REVIEW

Open Access

Check for updates

Clinical prognostic models for sarcomas: a systematic review and critical appraisal of development and validation studies

Philip Heesen^{1*}, Sebastian M. Christ², Olga Ciobanu-Caraus³, Abdullah Kahraman⁴, Georg Schelling⁵, Gabriela Studer^{6,7}, Beata Bode-Lesniewska⁸ and Bruno Fuchs^{5,7,9}

Abstract

Background Current clinical guidelines recommend the use of clinical prognostic models (CPMs) for therapeutic decision-making in sarcoma patients. However, the number and quality of developed and externally validated CPMs is unknown. Therefore, we aimed to describe and critically assess CPMs for sarcomas.

Methods We performed a systematic review including all studies describing the development and/or external validation of a CPM for sarcomas. We searched the databases MEDLINE, EMBASE, Cochrane Central, and Scopus from inception until June 7th, 2022. The risk of bias was assessed using the prediction model risk of bias assessment tool (PROBAST).

Results Seven thousand six hundred fifty-six records were screened, of which 145 studies were eventually included, developing 182 and externally validating 59 CPMs. The most frequently modeled type of sarcoma was osteosarcoma (43/182; 23.6%), and the most frequently predicted outcome was overall survival (81/182; 44.5%). The most used predictors were the patient's age (133/182; 73.1%) and tumor size (116/182; 63.7%). Univariable screening was used in 137 (75.3%) CPMs, and only 7 (3.9%) CPMs were developed using pre-specified predictors based on clinical knowledge or literature. The median c-statistic on the development dataset was 0.74 (interquartile range [IQR] 0.71, 0.78). Calibration was reported for 142 CPMs (142/182; 78.0%). The median c-statistic of external validations was 0.72 (IQR 0.68–0.75). Calibration was reported for 46 out of 59 (78.0%) externally validated CPMs. We found 169 out of 241 (70.1%) CPMs to be at high risk of bias, mostly due to the high risk of bias in the analysis domain.

Discussion While various CPMs for sarcomas have been developed, the clinical utility of most of them is hindered by a high risk of bias and limited external validation. Future research should prioritise validating and updating existing well-developed CPMs over developing new ones to ensure reliable prognostic tools.

Trial registration PROSPERO CRD42022335222.

Keywords Sarcoma, Bone neoplasms, Soft tissue neoplasms, Prognosis, Nomograms

*Correspondence: Philip Heesen philip.heesen@uzh.ch Full list of author information is available at the end of the article



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

Background

Sarcomas are a diverse group of malignant soft-tissue or bone tumors arising from mesenchymal tissue, classified as a rare disease with an estimated incidence of 5 cases per 100,000 [1, 2]. While sarcomas only make up 0.9% of adult cancers, they account for 15-20% of childhood cancers diagnosed in the USA [2, 3]. The 5-year survival rate is about 70%, but greatly depends on patient-specific and tumor-specific factors and can be as low as 16.7% in patients with metastasis [3, 4]. The Surveillance, Epidemiology, and End Result Program (SEER) estimates 13,590 new cases in the USA in 2024 with 5200 estimated deaths [3]. Due to their rarity, heterogeneous nature, and poorly predictable clinical course, the clinical management of sarcomas remains challenging. Notably, the survival probabilities of sarcoma patients vary between and within different sarcoma subtypes. For example, it has been shown that the 5-year overall survival (OS) rate of synovial sarcoma patients differs between genders, with a rate of 35.6% for males and a rate of 68.7% for females [5].

Clinical prognostic models (CPMs) may enable more precise outcome prediction and allow for risk-stratified clinical decision-making in patients with sarcomas [6, 7]. The current ESMO-EURACAN-GENTURIS clinical practice guideline published by the European Society of Medical Oncology (ESMO) states that risk-predicting tools have identified a threshold of risk (<60% 10-year OS) above which the administration of chemotherapy may provide statistically and clinically significant benefits and thereby recommend the use of such prediction models [7].

