BMC Medical Research Methodology

RESEARCH

Open Access

The reporting quality and methodological quality of dynamic prediction models for cancer prognosis

Peijing Yan[1,2†], Zhengxing Xu[3†], Xu Hui[4,5], Xiajing Chu[6], Yizhuo Chen[7], Chao Yang[1], Shixi Xu[8], Huijie Cui[1], Li Zhang[1], Wenqiang Zhang[1], Liqun Wang[9], Yanqiu Zou[1], Yan Ren[10], Jiaqiang Liao[1], Qin Zhang[11], Kehu Yang[4,5,12], Ling Zhang[13], Yunjie Liu[1], Jiayuan Li[1], Chunxia Yang[1], Yuqin Yao[9], Zhenmi Liu[14,16*], Xia Jiang[1,15,16*] and Ben Zhang[1,11,16*]

Abstract

Background To evaluate the reporting quality and methodological quality of dynamic prediction model (DPM) studies on cancer prognosis.

Methods Extensive search for DPM studies on cancer prognosis was conducted in MEDLINE, EMBASE, and the Cochrane Library databases. The Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) and the Prediction model Risk of Bias Assessment Tool (PROBAST) were used to assess reporting quality and methodological quality, respectively.

Results A total of 34 DPM studies were identified since the first publication in 2005, the main modeling methods for DPMs included the landmark model and the joint model. Regarding the reporting quality, the median overall TRIPOD adherence score was 75%. The TRIPOD items were poorly reported, especially the title (23.53%), model specification, including presentation (55.88%) and interpretation (50%) of the DPM usage, and implications for clinical use and future research (29.41%). Concerning methodological quality, most studies were of low quality ($n=30$) or unclear ($n=3$), mainly due to statistical analysis issues.

Conclusions The Landmark model and joint model show potential in DPM. The suboptimal reporting and methodological qualities of current DPM studies should be improved to facilitate clinical application.

Keywords Cancer prognosis, Dynamic prediction models, Reporting quality, Methodological quality, Methodological characteristics

†Peijing Yan and Zhengxing Xu contributed equally to this work.

*Correspondence:
Zhenmi Liu
zhenmiliu@scu.edu.cn
Xia Jiang
xiajiang@scu.edu.cn
Ben Zhang
wcsphwcfh@vip.163.com

Full list of author information is available at the end of the article

Introduction

The prognostic prediction model (commonly known as the "prognostic model") is an integral part of current clinical practice [1] and contributes valuable evidence for lifestyle modification and optimal therapeutic interventions [2, 3]. As global cancer prevalence grows, oncology remains a key area demanding useful prognostic models for patient prognosis and therapy guidance [4], as recommended by clinical guidelines such as National Comprehensive Cancer Network, the American Society of Clinical Oncology, and the European Society of Medical Oncology [5–7]. However, despite the rapid increase in the number of prognostic models for cancer patients in recent years [8], most are based on static data, which can only predict the prognosis at diagnosis, failing to reflect changes in patient survival probability during follow-up [9]. Furthermore, the effects of time-dependent covariates which may change over time are often overlooked in the static prognostic model. Thus, static models, outdated and inaccurate, are prone to calibration drift, which is one of the major pitfalls of using prognostic models in practice [10]. Worse still, the reporting quality and methodological quality of oncology predictive models were mostly poor [4, 11, 12], providing insufficiently accurate predictions [13]. These drawbacks collectively impede the applicability of prognostic models in clinical practice.

The dynamic prediction model (DPM) is a novel approach to addressing the problem of inaccuracy and calibration drift in static model [14]. The DPM acknowledges the real-time of each point and the time-dependent effect of prognostic factors, which are designed to evolve over time [15–17]. Thus, the prognosis prediction can be computed throughout cancer patient follow-up using dynamic models that consider the time elapsed from diagnosis or treatment, the event history, and the time-dependent effect of prognostic factors [15–17]. DPM studies on cancer prognosis have been increasingly published in recent years [18–20]. However, whether the reporting quality and methodological quality of the current DPMs for cancer prognosis have improved and the methodological characteristics of the novel models remain unclear, which affect the future development of DPMs and their applicability in clinical practice.

Therefore, this study aimed to systematically evaluate the reporting quality and methodological quality of DPM studies on cancer prognosis as well as summarize the methodological characteristics. This study is expected to serve as a useful reference of current status and help improve future development and application of DPM.

Materials and methods

The protocol for this study is publicly available on the Open Science Framework (OSF) at DOI: https://doi.org/10.17605/OSF.IO/Y7DGZ. All stages of the literature review, including study selection, appraisal, and data abstraction, were conducted in accordance with the protocol. However, we amended the protocol by introducing a test to compare overall TRIPOD adherence scores across different study types and publication years, aiming to explore potential variations in TRIPOD adherence scores across these categories.

Literature search

We systematically searched the Ovid MEDLINE, Ovid EMBASE, and the Cochrane Library databases from inception to November 11, 2021. No language restrictions were imposed. The relevant keywords used were based on search filters published from the Search Filters Resource website of the InterTASC Information Specialists' Sub-Group [21], Cochrane reviews [11, 22] of prediction models, and DPM reviews [14]. The detailed search strategies are shown in Table S1. The reference list of all included studies and relevant reviews were also manually checked to identify extra studies.

Study selection

Studies that aimed to develop, validate or compare a DPM for cancer prognosis were included. Studies were excluded if they were: (1) conference abstracts, reviews (including meta-analyses), comments, letters, protocols, or non-human studies; (2) articles published in non-English or non-Chinese languages; or (3) methodological studies emphasizing research methodology rather than clinical practice. When the same study provided more than one DPM for multiple outcomes, we selected the DPM corresponding to the primary outcome with the largest event number.

Two authors independently screened the titles and abstracts to exclude ineligible studies, followed by full-text assessment. The reasons for exclusion in full-text screening were documented. Any disagreement was resolved through discussion.

Data collection

Data collection consisted of four major sections: general characteristics, methodological characteristics, reporting quality, and methodological quality of the DPM studies. Four authors were trained to collect the data based on a predefined electronic data extraction form containing the items in the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) [23, 24], the Prediction model Risk of Bias Assessment Tool (PROBAST) [25, 26]. The form was also based on the Checklist for Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modeling Studies checklist [27]. The trained authors were divided into two groups for independent information extraction. Any discrepancies were resolved by

discussion. For studies that aimed to validate a previously published prediction model, we also retrieved information from the original papers, including comprehensive details of the model development, such as the full model (which presents the complete prediction model, including all regression coefficients and the model intercept or baseline survival at a given time point) specifications. Detailed information on the extraction is available in Text S1.

Study quality assessment

The TRIPOD was used to assess the reporting quality of each included study. The TRIPOD contains 22 main criteria with 37 items in six sections. The reporting quality was assessed at study level and item level, respectively. At study level, the reporting quality was quantified by calculating the overall TRIPOD adherence score, which was defined as the number of TRIPOD items with "yes" responses divided by the total number of TRIPOD items applicable to the particular study [28]. At specific item level, the reporting quality of was measured by calculating the overall adherence per TRIPOD item (the number of studies that adhered to a specific TRIPOD item divided by the number of studies in which the specific TRIPOD item was applicable) [28].

The PROBAST tool was designed to facilitate the assessment of risk of bias and provides an evaluation of methodological quality for studies reporting on the development, validation, or update of prediction models [29]. It has been widely utilized for assessing study methodological quality [29, 30]. The PROBAST contains 20 signaling questions in four domains: participants, predictors, outcomes, and statistical analyses. The answers to the signaling questions contained yes (Y), probably yes (PY), no (N), probably no (PN), or no information (NI). Each domain and the overall risk of bias were rated as having a low, high, or unclear risk of bias.

