Clin Orthop Relat Res (2022) 480:367-378
DOI 10.1097/CORR.0000000000001969

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

Clinical Research

OPEN

International Validation of the SORG Machine-learning Algorithm for Predicting the Survival of Patients with Extremity Metastases Undergoing Surgical Treatment

Ting-En Tseng MD[1], Chia-Che Lee MD[2], Hung-Kuan Yen MD[3], Olivier Q. Groot MD[4], Chun-Han Hou MD, PhD[2], Shin-Ying Lin MD[2], Michiel E. R. Bongers MD[4], Ming-Hsiao Hu MD, PhD[2], Aditya V. Karhade MD, MBA[4], Jia-Chi Ko MD[1], Yi-Hsiang Lai MD[1], Jing-Jen Yang MD[2], Jorrit-Jan Verlaan MD, PhD[5], Rong-Sen Yang MD, PhD[3], Joseph H. Schwab MD, MS[4], Wei-Hsin Lin MD, PhD[2]

Received: 2 May 2021 / Accepted: 17 August 2021 / Published online: 7 September 2021

Abstract

Background The Skeletal Oncology Research Group machine-learning algorithms (SORG-MLAs) estimate 90-day and 1-year survival in patients with long-bone metastases undergoing surgical treatment and have demonstrated good discriminatory ability on internal validation. However, the performance of a prediction model could potentially vary by race or region, and the SORG-MLA must be externally validated in an Asian cohort. Furthermore, the authors of the original developmental study did not consider the Eastern Cooperative Oncology Group (ECOG) performance status, a survival prognosticator repeatedly validated in other studies, in their algorithms because of missing data.

Questions/purposes (1) Is the SORG-MLA generalizable to Taiwanese patients for predicting 90-day and 1-year

Each author certifies that there are no 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 related to the author or any immediate family members.
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.
Ethical approval for this study was obtained from the National Taiwan University Hospital, Taipei City, Taiwan (number 201912022RIND).
This work was performed at National Taiwan University Hospital, Taipei City, Taiwan.

The first three authors contributed equally to this manuscript.

[1]Department of Medical Education, National Taiwan University Hospital, Taipei City, Taiwan

[2]Department of Orthopedic Surgery, National Taiwan University Hospital, Taipei City, Taiwan

[3]National Taiwan University, Taipei City, Taiwan

[4]Department of Orthopaedic Surgery, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA

[5]Department of Orthopedic Surgery, University Medical Center Utrecht, Utrecht University, Utrecht, the Netherlands

W-H. Lin ✉, Department of Orthopedic Surgery, National Taiwan University Hospital, No. 7, Zhongshan S. Road, Zhongzheng District, Taipei City 100, Taiwan Email: oweihsin@gmail.com

mortality? (2) Is the ECOG score an independent factor associated with 90-day and 1-year mortality while controlling for SORG-MLA predictions?

Methods All 356 patients who underwent surgery for long-bone metastases between 2014 and 2019 at one tertiary care center in Taiwan were included. Ninety-eight percent (349 of 356) of patients were of Han Chinese descent. The median (range) patient age was 61 years (25 to 95), 52% (184 of 356) were women, and the median BMI was 23 kg/m^2 (13 to 39 kg/m^2). The most common primary tumors were lung cancer (33% [116 of 356]) and breast cancer (16% [58 of 356]). Fifty-five percent (195 of 356) of patients presented with a complete pathologic fracture. Intramedullary nailing was the most commonly performed type of surgery (59% [210 of 356]), followed by plate screw fixation (23% [81 of 356]) and endoprosthetic reconstruction (18% [65 of 356]). Six patients were lost to follow-up within 90 days; 30 were lost to follow-up within 1 year. Eighty-five percent (301 of 356) of patients were followed until death or for at least 2 years. Survival was 82% (287 of 350) at 90 days and 49% (159 of 326) at 1 year. The model's performance metrics included discrimination (concordance index [c-index]), calibration (intercept and slope), and Brier score. In general, a c-index of 0.5 indicates random guess and a c-index of 0.8 denotes excellent discrimination. Calibration refers to the agreement between the predicted outcomes and the actual outcomes, with a perfect calibration having an intercept of 0 and a slope of 1. The Brier score of a prediction model must be compared with and ideally should be smaller than the score of the null model. A decision curve analysis was then performed for the 90-day and 1-year prediction models to evaluate their net benefit across a range of different threshold probabilities. A multivariate logistic regression analysis was used to evaluate whether the ECOG score was an independent prognosticator while controlling for the SORG-MLA's predictions. We did not perform retraining/recalibration because we were not trying to update the SORG-MLA algorithm in this study.

