acta orthopaedica belgica, 2024, 90, 493-501

doi.org/10.52628/90.3.12636

External validation of the SORG machine learning for 90-day and 1-year mortality in patients suffering from extremity metastatic disease in an European cohort of 174 patients

T.M. de GROOT[1,2,3,4], A.A. SOMMERKAMP[1,2,3], Q.C.B.S. THIO[4], A.V. KARHADE[4], O.Q. GROOT[4], J.H.F. OOSTERHOF[4], F.F.A. IJPMA[2], P.M.A. VAN OOIJEN[3], J.J.W. PLOEGMAKERS[1,2,3], P.C. JUTTE[1,2,3], J.H. SCHWAB[4], J.N. DOORNBERG[1,2,3,] — *On behalf of the Machine Learning Consortium:* K. AKSAKAL, B. BARVELINK, B. BEUKER, A.E. BULTRA, L. OLIVIERA E CARMO, J. COLARIS, A. DUCKWORTH, K. TEN DUIS, E. FENNEMA, M. GORDON, J. HARBERS, R. HENDRICKX, M. HENG, S. HOEKSEMA, M. HOGERVORST, B. JADAV, J. JIANG, G. KERKHOFFS, J. KUIPERS, C. LAANE, D. LANGERHUIZEN, B. LUBBERTS, W. MALLEE, H. MHMUD, M. EL MOUMNI, P. NIEBOER, K. OUDE NIJHUIS, J. OLCZAK, P. VAN OOIJEN, J. RAWAT, D. RING, S. SCHILSTRA, S. SPRAGUE, S. STUFKENS, E. TIJDENS, P. VAN DER VET, J.-P. DE VRIES, K. WENDT, M. WIJFFELS, D. WORSLEY.

[1]Department of Orthopaedics, Groningen, The Netherlands, University of Groningen, University Medical Center Groningen, The Netherlands; [2]Department of Surgery, Groningen, The Netherlands, University of Groningen, University Medical Center Groningen; [3]Data Science Center in Health, Groningen, University of Groningen, University Medical Center Groningen, The Netherlands; [4]Department of Orthopaedic Surgery, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA.

Correspondence at: Tom de Groot, BSc, Department of Orthopaedic Surgery, Groningen, University Medical Center Groningen, The Netherlands, Email: t.m.de.groot@umcg.nl

Accurate survival prediction of patients with long-bone metastases is challenging, but important for optimizing treatment. The Skeletal Oncology Research Group (SORG) machine learning algorithm (MLA) has been previously developed and internally validated to predict 90-day and 1-year survival. External validation showed promise in the United States and Taiwan. To ensure global generalizability, the algorithm remains to be validated in Europe. We therefore asked: does the SORG-MLA for long-bone metastases accurately predict 90-day and 1-year survival in a European cohort?

One-hundred seventy-four patients undergoing surgery for long-bone metastases between 1997-2019 were included at a tertiary referral Orthopaedic Oncology Center in the Netherlands. Model performance measures included discrimination, calibration, overall performance, and decision curve analysis.

The SORG-MLA retained reasonable discriminative ability, showing an area under the curve of 0.73 for 90-day mortality and 0.77 for 1-year mortality. However, the calibration analysis demonstrated overestimation of European patients' 90-day mortality (calibration intercept -0.54, slope 0.60). For 1-year mortality (calibration intercept 0.01, slope 0.60) this was not the case. The Brier score predictions were lower than their respective null model (0.13 versus 0.14 for 90-day; 0.20 versus 0.25 for 1-year), suggesting good overall performance of the SORG-MLA for both timepoints.

The SORG-MLA showed promise in predicting survival of patients with extremity metastatic disease. However, clinicians should keep in mind that due to differences in patient population, the model tends to underestimate survival in this Dutch cohort. The SORG model can be accessed freely at https://sorg-apps.shinyapps.io/extremitymetssurvival/

INTRODUCTION

Adequate treatment of bone metastasis is of importance for the mobility and overall quality of life of the patient, while keeping patients' preferences and values in mind in shared decision making. Optimal surgical intervention in bone metastasis in the extremities relies on accurate estimation of survival of the patient. In general, it is believed that patients with short life expectancy will not benefit from invasive surgery and therefore may be treated nonoperatively. In this decision-making context, it is well-established that patient survival at 90-day and 1-year after the surgical procedure are of primary importance. Patients with a life expectancy of more than 90 days are more likely to benefit from more extensive surgery and durable reconstruction[1,2].

Surgeons' prediction of survival for these timepoints is challenging and proves sub-optimal as it suffers from bias[3]. Over the past decades, numerous prediction tools have been developed to help physicians estimate the prognosis of their patients[4-7]. These models have been performing reasonably well, though it is important to keep them up to date. Individualized models would allow physicians to better discuss treatment options with patients and their relatives[8]. Thio et al.[9] have developed and internally validated the Skeletal Oncology Research Group machine learning algorithm (SORG-MLA) to predict one patient's postoperative 90-day and 1-year survival. Besides using clinical variables such as tumor characteristics and patient demographics, it also uses a wide range of laboratory values that were found to be predictive in statistical analysis[10,11], therefore providing an individual approach to determine the trajectory of care. The SORG-MLA showed great promise on external validation in Iowa (United States) and Taiwan[12,13]. However, external validation of a predictive tool should be mandatory to establish whether the tool works satisfactorily in different patient populations[14]. To this day, the SORG-MLA remains to be externally validated in a European cohort.

