Does the SORG algorithm predict 5-year survival in patients with chondrosarcoma? An external validation

Bongers, M.E.R.; Thio, Q.C.B.S.; Karhade, A.V.; Stor, M.L.; Raskin, K.A.; Calderon, S.A.L.; ... ; Schwab, J.H.

Clin Orthop Relat Res (2019) 477:2296-2303
DOI 10.1097/CORR.0000000000000748

Clinical Orthopaedics and Related Research®
A Publication of The Association of Bone and Joint Surgeons®

Clinical Research

Does the SORG Algorithm Predict 5-year Survival in Patients with Chondrosarcoma? An External Validation

Michiel E. R. Bongers MD, Quirina C. B. S. Thio MD, Aditya V. Karhade BE, Merel L. Stor BS, Kevin A. Raskin MD, Santiago A. Lozano Calderon MD, PhD, Thomas F. DeLaney MD, Marco L. Ferrone MD, Joseph H. Schwab MD, MS

Received: 21 November 2018 / Accepted: 8 March 2019 / Published online: 27 April 2019

Abstract

Background We developed a machine learning algorithm to predict the survival of patients with chondrosarcoma. The algorithm demonstrated excellent discrimination and calibration on internal validation in a derivation cohort based on data from the Surveillance, Epidemiology, and End Results (SEER) registry. However, the algorithm has not been validated in an independent external dataset.

Questions/purposes Does the Skeletal Oncology Research Group (SORG) algorithm accurately predict 5-year survival in an independent patient population surgically treated for chondrosarcoma?

Methods The SORG algorithm was developed using the SEER registry, which contains demographic data, tumor characteristics, treatment, and outcome values; and includes approximately 30% of the cancer patients in the United States. The SEER registry was ideal for creating the derivation cohort, and consequently the SORG algorithm, because of the high number of eligible patients and the availability of most (explanatory) variables of interest. Between 1992 to 2013, 326 patients were treated surgically for extracranial chondrosarcoma of the bone at two tertiary care referral centers. Of those, 179 were accounted for at a

Each author certifies that neither he, nor any member of his immediate family, has funding or commercial associations (consultancies, stock ownership, equity interest, patent/licensing arrangements, etc) that might pose a conflict of interest in connection with the submitted article.
Each author certifies that his or her institution approved the human protocol for this investigation and that all investigations were conducted in conformity with ethical principles of research.
This work was performed at Massachusetts General Hospital, Boston, MA, USA.

M. E. R. Bongers, Q. C. B. S. Thio, A.V. Karhade, M. L. Stor, K. A. Raskin, S. A. Lozano-Calderon, Department of Orthopaedic Surgery, Division of Orthopaedic Oncology, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA

T. F. DeLaney, Department of Radiation Oncology, Massachusetts General Hospital, Boston, MA, USA

M. L. Ferrone, Department of Orthopaedic Surgery, Orthopaedic Oncology Service, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA

J. H. Schwab, Department of Orthopaedic Surgery, Division of Orthopaedic Oncology, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA

M. E. R. Bongers (✉), Department of Orthopaedic Surgery, Division of Orthopaedic Oncology, Massachusetts General Hospital, Harvard Medical School, Room 3.550, Yawkey Building, Massachusetts General Hospital, 55 Fruit Street, Boston, MA, 02114, USA, Email: michielbongers@gmail.com

All ICMJE Conflict of Interest Forms for authors and *Clinical Orthopaedics and Related Research*® editors and board members are on file with the publication and can be viewed on request.

Wolters Kluwer

minimum of 5 years after diagnosis in a clinical note at one of the two institutions, unless they died earlier, and were included in the validation cohort. In all, 147 (45%) did not meet the minimum 5 years of followup at the institution and were not included in the validation of the SORG algorithm. The outcome (survival at 5 years) was checked for all 326 patients in the Social Security death index and were included in the supplemental validation cohort, to also ascertain validity for patients with less than 5 years of institutional followup. Variables used in the SORG algorithm to predict 5-year survival including sex, age, histologic subtype, tumor grade, tumor size, tumor extension, and tumor location were collected manually from medical records. The tumor characteristics were collected from the postoperative musculoskeletal pathology report. Predicted probabilities of 5-year survival were calculated for each patient in the validation cohort using the SORG algorithm, followed by an assessment of performance using the same metrics as used for internal validation, namely: discrimination, calibration, and overall performance. Discrimination was calculated using the concordance statistic (or the area under the Receiver Operating Characteristic (ROC) curve) to determine how well the algorithm discriminates between the outcome, which ranges from 0.5 (no better than a coin-toss) to 1.0 (perfect discrimination). Calibration was assessed using the calibration slope and intercept from a calibration plot to measure the agreement between predicted and observed outcomes. A perfect calibration plot should show a 45° upwards line. Overall performance was determined using the Brier score, ranging from 0 (excellent prediction) to 1 (worst prediction). The Brier score was compared with the null-model Brier score, which showed the performance of a model that ignored all the covariates. A Brier score lower than the null model Brier score indicated greater performance of the algorithm. For the external validation an F1-score was added to measure the overall accuracy of the algorithm, which ranges between 0 (total failure of an algorithm) and 1 (perfect algorithm).