Statistical analysis

Statistical analyses and figure preparation were performed using R, version 4.1.2. The data for continuous variables are reported either as the mean ± SD or the median with interquartile range (IQR) according to normality tests (Kolmogorov-Smirnov and Shapiro-Wilk). Categorical variables are expressed as frequencies and percentages. The Kruskal-Wallis rank test and two-sample Kolmogorov-Smirnov test were used to compare the overall TRIPOD adherence scores among study types. The Mann-Whitney U was used to compare the overall TRIPOD adherence scores between study types. The false discovery rate was used to correct for multiple comparisons. Spearman's rank correlation was used to assess the association between the reporting quality and methodological quality. A two-tailed α level of 0.05 was used for all statistical tests.

Results

Study selection

Of the 19,170 publications identified through searching, 15,740 were eligible for title and abstract reviews, and 15,398 publications were further excluded for being irrelevant. We then assessed 342 full texts, and 34 studies were finally included (Fig. 1). The reasons for the full-text exclusions are provided in Text S2, and the list of included studies is provided in Text S3.

Characteristics of included studies

The annual trend in the number of publications on DPMs for cancer prognosis is shown in Figure S1. The first study was published in 2005, and the number of DPM studies increased rapidly in 2018 and has been steadily increasing since then. Among the included DPM studies, eight studies were developed without internal validation, and 17 studies were developed with internal validation. The remaining nine studies were developed with external validation, of which one study contained two external validation sets [31] (Table 1 and Table S2).

The patients were mainly recruited from Europe, North America, and Asia. Most of the included studies ($n = 30$, 88.2%) were retrospective cohorts, and the data were mainly from hospitals. The studies included 19 cancer types, mainly focusing on soft tissue sarcomas ($n = 4$), breast cancer ($n = 4$), gastrointestinal cancer ($n = 3$), and prostate cancer ($n = 3$). The most common primary outcome was overall survival ($n = 23$, 67.6%), with follow-up durations ranging from 1.1 year to over 21 years. Other outcomes included disease-specific survival, progression-free survival, thromboembolism, relapse-free survival, and infection rates. The studies utilized a broad range of candidate predictors, including demographic details, clinical characteristics, treatment approaches, and biological markers, which are crucial for understanding the context of the models developed (Table 1 and Table S2).

Methodological characteristics

The methodological characteristics of the included studies are summarized in Fig. 2 and Table S3. Most DPM studies ($n = 24$) considered time-dependent variables in the model development. The landmark ($n = 12$), joint ($n = 6$), time-dependent Cox ($n = 6$), and conditional survival model ($n = 6$) were the most common modeling methods. All but five of the studies reported discrimination evaluations, mainly using C-statistic ($n = 16$), the area under the curve ($n = 10$), and Brier scores ($n = 5$). More than 35% of the studies ($n = 12$) did not report whether model calibration evaluation was performed. Of the remaining 22 studies, 12 used calibration plots, five

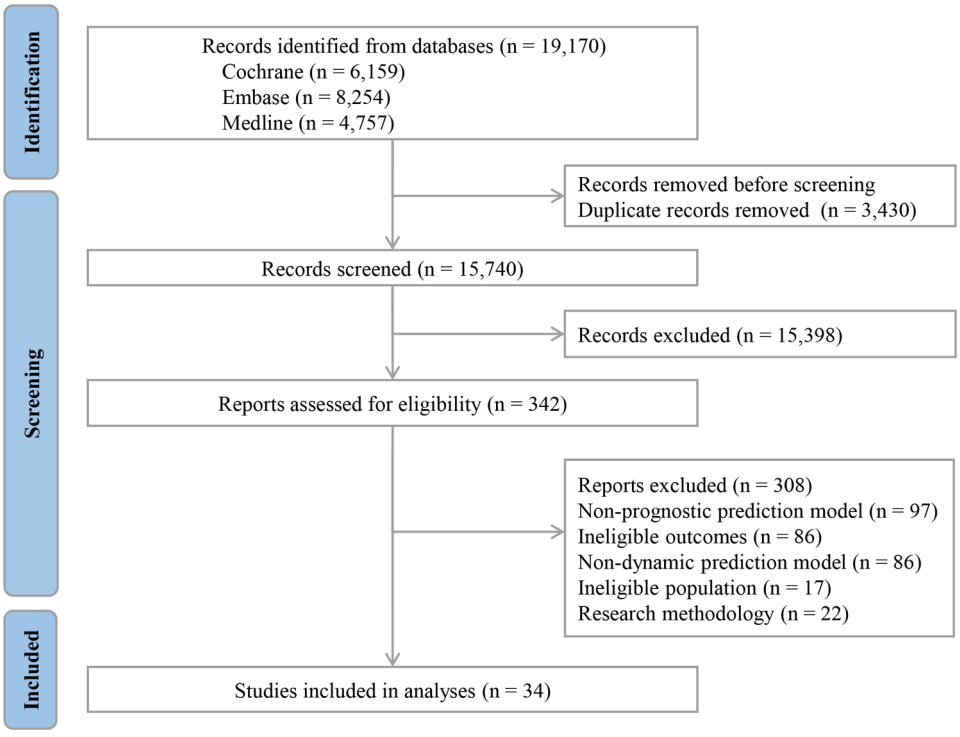

Fig. 1 Flow chart of study selection

used heuristic shrinkage factors, and the other five used Brier scores. Regarding internal validation of the model, 11 studies did not report, three studies used random splits, and the remaining 58.8% used cross-validation ($n = 12$) or bootstrapping ($n = 8$). All studies except one [32] reported the presentation format of models. The most common form was the full model ($n = 19$), followed by nomograms ($n = 11$) and web calculators ($n = 3$). However, only one study reported both the full model and a visual model (web calculator) [33].

Reporting quality

At study level, the overall TRIPOD adherence score of the 34 studies ranged from 41.38 to 85.71%, with a median of 75% (Table S4). The adherence scores of studies with development that included external validation (median, 78.12%; IQR, 11.55%) were higher than those of development-only studies. As for development-only studies, those with internal validation (median, 75.00%; IQR, 10.71%) had better adherence scores than those without internal validation (median, 70.20%; IQR, 6.5%). However, no significant difference was found in the adherence scores among study types (Table S4 and Fig. 3). Four studies [9, 19, 34, 35] reporting that the TRIPOD was considered or followed (median, 81.98%; IQR, 1.65%) had higher and more stable overall adherence scores than the other studies (median, 72.87%; IQR, 9.85%). Additionally, a total of 28 studies published after the release of TRIPOD (since 2015) showed higher adherence scores, with a median of 75.43% compared to 59.67% for earlier studies (Table S4). This trend continued with 20 studies published after the release of both TRIPOD and PROBAST (since 2019), where a median of 76.50% versus 67.25% for studies published before (Table S4) was observed. Similarly, 9 studies published since 2018 showed higher overall adherence scores, with a median of 75.86% compared to 65.52% for those published earlier (Table S4).

At item level, the overall adherence per TRIPOD item is shown in Fig. 4 and Table S5. Particularly problematic areas were title (item 1, 23.53%), model interpretation consisting of clinical and research implications (item 20, 29.41%) and explanation of the DPM usage (item 15b, 50%), and full model presentation (item 15a, 55.88%). Apart from this, other common poorly reported items were sample size justification (item 8, 8.82%) and details of patient treatment received (item 5c, 32.35%).

