additional details have been described in the study protocol and Supplementary methods [18].

2.2.2. Model utilization

Parallel to the citation search of the literature, we asked the authors of the 109 CPM development articles about utilization of their models via an online survey. We defined utilization of CPMs as consistent clinical use (current or past) in the population the model was intended for (based on target population and setting). If respondents reported that the CPM was used, we asked if this was in clinical practice outside the research context, an application in research, or both, to distinguish between clinical utilization and application in research. If the CPMs were used in clinical settings, we also asked whether this use was inside and/or outside the site of model development, and whether it was inside and/or outside the country of origin. In case of research application, we asked authors about the purpose (validation study, impact assessment, as inclusion criterion in another study, or used as a variable in analysis). We also inquired from the authors whether the CPM was recommended by local or international guidelines, if any training to apply the model had been organized, and whether there were any decision aids (as charts, nomograms, decision rules, or software) based on their CPM. Further details on survey design and piloting are described in Supplementary methods and elsewhere [18].

Surveys were first shared with the corresponding author of the CPM development article via email. In case of no response from the corresponding author, surveys were subsequently shared with the last author, the second to last, first and second author, in that order (unless no contact information was publicly available). All authors received two reminders to complete the survey. Nonresponse was defined as receiving no response from any of the contacted authors. All responding authors gave consent in an online form.

2.3. Statistical analysis

Characteristics of CPM development and follow-up articles from the citation search were described for the total cohort and stratified by model utilization. Next, we used Kaplan-Meier analysis to estimate five- and ten-year probabilities of external validation, impact assessment, guideline recommendation and inclusion in (systematic) reviews/meta-analyses for the total cohort, and stratified by publication year of the model development. Time to each outcome was the time difference between publication date of the model development article and the first publication for each outcome event. CPM observations without outcomes at the time of the forward citation search were censored. In post hoc analyses, we used 1 multivariable Cox regression to explore the association between publication year (per 10 years), sample size (per 1000 individuals), reporting of calibration or discrimination, and any model presentation in the development article with external validation. In a prespecified analysis [18], we used a multivariable logistic regression to study the association of publication year, external validation, and clinical use of

the model. We had also prespecified to study the association between impact assessment and use in the regression, but due to insufficient impact assessments, we investigated the association between having either guidelines/consensus or systematic review/meta-analysis (defined as a single binary variable) and use instead. Due to the limited sample size, stratified analyses of survey results were not performed although it was planned in the original study protocol.

3. Results

We retrieved over 5 million hits after searching for published prediction model development studies, of which over ten thousand were randomly selected for manual screening and 109 were eventually included as studies developing a CPM (Fig 1). Median publication year (interquartile range [IQR]) was 2013 (2006−2018). The top three medical fields in our random sample were cardiovascular disease (20%), gynecology and obstetrics (15%), and gastroenterology (13%) [5]. Of these articles, 61% developed prognostic models and 39% diagnostic models, with a median C-statistic of 0.8 (IQR: 0.73−0.97). Fourteen (13%) CPM development articles reported external validation within the same publication.

3.1. External validations, impact assessments, guidelines, and reviews after CPM development

The forward citation search of 109 CPM development articles yielded 5821 publications. After screening, we identified 66 external validation articles (48 by independent investigators), two impact assessment/intervention articles,