The 5-year survival was lower in the validation cohort than it was in the derivation cohort from SEER (61.5% [110 of 179] versus 76% [1131 of 1544] ; $p < 0.001$). This difference was driven by higher proportion of dedifferentiated chondrosarcoma in the institutional population than in the derivation cohort (27% [49 of 179] versus 9% [131 of 1544]; $p < 0.001$). Patients in the validation cohort also had larger tumor sizes, higher grades, and nonextremity tumor locations than did those in the derivation cohort. These differences between the study groups emphasize that the external validation is performed not only in a different patient cohort, but also in terms of disease characteristics. Five-year survival was not different for both patient groups between subpopulations of patients with conventional chondrosarcomas and those with dedifferentiated chondrosarcomas.

Results The concordance statistic for the validation cohort was 0.87 (95% CI, 0.80–0.91). Evaluation of the algorithm's calibration in the institutional population resulted in a calibration slope of 0.97 (95% CI, 0.68–1.3) and calibration intercept of -0.58 (95% CI, -0.20 to -0.97). Finally, on overall performance, the algorithm had a Brier score of 0.152 compared with a null-model Brier score of 0.237 for a high level of overall performance. The F1-score was 0.836. For the supplementary validation in the total of 326 patients, the SORG algorithm had a validation of 0.89 (95% CI, 0.85–0.93). The calibration slope was 1.13 (95% CI, 0.87–1.39) and the calibration intercept was -0.26 (95% CI, -0.57 to 0.06). The Brier score was 0.11, with a null-model Brier score of 0.19. The F1-score was 0.901.

Conclusions On external validation, the SORG algorithm retained good discriminative ability and overall performance but overestimated 5-year survival in patients surgically treated for chondrosarcoma. This internet-based tool can help guide patient counseling and shared decision making.

Level of Evidence Level III, prognostic study.

Introduction

Chondrosarcoma is the second most prevalent type of primary malignant bone tumor, accounting for 30% of all primary bone tumors [6, 18]. Management principally involves surgical excision because chondrosarcoma is insensitive to conventional-dose radiation therapy and chemotherapy [3, 5, 13, 15, 18]. Previous studies have identified multiple prognostic factors for local recurrence, distal metastasis, and survival [1, 4, 7, 9]. Machine learning algorithms have been created for other malignant tumors in the breast, lung, and prostate [14, 16, 17]. However, there are currently no validated prognostic algorithms for predicting postoperative survival in chondrosarcoma. This is probably due to the rarity of the disease as large numbers of patients are needed to create such algorithms.

Accurate survival prediction may be helpful in clinical decision making and risk stratification when comparing treatment efficacy, such as in randomized controlled trials or prospective matching studies. Few studies have attempted to make a prognostic nomogram or algorithm using postoperative factors to predict individualized survival in patients who undergo chondrosarcoma treatment [23, 27]. We developed a machine learning algorithm to postoperatively predict 5-year survival in patients with chondrosarcoma, using sex, age, histologic subtype, tumor grade, tumor size, tumor extension, and tumor location [27]. On internal validation, the algorithm performed very well with a c-statistic of 0.868, calibration slope of 1.025, calibration intercept of 0.001, and Brier score of 0.117 compared with a null-model Brier score of 0.182.

However, the generalizability of our findings to independent institutional populations has not been evaluated, and such validation is important [11].

In this study, therefore, we asked: Does the Skeletal Oncology Research Group (SORG) algorithm accurately predict 5-year survival in an independent patient population surgically treated for chondrosarcomas?

Patients and Methods

Study Design

This retrospective study was approved by our institutional review board.

For the development of the SORG algorithm, Thio et al. [27], included 1544 chondrosarcoma patients in the development cohort from the Surveillance, Epidemiology, and End Results (SEER) database, which is the cancer registry of the National Cancer Institute that covers approximately 30% of the US population. The SEER database consists of demographic data, tumor characteristics, treatment data, and patient survival data. Using four different machine learning techniques, four different algorithms were created for the prediction of 5-year survival in chondrosarcoma patients. These algorithms were compared by measures of discrimination, calibration, and overall performance. First, discrimination was graphically shown using a receiver operating characteristic curve and numerically by calculating the area under the curve, also known as the concordance (c-statistic) index for binary classification. The c-statistic ranges from 0.5 (no better than a coin-toss) to 1.0 (perfect discrimination). Second, a calibration plot was made, and the slope and intercept were calculated, which presents the agreement between the predicted probability of 5-year survival on the x-axis versus the actual observed proportions of 5-year survival on the y-axis [25, 26]. A perfect calibration plot should show a 45° upwards line, with an intercept of 0 and a slope of 1. Third, the Brier Score and the null-model Brier score were used to calculate the overall algorithm performance. The Brier score ranges from 0 (excellent prediction) to 1 (worst prediction). The Brier score was compared with the null-model Brier score, which shows the performance of the model that ignored all the covariates. A Brier score lower than the null model Brier score indicated greater performance of the algorithm. The model that performed the best was chosen to be the SORG algorithm, which is available at https://sorg-apps.shinyapps.io/chondrosarcoma.