Therefore, our study question was: does the SORG-MLA for long-bone metastases accurately predict 90-day and 1-year mortality in a European cohort?

PATIENTS AND METHODS

This study was conducted according to the Guidelines for Developing and Reporting Machine Learning Predictive Models in Biomedical Research[15] and the Transparent Reporting of Multivariable Prediction Models for Individual Prognosis or Diagnosis (TRIPOD) guidelines[16].

For safe multicenter data exchange and analysis, our Machine Learning Consortium adhered to World Healthcare regulations: "Policy on use and sharing of data collected in Member States by the World Health Organization outside the context of public health emergencies"[17].

The SORG-MLA was developed by Thio et al.[9], who included 1090 patients from the Massachusetts General Hospital in Boston, Massachusetts, USA. By using 5 different machine learning techniques five algorithms were created, each predicting the 90-day and 1-year mortality of patients suffering from long-bone metastatic disease. After development, these algorithms were then compared to one another by means of evaluating their efficacy through measuring discrimination, calibration, and overall performance.

This retrospective cohort design included patients with metastatic long-bone lesions that underwent surgery between 1997 and 2019 at a tertiary orthopaedic oncology center in the Netherlands. Medical records were manually assessed for patient demographics, operation notes, and follow-up data. Exclusion criteria were age below 18 years of age at the time of surgery; surgery conducted at a non-tertiary hospital; surgery less than 90 days ago; 90-day loss to follow-up; and proven or suspected primary bone tumors. In general, surgery was performed in patients considered fit for surgery based on a multidisciplinary assessment by a medical oncologist, anesthesiologist and orthopedic surgeon, and the presence of a pathological or impending fracture. An impending pathological fracture was diagnosed by an orthopedic surgeon, assessing clinical and radiographical features of the lesion.

The same inclusion and exclusion criteria used in the development study[9] were applied, resulting in 225

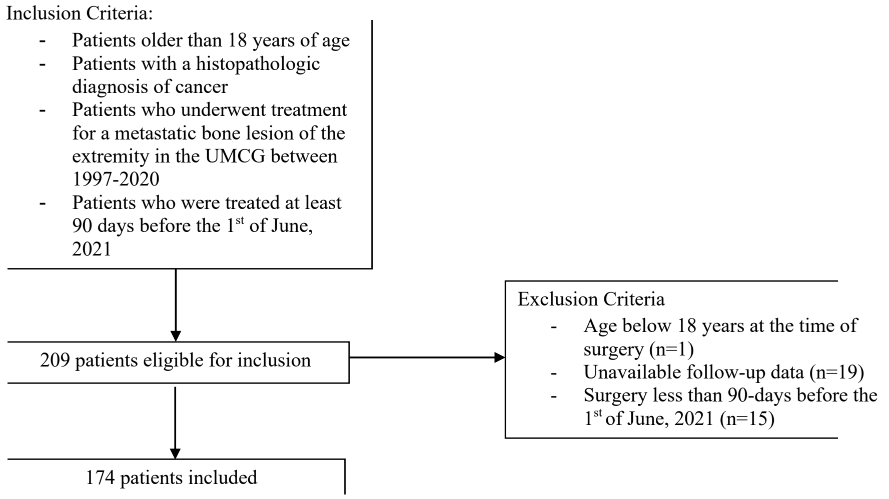

Inclusion Criteria:
- Patients older than 18 years of age
- Patients with a histopathologic diagnosis of cancer
- Patients who underwent treatment for a metastatic bone lesion of the extremity in the UMCG between 1997-2020
- Patients who were treated at least 90 days before the 1st of June, 2021
209 patients eligible for inclusion
Exclusion Criteria
- Age below 18 years at the time of surgery (n=1)
- Unavailable follow-up data (n=19)
- Surgery less than 90-days before the 1st of June, 2021 (n=15)
174 patients included

Fig. 1. — Flowchart visualizing the enrollment of the patients in the external validation cohort

Table I. — Patient baseline characteristics of the validation cohort in comparison with the development cohort