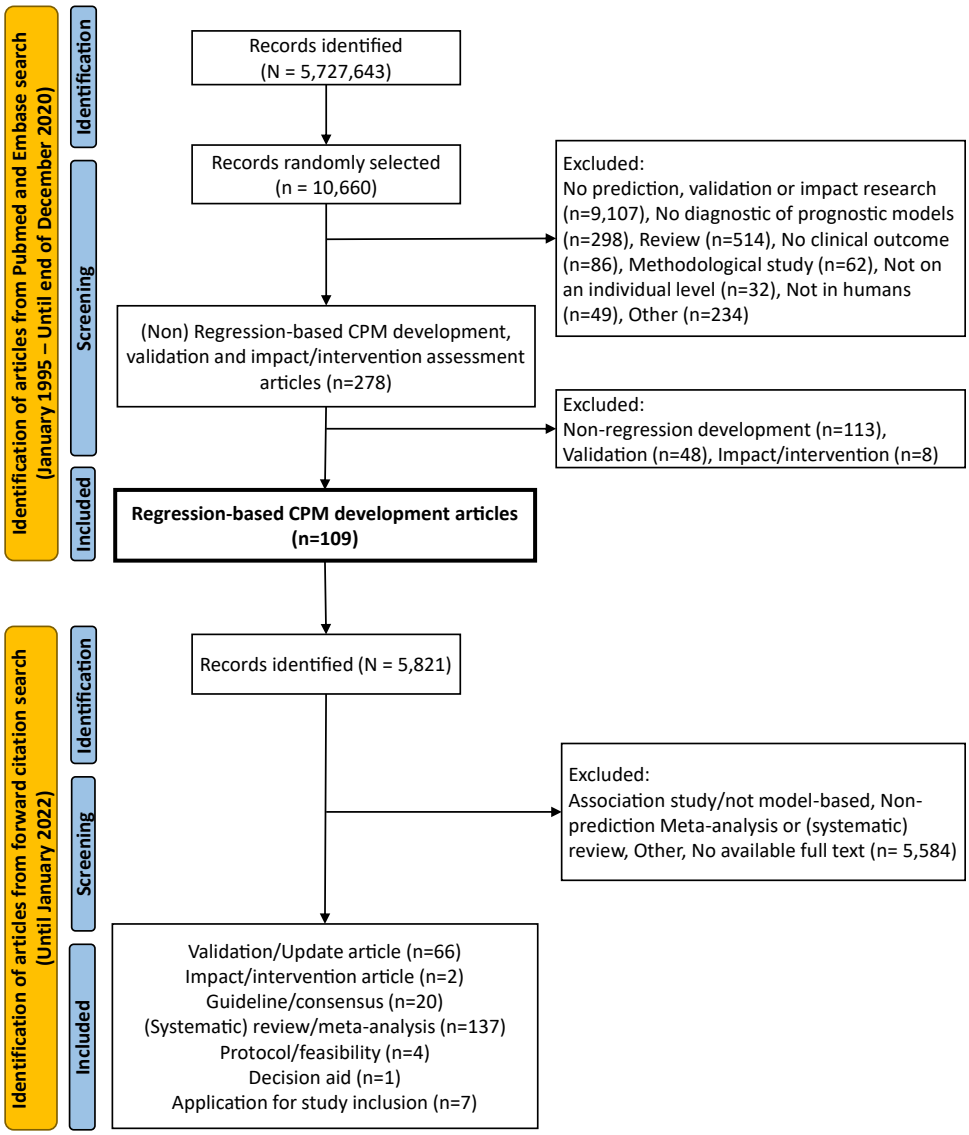

Figure 1. PRISMA chart for screening of regression-based CPM development articles and follow-up by forward citation search. PRISMA, Preferred Reporting Items for Systematic Reviews and Meta-Analyses; CPM, clinical prediction model.

20 guideline/consensus articles and 137 (systematic) reviews/meta-analyses (Fig 1). We also identified four published protocols for validation or impact assessment or feasibility studies, none of which had subsequent publications by the search date. Characteristics of these identified articles are presented in Supplementary Table 1. We also encountered two articles describing a decision aid and seven articles applying the CPM as inclusion criterion.

After their publication, 18 (17%) of the CPM development articles were externally validated. Eight (7%) had independent external validations by nonoverlapping investigators and 2 (2%) had both overlapping also independent investigators. Median (IQR) time-to-first validation was: 3 (2−6) years. The median number of external validation articles per CPM was 2 (IQR: 1−2, maximum: 25). Five- and 10-year probabilities of external validation for a CPM were 0.13 (0.06−0.19) and 0.16 (0.08−0.23), respectively (Fig 2). The probabilities of validation were somewhat similar among CPMs developed before and after 2010 (Supplementary Table 2). Characteristics of the external validation publication after development are presented in Supplementary Table 3. Overall, 25 (23%) CPMs had an external validation in the development article or during follow-up.

Among the 109 CPM development articles, 1 (1%) had two impact assessment articles that were published 6 and 8 years after the development publication. The impact articles had a pre-experimental design, had coupled the CPM with risk-based patient management recommendations, and reported favorable impact on the outcome. The 10-year probability of impact assessment was 0.01 (0−0.04) (Figure 2).