While the high predictive performance of a CPM does not necessarily improve therapeutic decision-making, the accuracy and performance of a CPM are essential when considering implementing a CPM in routine clinical practice. To ensure that CMPs are transportable from the development cohort to a different cohort of patients, e.g. in a different country, the use of high methodological standards for development and rigorous external validation are essential. However, a systematic review assessing the risk of bias in prediction models developed using supervised machine learning showed that most included studies used poor methodology and are at high risk of bias [8, 9]. A major contributing factor to the high risk of bias was a small sample size and too few events during the development phase [9]. Complexities of the collected data such as censoring were rarely accounted for in those models [9]. Moreover, more than half of the included models were not adequately reported and are therefore not available for independent validation [9]. These results were corroborated by a systematic review of prediction models for patients with chronic lymphocytic leukemia in which poor reporting standards and a high risk of bias were found for all included CPMs [10].

While various CPMs for sarcomas have been published over the past years, only a minority (e.g. "Sarculator" and "PERSARC") are used in clinical practice. This may be largely due to a lack of knowledge of the performance of developed CPMs and the number of external validations of these CPMs.

Therefore, the aims of this systematic review were to systematically review and critically assess the CPM landscape in the field of sarcomas.

Methods

This systematic review reported the following aspects of the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement [11]. This study was registered in PROSPERO on August 12th, 2022 (CRD42022335222).

Information sources and search strategy

We performed a literature search in Medline, EMBASE, Cochrane Central and Scopus on June 7th, 2022. No filters, date or language restrictions were applied. The search strategy was developed and conducted by an information specialist and details about the literature search are presented in the Supplementary Table S1.

Eligibility criteria

We included CPM studies for any kind of sarcoma that were developed or validated on cohort studies, randomised controlled trials, nested case-control studies, case-cohort studies, databases, or registries [12]. A CPM was defined as a prognostic model that includes only predictors readily available in clinical practice (e.g. patient's sex, patient's age, tumor size, tumor grade). Therefore, studies that developed or validated prognostic models that included pathological, biological or radiological variables not commonly used in clinical practice were excluded. The decision about which predictors are commonly available in clinical practice was independently made by two reviewers and the discrepancy was resolved by discussion or involvement of a third reviewer. According to Riley et al., a random split of the development dataset is not considered a true external validation [13]. Therefore, we defined that models whose performance was evaluated using a random split technique were only internally validated. Table 1 presents the PICOTS format used for eligibility assessment.

Study selection, data collection process and analysis

After the literature search, studies were de-duplicated using Endnote 20 and then imported into Rayyan [14]. The abstract screening was conducted on Rayyan [14],

Table 1 PICOTS format used for eligibility assessment

PICOTS framework	Eligibility criteria for this systematic review		
P—Participants	Patients with a histological diagnosis of soft tissue or bone sarcoma		
I—Index models	All clinical prognostic models predict overall survival, event-free survival, progression-free survival, disease-specific survival/sarcoma-specific survival, progression of disease, local recurrence rate, systemic recurrence rate and development of metastasis		
C—Comparator	Not applicable because we aim to identify all models that were developed and/or validated for sarcomas		
O—Outcome	Overall survival, event-free survival, progression-free survival, disease-specific survival/sarcoma-specific survival, progression of the disease, local recurrence rate, systemic recurrence rate and development of metastasis		
T—Timing	Any timing starting with the date of pathologically established diagnosis		
S—Setting	Any setting		

Page 3 of 9

subsequently followed by full-text screening. Data were extracted using Microsoft Excel. Title screening, data extraction and risk of bias assessment were done independently and in duplicate (PH, SMC). Discrepancies were resolved by discussion or by consultation of a third reviewer (BF, OCC). The collected variables are presented in Supplementary Table S5. For discrimination, the c-index and area under the receiver operating characteristic curve (AUC) were combined and are jointly referred to as AUC in this manuscript. The risk of bias of each developed and/or validated CPM was assessed using the prediction model risk of bias assessment tool (PROBAST) [15]. PROBAST assesses the risk of bias over four domains: participant selection, predictors, outcomes, and analysis. If at least two questions of one domain were rated as "probably no" or "no", the whole domain was rated to be at high risk of bias. If at least two questions of one domain were rated as "No information", the whole domain was rated to be of unclear risk of bias. Each domain is rated to be of low, high or unclear risk of bias. When investigators assessed more than one domain to be of high risk of bias, the whole study was ranked to be of high risk of bias, irrespective of the other two domains being of unclear or low risk of bias. At the time of writing this review, no complete guidance on the application of the GRADE system for CPM development/validation studies was published. Therefore, no GRADE assessment was conducted for this review. We qualitatively summarised the results of this review using descriptive statistics. Data were analyzed using R Statistical Software (version 4.3.1) [16].