Variable	**Validation cohort (n = 174)**	**Development cohort (n = 1090)**	**Missing values in validation cohort**	***p*-value**
	n (%); median (IQR)		**n (%)**	
Age (years)	63 (56-70)*	63 (54-72)*	-	0.75
Female sex	92 (53)	610 (56)	-	0.50
BMI (kg/m^2)	25 (23-29)	27 (23-30)	44 (25)	0.20
Charlson comorbidity	89 (51)	584 (54)	6 (3)	0.90
Primary tumor type			-	**0.04**
Slow growth	59 (34)	460 (42)		
Moderate growth	56 (32)	263 (24)		
Rapid growth	59 (34)	367 (34)		
Pathologic fracture	83 (48)	594 (55)	1 (1)	0.13
ECOG score			72 (41)	**<0.001**
0-2	100 (58)	360 (85)		
3-4	2 (1)	62 (15)		
Tumor location			-	0.30
Upper extremity	34 (20)	255 (23)		
Lower extremity	140 (80)	835 (77)		
Other bone metastases	129 (74)	845 (78)	4 (2)	0.71
Spine metastases	90 (52)	626 (57)	4 (2)	0.31
Visceral metastases	76 (44)	487 (45)	5 (3)	1.00
Brain metastases	9 (5)	175 (16)	4 (2)	**<0.001**
Previous systemic therapy	92 (53)	676 (62)	8 (5)	0.12
Local radiation	54 (31)	194 (18)	7 (4)	**<0.001**
Laboratory Values				
Absolute lymphocyte count (10^3/uL)	1.4 (0.8-1.6)	1 (1-2)	112 (64)	**0.03**
Absolute neutrophil count (10^3/uL)	5.7 (3.6-7.4)	5 (4-8)	112 (64)	0.82
Albumin level (g/dL)	4.0 (3.6-4.3)	4 (3-4)	58 (33)	**<0.001**
Alkaline phosphatase level (IU/L)	105 (81-146)	101 (74-146)	50 (29)	0.21
Calcium (mg/dL)	9.6 (9.3-10.0)	9 (9-10)	51 (29)	**<0.001**
Creatinine (mg/dL)	0.8 (0.6-1.0)	0.8 (0.7-1.1)	29 (17)	0.44
Hemoglobin level (g/dL)	11.9 (10.5-13.2)	11 (10-13)	22 (13)	**0.002**
Neutrophil-to-lymphocyte ratio	4.0 (2.5-6.7)	5(3-9)	116 (67)	**0.02**
Platelet count (10^3/uL)	278 (215-354)	251 (184-332)	30 (17)	**0.009**
Platelet-to-lymphocyte ratio	224 (131-281)	234 (158-374)	112 (64)	0.09
Sodium (mg/dL)	138 (136-141)	138 (136-140)	32 (18)	**0.03**
White blood cell count (10^3/uL)	8.2 (5.5-11.1)	7 (5-10)	34 (20)	0.05
Post-surgery mortality				
90-day	27 (16)	305 (29)	3 (2)	**<0.001**
1-year	81 (51)	639 (62)	14 (8)	**0.008**

IQR = Interquartile range; kg/m^2 = kilograms meter squared; ECOG = Eastern Cooperative Oncology Group; p-values were calculated using the student t test for parametric continuous variables and Mann-Whitney U test for non-parametric continuous variables, the Fisher's exact test for dichotomous variables and the Chi-squared test for ordinal data. Bold indicates a statistically significant difference of 0.05.

patients who underwent surgical treatment for a long-bone metastasis. A total of 174 patients were included in the analysis after excluding 16% (35/225) patients (Fig. 1).

In accordance with the developmental cohort of the SORG-MLA, the primary outcomes were 90-day and 1-year mortality, which was defined as the time between a patient's first surgical treatment for a long-bone metastasis and death by any cause. In patients without a recorded date of death in their medical record, minimum survival was derived from the date of the last recorded follow-up contact with the patient. In some cases, more accurate vital data could be obtained by contacting the patient's general practitioner. Loss to follow-up occurred in 2% (3/174) for 90-day survival, and in 8% (14/174) for 1-year survival. All other required predictive variables used in the SORG-MLA to predict 90-day and 1-year survival were manually obtained from the medical record: sex; age; body mass index (BMI) (kg/m^2); histologic subtype (classified into the following three groups: slow growth, moderate growth and rapid growth, when applying the definitions stated by Katagiri et al.[18]); tumor location; visceral metastases; presence of pathological fracture; preoperative laboratory values; and previous local radiation or systemic therapy. Systemic therapy was defined as having received at least one of the following: chemotherapy, targeted therapy, hormone therapy, and/ or immunotherapy. The presence of visceral metastases (liver and/or lung), brain-, spine-, and other bone metastases was confirmed by reviewing radiology reports in medical records. Preoperative laboratory values within 3 weeks of surgery were also manually obtained and consisted of absolute lymphocyte count (x 103/uL), absolute neutrophil count (x 103/uL), albumin level (g/dL), and alkaline phosphatase (IU/L), calcium (mg/dL), creatinine (mg/dL), hemoglobin level (g/dL),), neutrophil-to-lymphocyte ratio, platelet count (x 103/uL platelet-to-lymphocyte count, sodium (mg/dL) levels, and white blood cell count (x 103/ uL).

The median age was 63 years (interquartile range [IQR], 54-72; Table I). Ninety-two were female (%) and 82 (%) were male. The median BMI was 25 kg/ m2 (IQR 23-29). Regarding tumor characteristics, 34% (59/174) had a slow-growth tumor; 32% (56/174) had a moderate-growth tumor, and 34% (59/174) had a rapid-growth tumor. The most common primary tumors were breast cancer (26%; 45/174) and lung cancer (21%; 37/174). In most patients (80%), the operative procedure concerned the lower extremity. Forty-eight percent of patients (84/174) were treated with intramedullary nailing; followed by prosthetic reconstruction; 34% (59/174); plate screw fixation: 12% (20/174) or dynamic hip screw; 2% (3/174).