In addition, 9 (8%) of the 109 CPM development articles had a CPM mentioned in guidelines/consensus publications (median [IQR]: 2 [1−2] per CPM). Four CPMs received a positive recommendation, such as remarking that a score based on the CPM could help with differentiating or predicting malignancy by pathologists, or citing them among risk scores that could be incorporated in practice and used for communicating risk to patients. Five- and 10-year probabilities of mentions in guidelines/consensus articles were 0.04 (0.01−0.1) and 0.09 (0.02−0.15) (Fig 2, Supplementary Table 2). Lastly, 35 (32%) CPM development articles were included in at least 1 review after

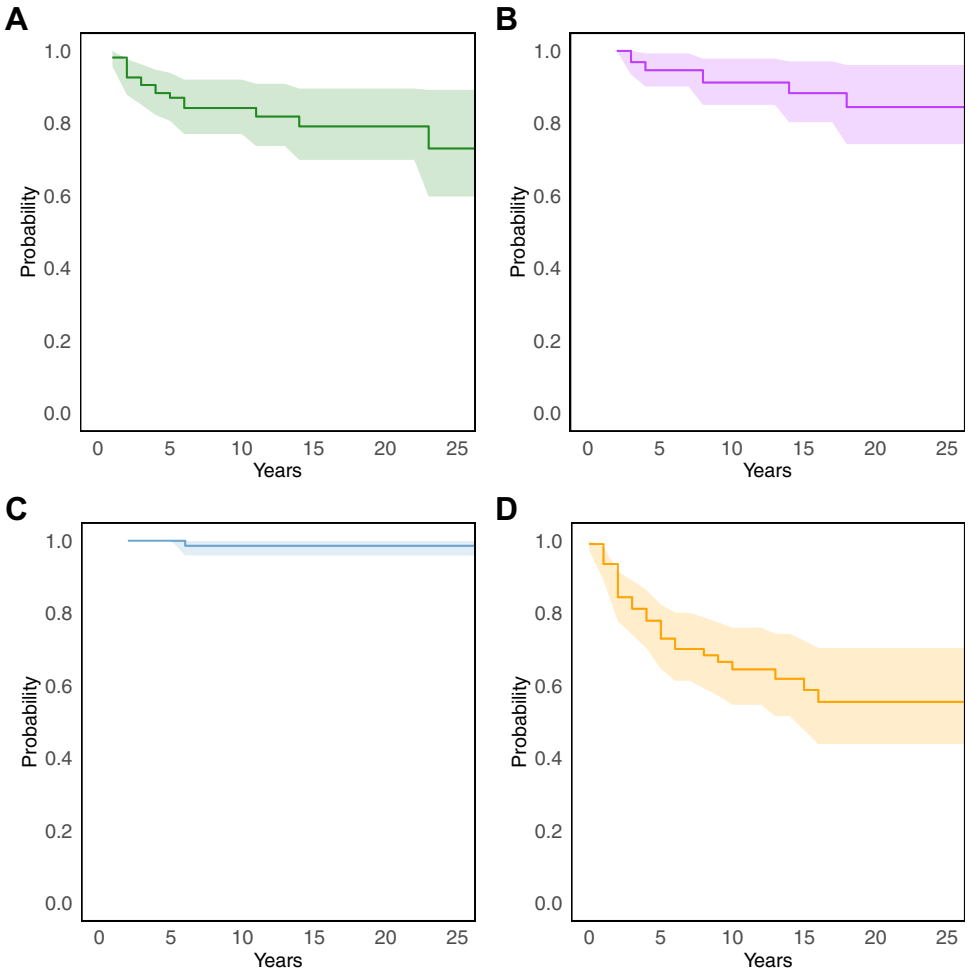

Figure 2. Kaplan-Meier plot for the time to first validation (A), guideline/consensus (B), impact/intervention (C), and (systematic) review/Meta-analysis (D) publication after CPM development (N = 109).

publication (5- and 10-year probabilities were 0.27 (0.18−0.35) and 0.35 (0.24−0.45), respectively).