Methodological quality

Figure 5 summarizes the methodological quality of the included studies. Only one study (2.94%) was of high quality, three studies (8.82%) were unclear, and the remaining 30 studies (88.24%) were of low quality. The high-quality study updated and externally validated the DPM for patients with soft tissue extremity sarcoma using the landmark model [35].

At domain level, 33 studies were rated as high quality for the participants, predictors, and outcome domains,

Table 1 Characteristics of the included studies by study type

Characteristics	Study types, *n* (%)		
	D-IV (*n* = 8)	D (*n* = 17)	D + EV (*n* = 9)
Participants			
Mean age, years, median (IQR)	63.7 (59.0, 64.2)	60.9 (47.0, 63.0)	D*: 60.0 (59.4, 66.0) EV: 60.7 (57.4, 62.5)
Male sex, %, median (IQR)	76.5 (0.0, 87.2)	54.2 (0.0, 72.6)	D*: 56.9 (52.8, 62.3) EV: 55.3 (52.6, 75.9)
Median follow-up duration, Years, median (IQR)	8.6 (5.6, 13.2)	4.9 (2.8, 8.5)	D*: 6.0 (5.3, 6.6) EV: 5.1 (2.5, 6.9)
Sample size	33,409	24,557	D*: 37,295; EV: 40,273
Publication year			
< 2005	0 (0.0)	1 (5.9)	0 (0.0)
2006–2010	0 (0.0)	0 (0.0)	1 (11.1)
2011–2015	2 (25.0)	3 (17.7)	1 (11.1)
2016–2020	6 (75.0)	9 (52.9)	3 (33.3)
2021	0 (0.0)	4 (23.5)	4 (44.4)
Study designs			
Retrospective cohort	5 (62.5)	17 (100.0)	8 (88.9)
Prospective cohort	3 (37.5)	0 (0.0)	1 (11.1)
Median follow-up duration			
≤ 10 years	5 (62.5)	12 (70.6)	D*: 7 (77.8); EV: 7 (70.0)
> 10 years	3 (37.5)	2 (11.8)	D*: 2 (22.2); EV: 1 (10.0)
Not specified	0 (0.0)	3 (17.7)	D*: 0 (0.0); EV: 2 (20.0)
Regions			
Europe	2 (25.0)	6 (35.3)	D*: 3 (33.3); EV: 5 (50.0)
North America	3 (37.5)	4 (23.5)	D*: 2 (22.2); EV: 0 (0.0)
Asia	1 (12.5)	6 (35.3)	D*: 1 (11.1); EV: 3 (30.0)
Oceania	0 (0.0)	0 (0.0)	D*: 1 (11.1); EV: 1 (10.0)
Multiple countries	1 (12.5)	0 (0.0)	D*: 2 (22.2); EV: 1 (10.0)
Not specified	1 (12.5)	1 (5.9)	D*: 0 (0.0); EV: 0 (0.0)
Data Sources			
Hospital	6 (75.0)	11 (64.7)	D*: 5 (55.6); EV: 7 (70.0)
Registry data	1 (12.5)	6 (35.3)	D*: 4 (44.4); EV: 3 (30.0)
Cohort	1 (12.5)	0 (0.0)	D*: 0 (0.0); EV: 0 (0.0)
Cancer types			
Soft tissue sarcomas	0 (0.0)	1 (5.9)	D*: 3 (33.3); EV: 3 (30.0)
Breast cancer	1 (12.5)	2 (11.8)	D*: 1 (11.1); EV: 2 (20.0)
Gastrointestinal cancer	1 (12.5)	0 (0.0)	D*: 1 (11.1); EV: 2 (20.0)
Prostate cancer	1 (12.5)	2 (11.8)	D*: 0 (0.0); EV: 0 (0.0)
Non-small cell lung cancer	0 (0.0)	1 (5.9)	D*: 1 (11.1); EV: 0 (0.0)
Nasopharyngeal carcinoma	0 (0.0)	1 (5.9)	D*: 1 (11.1); EV: 1 (10.0)
Cervical cancer	0 (0.0)	2 (11.8)	D*: 0 (0.0); EV: 0 (0.0)
Hematological malignancies	0 (0.0)	2 (11.8)	D*: 0 (0.0); EV: 0 (0.0)
Hepatocellular carcinoma	1 (12.5)	1 (5.9)	D*: 0 (0.0); EV: 0 (0.0)
Ovarian Cancer	1 (12.5)	1 (5.9)	D*: 0 (0.0); EV: 0 (0.0)
Myelodysplastic syndromes	0 (0.0)	0 (0.0)	D*: 1 (11.1); EV: 1 (10.0)
Esophageal cancer	0 (0.0)	0 (0.0)	D*: 1 (11.1); EV: 1 (10.0)
Primary central nervous system lymphoma	0 (0.0)	1 (5.9)	D*: 0 (0.0); EV: 0 (0.0)
Ewing sarcoma	0 (0.0)	1 (5.9)	D*: 0 (0.0); EV: 0 (0.0)
Head and neck squamous cell carcinoma	0 (0.0)	1 (5.9)	D*: 0 (0.0); EV: 0 (0.0)
Kidney renal clear cell carcinoma	0 (0.0)	1 (5.9)	D*: 0 (0.0); EV: 0 (0.0)
Bladder Cancer	1 (12.5)	0 (0.0)	D*: 0 (0.0); EV: 0 (0.0)
Laryngeal Cancer	1 (12.5)	0 (0.0)	D*: 0 (0.0); EV: 0 (0.0)
Metastatic colorectal cancer	1 (12.5)	0 (0.0)	D*: 0 (0.0); EV: 0 (0.0)
Primary outcomes			

Table 1 (continued)

Characteristics	Study types, *n* (%)		
	D-IV (*n* = 8)	D (*n* = 17)	D + EV (*n* = 9)
Overall survival	5 (62.5)	10 (58.8)	8 (88.9)
Progression-free survival	0 (0.0)	4 (23.5)	0 (0.0)
Thromboembolism	0 (0.0)	0 (0.0)	1 (11.1)
Biochemical recurrence-free	0 (0.0)	1 (5.9)	0 (0.0)
Relapse-free survival	0 (0.0)	1 (5.9)	0 (0.0)
Disease-specific survival	1 (12.5)	0 (0.0)	0 (0.0)
Progression	1 (12.5)	0 (0.0)	0 (0.0)
Systemic progression	1 (12.5)	0 (0.0)	0 (0.0)
The infection rate	0 (0.0)	1 (5.9)	0 (0.0)

Fig. 2 Methodological characteristics of dynamic prediction models for cancer prognosis. AUC, area under the curve

but only one study (2.94%) was high-quality for the statistical analysis domain.