Results

After deduplication, we retrieved 7656 records. After the abstract screening, 181 full-text articles were assessed for eligibility. Of those, 31 were excluded because the study population included non-sarcoma patients. Five studies were excluded since some of the variables that were

included in the CPM were judged to not be available in routine clinical practice. A total of 145 studies developing 182 CPMs and externally validating 59 CPMs were included. Figure 1 presents the flow diagram for study selection.

Datasets for CPM development

One hundred seventy models (170/182; 93.4%) were developed using retrospective data, 10 models (10/182; 5.5%) were developed on prospectively collected data and 2 models (2/182; 1.1%) were developed using a combination of retrospectively and prospectively collected data. Most models (109/182, 59.9%) were developed using the Surveillance, Epidemiology, and End Results Program (SEER) dataset. Of the models not developed on the SEER dataset, single-institution datasets were used in the development of 50 models (50/182; 27.5%), multiinstitutional datasets in 22 models (22/182; 12.1%), and one model (1/182; 0.5%) utilised data from the National Cancer Database. Only 15 (15/182; 8.2%) CPMs were developed on a multi-national dataset. None of the models developed on multi-institutional datasets accounted for clustering between institutions.

The median sample size was 635 (IQR 317.0–1110.5), but no study provided justification for the sample size. The median number of events was 201.5 (IQR 119.2–316.8), however, the number of events was not provided for 102 CPMs (102/182; 56.0%).

CPM development

In the development process of 135 models (135/182; 74.2%), continuous variables were categorised, with data-driven methods being the most common approach (47/182, 25.8%). Twelve models (12/182; 6.6%) included continuous predictors using restricted cubic splines as a method of modelling non-linear relationships between continuous variables and the outcome, while 28 models (28/182; 15.4%) included continuous predictors as linear



Fig. 1 Flow diagram of selected articles

variables. Table 2 describes the handling of missing data during the development and external validation of the included CPMs.

Table 3 describes the histological entities that CPMs were developed for, the predicted outcomes and included predictors. The median number of candidate predictors was 10.5 (IQR 8.8–13.0), though the number of candidate predictors was not available for 18 models (9.9%). The median number of events per variable was 20.0 (IQR 11.4–39.9), and the median number of predictors included in the final model was 5.0 (IQR 4.0–6.0). Supplementary Figure S1 presents a bar chart of the 10 most frequently considered and included predictors.

Cox proportional hazards regression (146/182, 80.2%) was the most widely used statistical model, followed by logistic regression (14/182, 7.7%) and the Fine and Gray competing risks model (12/182, 6.6%). Univariable selection of predictors was performed in 137 of 182 models (75.3%), multivariable selection without prior univariable selection in 11 models (11/182; 6.0%), and predictors

Table 2Missing data handling during development and
external validation

	Overall N = 241	Development only <i>n</i> = 182	External validation only n=59
Missing data reporti	ng		
None	178 (74.2%)	138 (75.8%)	40 (67.8%)
Overall	40 (16.7%)	32 (17.6%)	8 (13.6%)
Detail ^a	23 (9.6%)	12 (6.6%)	11 (18.6%)
Missing data method	b		
Complete case	161 (67.1%)	129 (70.9%)	32 (54.2%)
Multiple imputa- tion	6 (2.5%)	6 (3.3%)	0
Unclear	73 (30.4%)	47 (25.8%)	27 (45.8%)
Assumed mech- anism of missing- ness reported	1 (0.4%)	1 (0.5%)	0