Results The SORG-MLA had good discriminatory ability at both timepoints, with a c-index of 0.80 (95% confidence interval 0.74 to 0.86) for 90-day survival prediction and a c-index of 0.84 (95% CI 0.80 to 0.89) for 1-year survival prediction. However, the calibration analysis showed that the SORG-MLAs tended to underestimate Taiwanese patients' survival (90-day survival prediction: calibration intercept 0.78 [95% CI 0.46 to 1.10], calibration slope 0.74 [95% CI 0.53 to 0.96]; 1-year survival prediction: calibration intercept 0.75 [95% CI 0.49 to 1.00], calibration slope 1.22 [95% CI 0.95 to 1.49]). The Brier score of the 90-day and 1-year SORG-MLA prediction models was lower than their respective null model (0.12 versus 0.16 for 90-day prediction; 0.16 versus 0.25 for 1-year prediction), indicating good overall performance of SORG-MLAs at these two timepoints. Decision curve analysis showed SORG-MLAs provided net benefits when threshold probabilities ranged from 0.40 to 0.95 for 90-day survival prediction and from 0.15 to 1.0 for 1-year prediction. The ECOG score was an independent factor associated with 90-day mortality (odds ratio 1.94 [95% CI 1.01 to 3.73]) but not 1-year mortality (OR 1.07 [95% CI 0.53 to 2.17]) after controlling for SORG-MLA predictions for 90-day and 1-year survival, respectively.

Conclusion SORG-MLAs retained good discriminatory ability in Taiwanese patients with long-bone metastases, although their actual survival time was slightly underestimated. More international validation and incremental value studies that address factors such as the ECOG score are warranted to refine the algorithms, which can be freely accessed online at https://sorg-apps.shinyapps.io/extremitymetssurvival/.

Level of Evidence Level III, therapeutic study.

Introduction

The incidence of long-bone metastases has been rising because of increased survival rates among patients with cancer [25, 47]. Without proper treatment, a long-bone metastasis may cause skeleton-related events such as pain, disability, and pathologic fracture. These adverse events often lead to worse quality of life and are associated with higher mortality rates [10, 35]. Commonly used nonoperative treatments for bone metastases include systemic chemotherapeutics, various types of radiation therapy, and bone-targeting agents such as bisphosphonates or denosumab. However, these treatment modalities rarely cure metastatic bone disease because of the aggressive nature of advanced-stage cancer [40, 46, 47, 51, 55], and surgical procedures may be indicated to address an impending or actual fracture of the involved bone. It is challenging for clinicians to decide whether to offer surgical interventions for patients whose lifespans may be limited. Aside from the location of the metastasis, the extent of tumor involvement, response to adjuvant therapies, and severity of symptoms, the surgeon must also weigh the benefits, risks, and potential complications associated with surgery against the patient's expected survival [2, 13]. Generally, patients with a short life expectancy may be treated nonoperatively if other means exist to properly control the local symptoms and maintain quality of life, or they may be treated surgically with less invasive palliative techniques if they are not expected to have enough time to recover from a more extensive surgical procedure. Patients with longer expected survival are often given the choice of surgical procedures if other adjuvant therapies are deemed unlikely to relieve symptoms or prevent fracture. Patients with longer expected survival may also benefit from tumor resection and more durable limb reconstruction, which achieves better local tumor control and sustained functional improvement. Two clinically practical time thresholds, namely 90-day (intermediate-term) and 1-year (long-term) survival, have been proposed for treatment decisions in patients with long-bone metastases [2, 13, 41]. Although patients who have not sustained a pathologic fracture and are expected to live less than 90 days are less likely to benefit from surgery, patients with an estimated survival of more than 1 year are candidates for more extensive surgery and durable reconstruction, such as prosthetic replacement [4, 17, 36, 38, 42, 50]. An accurate survival estimation can thus help clinicians and patients in the shared decision-making process.