The missForest method[19] was used to impute missing values for variables with missing data: pathologic fracture (1%), other bone metastases (2%), brain metastases (2%), spine metastases (2%), visceral metastases (3%), Charlson comorbidity (3%), previous systemic therapy (5%), local radiation (4%), hemoglobin level (13%), creatinine level (17%), white blood platelet count (17%), sodium level (18%). cell count (20%), BMI (25%) and alkaline phosphatase level (29%). No imputation was performed for variables of which more than 30% of the data was missing: albumin level (33%), absolute lymphocyte count (64%), absolute neutrophil count (64%), platelet-to-lymphocyte ratio (64%) and neutrophil-to-lymphocyte ratio (67%).

To evaluate the performance of the SORG-MLA for extremity metastatic disease, we used the same metrics as Thio et al.[9] from the development study including discrimination using the AUC, calibration

Table II. — Performance of SORG machine learning algorithms for predicting 90-day survival in patients with extremity metastases on external validation (n = 174)

Metric	**Institution**			
Cohort	Groningen[a]	Boston[b]	Taiwan[c]	Iowa[d]
	Discrimination			
AUC	0.73 (0.62, 0.82)	0.87 (0.86, 0.88)	0.80 (0.74, 0,86)	0.83 (0.76, 0.88)
	Calibration			
Intercept	-0.54 (-1.02, -0.05)	0.01 (-0.06, 0.08)	0.78 (0.46, 1.10)	-0.21(-0.58, 0.37)
Slope	0.60 (0.30, 0.89)	1.03 (0.96, 1.12)	0.74 (0.53, 0.96)	0.84 (0.59, 1.09)
	Overall performance			
Brier score	0.13 (0.10, 0.17)	0.13 (0.12, 0.14)	0.12	0.12 (0.10, 0.15)
Null model	0.14	0.21	0.16	0.16

AUC = Area under the curve. [a]University Medical Center Groningen, Groningen, The Netherlands. n = 174. [b]Massachusetts General Hospital, Boston, MA, USA. n = 1090. [c]National Taiwan University Hospital, Taipei City, Taiwan. n = 356. [d]University of Iowa Hospitals and Clinics, Iowa City, IA, USA. n = 264.

using the calibration plot, overall performance using the Brier score; and decision curve analysis. The AUC ranges from 0.5 to 1.0, with 0.5 indicating pure chance and 1.0 indicating the highest discriminating score. Graphically, discrimination was visualized with receiver operating characteristic curve plots. Calibration indicates agreement between the predicted outcome and the actual outcome, and perfect calibra-tion has an intercept of 0 and a slope of 1[20,21]. The Brier score refers to overall performance, with 0 as a perfect Brier score. However, the prevalence of the outcome had to be considered; therefore, the Brier score of the null model

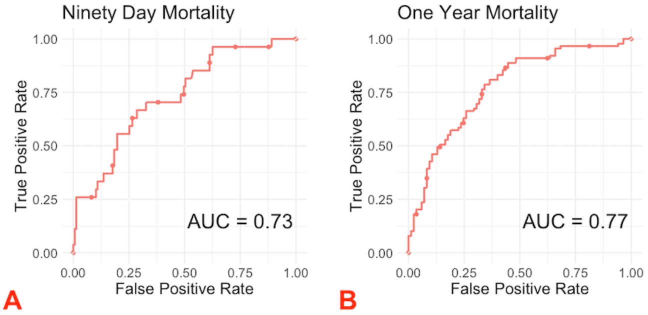


Fig. 2. — A-B. Discrimination of SORG machine learning algorithm for extremity metastasis on external validation with imputation using the missForest method, n = 174. AUC= Area under the curve.

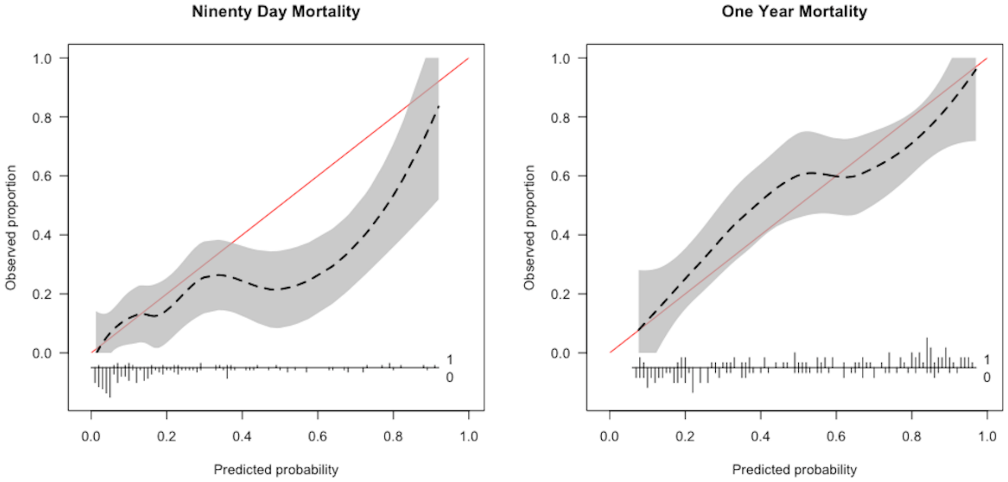


Fig. 3. — A-B Calibration plots representing the predictions of the SORG-MLAs are shown for (A) 90-day and (B) 1-year survival, n = 174. The calibration plot visualizes how accurate the predictions are for different probabilities. The diagonal dashed line represents the perfect calibration in which (predicted probabilities = observed probabilities). SORG-MLA = 174.