For the external validation an F1-score was added to measure the overall accuracy of the algorithm, which ranges between 0 (total failure of an algorithm) and 1 (perfect algorithm). The F1-score computes the harmonic mean of precision and recall [10], with precision being the proportion of the true positives on all positive predictions, and where a precision of 1 means that there are no false positives. The recall is the proportion of true positives on all actual positive elements; a recall of 1 means that are no false negative predictions.

For the development of the institutional cohort, the electronic medical records of patients who underwent surgery for a primary histologically confirmed chondrosarcoma at one of two tertiary referral centers from January 1992 to August 2013 were reviewed. Overall, 1546 electronic medical records were manually reviewed to ascertain whether the inclusion criteria were met. A minimum followup duration of 5 years was required to establish certainty of survival. We included patients with the word "chondrosarcoma" in the pathology report at one of our two tertiary care centers. We excluded patients treated outside of the two tertiary care centers or whose first surgery for chondrosarcoma was not performed at these institutions (n = 807), patients who declined surgery (n =5), patients with a chondrosarcoma of the soft tissue, brain, or skull (n = 400), and international patients for whom survival status could not be determine through the Social Security death index (n = 7).

Demographics, Description of Study Population

Between 1992 and 2013, 326 patients were treated surgically for any type of extracranial chondrosarcoma of the bone at two tertiary care referral centers. Of those, 179 were accounted for at a minimum of 5 years after diagnosis in a clinical note at one of the two institutions, unless they died earlier, and were included in the validation cohort; a total of 147 (45%) did not meet the minimum of 5 years of followup at the institution and were not included in the validation cohort of the SORG algorithm. The outcome (survival at 5 years) was checked for all 326 patients in the Social Security death index and all were included in a supplementary validation cohort, to also ascertain validity for patients with less than 5 years of institutional followup. The median age at surgery was 54 years (interquartile range [IQR], 41–66) and 102 of 179 patients (57%) were men. These variables were not different between the validation and derivation cohorts (Table 1). The 5-year survival was lower in the validation cohort than it was in the derivation cohort from SEER (61% [110 of 179] versus 76% [1131 of 1544]; $p < 0.001$) The institutional cohort had more patients with dedifferentiated chondrosarcoma compared with the derivation cohort (27% [49 of 179] versus 9% [131 of 1544]; $p < 0.001$). Five-year survival for patients with conventional chondrosarcomas was not different from the derivation cohort (76% [99 of 130] versus 81% [1104 of 1357]; $p < 0.161$). Also, for dedifferentiated chondrosarcoma, 5-year survival was not different between the

Table 1. Baseline characteristics of patients from the developmental data set and institutional external validation data set

Variable	Development cohort (n = 1544); n (%); median (IQR)	Institutional validation cohort (n = 179); n (%); median (IQR)	p value	Supplementary validation cohort (n = 326); n (%); median (IQR)	p value
Sex					
Female	731 (47)	77 (43)	0.272	150 (46)	0.661
Male	813 (53)	102 (57)		176 (54)	
Age (years)	52 (40-64)	54 (41-66)	0.105	53 (40-63)	0.990
Histologic subtype					
Conventional chondrosarcoma	1413 (91)	130 (73)	< 0.001*	269 (82)	< 0.001*
Dedifferentiated chondrosarcoma	131 (9)	49 (27)		57 (18)	
Size†	70 (42-105)	82.0 (55-128)	< 0.001*	69 (41-110)	0.822
Grade†			0.024*		0.343
1 (well differentiated)	592 (41)	54 (30)		133 (41)	
2 (moderately differentiated)	588 (41)	83 (46)		141 (44)	
3 (poorly differentiated)	276 (19)	42 (24)		50 (15)	
Tumor extension†					
Localized	851 (57)	153 (86)	< 0.001*	192 (88)	< 0.001*
Regional extension	557 (37)	13 (7)		14 (6)	
Distant metastasis	95 (6)	13 (7)		12 (6)	
Location					
Extremities	915 (59)	66 (37)	< 0.001*	140 (43)	< 0.001*
Rib, sternum, clavicle	289 (19)	47 (26)		78 (24)	
Pelvis	264 (17)	50 (28)		74 (23)	
Spine	76 (5)	16 (9)		34 (10)	
5-year survival	1131 (76)	110 (61)	< 0.001*	242 (74)	0.486
Conventional chondrosarcoma†	1104 (81)	99 (76)	0.161	227 (84)	0.239
Dedifferentiated chondrosarcoma	27 (21)	11 (22)	0.839	15 (26)	0.403