^a Missing data was provided for each variable

	Overall N = 241	Development <i>n</i> = 182	External validation n=59
Sarcoma entity			
Osteosarcoma	54 (22.4%)	46 (25.3%)	8 (13.6%)
Ewing sarcoma	29 (12.0%)	25 (13.7%)	4 (6.8%)
Soft-tissue sarcoma	61 (25.3%)	21 (11.5%)	40 (67.8%)
Chondrosarcoma	26 (10.8%)	19 (10.4%)	7 (11.9%)
Leiomyosarcoma	13 (5.4%)	13 (7.1%)	0 (0.0%)
Angiosarcoma	10 (4.1%)	10 (5.5%)	0 (0.0%)
Liposarcoma	8 (3.3%)	8 (4.4%)	0 (0.0%)
Rhabdomyosarcoma	7 (2.9%)	7 (3.8%)	0 (0.0%)
Outcomes			
Overall survival	101 (41.9%)	81 (44.5%)	20 (33.9%)
Cancer-specific survival	70 (29.0%)	56 (30.7%)	14 (23.7%)
Distant metastasis	14 (5.8%)	10 (5.5%)	4 (6.8%)
Local recurrence	10 (4.1%)	8 (4.4%)	2 (3.4%)
Disease-free survival	9 (3.7%)	4 (2.2%)	5 (8.5%)
Lung metastasis	3 (1.2%)	3 (1.6%)	0 (0.0%)
Metastasis-free survival	3 (1.2%)	3 (1.6%)	0 (0.0%)
Progression-free survival	3 (1.2%)	3 (1.6%)	0 (0.0%)
Other	28 (11.6%)	14 (7.7%)	14 (23.7%)
Ten most included predictors			
Age	179 (74.3%)	133 (73.1%)	46 (78.0%)
Size	149 (61.8%)	116 (63.7%)	36 (61.0%)
Grade	103 (42.7%)	71 (39.0%)	32 (54.2%)
Surgery	77 (32.0%)	69 (37.9%)	8 (13.6%)
Site	71 (29.5%)	58 (31.9%)	13 (22.0%)
Stage	61 (25.3%)	56 (30.8%)	5 (8.5%)
Sex	41 (17.0%)	34 (18.7%)	7 (11.9%)
Chemotherapy	39 (16.2%)	33 (18.1%)	6 (10.2%)
Histological type	63 (26.1%)	29 (15.9%)	34 (57.6%)
Race	24 (9.9%)	21 (11.5%)	3 (5.1%)

Table 3 Outcomes and predictors of included CPMs

were pre-specified using expert opinion or literature in only 7 models (7/182; 3.9%).

No performance measure was reported for 9 (9/182; 5.0%) CPMs. A measure of discrimination was reported for 89.6% (163/182) of developed CPMs. The median reported AUC was 0.74 (IQR 0.71–0.78). Calibration was reported for 142 (142/182; 78.0%) CPMs and the only reported measure of the calibration was the calibration plot (142/142; 100%). DCA was only rarely reported (41/182; 22.5%).

CPM validation

Of 182 developed models, internal validation was performed for 116 models (63.7%). The most common form of internal validation was split sample (67/116; 57.8%), followed by bootstrapping (30/116; 25.9%), cross-validation (10/116; 8.6%) and a combined approach of split sample and bootstrapping (13/116; 11.2%). Forty-two (62.7%) of the 67 CPMs that were internally validated using a random split sample were classified as external validation by the authors of the respective studies.

Fifty-nine models (59/182; 32.4%) were externally validated, of which 29 models (29/59; 49.2%) were developed in the same study. Thirty models (30/59; 50.8%) were externally validated after their development was described in a previously published paper. Nine models (9/59; 15.3%) were externally validated based on prospectively collected data. Only 8 CPMS (8/59; 13.6%) were validated on a multi-national dataset. The median sample size was 307.0 (IQR 115.5–631.0).

Fifty-six out of 59 external validations (94.9%) reported at least one measure of discrimination. The median AUC of all external validations was 0.72 (IQR 0.68–0.75). Calibration was only reported for 46 out of 59 (78.0%) externally validated CPMs. A DCA was only provided for 7 out of 59 (11.9%) external validations.

Risk of bias

Of the 241 evaluated models, 169 (70.1%) were at high risk of bias, 45 (18.7%) were at unclear risk of bias and only 27 models (11.2%) were at low risk of bias. A summary of the risk of bias for each dimension is presented in Fig. 2 and Supplementary Table S7.

Discussion

In this systematic review of CPMs for sarcomas, we identified 182 CPMs. Despite the abundance of available CPMs in the current literature, a considerably high proportion suffered from poor methodological conduct (70.8% had a high risk of bias), lack of independent external validation (67.6% of developed CPMs were not externally validated), inadequate reporting of performance and in particular calibration metrics (22.0% reported no calibration metric), lack of sample size calculation (none presented a sample size calculation) and inefficient handling of missing data (70.9% applied complete case analysis). The most common dataset was the SEER cancer dataset which hosts freely available cancer data maintained by the National Cancer Institute in the USA. This dataset encompasses roughly 48% of the total cancer

population in the USA [17]. While this dataset contains a large number of observations, certain predictors are not reported in enough detail (such as pathological information or type of chemotherapy).