Concerning the signaling question level, low quality was mainly caused by the inappropriate handling of predictors (N/PN, 35.29%) and using univariable analysis to select predictors before multivariable modelling (N/PN, 41.18%; NI, 23.53%). Inadequate handling of missing data (N/PN, 44.12%; NI, 29.41%), inappropriate evaluation of the model's performance (N/PN, 44.12%), and the lack

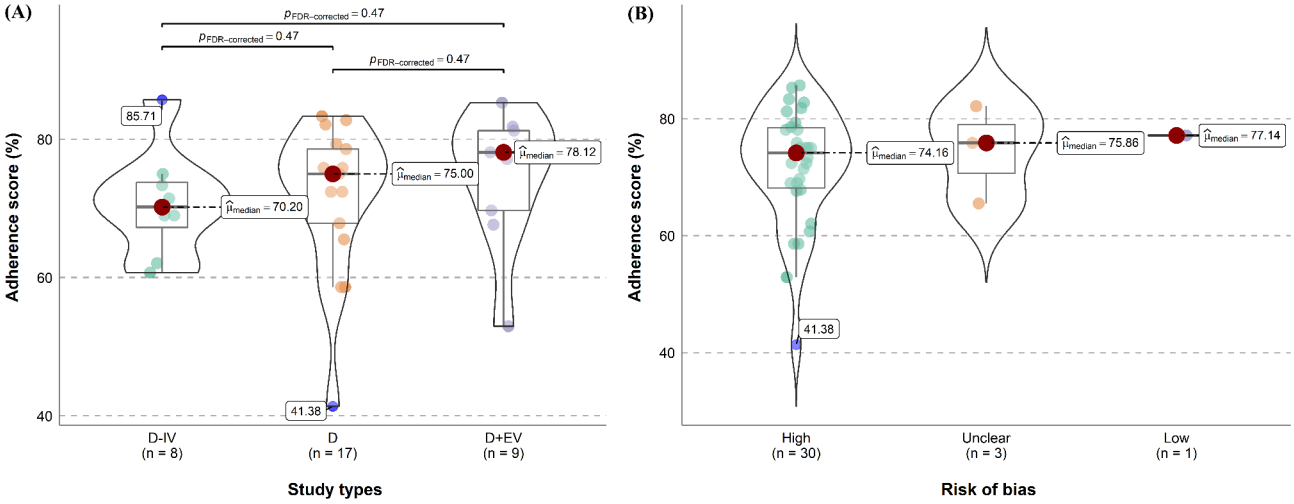

Fig. 3 The overall TRIPOD adherence score of individual studies for (**A**) different study types and (**B**) different risks of bias. D-IV, development without internal validation; D, development with internal validation; D + EV, development with external validation. FDR, false discovery rate

of adjustment for overfitting (N/PN, 44.12%) were also common reasons (Table S6).

The correlation between methodological quality and reporting quality

Figure 3 and Table S4 show the overall TRIPOD adherence scores for studies with different methodological quality. The adherence score of one high-quality study (77.14%) was slightly higher than studies with unclear risk of bias (median, 75.86%; IQR, 8.31%), while the studies with unclear risk of bias had a higher adherence score than low-quality studies (median, 74.17%; IQR, 10.32%). However, no correlation was detected between the two qualities (r_s = -0.11, P = 0.536).

Discussion

Principal findings

The DPMs included in this review primarily used the landmark model and the joint model to dynamically predict the overall survival of cancer patients. However, most DPMs were of poor-quality owing to (1) suboptimal reporting of the model presentation format, explanation, and title, and (2) poor methodology for selecting and handling predictors, in addition to the familiar problems of reporting and developing predictive models (e.g., sample size, missing data, and model performance).

Study implications for research and practice

This study found that the reporting quality of DPM has improved over time, especially after the publication of TRIPOD and PROBAST. However, DPMs still had a poor model presentation, interpretation, and titling, making them difficult to apply in practice. Approximately 60% of the DPMs lacked manipulatable presentation formats, 50% lacked explanations on how to use the model, and over 70% failed to provide a proper title (as recommended by TRIPOD), which impeded model operation and information acquisition for users. Presentation formats that are easy to understand and manipulate (e.g., nomograms, websites, etc.) based on the intended users, timing, and settings of use, supplied with clear instructions, will be highly valuable [36]. We also strongly recommend reporting the full regression model, which would benefit the quick expansion and updating of approved models. Moreover, to further understand and promote the application of DPMs, researchers are encouraged to give a working example of the DPM and discuss the potential clinical use and implications for future research of the DPM [24].

It is important to recognize that different modeling methods can influence modeling results [37]. Consequently, researchers should take into account the characteristics of various modeling methods when making design and selections [38]. Among the included DPMs, there was a noticeable preference for landmark and joint models. Landmark models are statistical approaches that evaluate the effect of time-dependent covariates by considering a fixed time point, known as the "landmark time", after which the risk of an event is reassessed. These models are particularly useful when the risk of an event changes over time and can be influenced by events that occur after the landmark time [39]. Joint models, on the other hand, combine the analysis of longitudinal and time-to-event data within a single statistical framework. They are designed to handle the complex relationships between repeated measurements and the time-to-event outcomes, providing a comprehensive view of the data structure [40]. Generally, time-dependent Cox, landmark, and joint models outperformed the conditional survival (CS) model when time-dependent covariates were

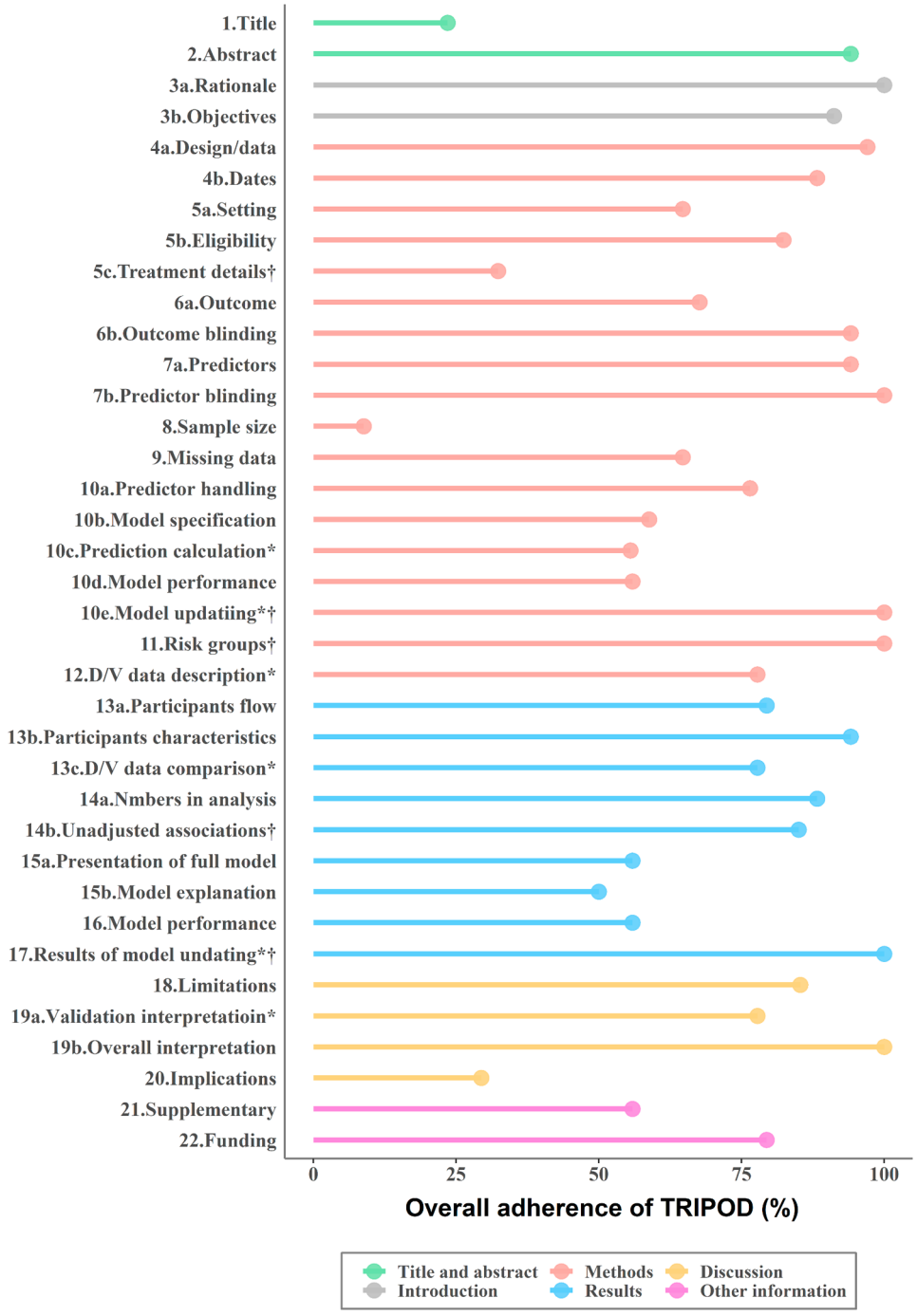

Fig. 4 Overall adherence per TRIPOD item. *items not applicable for a development study; †items might not be applicable for a specific study