*Indicates significance (two-tailed p value below 0.05);
†there were no missing values in the institutional validation cohort; in the development set, values for tumor size were available for 1201 (78%) of patients, tumor grade for 1456 (94%), tumor extension for 1503 (97%), and survival rates for 1488 with conventional chondrosarcomas (96%) [17]. In the supplementary validation cohort values were missing in: one patient for size, two patients for tumor grade, and 108 for tumor extension.

validation cohort compared to the derivation cohort (22% [11 of 49] versus 21% [27 of 131]; p = 0.839). The median tumor size was larger in the institutional population than in the derivation cohort (82 mm (IQR, 55–128) versus 70 mm (IQR, 42–105); p < 0.001). Similarly, the institutional population had a greater proportion of patients with Grades 2 and 3 chondrosarcoma than the derivation cohort (46% [83 of 179 Grade 2] and 24% [42 of 179 Grade 3] versus 41% [588 of 1544 Grade 2] and 19% [276 of 1544 Grade 3]; p = 0.024). A greater percentage of patients in the institutional population had localized disease than did those in the derivation cohort (86% [153 of 179] versus 57% [851 of 1544]; p < 0.001). Finally, the institutional cohort had more patients with tumors in the rib, sternum, clavicle, pelvis, and spine relative to the extremities than the derivation cohort did (63% [113 of 179] versus 41% [629 of 1544]; p < 0.001). These differences between the study groups emphasize that that the external validation is performed not only in a different patient cohort, but in a cohort with different disease characteristics.

Variables, Outcome Measures, Data Sources, and Bias

The outcome variable was survival at 5 years, defined as the period from the date of surgery to the date of death of any cause. The Social Security Death Index was used to identify a date of death for each patient if this was not explicitly noted in the electronic medical record [22]. The minimum 5-year followup period for survivors was checked by noting the date of the last clinical report in the electronic medical record. The following explanatory variables were collected manually (MB, MS) from the patients' electronic medical records and postoperative musculoskeletal pathology

reports, which were needed to complete the SORG algorithm [27]: sex (male or female), age (years at surgery), histologic subtype (conventional chondrosarcoma or dedifferentiated chondrosarcoma), tumor grade reported by one of the two institutional surgical pathology departments (well differentiated, moderately differentiated, or poorly differentiated), tumor size (mm), tumor extension (localized, extraosseous extension, or distant metastasis), and location (extremities; spine, pelvic bones, and sacrum; and rib, sternum, and clavicle). Explanatory variables and outcomes were collected at different times in different databases.

Statistical Analysis, Study Size

Categorical variables are presented as frequencies and percentages, and continuous variables are presented as medians and IQRs. We used the chi-square and Mann-Whitney U tests to compare baseline characteristics between the SEER database and the validation database.

The validation of the institutional cohort was a complete case analysis, only in the supplementary validation cohort missing data were imputed for size, grade, and tumor extension using the nonparametric missForest method [24].

We estimated the predicted probabilities of 5-year survival by applying the SORG algorithm to each patient in the validation set.

To measure the performance of the SORG algorithm in the independent validation cohort, we used the same metrics as those used during the development of the SORG algorithm; discrimination using the c-statistic, calibration using the calibration plot, and overall performance using the Brier score and the null model Brier score as explained above [27].

The following software programs were used for data analysis and model validation: Microsoft Excel and Microsoft Azure (Redmond, WA, USA), the Anaconda Distribution (Continuum Analytics, Austin, TX, USA) with RStudio (Version 1.0.153, Boston, MA, USA), Python Version 3.6 (Python Software Foundation, Wilmington, DE, USA), and StataCorp 2013 (Stata Statistical Software: Release 13; StataCorp LP, College Station, TX, USA).

Results

On external validation, the SORG machine learning algorithm proved to be accurate but systematically overestimated 5-year survival. The algorithm had a validation c-statistic of 0.87 (95% CI, 0.80–0.91) (Fig. 1A). Calibration using calibration plots resulted in a calibration slope of 0.97 (95% CI, 0.68–1.3) and a calibration intercept of -0.58 (95% CI, -0.20 to -0.97) (Fig. 2A). The Brier score for the overall algorithm performance was 0.152, with a null-model Brier score of 0.237. The F1-score was 0.836.