We found that only 32.4% of identified CPMs were externally validated. This is largely in line with a previous review of prognostic prediction models using machine learning in oncology where a proportion of 24% was found [9]. Furthermore, 62.7% of internal validations that used the random split-sample approach, were wrongly classified as external validations. However, external validations are a prerequisite for the use of a CPM in clinical practice [18]. As a CPM cannot be validated on every patient population (e.g. each individual geographical region), it has been argued that a validated CPM may not exist [19]. Nevertheless, external validations are indispensable to assess the transportability of a CPM [18, 20]. As we found that only a minority of CPMs were externally validated, future research should prioritise external validation over the development of CPMs. Published recommendations for the validation of CPMs should be followed [13, 18, 20–25].

Since prediction models are used to guide physicians and patients in making a joint decision about future therapy, the performance of a CPM needs to be rigorously evaluated [26]. However, we found that 5% of included CPMs did not report any performance measures. In 22% of the



Fig. 2 Risk of Bias of all included CPMs (A), of all developed CPMs only (B), of all externally validated CPMs only (C)

included models, no metric of calibration was reported. This figure is higher than the number that has been reported in a systematic review of prognostic prediction models in oncology (7%) published in 2022 [9]. A potential explanation is that studies included in this systematic review might have adhered more closely to reporting guidelines. However, neither this study nor the study by Dhiman et al. assessed the adherence to a reporting guideline [9]. The reporting of calibration measures is important as the accuracy of the point-estimate of the predicted risk may have significant implications for treatment decisions and ultimately impact patient outcomes [27, 28].

None of the studies included in this review described an adequate sample size calculation prior to the development or external validation of their CPM. This finding is in line with the results from other systematic reviews that critically assessed the methodological conduct of prognostic modelling studies [29]. However, an adequate sample size during model development determines the stability of a CPM, while during model validation it influences the precision of estimates and width of confidence intervals of performance statistics [30]. Extensive guidance on sample size considerations for CPM development (Riley et al., 2019 [31, 32]) and external validation (Collins et al., 2015 [33]) has been published.

We found that the most frequently (70.9%) used method for handling missing data was complete case analysis. This is higher than reported in the previous systematic review of prognostic models for sarcomas [9]. The complete case approach is only valid when data can be assumed to be missing completely at random [34]. However, in the majority of cases, it is questionable whether this assumption is correct. Only one of the studies included in this review described the assumed mechanism of data missingness. The problem of poor handling of missing data in prediction modelling research has been described previously [35]. Missing data can lead to biased estimates of model parameters, resulting in inaccurate predictions, overfitting and eventually a low generalisability of the CPM [36]. Modern imputation methods may be used to mitigate the issue of missing data [37]. Sensitivity analyses should be conducted to quantify the effect of missing data and different imputation methods on predicted risks and performance of the CPM.

A model that was developed and subsequently validated on a multi-institutional and multi-national dataset will likely be more generalisable than a model developed on a dataset from a single institution [38]. In this systematic review, we found that 27.5% of CPMs were developed on a dataset from a single institution, only 8.2% were developed on a multi-national dataset, and only 13.6% were validated on a multi-national dataset.

To mitigate this issue, data harmonisation on an international level is necessary to be able to merge data from multiple institutions and multiple countries. The development of a unified dataset is also necessary as the rapid changes in the histopathological classification of sarcomas due to advances in molecular tumor research pose a particular challenge for data harmonisation. The use of different hospital information systems, storage of data in different databases and ultimately the big quest for data privacy further complicate international data sharing and migration.(38) In this respect, the use of federated learning may represent a viable solution that allows to development of a machine learning or classical statistical model without sharing patient data beyond their respective institution [38, 39]. Federated learning could thereby enable multi-institutional collaborations to develop and validate CPMs [38].