In post hoc multivariable Cox regression, presentation of the final model in the development article (formula with intercept or nomograms, sum score, etc) was positively associated with external validation (hazard ratio [HR; 95% CI]: 4.25 [1.60−11.2]). However, publication year, sample size and reporting calibration/discrimination were not meaningfully associated with external validation (HR per 10 years [95% CI]:1.05 [0.51−2.16], per 1000 individuals: 1.01 [0.99−1.04] and 0.95 (0.27−3.33), respectively; Supplementary Table 4). Finally, classifying 1 external validation that described decision-curve analysis (published 23 years after the development publication) as impact assessment did not change our results (Supplementary Fig 2).

3.2. Model utilization

We received survey responses for 34 (31%) of the 109 development articles. Authors of 17 (50%) articles reported use of the developed CPM in clinical practice (ie, outside the research context) in a median of 5 (IQR: 1−347) sites. Clinical use was supported by practical training for health care professionals (29%), model-based decision aids (29%), or software (18%). Thirteen of these clinically used CPMs were also reported to be used in patients for research purposes, most frequently as a confounder or stratification factor (41%), or for validation (35%). Further survey results are reported in Supplementary Table 5.

Based on the forward citation search, four (24%) of the 17 used models had an external validation (Table 1). External validation was associated with a 2.7 higher odds ratio (OR) of use (95% CI 0.40−22.0, Supplementary Table 6). None of the CPMs with clinical use had an impact assessment publication based on the citation search, 23% percent had a mention in published clinical guidelines or consensus statements, and 29% were included in reviews. CPMs that were more recently published were used less often (OR 0.40 per 10 years, 95% CI 0.12−1.04). Mentions in guidelines or reviews were not positively associated with use (OR 0.88, 95% CI 0.19−4.00).

4. Discussion

Among a representative random sample of 109 regression-based clinical prediction models developed between 1995 and 2020, 17% had a published external validation and 1% had impact assessment during follow-up. While half (17) of the 34 articles for which we had a survey response had CPMs that were used in clinical settings, only four had a published external validation after development; none had a published impact assessment.

Models often perform poorly outside of their development sample, due to differences between populations or changes in population characteristics over time [21]. Also, the reported performance at model development is commonly too optimistic due to methodological bias, overfitting and area under the curve-hacking (reanalyzing data until a favorable area under the curve is achieved) [12,22]. Yet, external validations are infrequently published [7,15,23]. In another study, the 10-year validation rate of selected newly developed CPMs before 2010 was higher (38%) [4]. Articles followed up in our study were randomly selected, more recent and had smaller size but similar performance metrics. Our results suggest that the probability of validation since 2010 has not changed, which could be due to the noise created by the surge in model development [5,6,24]. Improvements in reporting may increase future validation rates.

In our study, presenting the CPM— model equation, nomogram, score, etc— was linked to an increased likelihood of validation during follow-up while sample size or reporting of model performance were not. We believe barriers to validation could also be linked to lack of a targeted dissemination and utilization plan at model development and emphasis on novelty by funding agencies and stakeholders which may encourage development of new models rather than further evaluation of existing ones.

To our knowledge, the proportion of regression-based CPMs that undergo impact assessment after development has not been quantified until now [12]. Published impact assessments are exceedingly rare [10,12,16], also for nonregression based models [25,26]. Impact assessment investigates whether applying a model in clinical practice is better than standard care in terms of decision-making process and/or cost outcomes [10]. Without this crucial information, clinicians and policy makers will not be able to make informed decisions about adoption of CPMs [4,16]. Lack of guidance on the eligibility of CPMs for actual impact assessment leaves this a gray area in current practice. Currently, impact analysis is warranted only when a CPM is externally validated (at least once) and is ready for implementation [13,27,28]. Models that do not indicate potential impact (eg, net benefit, decision curve analysis), do not warrant an interventional impact study [23]. We recommend that the formulation of an intervention based on the CPM should already be considered at development and validation by researchers and stakeholders, to make sure the model being developed will be fit for purpose. Such considerations include feasibility, acceptability and integration of the model in the clinical workflow [10].