Several preoperative scoring systems have been developed to estimate patients' postoperative survival [2, 3, 13, 19, 25, 30, 34, 41, 52]. However, some of the scoring systems, such as the revised Katagiri score, did not achieve acceptable discriminatory ability in external validation [25, 28, 30]. Recently, Thio et al. [47] capitalized on the novel machine-learning concept and developed the Skeletal Oncology Research Group machine learning algorithm (SORG-MLA) to evaluate the intermediate-term and long-term survival probability of patients with extremity metastases. Although it has shown good discriminatory ability in the internal validation cohort of the developmental study, the SORG-MLA has not been externally validated [16]. Several studies suggested that racial distinctions among regions could influence the discriminatory ability of preoperative scoring systems because of differences in racial compositions, dominant cancer types, healthcare systems, and socioeconomic environments [5, 22, 23, 39, 54]. Han Chinese people account for 18% of the global population [56] but constitute less than 5% of the US population. In addition, several studies found that Chinese patients with certain types of malignancies had a better prognosis than their Western counterparts [6-9, 32, 40, 53]. In a world where international travel, education, and migration have become the norm, physicians in many countries could see an increasing racially diverse patient population in their practices. Therefore, it is important to understand whether a clinical tool such as SORG-MLAs can be generalized to different racial groups or used in regions outside of the United States.

The authors of the original SORG-MLA development study reported a lack of functional status data as one of their research limitations, and they suggested future studies should include these factors to improve algorithm performance. The Eastern Cooperative Oncology Group (ECOG) performance scale is widely used by oncologists in clinical practice due to its simplicity, but it was not considered in the original development study due to missing data. It has also been shown to be associated with survival in cancer patients [13, 20, 41, 52], and a number of preoperative scoring systems consider it as a prognosticator [13, 25]. It would be of interest to know whether the ECOG should be investigated as a potential factor to be added into SORG-MLA to enhance the model's performance.

Therefore, in this study, we asked: (1) Is the SORG-MLA generalizable to a Taiwanese cohort for predicting 90-day and 1-year survival? (2) Is the ECOG score an independent factor associated with 90-day and 1-year survival while controlling for SORG-MLA predictions?

Patients and Methods

Study Design and Setting

The selection criteria used in the development study [47] were applied, resulting in 356 patients who underwent surgical treatment for long-bone metastases between 2014 and 2019 at the National Taiwan University Hospital (Fig. 1). In general, the indications for surgery were patients with an American Society of Anesthesiologists classification of IV or below or patients considered fit for surgery based on a multidisciplinary assessment by a medical oncologist, anesthesiologist, and orthopaedic surgeon (C-CL, C-HH, S-YL, R-SY, W-HL) and the occurrence of a complete pathologic fracture or an impending pathologic fracture deemed unlikely to resolve with nonoperative treatment alone. Surgery was often offered for actual pathologic femur fractures unless clear medical contraindications existed, such as ongoing shock, comatose state, acute respiratory failure, decompensated hepatic failure, and severe heart dysfunction, because femoral fractures tend to profoundly impact the patient's quality of life. An impending fracture was diagnosed if the lesion in question had a Mirels score of 9 or more and caused pain or weakness in the involved limb [29]. We excluded patients diagnosed with primary bone sarcoma or sarcoma bone metastasis because these tumors include various histologic types and tend to behave differently than carcinomas [31, 37, 55].

Participants' Baseline Characteristics

More than 98% (349 of 356) of patients were of Han Chinese descent. The median (range) age was 61 years (25 to 95), and 52% (184 of 356) of patients were women (Table 1). The median BMI was 23 kg/m^2 (13 to 39 kg/m^2). In this study, 27% (97 of 356) of patients had a slow-growth tumor, 33% (118 of 356) had a moderate-growth tumor, and 40% (141 of 356) had a rapid-growth tumor according to the definition proposed by Katagiri et al. [25] and later adopted in the original SORG-MLA development study [47]. In summary, hormone-dependent breast cancer, hormone-dependent prostate cancer, malignant lymphoma, malignant myeloma, and thyroid cancer were referred to as slow-growth tumors; non–small cell lung cancer with molecularly targeted therapy,