This systematic review has several limitations. Our literature search might have missed studies that did not mention the development or external validation of a CPM in the title or abstract of their manuscript. Furthermore, the literature search focused on the term "survival", "progression" and "recurrence". Therefore, studies of CPMs predicting other outcomes might have been missed. A scoping review during the specification of the literature search has shown that the vast majority of studies developing or validating a CPM for sarcomas focused on such an outcome which is the reason that the final literature search only included these search terms. As sarcomas are a group of cancers with an expected life expectancy of several years, short-term outcomes (such as time until discharge from hospital) might be less reported [40]. Moreover, some studies used a CPM to stratify patients into risk categories and reported the predicted as well as observed risk in their study group, thereby unintentionally performing an external validation of the respective CPM [41]. These studies are inherently difficult to identify because the use of a CPM for stratification might not have been mentioned in the title or abstract. Our search strategy was performed nearly 2 years ago and more CPMs for sarcomas have been published in the meantime. However, given the large amount of CPMs included in our review, including studies published after June 7, 2022 would likely not change our conclusions or recommendations.

In conclusion, we found that the majority (70.8%) of included models were at high risk of bias and suffered from poor methodological conduct. Therefore, use in routine clinical practice may not be recommended for most published CPMs of sarcomas. As most risk of bias stemmed from the analysis domain, researchers should consider published guidance on the development and external validation of CPMs. As only 32.4% of CPMs were externally validated, future research efforts should concentrate on the external validation of existing CPMs rather than on the development of new ones.

Supplementary Information

The online version contains supplementary material available at https://doi. org/10.1186/s41512-025-00186-8.

Supplementary Material 1: Table S1. Search strategy in Cochrane Central.

Supplementary Material 2: Table S2. Search strategy in Embase.

Supplementary Material 3: Table S3. Search strategy in Medline.

Supplementary Material 4: Table S4. Search strategy in Scopus.

Supplementary Material 5: Table S5. Table of extracted variables.

Supplementary Material 6: Table S6. Table of included studies.

Supplementary Material 7: Table S7. Risk of Bias using the Prediction model Risk Of Bias Assessment Tool (PROBAST).

Supplementary Material 8: Figure S1. Bar chart of the (a) most considered and (b) included predictor variables.

Acknowledgements

We thank Dr. Martina Gosteli for creating the literature search strategy and running the systematic literature search.

Authors' contributions

PH and BF conceptualised the study. PH, SMC, OCC and BF curated the data. PH and OCC did the formal analysis. PH, OCC and BF wrote the first draft of the manuscript and all authors were involved in the writing, reviewing, and editing of the manuscript. PH, OCC and BF accessed and verified the data. All authors had access to the data and had final responsibility for the decision to submit for publication.

Funding

None.

Data availability

No datasets were generated or analysed during the current study.

Declarations

Ethics approval and consent to participate Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹ Faculty of Medicine, University of Zurich, Raemistrasse 71, Zurich 8006, Switzerland. ²Department of Radiation Oncology, University Hospital Zurich and University of Zurich, Raemistrasse 100, Zurich 8091, Switzerland. ³ Medical University of Vienna, Spitalgasse 23, Vienna 1090, Austria. ⁴School of Life Sciences, University of Applied Sciences and Arts Northwestern Switzerland, Hofackerstrasse 30, Muttenz 4132, Switzerland. ⁵Department of Orthopaedics and Trauma, University Teaching Hospital LUKS, Sarcoma Service, Spitalstrasse, 6000, Lucerne, Switzerland. ⁶Department of Radiation Oncology, University Teaching Hospital LUKS, Spitalstrasse, 6000, Lucerne, Switzerland. ⁷Faculty of Health Sciences and Medicine, University of Lucerne, Frohburgstrasse 3, Lucerne 6002, Switzerland. ⁸Pathology Institute Enge and University of Zurich, Museumstrasse 135, Zurich 8005, Switzerland. ⁹Department of Orthopaedics and Trauma, Kantonsspital Winterthur, Sarcoma Service, Brauerstrasse 15, Winterthur 8400, Switzerland.

Received: 22 September 2024 Accepted: 28 February 2025 Published online: 07 April 2025

References

 Vibert J, Watson S. The molecular biology of soft tissue sarcomas: current knowledge and future perspectives. Cancers (Basel). 2022;14(10):2548.