Little is known about the actual use of CPMs in clinical settings [8]. A study of models for prediction of clinical response to COVID-19 based on academic and gray literature reported that some models were used in hundreds of thousands of patients, while others were used to an unknown or limited extent [29]. The authors raised concern about the limited evidence supporting the use of these models. We also found that CPMs with a response to our survey were used in sites outside the development population which points to a wider range of use than an isolated

Table 1. Characteristics of CPM development articles and forward search results for the total cohort and by clinical utilization

Study characteristics	Total cohort (*N* = 109)	Used in clinical practice (*N* = 17)
CPM development publication		
Publication year, *median (IQR, range)*	2013 (2006–2018; 1995–2020)	2012 (2002–2017; 1997–2020)
Type of model, *n* (%)		
Prognostic	67 (61)	13 (76)
Diagnostic	42 (39)	4 (24)
Development centers, *n (%)*		
Multicenter	41 (38)	5 (29)
Monocenter	63 (58)	11 (65)
Not reported	5 (5)	1 (6)
Sample size, *median (IQR; range)*	305 (117–794; 20–82,359)	515 (293–1299; 67–2618)
Not reported, *n* (%)	3 (3)	-
Number of events, *median (IQR; range)*	76 (50–133; 7–6139)	76 (51–102; 45–136)
Not reported, *n* (%)	47 (43)	8 (47)
Missing data treatment reported, *n (%)*	41 (38)	6 (35)
Complete case analysis	33 (30)	6 (35)
Imputation	8 (7)	0
Calibration information reported, *n* (%)	31 (28)	4 (24)
Calibration plot	15 (14)	2 (12)
Hosmer-Lemeshow test	19 (17)	3 (18)
Other[a]	2 (2)	0
Discriminative performance reported, *n* (%)[b]	87 (80)	13 (76)
C-statistic reported, *n* (%)	65 (60)	7 (41)
C-statistic, *median (IQR; range)*	0.8 (0.73–0.97; 0.55–0.97)	0.75 (0.68–0.79; 0.55–0.85)
Final model presented, *n (%)*[c]	37 (34)	5 (29)
Other model presentations, *n (%)*	27 (25)	5 (29)
Internal validation in the article, *n (%)*	27 (25)	5 (29)
External validation in the article, *n (%)*	14 (13)	2 (12)
After CPM development publication		
External validation after development, *n* (%)	18 (17)	4 (24)
Articles per CPM, *median (IQR; range)*[d]	2 (1–2; 1–25)	3 (1–2; 1–25)
Overlapping investigators	8 (7)	2 (2)
Independent investigators	8 (7)	2 (2)
Both	2 (2)	0
Impact assessment/intervention, *n* (%)	1 (1)	0
Articles per CPM, *median (IQR; range)*[d]	2 (2–2; 2–2)	-
Guideline/consensus statement, *n* (%)	9 (8)	4 (24)
Articles per CPM, *median (IQR; range)*[d]	2 (1–2; 1–5)	2 (1–3; 1–5)
Positive recommendation	4 (4)	2 (12)
(Systematic) review/meta-analysis intervention, *n* (%)	35 (32)	5 (29)
Articles per CPM, *median (IQR; range)*[d]	1 (1–5; 1–23)	5 (1–2; 1–17)

CPM, clinical prediction model; IQR, interquartile range.

Data are median (IQR; range) for continuous variables and N (%) for categorical variables.

[a] Harrel's E, Chi-square goodness of fit test, observed to estimated ratios, calibration slope, tabulation of observed vs predicted risk.

[b] Includes C-statistic, sensitivity, specificity, negative predictive value and positive predictive value.

[c] Final presentation was defined as the model formula with intercept.

[d] Reported among CPMS with external validation, Impact assessment/intervention, guidelines/consensus mentions and (Systematic) reviews/meta-analysis after publication, accordingly.

clinician-investigator. This is also supported by clinical training offered, incorporation in decision aids, and mentions in clinical guidelines and consensus statements. However, the majority of models with clinical use were never externally validated and none had a published impact assessment. This suggests that the generalizability, harms, and benefits of most models used in clinical practice are unknown or unpublished. Our findings show that the practical

pathway to clinical utilization does not align with the ideal evidence-based pathway.