For the supplementary validation in the total of 326 patients, the SORG algorithm had similar results and proved to be accurate and showed better results on overall performance and accuracy, which resulted in less

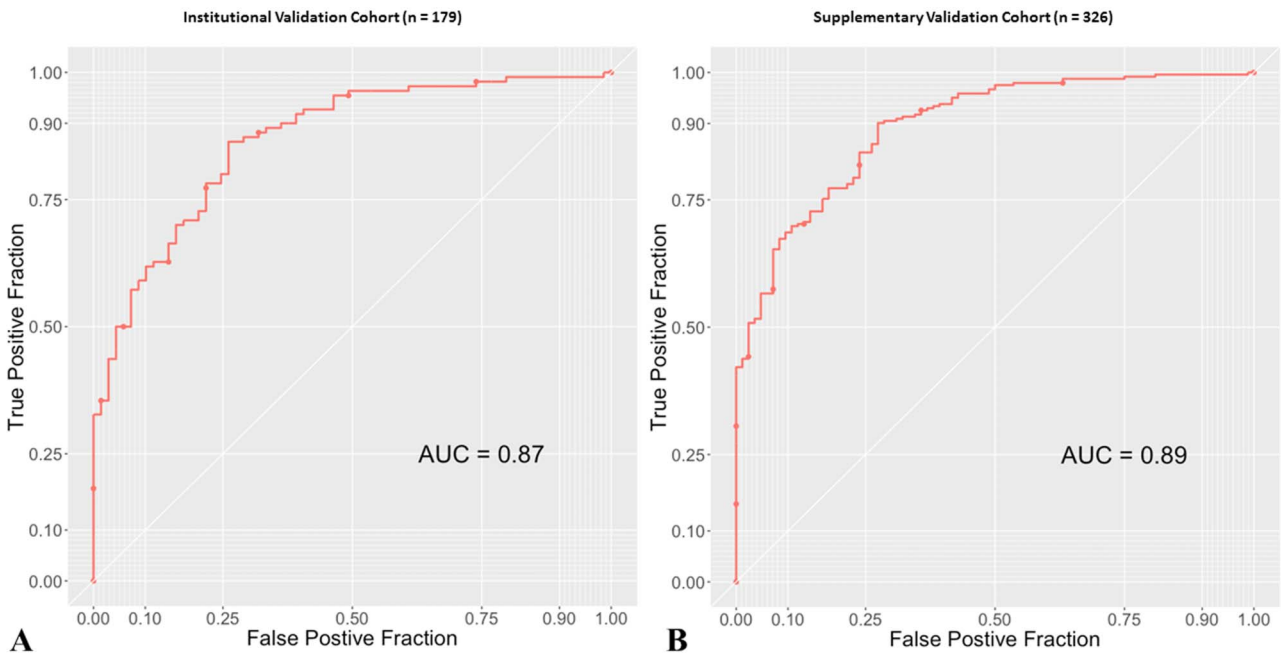

Fig. 1 A-B This figure shows the receiver operating curve for 5-year survival for the performance of the SORG algorithm using **(A)** the independent validation cohort (n = 179) and **(B)** the supplemental validation cohort (n = 326).