considered as potential predictors in time-to-event data, although the CS model provided more accurate information about the patient's prognosis, especially when the patient exceeded a pre-specified landmark survival time [41, 42]. Compared to time-dependent Cox models, the landmark models were more transparent, especially in the present context of a binary time-dependent covariate [39], while the joint models provided a better fit and were more robust [43]. The landmark model was more widely used than the joint model probably because it does not require the specification of a longitudinal model [44] and is easier to implement [45, 46]. So far, there is no clear conclusion on the comparison of the predictive performance of these two approaches [44, 45, 47]. More studies are needed to systematically compare them to facilitate

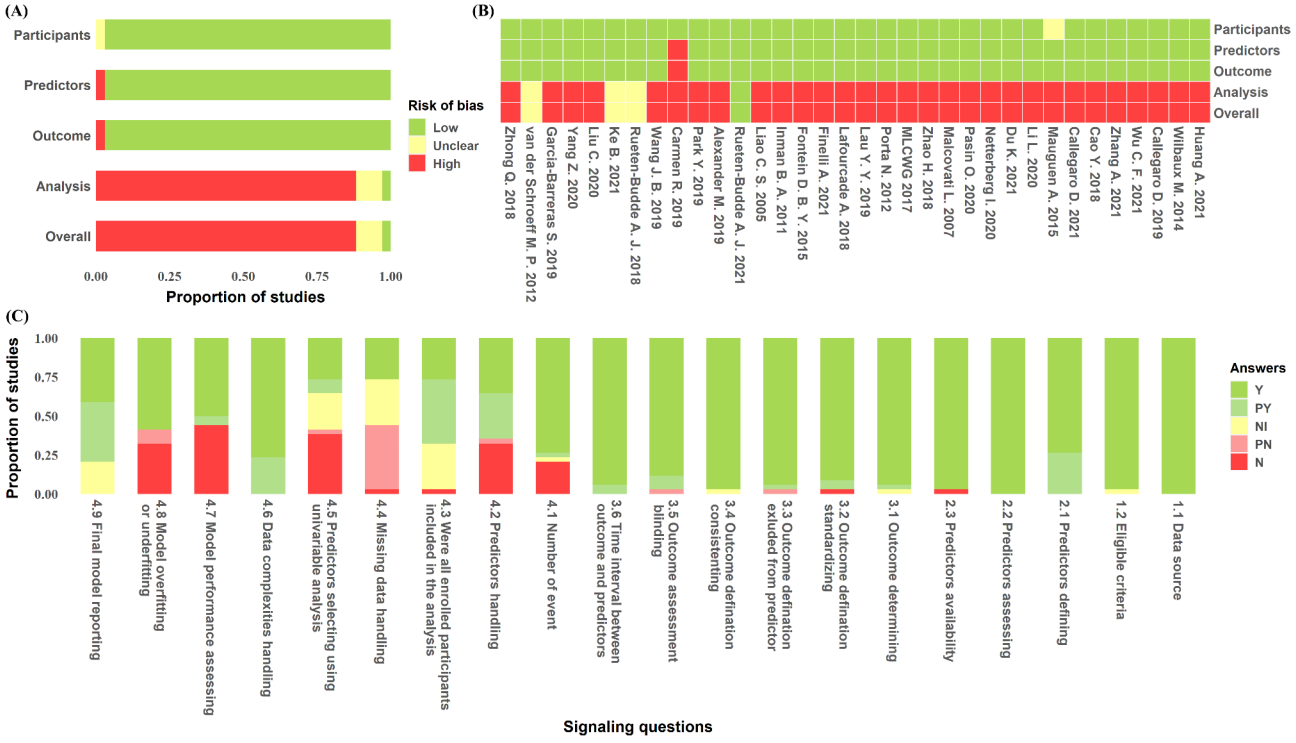

Fig. 5 Risk of bias assessment of the included studies. (**A**) The proportion of the risk of bias across all included studies for four domains and overall, (**B**) the risk of bias for each included study, and (**C**) the proportion of the risk of bias types across all included studies for each signaling question. MLCWG, Multidisciplinary Larynx Cancer Working Group; Y, yes; PY, probably yes; NI, no information; PN, probably no; N, no

researchers' choices in the development and validation of DPMs.

Finally, we suggest that the selection and handling of DPM predictors be improved in future studies. About 24% of the DPMs failed to report how the predictors were selected, and >33% selected predictors based solely on univariate analysis. This practice can lead to the model overfitting, especially when combined with small sample sizes or collinearity among variables [27, 48]. Thus, additional information, such as literature, expert opinions, clinical knowledge, reliability, consistency, applicability, usability, and cost of the predictor's measurements, should be considered [26, 49]. Furthermore, more than half of the DPM studies converted continuous predictors into binary or multi-categorical variables for modeling, which led to information loss and reduced predictive performance [50]. Hence, it is not recommended to convert continuous predictors to categorical variables unless the predictor has a solid, constant risk of the outcome over a range of values; and nonlinear relationships could be explored [51].

Comparison with other studies

To our knowledge, no previous study has specifically evaluated the reporting and methodological quality of DPMs. Consequently, we have benchmarked our findings against those from studies on static predictive models.

Regarding reporting quality, we identified significant shortcomings in how DPMs report certain aspects of the TRIPOD guidelines, particularly in the title (item 1), which was adequately addressed in less than 25% of the models. This aligns with a previous article [52], indicating a broader issue with the reporting of prognostic model development in published literature. There is a clear need for researchers to employ more rigorous methods and to enhance the reporting of their studies.

Furthermore, our study observed an improvement in the reporting of full models compared to earlier studies [53, 54]. The adoption of innovative methodologies, such as landmark and joint models, appears to contribute positively to this improvement in reporting full models to some extent.

Regarding methodological quality, our research revealed that the methodological quality of current DPMs in cancer prognosis is suboptimal, particularly in areas such as model performance evaluation, handling of missing data, and selection of predictors. These findings align with the methodological shortcomings identified in previous reviews of static predictive models for cancer prognosis [52, 55, 56].

Strengths and limitations

To our knowledge, this is the first investigation to assess the reporting quality and methodological quality of DPM

studies and summarize their methodological characteristics. Our findings could provide a valuable reference for the development and application of DPMs. Nevertheless, the DPM is a newly emerged field, and thus, only 34 current DPM studies on cancer prognosis (except the methodological studies) were identified, so future improvements are needed.