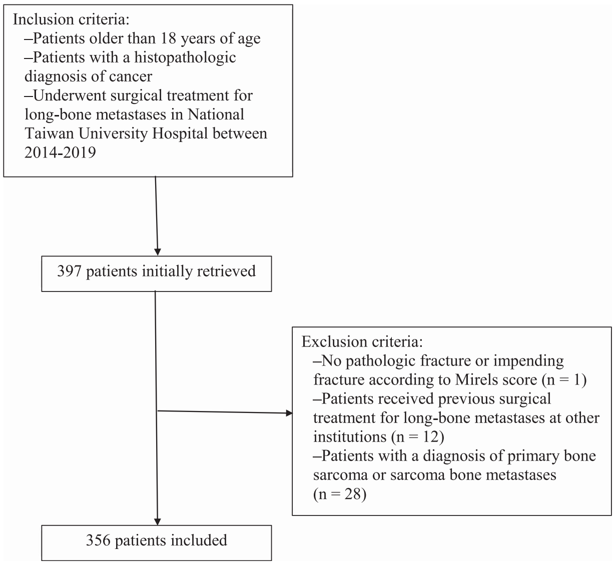
Inclusion criteria:
–Patients older than 18 years of age
–Patients with a histopathologic diagnosis of cancer
–Underwent surgical treatment for long-bone metastases in National Taiwan University Hospital between 2014-2019
397 patients initially retrieved
Exclusion criteria:
–No pathologic fracture or impending fracture according to Mirels score (n = 1)
–Patients received previous surgical treatment for long-bone metastases at other institutions (n = 12)
–Patients with a diagnosis of primary bone sarcoma or sarcoma bone metastases (n = 28)
356 patients included

Fig. 1 The flow diagram showing the enrolled patients.

hormone-independent breast cancer, hormone-independent prostate cancer, renal cell carcinoma, other gynecological cancer, and other cancers were referred to as moderate-growth tumors; and other lung cancer, colon and rectal cancer, gastric cancer, hepatocellular carcinoma, pancreatic cancer, head and neck cancer, other urological cancer, esophageal cancer, malignant melanoma, gallbladder cancer, cervical cancer, and cancer of unknown origin were referred to as rapid-growth tumors. The most common primary tumors were lung cancer (33% [116 of 356]) and breast cancer (16% [58 of 356]). A pathologic fracture occurred in 55% (195 of 356) of patients, other-bone metastases were identified in 72% (256 of 356) of patients, visceral metastases were present in 51% (180 of 356) of patients, and brain metastases were found in 17% (60 of 356) of patients. Twenty-one percent (73 of 356) of patients had an ECOG score of 3 or 4. The most common surgical site was the lower extremities in 76% (269 of 356) of patients. A total of 79% (281 of 356) of patients had preoperative systemic medical therapy (defined as having at least one type of the following treatment: chemotherapy, targeted therapy, hormone therapy, or immunotherapy), and 60% (214 of 356) had local radiation. Six patients were lost to follow-up within 90 days; 30 were lost to follow-up within 1 year. Mortality at 90 days was 18% (63 of 350) and at 1 year was 51% (167 of 326).

Baseline characteristics in the validation cohort differed from those in the original SORG-MLA development cohort reported by Thio et al. [47] in several regards (Table 1). The Taiwanese cohort had more patients with other Charlson comorbidities, moderate and rapid primary tumor growth, ECOG score of 3 or 4, preoperative systemic therapy, preoperative local radiation, and fewer other-bone metastases (all $p < 0.05$). The 90-day and 1-year mortality rates were higher in the developmental cohort than in the validation cohort (29% versus 18% and 62% versus 51%, respectively).

Surgical Treatment

In general, stabilization with a nail or plate-and-screws construct followed by adjuvant radiotherapy was recommended for metastases from radiosensitive tumors such as breast, prostate, lung cancer, and hematologic malignancies. Metastatectomy and cement augmentation was typically performed for radioresistant tumors such as renal cell carcinoma and hepatocellular carcinoma. Endoprosthetic replacement was considered for patients with an unsalvageable joint or extensive metaphyseal bone loss if they had a reasonably long survival and for those who had oligometastatic disease and may benefit from wide excision of metastatic tumor. We tended to offer surgery to patients with actual femoral pathologic fractures even if their expected survival was shorter than 6 weeks, as nonsurgical treatment in this setting rarely resulted in satisfactory pain control and improvement in quality of life. Fifty-nine percent (210 of 356) of patients were treated with intramedullary nailing, followed by plate-and-screws fixation in 23% (81 of 356) and endoprosthetic reconstruction in 18% (65 of 356).