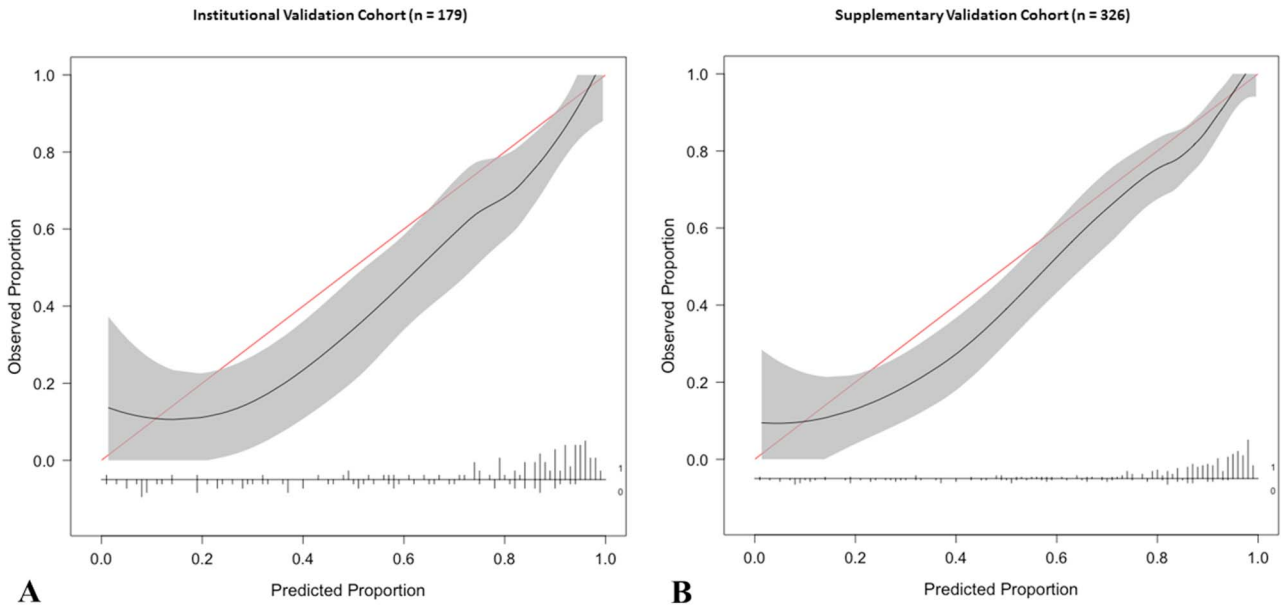

Fig. 2 A-B This figure shows the calibration with 95% CIs for 5-year survival for the SORG algorithm using **(A)** the independent validation cohort (n = 179) and **(B)** the supplemental validation cohort (n = 326).

overestimation of 5-year survival. The supplementary validation had a c-statistic of 0.89 (95% CI, 0.85–0.93) (Fig. 1B). The calibration slope was 1.13 (95% CI, 0.87-1.39) and the calibration intercept was -0.26 (95% CI, -0.57 to 0.06) (Fig 2B). The Brier score was 0.11, with a null-model Brier score of 0.19. The F1-score was 0.901. To check if the algorithm is accurate for the subpopulations of grade, tumor location, and tumor histology, the c-statistics were calculated separately in the supplementary cohort. This showed good accuracy for all subpopulations except for tumor grade 3, which was slightly inaccurate. The c-statistic for Grade 1, Grade 2, and Grade 3 was: 0.86; 0.76; and 0.63 respectively. For tumor location the c-statistic for extremities; rib/sternum/clavicle; pelvis; and spine was: 0.93; 0.83; 0.83; and 0.93 respectively. For tumor histology the c-statistic for conventional, and dedifferentiated was: 0.84, and 0.80 respectively.

Discussion

Survival prognostication in patients with chondrosarcoma, the second most prevalent primary bone tumor [6, 18], may aid both patients and clinicians in decision making, such as when deciding between treatment options or whether to enroll in randomized, controlled trials with prospective matching. SORG created a prognostication algorithm using a machine learning technique that can help predict 5-year mortality in patients with chondrosarcoma [27]. However, this algorithm has not yet been externally validated. This external validation is of utmost importance because even though a model performs accurately on internal validation, the external validation results can be disappointing and may be misleading when applied to different populations [2, 11, 19]. We therefore sought to assess how the SORG algorithm performs in an independent dataset. We found that the results from validating the SORG algorithm in our institutional cohort match the results from the development cohort and noted that the SORG algorithm retained a good discriminative ability and overall performance but overestimated survival. In the subgroup analysis, we found that the algorithm kept good discriminative values except for Grade 3 tumors. We performed a secondary validation in a cohort that also includes patients who did not meet 5 years of followup in our institution's data, but for whom survival status at 5 years was available in the Social Security death index. The reason for this analysis was to ascertain whether or not excluding patients with less than 5 years of institutional followup biased the study. This supplemental validation showed similar results as the institutional validation analysis but with less overestimation of survival.

There are a number of limitations to our study. First, this study was performed using a dataset approximately 10-fold smaller than the development dataset [27], with numerous baseline characteristics that were different from the derivation cohort except age and sex. This is largely because the institutional population was drawn from two tertiary care centers where patients with more severe disease are often referred for multidisciplinary management, as witnessed by a higher proportion of patients with dedifferentiated chondrosarcoma, larger tumor sizes, and higher