Several limitations in this study should be noted. First, currently no commonly accepted reporting or methodological tools for DPM studies are available, so we used TRIPOD and PROBAST, which are designed for predictive models. Nearly all items in the assessment tools were applicable for DPM studies, although some DPM characteristics were not considered in the existing assessment tools. For example, the report should be identified as a DPM in the title, explain why dynamic predictions are made, how dynamically changing variables (e.g., repeated measurement data and time-dependent covariates) are measured, and how such variables are handled in the model. Second, to better evaluate the qualities of DPM studies, non-clinical methodology research was excluded, which may have affected the characterization of the methodological characteristics. Third, due to the current status of the DPM field, it is highly time-consuming to search for related reports, and certain DPM reports may have been missed. In summary, the DPM is a highly valuable prognostic tool with great potential. However, establishing well-acknowledged modeling and reporting guidance and developing relevant tools for DPMs are urgently needed.

Conclusion

DPM studies on cancer prognosis involving the time-dependent effects of predictors or repeated measurement data showed great potential, especially those based on landmark and joint models. However, the reporting quality and methodological quality are suboptimal. An informative title and full model equation should always be presented to allow the user to retrieve and correctly validate the model, and the manipulatable presentation and interpretation of DPMs are urgently needed for their application. The predictors should be properly selected and handled to enhance the quality of DPMs.

Abbreviations

DPM(s)	Dynamic prediction model(s)
TRIPOD	Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis
PROBAST	Prediction model Risk Of Bias Assessment Tool
NCCN	National Comprehensive Cancer Network
ASCO	American Society of Clinical Oncology
ESMO	European Society of Medical Oncology
OSF	Open Science Framework
CHARMS	Checklist for Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies
Y	Yes
PY	Probably yes
N	No
PN	Probably no
NI	No information
SD	Standard deviation
IQR	Interquartile range
D-IV	Development without internal validation
D	Development with internal validation
D + EV	Development with external validation
AUC	Area under the curve
CS	Conditional survival
EPV	Events per variable

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12874-025-02516-2.

Supplementary Material 1: Figure S1. The annual trend of dynamic prediction model studies on cancer prognosis. Text S1. Information on extracted data of included studies. Text S2. Excluded reasons for the full text. Table S1. Search strategies. Table S2. Characteristics of the included studies. Table S3. Methodological characteristics of dynamic prediction models for cancer prognosis. Table S4. Overall TRIPOD adherence score. Table S5. The overall adherence per TRIPOD item. Table S6. The risk of bias across all included studies for each signaling question. Table S7. PRISMA 2020 checklist.

Acknowledgements

Not applicable.

Author contributions

B Z, X J, Z L, P Y and Z X conceive and designed the study; P Y and Z Xu conducted the literature searching; P Y, Z X, X H, X C and Y C screened the articles and extracted the data; P Y, Z X and X H analyzed the data; all authors contributed to the interpretation of data. P Y, Z X, X H, X C drafted the manuscript; all authors contributed to the advanced draft and revise the manuscript; all authors read and approved the final manuscript.

Funding

This work was supported by the National Key R&D Program of China (2022YFC3600600, 2022YFC3600604), the National Natural Science Foundation of China (U22A20359, 81874283, 81673255), the Recruitment Program for Young Professionals of China, the Science Fund for Creative Research Groups of Science and Technology Bureau of Sichuan Province (2024NSFT0030), the Promotion Plan for Basic Medical Sciences and the Development Plan for Cutting-Edge Disciplines, Sichuan University, and other Projects from West China School of Public Health and West China Fourth Hospital, Sichuan University. The funder had no role in the study design, data analysis and preparation of the manuscript.

Data availability

No datasets were generated or analysed during the current study.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

[1]Department of Epidemiology and Biostatistics, Institute of Systems Epidemiology and West China-PUMC C. C. Chen Institute of Health, West China School of Public Health and West China Fourth Hospital, Sichuan University, Chengdu, Sichuan, China

[2]Clinical Research Center, Sichuan Provincial People's Hospital, University of Electronic Science and Technology of China, Chengdu, Sichuan, China
[3]Department of Epidemiology and Health Statistics, School of Public Health, Southwest Medical University, Luzhou, Sichuan, China
[4]Evidence Based Social Science Research Center, School of Public Health, Lanzhou University, Lanzhou, Gansu, China
[5]Evidence-Based Medicine Center, School of Basic Medical Sciences, Lanzhou University, Lanzhou, Gansu, China
[6]Department of Health Research Methods, Evidence & Impact, McMaster University, Hamilton, ON, Canada
[7]The Second Clinical Medical Hospital, Lanzhou University, Lanzhou, Gansu, China
[8]Department of Preventive Medicine, West China School of Public Health and West China Fourth Hospital, Sichuan University, Chengdu, Sichuan, China
[9]Department of Hygienic Toxicology, West China School of Public Health and West China Fourth Hospital, Sichuan University, Chengdu, Sichuan, China
[10]Chinese Evidence-based Medicine Center, West China Hospital, Sichuan University, Chengdu, Sichuan, China
[11]Department of Occupational and Environmental Health, West China School of Public Health and West China Fourth Hospital, Sichuan University, Chengdu, Sichuan, China
[12]Key Laboratory of Evidence Based Medicine and Knowledge Translation of Gansu Province, Lanzhou, Gansu, China
[13]Department of Iatrical Polymer Material and Artificial Apparatus, School of Polymer Science and Engineering, Sichuan University, Chengdu, Sichuan, China
[14]Department of Maternal, Child and Adolescent Health, West China School of Public Health and West China Fourth Hospital, Sichuan University, Chengdu, China
[15]Department of Nutrition and Food Hygiene, West China School of Public Health and West China Fourth Hospital, Sichuan University, Chengdu, Sichuan, China
[16]West China School of Public Health and West China Fourth Hospital, Sichuan University, No. 16, Section 3, South Renmin Road, Wuhou District, Chengdu 610041, China

Received: 27 January 2024 / Accepted: 20 February 2025
Published online: 01 March 2025