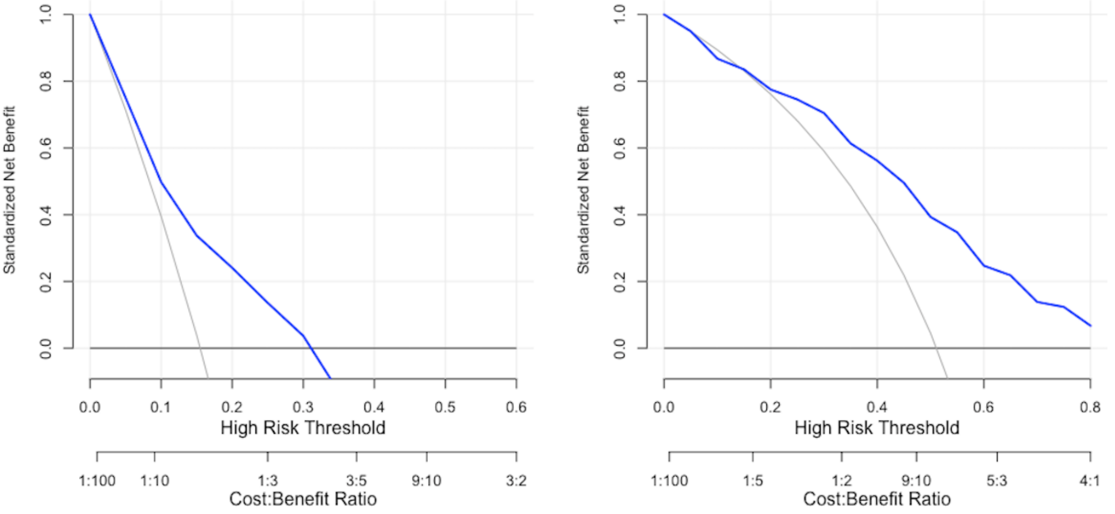
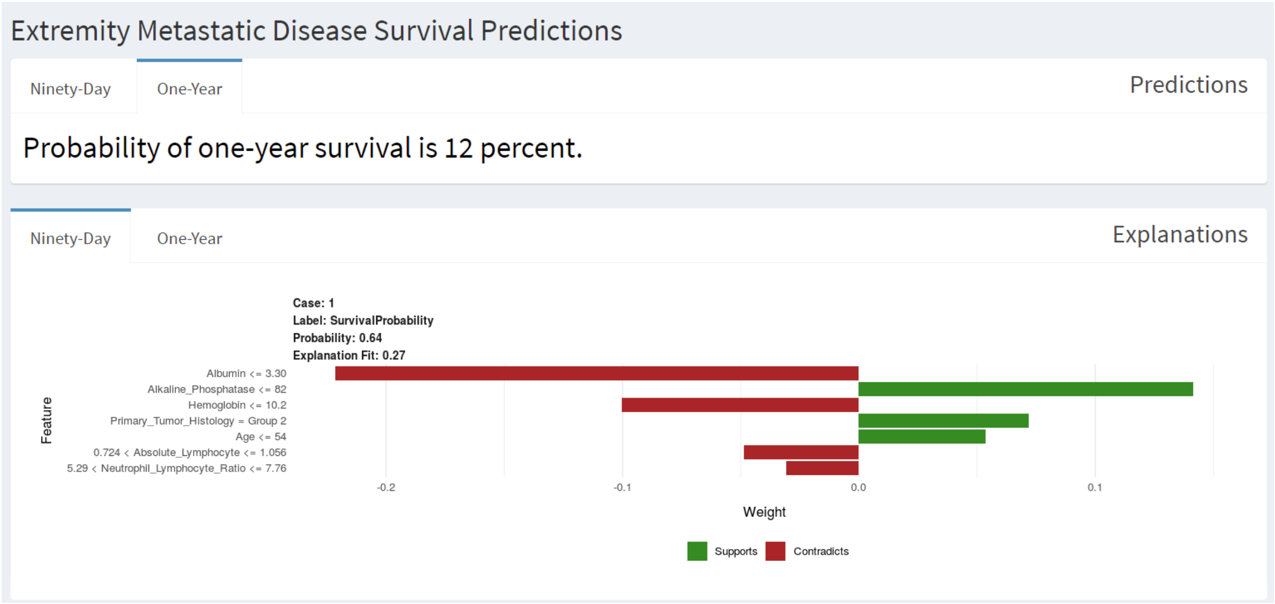


Fig. 4 . — A-B Decision curve analysis representing the cost-benefit ratio. (A) 90-day and (B) 1-year survival, n = 174. Above the threshold of 0.5, the predictions of the SORG ML algorithm resulted in a larger net (survival) for benefit compared to changing the treatment for all patients or for no patients.



Extremity Metastatic Disease Survival Predictions

Probability of one-year survival is 12 percent.