tumor grades than patients in the derivation cohort. This difference was likely because of the size of the validation cohort and the proportional differences in the baseline characteristics of the two populations. These differences between the study groups emphasize that the external validation is performed not only in a different patient cohort but also in a cohort with different disease characteristics. Second, because of the small effective sample size of the institutional cohort, the suggested minimum of 100 events for external validation by Vergouwe et al. [28] for both outcome groups was met for the survivors (n = 110) but could not be met by the nonsurvivors (n = 69) in the validation cohort. Despite the differences in tumor characteristics, and the small effective sample size, the algorithm performed well in discriminating between survivors and nonsurvivors and on overall performance, but plotting the calibration plot resulted in a calibration intercept of -0.58, compared with 0.001 in the original dataset, which caused the algorithm to systematically underestimate 5-year mortality in the institutional population, as shown by the calibration plot (Fig. 2A). However, by increasing the study size in the supplemental validation cohort, the calibration intercept improved to -0.26, which reduces the underestimation of the algorithm. Third, another limitation of the small effective sample size is that improving the model by recalibration was not possible [12, 21]; also assessment of possible prognostic factors, subsequently creating another machine-learning algorithm using our institutional dataset, could not be achieved because it would not be possible to control for confounding variables. Due to the rarity of the disease [20], a larger validation group with available tumor characteristics and a minimum of 5 years of followup is difficult to achieve. In the future, larger studies can consider recalibrating the SORG algorithm to further improve performance on calibration. Fourth, the grading system used in the SORG algorithm, which grades both differentiated and dedifferentiated from Grade 1 to 3, does not reflect the grading system used in practice, which only grades differentiated chondrosarcoma from Grade 1 to 3; dedifferentiated chondrosarcoma is not graded. Future studies can also seek to improve this issue. Fifth, because the timeframe from 1992 to 2013 in which patients were included is very broad, one might ask if the cohort on which this algorithm was built and externally validated may have been treated differently than the (future) population in which the algorithm will be used. However, because treatment options for chondrosarcoma have changed very little in the past 30 years, and because of chondrosarcomas' resistance to chemotherapy [5, 9, 13, 18], we believe that the SORG algorithm validation is still applicable and can be used for patients primarily treated with surgery. Nonetheless, future research with prospective validation of the algorithm is desirable. Sixth, despite being an independent validation cohort, both the original database and the validation database primarily consist of patients from the United States; hence, validation in a nonAmerican population would also be warranted. Tumors may behave differently among different races/ethnicities, and (surgical) treatment in other parts of the world may be very different. Such patients will be very valuable to include in future studies to train and recalibrate this algorithm, suggesting the need for future multiinstitutional collaboration.

Seventh, the SORG algorithm does not include patients who choose nonoperative treatment. Because the algorithm is based on information that would not have been available without surgery, it should not be used for patients in a nonsurgical setting or used with information derived from imaging studies and biopsies alone. Because surgery is the primary treatment option for chondrosarcomas not involving the skull base [3, 5, 13, 15, 18], a machine learning algorithm to predict the outcomes of the disease in a patient who opts for nonoperative treatment may be less useful in general practice. Additionally, surgical margins were not used as a variable in developing the SORG algorithm because this variable was not available in the national database that was used for creating the original algorithm. Surgical margins had prognostic value for survival in some series [9, 15], although others did not find this to be true [1, 8]. Future research on this topic, most likely multicenter, retrospective studies using a larger number of patients, can focus on identifying and including other predictors such as surgical margins, and training this model with the addition of new data may improve it.

The SORG machine learning algorithm to predict 5-year survival in patients who underwent surgical resection of primary chondrosarcoma of the bone shows a good discriminative ability and overall performance but overestimated survival on external validation.

The algorithm published online by the authors of the SORG machine learning algorithm may be useful in clinical practice for decision making by both patients and physicians. This application can be accessed at https://sorg-apps.shinyapps.io/chondrosarcoma/. External validation is also important because discrimination, calibration, and overall performance might differ across populations [9], making an unvalidated algorithm less reliable. Further external validation, especially in a nonAmerican population, is encouraged to establish better understanding of the performance, strengths, and weaknesses of the algorithm.

Acknowledgments We thank the contributors to the Surveillance, Epidemiology, and End Results Program of the National Cancer Institute for use of their data for the development of the SORG algorithm.

References

1. Andreou D, Ruppin S, Fehlberg S, Pink D, Werner M, Tunn PU. Survival and prognostic factors in chondrosarcoma: results in