Our study benefits from use of a representative cohort of CPMs based on a validated search strategy across all medical fields and a prospective design and follow-up. However, there are limitations. The search strategy used for selection of our cohort was developed prior to the popularity of artificial intelligence models. Therefore, our study sample was not representative of CPMs using nonregression-based techniques. Validation, impact assessment and use of machine learning and AI models are relevant topics for future research. Second, we used citation searches in Scopus and Web of Science—with 20% to 54% reported sensitivities [30,31]. To maximize coverage, we searched both complementing databases [32] and manually checked references during screening to identify missed citations. Third, clinical use was assessed via a survey of CPM authors to ensure specificity and precision, though some authors may not know the extent of their model's use or may overstate uptake. In addition, authors of CPMs that are actively implemented may have been more likely to respond. CPMs with and without survey responses were similar in methodological quality, external validation, and C-statistics (see Supplementary Table 7). Future qualitative research is warranted to explore authors' perspectives on model evaluation and utilization, and to identify barriers and facilitators to evidence-based implementation.

We recommend that researchers adopt a targeted approach in development, external validations and impact assessments based on the intended settings for use. The intended context for use determines the appropriate methodology, including the target population and setting for validation [5] and impact assessment. In light of absence of supporting evidence for utilization in the intended settings, we advise that health-care professionals and other stakeholders abstain from using models, or use them with caution.

5. Conclusion

One in six clinical prediction models is externally validated after initial publication, but impact assessment is very rare. CPMs may be in use in clinical settings without published evidence of validity or positive impact supporting their use. To bring prediction research closer to clinical practice and ensure patient safety, targeted conduct and publication of external validations and impact assessments and careful interpretation of their findings are warranted.

CRediT authorship contribution statement

Banafsheh Arshi: Writing − original draft, Software, Methodology, Formal analysis, Conceptualization, Writing − review & editing, Visualization, Project administration, Investigation, Data curation. **Laura Elizabeth Cowley:** Writing − review & editing, Data curation, Investigation. **Eline Rijnhart:** Validation, Writing − review & editing, Data curation. **Kelly Reeve:** Data curation, Writing − review & editing. **Luc J. Smits:** Supervision, Project administration, Conceptualization, Writing − review & editing, Resources, Methodology. **Laure Wynants:** Writing − original draft, Project administration, Investigation, Conceptualization, Writing − review & editing, Supervision, Methodology, Data curation.

Declaration of competing interest

There are no competing interests for any author.

Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.jclinepi.2025.111902.

Data availability

Data will be made available on request.

References

[1] Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) the TRIPOD statement. Circulation 2015;131:211−9.
[2] 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.
[3] Cowley LE, Farewell DM, Maguire S, Kemp AM. Methodological standards for the development and evaluation of clinical prediction rules: a review of the literature. Diagn Progn Res 2019;3:1−23.
[4] Siontis GCM, Tzoulaki I, Castaldi PJ, Ioannidis JPA. External validation of new risk prediction models is infrequent and reveals worse prognostic discrimination. J Clin Epidemiol 2015;68:25−34.
[5] Arshi B, Smits LJ, Wynants L, Cowley LE, Reeve K, Rijnhart E. Number of publications on new clinical prediction models: a systematic literature search. JMIR Med Inform 2023;13:e62710.
[6] Wessler BS, Paulus J, Lundquist CM, Ajlan M, Natto Z, Janes WA, et al. Tufts PACE clinical predictive model registry: update 1990 through 2015. Diagn Progn Res 2017;1:1−8.
[7] Yang C, Kors JA, Ioannou S, John LH, Markus AF, Rekkas A, et al. Trends in the conduct and reporting of clinical prediction model development and validation: a systematic review. J Am Med Inform Assoc 2022;29:983−9.
[8] Reilly BM, Evans AT. Translating clinical research into clinical practice: impact of using prediction rules to make decisions. Ann Intern Med 2006;144:201−9.
[9] Moons KGM, Kengne AP, Grobbee DE, Royston P, Vergouwe Y, Altman DG, et al. Risk prediction models: II. External validation, model updating, and impact assessment. Heart 2012;98:691−8.
[10] Wallace E, Smith SM, Perera-Salazar R, Vaucher P, McCowan C, Collins G, et al. Framework for the impact analysis and implementation of Clinical Prediction Rules (CPRs). BMC Med Inform Decis Mak 2011;11:1−7.