Case: 1
Label: SurvivalProbability
Probability: 0.64
Explanation Fit: 0.27

Fig. 5. — Interface of the SORG web application for explanation of variables that either support (green) or contradict (red) 1-year survival for an individual patient. This patient is a 49-year-old woman who received prosthetic surgery for a metastatic hip lesion. She had a moderate-growing primary tumor with visceral metastases. She received systemic therapy prior to surgery. Her laboratory values were as follows: hemoglobin level of 9.6 g/dL, platelet count of 390 x 103/uL, absolute lymphocyte count of 1.57 103/uL, absolute neutrophil count of 4.8 103/uL, creatinine of level of 0.64 mg/dL, white blood cell count of 5.1 103/uL, albumin level of 4.3 g/dL, alkaline phosphatase level of 67 IU/L, sodium level of 135 mg/dL, and calcium level of 9.4 mg/dL. Factors that support survival are visualized by the green bars. These include alkaline phosphatase level, primary tumor group, and age. Factors that contradict survival are visualized by the red bars, which represent the albumin levels, hemoglobin levels absolute lymphocyte count and lymphocyte to neutrophil ratio. The prediction model shows a 1-year survival probability of 12%. In hindsight, the choice for prosthetic surgery was not optimal, as she passed away before total recovery.

Table III. — Performance of SORG machine learning algorithms for predicting 1-year survival in patients with extremity metastases on external validation (n = 174)

Metric	**Institution**			
Cohort	Groningen[a]	Boston[b]	Taiwan[c]	Iowa[d]
	Discrimination			
AUC	0.78 (0.70-0.84)	0.85 (0.83, 0.87)	0.84 (0.80, 0.89)	0.84 (0.79, 0.88)
	Calibration			
Intercept	0.01 (-0.36, 0.37)	-0.04 (-0.12, 0,03)	0.75 (0.49-1.10)	-0.73 (-1.02, 0.44)
Slope	0.76 (0.50, 1.02)	1.12 (1.02, 1.21)	1.22 (0.95, 1.49)	1.08 (0.81, 1.35)
	Overall performance			
Brier score	0.20	0.18	0.16	0.18
Null model	0.25	0.24	0.25	0.25

AUC = Area under the curve; [a]University Medical Center Groningen, Groningen, The Netherlands. n = 174; [b]Massachusetts General Hospital, Boston, MA, USA. n = 1090; [c]National Taiwan University Hospital, Taipei City, Taiwan. n = 356; [d]University of Iowa Hospitals and Clinics, Iowa City, IA, USA. n = 264.

was calculated by assigning a probability equal to the prevalence of the outcome to each patient[4,21]. Decision curve analysis was used to provide a framework to judge the relative value of benefits (treating a true positive case) and harms (treating a false positive case) associated with the prediction model[22].

To evaluate differences in baseline characteristics, the validation cohort was compared to the original cohort with use of the student t test for parametric continuous variables and Mann-Whitney U test for non-parametric continuous variables, the Fisher's exact test for dichotomous variables and the Chi-squared test for ordinal data. Two clinical characteristics were different between the validation and developmental cohort in: less brain metastases and less local radiation occurred in the validation cohort. Also, the following laboratory values were higher in the validation cohort as compared with the development cohort: albumin, calcium, hemoglobin, neutrophil-lymphocyte ratio, platelets, and white blood cell count.

No sample size was calculated since all eligible patients between 1997 and 2019 were included, limited only by the size of the database itself. Associations with a p-value of <0.05 were considered significant. Statistical software used for data analysis and model validation was SPSS (IBM SPSS Statistics for Windows, Version 23.0. Armonk, NY: IBM Corp.) and R version 3.5.1 (The R Foundation, Vienna, 211 Austria).

RESULTS

Does the SORG-MLA for long-bone metastases accurately predict 90-day and 1-year survival in a European cohort?

The SORG-MLA showed promising results in predicting the postoperative 90-day and 1-year survival with an area under the curve (AUC) of 0.73 (95% CI: 0.62,0.82) for 90-day survival and an AUC of 0.78 (95% CI: 0.70-0.84) for 1-year survival (Table II). The calibration analysis provided an intercept of -0.54 (95% CI: -1.02,0.05) and slope of 0.60, (95% CI: 0.30, 0.89) for 90-day survival prediction. For 1-year survival, calibration showed an intercept of 0.01 (95% CI: -0.36, 0.37) and a slope of 0.76 (95% CI: 0.50,1.02) (Fig. 2). The actual 90-day survival rate in our cohort was higher than the predicted value (84% versus 71%; dependent t-test p < 0.01) which is represented by the negative calibration intercept. The actual 1-year survival rate was also higher than the predicted 1-year survival rate (49% versus 38%; dependent t-test p < 0.01). The Brier score for overall algorithm performance was 0.13 (0.10-0.17) compared with a higher null-model Brier score of 0.14 indicating greater performance of the SORG-MLA. Decision curve analysis showed greater net benefit when selectively changing the management for patients based on both of the clinical prediction models, compared to the default assumption that all patients would be deceased at the different timepoints (grey continuous line) or all patients would be alive at both 90-days and 1 year (Fig. 3 & 4). In our opinion, the results of this study show that the developed MLA could be of benefit for both physicians and patients when selecting the treatment strategy of long bone metastases.