115 patients with long-term follow-up. *Acta Orthop.* 2011;82: 749–755.
2. Angelini A, Guerra G, Mavrogenis AF, Pala E, Picci P, Ruggieri P. Clinical outcome of central conventional chondrosarcoma. *J Surg Oncol.* 2012;106:929–937.
3. Bhatt AD, Jacobson A, Lee RY, Giraud C, Schwab JH, Hornicek FJ, Nielsen P, Choy E, Harmon D, DeLaney TF, Chen Y-LE. High-dose proton beam–based radiation therapy in the management of extracranial chondrosarcomas. *Int J Part Ther.* 2016;3: 373–381.
4. Bindiganavile S, Han I, Yun JY, Kim H-S. Long-term outcome of chondrosarcoma: a single institutional experience. *Cancer Res Treat.* 2015;47:897–903.
5. Dickey ID, Rose PS, Fuchs B, Wold LE, Okuno SH, Sim FH, Scully SP. Dedifferentiated chondrosarcoma: the role of chemotherapy with updated outcomes. *J Bone Joint Surg Am*. 2004; 86:2412–2418.
6. Dorfman HD, Czerniak B. Bone cancers. *Cancer.* 1995;75: 203–210.
7. Evans HL, Ayala AG, Romsdahl MM. Prognostic factors in chondrosarcoma of bone. a clinicopathologic analysis with emphasis on histologic grading. *Cancer*. 1977;40:818–831.
8. Fiorenza F, Abudu A, Grimer RJ, Carter SR, Tillman RM, Ayoub K, Mangham DC, Davies a M. Risk factors for survival and local control in chondrosarcoma of bone. *J Bone Joint Surg Br.* 2002; 84:93–99.
9. Frezza AM, Cesari M, Baumhoer D, Biau D, Bielack S, Campanacci DA, Casanova J, Esler C, Ferrari S, Funovics PT, Gerrand C, Grimer R, Gronchi A, Haffner N, Hecker-Nolting S, Höller S, Jeys L, Jutte P, Leithner A, San-Julian M, Thorkildsen J, Vincenzi B, Windhager R, Whelan J. Mesenchymal chondrosarcoma: prognostic factors and outcome in 113 patients. A European musculoskeletal oncology society study. *Eur J Cancer*. 2015;51:374–381.
10. Goutte C, Gaussier E. A probabilistic interpretation of precision, recall and F-score, with implication for evaluation. *Lecture notes in computer science.* 2005:345–359.
11. Healey JH. CORR Insights(R): Can machine-learning techniques be used for 5-year survival prediction of patients with chondrosarcoma? *Clin Orthop Relat Res.* 2018;476:2049–2051.
12. van Houwelingen HC. Validation, calibration, revision and combination of prognostic survival models. *Stat Med.* 2000;19: 3401–3415.
13. Italiano A, Mir O, Cioffi A, Palmerini E, Piperno-Neumann S, Perrin C, Chaigneau L, Penel N, Duffaud F, Kurtz JE, Collard O, Bertucci F, Bompas E, Le Cesne A, Maki RG, Ray I, Blay JY. Advanced chondrosarcomas: role of chemotherapy and survival. *Ann Oncol.* 2013;24:2916–2922.
14. Jović S, Miljković M, Ivanović M, Šaranović M, Arsić M. Prostate cancer probabilityprediction by machine learning technique. *Cancer Invest.* 2017;35:647–651.
15. Lee FY, Mankin HJ, Fondren G, Gebhardt MC, Springfield DS, Rosenberg a E, Jennings LC. Chondrosarcoma of bone: an assessment of outcome. *J Bone Joint Surg Am.* 1999;81: 326–338.
16. Lynch CM, Abdollahi B, Fuqua JD, de Carlo AR, Bartholomai JA, Balgemann RN, van Berkel VH, Frieboes HB. Prediction of lung cancer patient survival via supervised machine learning classification techniques. *Int J Med Inform.* 2017;108:1–8.
17. Montazeri M, Montazeri M, Montazeri M, Beigzadeh A. Machine learning models in breast cancer survival prediction. *Technol Health Care*. 2016;24:31–42.
18. Nazeri E, Gouran Savadkoohi M, Majidzadeh -A K, Esmaeili R. Chondrosarcoma: an overview of clinical behavior, molecular mechanisms mediated drug resistance and potential therapeutic targets. *Crit Rev Oncol Hematol.* 2018;131:102–109.
19. Ohnuma T, Uchino S. Prediction models and their external validation studies for mortality of patients with acute kidney injury: a systematic review. *PLoS One*. 2017;12:1–14.
20. van Praag VM, Rueten-Budde AJ, Ho V, Dijkstra PDS, van der Geest IC, Bramer JA, Schaap GR, Jutte PC, Schreuder HB, Ploegmakers JJW, Fiocco M, van de Sande MAJ. Incidence, outcomes and prognostic factors during 25 years of treatment of chondrosarcomas. *Surg Oncol.* 2018;27:402–408.
21. Royston P, Altman DG. External validation of a cox prognostic model: principles and methods. *BMC Med Res Methodol.* 2013; 13:33.
22. Social Security Administration. *Social Security Death Index*. 2014. Available at: http://www.death-records.net/ssdi.htm.
23. Song K, Shi X, Wang H, Zou F, Lu F, Ma X, Xia X, Jiang J. Can a nomogram help to predict the overall and cancer-specific survival of patients with chondrosarcoma? *Clin Orthop Relat Res.* 2018; 476:987–996.
24. Stekhoven DJ, Bühlmann P. MissForest–non-parametric missing value imputation for mixed-type data. *Bioinformatics*. 2012;28: 112–118.
25. Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. *Eur Heart J.* 2014;35:1925–1931.
26. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, Pencina MJ, Kattan MW. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology*. 2010;21:128–138.
27. Thio QCBS, Karhade A V, Ogink PT, Raskin KA, De Amorim Bernstein K, Lozano Calderon SA, Schwab JH. Can machine-learning techniques be used for 5-year survival prediction of patients with chondrosarcoma? *Clin Orthop Relat Res.* 2018;476: 2040–2048.
28. Vergouwe Y, Steyerberg EW, Eijkemans MJC, Habbema JDF. Substantial effective sample sizes were required for external validation studies of predictive logistic regression models. *J Clin Epidemiol.* 2005;58:475–483.