- International Agency for Research on Cancer, World Health Organization, International Academy of Pathology. WHO classification of tumours of soft tissue and bone tumours. 5th ed. Fletcher CDM, editor. IARC; 2020.
 607 p. (World Health Organization Classification of Tumours). Accessed 3 Jan 2025.
- Cancer of soft tissue including heart Cancer Stat Facts. SEER. Available from: https://seer.cancer.gov/statfacts/html/soft.html. Accessed 10 Dec 2024.
- Tichanek F, Försti A, Hemminki O, Hemminki A, Hemminki K. Steady survival improvements in soft tissue and bone sarcoma in the Nordic countries through 50 years. Cancer Epidemiol. 2024;92(102449):102449.
- Bacon A, Wong K, Fernando MS, Rous B, Hill RJW, Collins SD, et al. Incidence and survival of soft tissue sarcoma in England between 2013 and 2017, an analysis from the National Cancer Registration and Analysis Service. Int J Cancer. 2023;152(9):1789–803.
- Pasquali S, Pizzamiglio S, Touati N, Litiere S, Marreaud S, Kasper B, et al. The impact of chemotherapy on survival of patients with extremity and trunk wall soft tissue sarcoma: revisiting the results of the EORTC-STBSG 62931 randomised trial. Eur J Cancer. 2019;109:51–60.
- Gronchi A, Miah AB, Dei Tos AP, Abecassis N, Bajpai J, Bauer S, et al. Soft tissue and visceral sarcomas: ESMO-EURACAN-GENTURIS Clinical Practice Guidelines for diagnosis, treatment and follow-up*. Ann Oncol. 2021;32(11):1348–65.
- Andaur Navarro CL, Damen JAA, Takada T, Nijman SWJ, Dhiman P, Ma J, et al. Risk of bias in studies on prediction models developed using supervised machine learning techniques: systematic review. BMJ. 2021;375:n2281.
- Dhiman P, Ma J, Andaur Navarro CL, Speich B, Bullock G, Damen JAA, et al. Methodological conduct of prognostic prediction models developed using machine learning in oncology: a systematic review. BMC Med Res Methodol. 2022;22(1):101.
- Kreuzberger N, Damen JA, Trivella M, Estcourt LJ, Aldin A, Umlauff L, et al. Prognostic models for newly-diagnosed chronic lymphocytic leukaemia in adults: a systematic review and meta-analysis. Cochrane Libr. 2020;2020(7). Available from: https://pubmed.ncbi.nlm.nih.gov/32735 048/. Accessed 26 Apr 2024.
- Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, The PRISMA, et al. statement: an updated guideline for reporting systematic reviews. BMJ. 2020;2021:n71.
- Moons KGM, de 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.
- Riley RD, Van Der Windt DP, Moons K. Prognosis research in healthcare: concepts, methods, and impact. Oxford University Press; 2019. Accessed 3 Jan 2025.
- 14. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan—a web and mobile app for systematic reviews. Syst Rev. 2016;5(1). Available from: https://doi.org/10.1186/s13643-016-0384-4.
- 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(1):51.
- The R project for statistical computing. Available from: https://www.Rproject.org/. Accessed 10 Dec 2024.
- Che W-Q, Li Y-J, Tsang C-K, Wang Y-J, Chen Z, Wang X-Y, et al. How to use the Surveillance, Epidemiology, and End Results (SEER) data: research design and methodology. Mil Med Res. 2023;10(1):50.
- 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.
- Van Calster B, Steyerberg EW, Wynants L, van Smeden M. There is no such thing as a validated prediction model. BMC Med. 2023;21(1). Available from: https://doi.org/10.1186/s12916-023-02779-w.
- Steyerberg EW, Harrell FE Jr. Prediction models need appropriate internal, internal–external, and external validation. J Clin Epidemiol. 2016;69:245–7.
- Riley RD, Archer L, Snell KIE, Ensor J, Dhiman P, Martin GP, et al. Evaluation of clinical prediction models (part 2): how to undertake an external validation study. BMJ. 2024;384:e074820.
- 22. Riley RD, Snell KIE, Archer L, Ensor J, Debray TPA, van Calster B, et al. Evaluation of clinical prediction models (part 3): calculating the sample size required for an external validation study. BMJ. 2024;384:e074821.