An example of the SORG-MLA's prediction of one of our patients is illustrated (Fig. 5). Variables that resulted in an increased probability of 1-year mortality are visualized by the red bars: albumin below 3.3, hemoglobin below 10.2, absolute lymphocyte count and a neutrophil to lymphocyte ratio of 7.76. Variables that supported survival are illustrated by the green bars: alkaline phosphatase levels above 82, primary tumor histology group and age. The model can be accessed freely at https://sorg-apps.shinyapps.io/extremitymetssurvival/ where clinical characteristics of every individual patient can be filled in to provide a prediction for 90-day and 1-year survival.

DISCUSSION

Adequate treatment of bone metastasis is of key importance to the mobility and overall quality of life of the patient[2]. In short, patients may not benefit from surgery if their life expectancy is less than 90 days, while prosthetic surgery aims to preserve mobility for many years, and intramedullary nailing provides limited durability but faster recovery for patient with life expectancy up to one year. Therefore, when counselling our patients in shared -surgical- decision making, patient's unique life expectancy should be taken into consideration to prevent over- and undertreatment[6]. Recently, Thio et al.[9] developed the SORG-MLA to predict 90-day and 1-year survival which can aid the shared decision-making process. Despite promising validation results in Iowa and Taiwan, the prediction tool remained to be validated in Europe.

In this study, we found that SORG-MLA showed promising discriminatory ability when predicting 90 days (AUC = 0.73, 95%CI = (0.58-0.88) and 1 year mortality (AUC=0.77, 95% CI = 0.62-0.91) in a Dutch cohort of patients with long-bone metastases, providing benefits to surgical decision-making. However, the negative calibration intercept suggests that the SORG-MLA tended to overestimate mortality for patients treated in this Dutch cohort. Clinicians should keep this in mind when they use the SORG-MLA for survival prediction of their patients from these geographic regions. This study is, to our knowledge, the first validation using a European cohort. The SORG-MLA can be accessed online at https://sorg-apps.shinyapps.io/extremitymetssurvival/.

This study has several limitations. First, due to the retrospective design, data collection was limited to archival information. Secondly, the sample group of 174 patients is relatively small for machine learning standards. As opposed to the United States, in the

Netherlands it is not standard that a total blood count is obtained prior to surgery, leading to large amounts of missing data, this poses a practical limitation to our study. However, these exact differences in day-to-day clinical practice are the primary reason to perform external validation studies. Missing data was imputed using statistical techniques. As a result, the analysis of the current study was partly based on assumptions instead of measurements. Also, several values were collected from measurement moments within a certain perioperative time range as compared to a standardized measurement occasion time-locked to the date of surgery. For example, the available lab values could have been taken the day before the surgery or up to three weeks in advance. Regarding the classification of metastatic spread, available radiology records could have been taken several weeks before surgery or even months in advance. This variation of measurement occasions in relation to the date of surgery might have induced inaccuracy in the results. This would however have led to overestimation of survival, the opposite of what was observed in this study.

Does the SORG-MLA for long-bone metastases accurately predict 90-day and 1-year survival in a Dutch cohort?

In this study, we found that the SORG-MLA performed well in a European cohort. With 9.5 percent, Europe repre-sents a substantial portion of the world's population[23]. This tool can aid both patients and physicians in their shared decision-making process. However, users should be aware that the SORG-MLA overestimates mortality rates in this patient population. A possible explanation for these results could be different approaches when treating long-bone metastases. Dutch surgeons tend to favor non-operative care when deciding treatment options for patients with a poor prognosis. This could possibly have led to more healthier patients being operated in the Netherlands, explaining both the observed longer survival in the baseline comparison and the overestimated mortality of patients in this Dutch cohort.

In contrast to other survival prediction models , the SORG-MLA have only been validated on two timepoints (90-day and 1-year survival). It also remains to be validated in nonoperatively treated patients.

Moreover, cancer research has made great advances in the recent years, leading to prolonged survival of patients in most primary tumor types[23,24]. The SORG-MLA need to therefore be updated and retrained to retain their accuracy. Furthermore, as databases in cancer research are ever-growing, researchers should focus on collecting high quality, tumor-specific data such as receptor status, mutation status, and histologic subtype to improve predictive abilities on a tumor-specific level. Ultimately, prospective testing and continuous improvement of these algorithms is warranted before they can be implemented in daily clinical practice.

CONCLUSION

The SORG survival prediction tool for patients with long-bone metastases showed promise in this Dutch cohort in terms of both discrimination and decision curve analysis. However, 90-day survival tended to be underestimated. To bridge the gap from development to implementation in clinical practice, future validation in larger, preferably prospective datasets, is warranted to further validate or refute these algorithms.

Conflicts of interest: None, the authors report no funding disclosure for this study.

Ethical Review Committee Statement: Institutional review board approval (METC) was granted for retrospective review of electronic health records at Groningen, the Netherlands.