[11] Collins GS, Dhiman P, Ma J, Schlussel MM, Archer L, Van Calster B, et al. Evaluation of clinical prediction models (part 1): from development to external validation. BMJ 2024;384:e074819.
[12] Ban J-W, Chan MS, Muthee TB, Paez A, Stevens R, Perera R. Design, methods, and reporting of impact studies of cardiovascular clinical prediction rules are suboptimal: a systematic review. J Clin Epidemiol 2021;133:111−20.
[13] Binuya MAE, Engelhardt EG, Schats W, Schmidt MK, Steyerberg EW. Methodological guidance for the evaluation and updating of clinical prediction models: a systematic review. BMC Med Res Methodol 2022;22:316.
[14] Wynants L, Van Calster B, 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.
[15] Wessler BS, Nelson J, Park JG, McGinnes H, Gulati G, Brazil R, et al. External validations of cardiovascular clinical prediction models: a large-scale review of the literature. Circ Cardiovasc Qual Outcomes 2021;14:e007858.
[16] Wallace E, Uijen MJM, Clyne B, Zarabzadeh A, Keogh C, Galvin R, et al. Impact analysis studies of clinical prediction rules relevant to primary care: a systematic review. BMJ Open 2016;6:e009957.
[17] Ingui BJ, Rogers MAM. Searching for clinical prediction rules in MEDLINE. J Am Med Inform Assoc 2001;8:391−7.
[18] Arshi B, Wynants L, Rijnhart E, Reeve K, Cowley LE, Smits LJ. What proportion of clinical prediction models make it to clinical practice? Protocol for a two-track follow-up study of prediction model development publications. BMJ open 2023;13:e073174.
[19] Knottnerus A, Tugwell P. STROBE—a checklist to strengthen the reporting of observational studies in epidemiology. J Clin Epidemiol 2008;61:323.
[20] 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. Syst Rev 2021;10:1−11.
[21] Van Calster B, McLernon DJ, Van Smeden M, Wynants L, Steyerberg EW. Calibration: the Achilles heel of predictive analytics. BMC Med 2019;17:1−7.
[22] Helmrich IRAR, Mikolić A, Kent DM, Lingsma HF, Wynants L, Steyerberg EW, et al. Does poor methodological quality of prediction modeling studies translate to poor model performance? An illustration in traumatic brain injury. Diagn Progn Res 2022;6:8.
[23] Steyerberg EW, Moons KGM, van der Windt DA, Hayden JA, Perel P, Schroter S, et al. Prognosis Research Strategy (PROGRESS) 3: prognostic model research. PLoS Med 2013;10:e1001381.
[24] Bonnett LJ, Snell KIE, Collins GS, Riley RD. Guide to presenting clinical prediction models for use in clinical settings. bmj 2019;365.
[25] Voets MM, Veltman J, Slump CH, Siesling S, Koffijberg H. Systematic review of health economic evaluations focused on artificial intelligence in healthcare: the tortoise and the cheetah. Value Health 2022;25:340−9.
[26] Lam TYT, Cheung MFK, Munro YL, Lim KM, Shung D, Sung JJY. Randomized controlled trials of artificial intelligence in clinical practice: systematic review. J Med Internet Res 2022;24:e37188.
[27] Toll DB, Janssen KJM, Vergouwe Y, Moons KGM. Validation, updating and impact of clinical prediction rules: a review. J Clin Epidemiol 2008;61:1085−94.
[28] Kappen TH, van Klei WA, van Wolfswinkel L, Kalkman CJ, Vergouwe Y, Moons KGM. Evaluating the impact of prediction models: lessons learned, challenges, and recommendations. Diagn Progn Res 2018;2:11.
[29] Mann S, Berdahl CT, Baker L, Girosi F. Artificial intelligence applications used in the clinical response to COVID-19: a scoping review. PLOS Digital Health 2022;1:e0000132.
[30] Linder SK, Kamath GR, Pratt GF, Saraykar SS, Volk RJ. Citation searches are more sensitive than keyword searches to identify studies using specific measurement instruments. J Clin Epidemiol 2015;68:412−7.
[31] Wright K, Golder S, Rodriguez-Lopez R. Citation searching: a systematic review case study of multiple risk behaviour interventions. BMC Med Res Methodol 2014;14:1−8.
[32] Burnham JF. Scopus database: a review. Biomed Digit Libr 2006;3:1−8.