- Ramspek CL, Jager KJ, Dekker FW, Zoccali C, van Diepen M. External validation of prognostic models: what, why, how, when and where? Clin Kidney J. 2021;14(1):49–58.
- 24. Steyerberg EW. Clinical prediction models: a practical approach to development, validation, and updating. Springer; 2019. Accessed 3 Jan 2025.
- Harrell FE. Regression Modeling Strategies: With applications to linear models, logistic and ordinal regression, and survival analysis. Springer; 2015. Accessed 3 Jan 2025.
- 26 Moons KGM, Royston P, Vergouwe Y, Grobbee DE, Altman DG. Prognosis and prognostic research: what, why, and how? BMJ. 2009;338(feb23 1):b375–b375.
- Van Calster B, On behalf of Topic Group 'Evaluating diagnostic tests and prediction models' of the STRATOS initiative, McLernon DJ, van Smeden M, Wynants L, Steyerberg EW. Calibration: the Achilles heel of predictive analytics. BMC Med. 2019;17(1). Available from: https://doi.org/10.1186/ s12916-019-1466-7.
- Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, et al. Assessing the performance of prediction models. Epidemiology. 2010;21(1):128–38.
- Collins GS, De Groot JA, Dutton S. External validation of multivariable prediction models: a systematic review of methodological conduct and reporting. BMC Med Res Methodol. 2014;14. Available from: https://doi. org/10.1186/1471-2288-14-40pmid:24645774.
- Pate A, Emsley R, Sperrin M, Martin GP, van Staa T. Impact of sample size on the stability of risk scores from clinical prediction models: a case study in cardiovascular disease. Diagn Progn Res. 2020;4(1). Available from: https://doi.org/10.1186/s41512-020-00082-3.
- Riley RD, Snell KIE, Ensor J, Burke DL, Harrell FE Jr, Moons KGM, et al. Minimum sample size for developing a multivariable prediction model: Part I - continuous outcomes. Stat Med. 2019;38(7):1262–75.
- Riley RD, Snell KIE, Ensor J, Burke DL, Harrell FE Jr, Moons KGM, et al. Minimum sample size for developing a multivariable prediction model: PART II - binary and time-to-event outcomes. Stat Med. 2019;38(7):1276–96.
- Collins GS, Ogundimu EO, Altman DG. Sample size considerations for the external validation of a multivariable prognostic model: a resampling study. Stat Med. 2016;35(2):214–26.
- Ross RK, Breskin A, Westreich D. When is a complete-case approach to missing data valid? The importance of effect-measure modification. Am J Epidemiol. 2020;189(12):1583–9.
- Nijman SWJ, Leeuwenberg AM, Beekers I, Verkouter I, Jacobs JJL, Bots ML, et al. Missing data is poorly handled and reported in prediction model studies using machine learning: a literature review. J Clin Epidemiol. 2022;142:218–29.
- Vergouw D, Heymans MW, Peat GM, Kuijpers T, Croft PR, de Vet HCW, et al. The search for stable prognostic models in multiple imputed data sets. BMC Med Res Methodol. 2010;10(1). Available from: https://doi.org/10. 1186/1471-2288-10-81.
- Sisk R, Sperrin M, Peek N, van Smeden M, Martin GP. Imputation and missing indicators for handling missing data in the development and deployment of clinical prediction models: a simulation study. Stat Methods Med Res. 2023;32(8):1461–77.
- Sheller MJ, Edwards B, Reina GA, Martin J, Pati S, Kotrotsou A, et al. Federated learning in medicine: facilitating multi-institutional collaborations without sharing patient data. Sci Rep. 2020;10(1):1–12.
- Rieke N, Hancox J, Li W, Milletari F, Roth HR, Albarqouni S, et al. The future of digital health with federated learning. NPJ Digit Med. 2020;3(1):1–7.
- Hayes AJ, Nixon IF, Strauss DC, Seddon BM, Desai A, Benson C, et al. UK guidelines for the management of soft tissue sarcomas. Br J Cancer. 2024;1–21. Accessed 3 Jan 2025.
- Pasquali S, Palmerini E, Quagliuolo V, Martin-Broto J, Lopez-Pousa A, Grignani G, et al. Neoadjuvant chemotherapy in high-risk soft tissue sarcomas: a sarculator-based risk stratification analysis of the ISG-STS 1001 randomized trial. Cancer. 2022;128(1):85–93.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.