References

1. Steyerberg EW, Moons KG, van der Windt DA, Hayden JA, Perel P, Schroter S, Riley RD, Hemingway H, Altman DG, Group P. Prognosis research strategy (PROGRESS) 3: prognostic model research. PLoS Med. 2013;10(2):e1001381. https://doi.org/10.1371/journal.pmed.1001381.
2. Moons KG, Royston P, Vergouwe Y, Grobbee DE, Altman DG. Prognosis and prognostic research: what, why, and how? BMJ (Clinical Res ed). 2009;338(b375). https://doi.org/10.1136/bmj.b375.
3. Beulens JWJ, Yauw JS, Elders PJM, Feenstra T, Herings R, Slieker RC, Moons KGM, Nijpels G, van der Heijden AA. Prognostic models for predicting the risk of foot ulcer or amputation in people with type 2 diabetes: a systematic review and external validation study. Diabetologia. 2021;64(7):1550–62. https://doi.org/10.1007/s00125-021-05448-w.
4. Dhiman P, Ma J, Navarro CA, Speich B, Bullock G, Damen JA, Kirtley S, Hooft L, Riley RD, Van Calster B, et al. Reporting of prognostic clinical prediction models based on machine learning methods in oncology needs to be improved. J Clin Epidemiol. 2021;138:60–72. https://doi.org/10.1016/j.jclinepi.2021.06.024.
5. Streiff MB, Holmstrom B, Angelini D, Ashrani A, Bockenstedt PL, Chesney C, Fanikos J, Fenninger RB, Fogerty AE, Gao S, et al. NCCN guidelines insights: Cancer-Associated venous thromboembolic disease, version 2.2018. J Natl Compr Cancer Netw J Natl Compr Canc Netw. 2018;16(11):1289–303. https://doi.org/10.6004/jnccn.2018.0084.
6. Lyman GH, Bohlke K, Khorana AA, Kuderer NM, Lee AY, Arcelus JI, Balaban EP, Clarke JM, Flowers CR, Francis CW, et al. Venous thromboembolism prophylaxis and treatment in patients with cancer: American society of clinical oncology clinical practice guideline update 2014. J Clin Oncol. 2015;33(6):654–6. https://doi.org/10.1200/JCO.2014.59.7351.
7. Mandalà M, Falanga A, Roila F. Management of venous thromboembolism (VTE) in cancer patients: ESMO clinical practice guidelines. Ann Oncol. 2011;22:vi85–92. https://doi.org/10.1093/annonc/mdr392.
8. de Jong Y, Ramspek CL, Zoccali C, Jager KJ, Dekker FW, van Diepen M. Nephrol (Carlton Vic). 2021;26(12):939–47. https://doi.org/10.1111/nep.13913. Appraising prediction research: a guide and meta-review on bias and applicability assessment using the Prediction model Risk Of Bias ASsessment Tool (PROBAST).
9. Du K, Li L, Wang Q, Zou J, Yu Z, Li J, Zheng Y. Development and application of a dynamic prediction model for esophageal cancer. Annals Translational Med 2021, 9(20):1546. https://doi.org/10.21037/atm-21-4964
10. Davis SE, Greevy RA Jr., Lasko TA, Walsh CG, Matheny ME. Detection of calibration drift in clinical prediction models to inform model updating. J Biomed Inf. 2020;112:103611. https://doi.org/10.1016/j.jbi.2020.103611.
11. Kreuzberger N, Damen J, Trivella M, Estcourt LJ, Aldin A, Umlauff L, Vazquez-Montes M, Wolff R, Moons KG, Monsef I, et al. Prognostic models for newly-diagnosed chronic lymphocytic leukaemia in adults: a systematic review and meta-analysis. Cochrane Database Syst Reviews. 2020;7(7):CD012022. https://doi.org/10.1002/14651858.CD012022.pub2.
12. Harrison H, Thompson RE, Lin Z, Rossi SH, Stewart GD, Griffin SJ, Usher-Smith JA. Risk prediction models for kidney cancer: A systematic review. Eur Urol Focus. 2021;7(6):1380–90. https://doi.org/10.1016/j.euf.2020.06.024.
13. Damen JA, Pajouheshnia R, Heus P, Moons KGM, Reitsma JB, Scholten R, Hooft L, Debray TPA. Performance of the Framingham risk models and pooled cohort equations for predicting 10-year risk of cardiovascular disease: a systematic review and meta-analysis. BMC Med. 2019;17(1):109. https://doi.org/10.1186/s12916-019-1340-7.
14. Jenkins DA, Sperrin M, Martin GP, Peek N. Dynamic models to predict health outcomes: current status and methodological challenges. Diagn Progn Res. 2018;2:23. https://doi.org/10.1186/s41512-018-0045-2.
15. Callegaro D, Barretta F, Swallow CJ, Strauss DC, Bonvalot S, Honore C, Stoeckle E, van Coevorden F, Haas R, Rutkowski P, et al. Longitudinal prognostication in retroperitoneal sarcoma survivors: development and external validation of two dynamic nomograms. Eur J Cancer. 2021;157:291–300. https://doi.org/10.1016/j.ejca.2021.08.008.
16. Fontein DBY, Klinten Grand M, Nortier JWR, Seynaeve C, Meershoek-Klein Kranenbarg E, Dirix LY, van de Velde CJH, Putter H. Dynamic prediction in breast cancer: proving feasibility in clinical practice using the TEAM trial. Annals Oncology: Official J Eur Soc Med Oncol. 2015;26(6):1254–62. https://doi.org/10.1093/annonc/mdv146.
17. Van Houwelingen HC. Dynamic prediction by landmarking in event history analysis. Scand J Stat. 2007;34(1):70–85. https://doi.org/10.1111/j.1467-9469.2006.00529.x.
18. Alexander M, Ball D, Solomon B, MacManus M, Manser R, Riedel B, Westerman D, Evans SM, Wolfe R, Burbury K. Dynamic thromboembolic risk modelling to target appropriate preventative strategies for patients with non-small cell lung cancer. Cancers. 2019;11(1):50. https://doi.org/10.3390/cancers11010050.
19. Finelli A, Beer TM, Chowdhury S, Evans CP, Fizazi K, Higano CS, Kim J, Martin L, Saad F, Saarela O. Comparison of joint and landmark modeling for predicting Cancer progression in men with Castration-Resistant prostate cancer: a secondary post hoc analysis of the PREVAIL randomized clinical trial. JAMA Netw. 2021;4(6):e2112426. https://doi.org/10.1001/jamanetworkopen.2021.12426.
20. Wu CF, Lv JW, Lin L, Mao YP, Deng B, Zheng WH, Wen DW, Chen Y, Kou J, Chen FP, et al. Development and validation of a web-based calculator to predict individualized conditional risk of site-specific recurrence in nasopharyngeal carcinoma: analysis of 10,058 endemic cases. Cancer Commun (Lond). 2021;41(1):37–50. https://doi.org/10.1002/cac2.12113.
21. ISSG Search Filter Resource. [Internet]: Prognosis and Prediction studies [https://sites.google.com/a/york.ac.uk/issg-search-filters-resource/home/prognosis]
22. Moriarty AS, Meader N, Snell KIE, Riley RD, Paton LW, Chew-Graham CA, Gilbody S, Churchill R, Phillips RS, Ali S, et al. Prognostic models for predicting relapse or recurrence of major depressive disorder in adults. Cochrane Database Syst Reviews. 2021;5(5):CD013491. https://doi.org/10.1002/14651858.CD013491.pub2.
23. 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(2):211–9. https://doi.org/10.1161/circulationaha.114.014508.
24. Moons KGM, Altman DG, Reitsma JB, Ioannidis JPA, Macaskill P, Steyerberg EW, Vickers AJ, Ransohoff DF, Collins GS. Transparent reporting of a