REFERENCES

1. Coleman RE. Clinical features of metastatic bone disease and risk of skeletal morbidity. Clin Cancer Res Off J Am Assoc Cancer Res. 2006 Oct;12(20 Pt 2):6243s-9s.
2. Wedin R. Surgical treatment for pathologic fracture. Acta Orthop Scand Suppl. 2001 Jun;72(302):2p., 1-29.
3. Surgery, The Ultimate Placebo: A Surgeon Cuts through the Evidence.
4. Bilimoria KY, Liu Y, Paruch JL, Zhou L, Kmiecik TE, Ko CY, et al. Development and evaluation of the universal ACS NSQIP surgical risk calculator: a decision aid and informed consent tool for patients and surgeons. J Am Coll Surg. 2013 Nov;217(5):833-42.e1-3.
5. Anderson AB, Wedin R, Fabbri N, Boland P, Healey J, Forsberg JA. External Validation of PATHFx Version 3.0 in Patients Treated Surgically and Nonsurgically for Symptomatic Skeletal Metastases. Clin Orthop. 2020 Apr;478(4):808-18.
6. Forsberg JA, Eberhardt J, Boland PJ, Wedin R, Healey JH. Estimating survival in patients with operable skeletal metastases: an application of a bayesian belief network. PloS One. 2011;6(5):e19956.
7. Willeumier JJ, Linden YM van der, Wal CWPG van der, Jutte PC, Velden JM van der, Smolle MA, et al. An Easy-to-Use Prognostic Model for Survival Estimation for Patients with Symptomatic Long Bone Metastases. J Bone Joint Surg Am. 2018 Feb;100(3):196-204.
8. Janssen SJ, Heijden AS van der, Dijke M van, Ready JE, Raskin KA, Ferrone ML, et al. 2015 Marshall Urist

Young Investigator Award: Prognostication in Patients With Long Bone Metastases: Does a Boosting Algorithm Improve Survival Estimates? Clin Orthop. 2015 Oct;473(10):3112-21.

9. Thio QCBS, Karhade AV, Ogink PT, Bramer JAM, Ferrone ML, Calderón SL, et al. Development and Internal Validation of Machine Learning Algorithms for Preoperative Survival Prediction of Extremity Metastatic Disease. Clin Orthop. 2020 Feb;478(2):322-33.
10. Thio QCBS, Karhade AV, Notman E, Raskin KA, Lozano-Calderón SA, Ferrone ML, et al. Serum alkaline phosphatase is a prognostic marker in bone metastatic disease of the extremity. J Orthop. 2020 Dec;22:346-51.
11. Thio QCBS, Goudriaan WA, Janssen SJ, Paulino Pereira NR, Sciubba DM, Rosovksy RP, et al. Prognostic role of neutrophil-to-lymphocyte ratio and platelet-to-lymphocyte ratio in patients with bone metastases. Br J Cancer. 2018 Sep;119(6):737-43.
12. Skalitzky MK, Gulbrandsen TR, Groot OQ, Karhade AV, Verlaan JJ, Schwab JH, et al. The preoperative machine learning algorithm for extremity metastatic disease can predict 90-day and 1-year survival: An external validation study. J Surg Oncol. 2021 Oct.
13. Tseng TE, Lee CC, Yen HK, Groot OQ, Hou CH, Lin SY, et al. International Validation of the SORG Machine-learning Algorithm for Predicting the Survival of Patients with Extremity Metastases Undergoing Surgical Treatment. Clin Orthop. 2022 Feb 1;480(2):367-78.
14. Groot OQ, Bindels BJJ, Ogink PT, Kapoor ND, Twining PK, Collins AK, et al. Availability and reporting quality of external validations of machine-learning prediction models with orthopedic surgical outcomes: a systematic review. Acta Orthop. 2021 Aug;92(4):385-93.
15. Luo W, Phung D, Tran T, Gupta S, Rana S, Karmakar C, et al. Guidelines for Developing and Reporting Machine Learning Predictive Models in Biomedical Research: A Multidisciplinary View. J Med Internet Res. 2016 Dec 16;18(12):e323.
16. Moons KGM, Altman DG, Reitsma JB, Ioannidis JPA, Macaskill P, Steyerberg EW, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): Explanation and Elaboration. Ann Intern Med. 2015 Jan;162(1):W1-73.
17. WHO. WHO Data Policy. 2018.
18. Katagiri H, Okada R, Takagi T, Takahashi M, Murata H, Harada H, et al. New prognostic factors and scoring system for patients with skeletal metastasis. Cancer Med. 2014 Oct;3(5):1359-67.
19. Stekhoven DJ, Bühlmann P. MissForest – non-parametric missing value imputation for mixed-type data. Bioinforma Oxf Engl. 2012 Jan;28(1):112-8.
20. Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. Eur Heart J. 2014 Aug;35(29):1925-31.
21. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. Epidemiol Camb Mass. 2010 Jan;21(1):128-38.
22. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. Med Decis Mak Int J Soc Med Decis Mak. 26(6):565-74.
23. United Nations. Department of Economic and Social Affairs, Population Division. World Population Prospects 2019: Data Booket; 2019.
24. Michielin O, Atkins MB, Koon HB, Dummer R, Ascierto PA. Evolving impact of long-term survival results on metastatic melanoma treatment. J Immunother Cancer. 2020 Oct;8(2):e000948.
25. Herbst RS, Morgensztern D, Boshoff C. The biology and management of non-small cell lung cancer. Nature. 2018 Jan 24;553(7689):446-54.