multivariable prediction model for individual prognosis or diagnosis (TRIPOD): explanation and elaboration. Ann Intern Med. 2015;162(1):W1–73. https://doi.org/10.7326/m14-0698.
25. Wolff RF, Moons KGM, Riley RD, Whiting PF, Westwood M, Collins GS, Reitsma JB, Kleijnen J, Mallett S, Altman D, et al. PROBAST: A tool to assess the risk of Bias and applicability of prediction model studies. Ann Intern Med. 2019;170(1):51–8. https://doi.org/10.7326/m18-1376.
26. Moons KGM, Wolff RF, Riley RD, Whiting PF, Westwood M, Collins GS, Reitsma JB, Kleijnen J, Mallett S. PROBAST: A tool to assess risk of Bias and applicability of prediction model studies: explanation and elaboration. Ann Intern Med. 2019;170(1):W1–33. https://doi.org/10.7326/m18-1377.
27. Moons KG, de Groot JA, Bouwmeester W, Vergouwe Y, Mallett S, Altman DG, Reitsma JB, Collins GS. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. PLoS Med. 2014;11(10):e1001744. https://doi.org/10.1371/journal.pmed.1001744.
28. Heus P, Damen J, Pajouheshnia R, Scholten R, Reitsma JB, Collins GS, Altman DG, Moons KGM, Hooft L. Uniformity in measuring adherence to reporting guidelines: the example of TRIPOD for assessing completeness of reporting of prediction model studies. BMJ Open. 2019;9(4):e025611. https://doi.org/10.1136/bmjopen-2018-025611.
29. Andaur Navarro CL, Damen JAA, Takada T, Nijman SWJ, Dhiman P, Ma J, Collins GS, Bajpai R, Riley RD, Moons KGM, et al. Risk of bias in studies on prediction models developed using supervised machine learning techniques: systematic review. BMJ (Clinical Res ed). 2021;375:n2281. https://doi.org/10.1136/bmj.n2281.
30. Slieker RC, van der Heijden A, Siddiqui MK, Langendoen-Gort M, Nijpels G, Herings R, Feenstra TL, Moons KGM, Bell S, Elders PJ, et al. Performance of prediction models for nephropathy in people with type 2 diabetes: systematic review and external validation study. BMJ (Clinical Res ed). 2021;374:n2134. https://doi.org/10.1136/bmj.n2134.
31. Mauguen A, Rachet B, Mathoulin-Pelissier S, Lawrence GM, Siesling S, MacGrogan G, Laurent A, Rondeau V. Validation of death prediction after breast cancer relapses using joint models. BMC Med Res Methodol. 2015;15:27. https://doi.org/10.1186/s12874-015-0018-x.
32. Malcovati L, Germing U, Kuendgen A, Della Porta MG, Pascutto C, Invernizzi R, Giagounidis A, Hildebrandt B, Bernasconi P, Knipp S, et al. Time-dependent prognostic scoring system for predicting survival and leukemic evolution in myelodysplastic syndromes. J Clin Oncol. 2007;25(23):3503–10. https://doi.org/10.1200/JCO.2006.08.5696.
33. Huang A, Chen Q, Fei Y, Wang Z, Ni X, Gao L, Chen L, Chen J, Zhang W, Yang J, et al. Dynamic prediction of relapse in patients with acute leukemias after allogeneic transplantation: joint model for minimal residual disease. Int J Lab Hematol. 2021;43(1):84–92. https://doi.org/10.1111/ijlh.13328.
34. Rueten-Budde AJ, van Praag VM, studygroup, van de Sande P, Fiocco MAJ. M: Dynamic prediction of overall survival for patients with high-grade extremity soft tissue sarcoma. Surg Oncol 2018, 27(4):695–701. https://doi.org/10.1016/j.suronc.2018.09.003
35. Rueten-Budde AJ, van Praag VM, van de Sande MAJ, Fiocco M, Group PS. External validation and adaptation of a dynamic prediction model for patients with high-grade extremity soft tissue sarcoma. J Surg Oncol. 2021;123(4):1050–6. https://doi.org/10.1002/jso.26337.
36. Bonnett LJ, Snell KIE, Collins GS, Riley RD. Guide to presenting clinical prediction models for use in clinical settings. BMJ (Clinical Res ed). 2019;365:l737. https://doi.org/10.1136/bmj.l737.
37. Adams ST, Leveson SH. Clinical prediction rules. BMJ (Clinical research ed) 2012, 344:d8312. https://doi.org/10.1136/bmj.d8312
38. Feng Q, May MT, Ingle S, Lu M, Yang Z, Tang J. Prognostic models for predicting overall survival in patients with primary gastric cancer: A systematic review. Biomed Res Int 2019, 2019:5634598. https://doi.org/10.1155/2019/5634598
39. Putter H, van Houwelingen HC. Understanding landmarking and its relation with Time-Dependent Cox regression. Stat Biosci. 2017;9(2):489–503. https://doi.org/10.1007/s12561-016-9157-9.
40. Asar Ö, Ritchie J, Kalra PA, Diggle PJ. Joint modelling of repeated measurement and time-to-event data: an introductory tutorial. Int J Epidemiol. 2015;44(1):334–44. https://doi.org/10.1093/ije/dyu262.
41. Zabor EC, Gonen M, Chapman PB, Panageas KS. Dynamic prognostication using conditional survival estimates. Cancer. 2013;119(20):3589–92. https://doi.org/10.1002/cncr.28273.
42. Hieke S, Kleber M, Konig C, Engelhardt M, Schumacher M. Conditional survival: A useful concept to provide information on how prognosis evolves over time. Clin Cancer Res. 2015;21(7):1530–6. https://doi.org/10.1158/1078-0432.CCR-14-2154.
43. Gupta R, Khoury JC, Altaye M, Jandarov R, Szczesniak RD. Assessing the relationship between gestational glycemic control and risk of preterm birth in women with type 1 diabetes: A joint modeling approach. J Diabetes Res 2020. 2020;3074532. https://doi.org/10.1155/2020/3074532.
44. Suresh K, Taylor JMG, Spratt DE, Daignault S, Tsodikov A. Comparison of joint modeling and landmarking for dynamic prediction under an illness-death model. Biom J. 2017;59(6):1277–300. https://doi.org/10.1002/bimj.201600235.
45. Putter H, van Houwelingen HC. Landmarking 2.0: bridging the gap between joint models and landmarking. Stat Med. 2022;41(11):1901–17. https://doi.org/10.1002/sim.9336.
46. van Houwelingen H, Putter H. Dynamic prediction in clinical survival analysis. Boca Raton: CRC; 2011.
47. Rizopoulos D, Molenberghs G, Lesaffre E. Dynamic predictions with time-dependent covariates in survival analysis using joint modeling and landmarking. Biom J. 2017;59(6):1261–76. https://doi.org/10.1002/bimj.201600238.
48. Bouwmeester W, Zuithoff NP, Mallett S, Geerlings MI, Vergouwe Y, Steyerberg EW, Altman DG, Moons KG. Reporting and methods in clinical prediction research: a systematic review. PLoS Med. 2012;9(5):1–12. https://doi.org/10.1371/journal.pmed.1001221.
49. Du M, Haag D, Song Y, Lynch J, Mittinty M. Examining Bias and reporting in oral health prediction modeling studies. J Dent Res. 2020;99(4):374–87. https://doi.org/10.1177/0022034520903725.
50. Collins GS, Ogundimu EO, Cook JA, Manach YL, Altman DG. Quantifying the impact of different approaches for handling continuous predictors on the performance of a prognostic model. Stat Med. 2016;35(23):4124–35. https://doi.org/10.1002/sim.6986.
51. Bellou V, Belbasis L, Konstantinidis AK, Tzoulaki I, Evangelou E. Prognostic models for outcome prediction in patients with chronic obstructive pulmonary disease: systematic review and critical appraisal. BMJ (Clinical Res ed). 2019;367:l5358. https://doi.org/10.1136/bmj.l5358.
52. Mallett S, Royston P, Dutton S, Waters R, Altman DG. Reporting methods in studies developing prognostic models in cancer: a review. BMC Med. 2010;8:20. https://doi.org/10.1186/1741-7015-8-20.
53. Xu L, He B, Zhang Y, Chen L, Fan D, Zhan S, Wang S. Prognostic models for amyotrophic lateral sclerosis: a systematic review. J Neurol. 2021;268(9):3361–70. https://doi.org/10.1007/s00415-021-10508-7.
54. Dhiman P, Ma J, Andaur Navarro CL, Speich B, Bullock G, Damen JAA, Hooft L, Kirtley S, Riley RD, Van Calster B, et al. Methodological conduct of prognostic prediction models developed using machine learning in oncology: a systematic review. BMC Med Res Methodol. 2022;22(1):101. https://doi.org/10.1186/s12874-022-01577-x.
55. Bradley A, Van Der Meer R, 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. https://doi.org/10.1136/bmjopen-2018-027192.
56. Pinart M, Kunath F, Lieb V, Tsaur I, Wullich B, Schmidt S, German Prostate Cancer C. Prognostic models for predicting overall survival in metastatic castration-resistant prostate cancer: a systematic review. World J Urol. 2020;38(3):613–35. https://doi.org/10.1007/s00345-018-2574-2.

Publisher's note