Explanatory Variables and Outcomes

The following preoperative data were extracted: age; sex; BMI (in kg/m^2); any Charlson comorbidity in addition to

Table 1. Comparison of the external validation cohort with the developmental cohort

Variable	Validation cohort (n = 356)	Developmental cohort (n = 1090)	p value
Age in years	61 (25-95)	63 (54-72)	0.08
Female sex	52 (184)	56 (610)	0.16
BMI in kg/m^2	23 (13-39)	27 (23-30)[a]	
Other comorbidities	60 (215)	54 (584)	0.03
Histologic groupings of the primary tumor			< 0.01
Slow growth	27 (97)	42 (460)	
Moderate growth	33 (118)	24 (263)	
Rapid growth	40 (141)	34 (367)	
Primary tumor histology			
Lung	33 (116)	23 (247)	< 0.01
Breast	16 (58)	24 (257)	0.04
Myeloma	5 (18)	15 (162)	< 0.01
Renal	6 (21)	11 (117)	< 0.01
Prostate	5 (19)	5 (58)	0.99
Lymphoma	1 (5)	4 (44)	0.02
Melanoma	1 (2)	3 (30)	0.02
Esophageal	2 (7)	2 (24)	0.79
Colon	3 (10)	2 (18)	0.17
Head and neck	4 (16)	2 (18)	< 0.01
Thyroid	2 (6)	2 (18)	0.97
Other	1 (5)	2 (16)	0.94
Unknown	2 (6)	2 (16)	0.77
Pancreas	2 (6)	1 (7)	0.07
Sarcoma	1 (1)	1 (14)	0.11
Cervical	1 (2)	1 (1)	0.09
Other gynecologic	1 (3)	1 (13)	0.58
Other urologic	3 (9)	1 (12)	0.05
Hepatocellular carcinoma	10 (36)	1 (16)	< 0.01
Stomach	1 (4)	1 (2)	0.02
Gallbladder	2 (6)	0 (0)	< 0.01
Pathologic fracture	55 (195)	54 (594)	0.93
ECOG score			
0-2	79 (283)	85 (360 of 422)[b]	0.03
3-4	21 (73)	15 (62 of 422)[b]	
Tumor location			0.69
Upper extremity	24 (87)	23 (255)	
Lower extremity	76 (269)	77 (835)	
Other bone metastases	72 (256)	78 (845)	0.03
Visceral metastases	51 (180)	45 (487)	0.05
Brain metastases	17 (60)	16 (175)	0.72
Previous systemic therapy	79 (281)	62 (676)	< 0.01
Local radiation	60 (214)	18 (194)	< 0.01
Preoperative laboratory values			
Hemoglobin level in g/dL	11 (6-18)	11 (10-13)[c]	0.18
White blood cell count in 10^3/uL	7 (1-90)	7 (5-10)[d]	0.93
Platelet count in 10^3/uL	234 (36-651)	251 (184-332)[e]	0.06
Absolute lymphocyte count in 10^3/uL	1 (1-8)[f]	1 (1-2)[f]	0.48

Table 1. continued

Variable	Validation cohort (n = 356)	Developmental cohort (n = 1090)	p value
Absolute neutrophil count in 10^3/uL	5 (1-77)[g]	5 (4-8)[g]	0.86
Neutrophil-to-lymphocyte ratio	5 (1-67)	5 (3-9)	0.18
Platelet-to-lymphocyte ratio	216 (14-2776)	234 (158-374)	0.11
Albumin level in g/dL	4 (1-5)[h]	4 (3-4)[h]	< 0.01
ALP level in IU/L	98 (23-2531)[i]	101 (74-146)[i]	0.10
Calcium level in mg/dL	9 (4-18) [j]	9 (9-10) [j]	< 0.01
Creatinine level in mg/dL	0.7 (0.3-8.1)	0.8 (0.7-1.1) [k]	< 0.01
Sodium level in mg/dL	137 (118-149)[l]	138 (136-140)[l]	< 0.01
Outcomes			
90-day mortality	18 (63 of 350)	29 (305 of 1052)	< 0.01
1-year mortality	51 (167 of 326)	62 (639 of 1031)	< 0.01

Data presented as % (n) or median (range).

[a] BMI was missing for 22% (237 of 1090) of patients in the developmental cohort.

[b] ECOG scale was missing for 61% (668 of 1090) of patients in the developmental cohort.

[c] Hemoglobin level was missing for 13% (146 of 1090) of patients in the developmental cohort.

[d] White blood cell count was missing for 13% (146 of 1090) of patients in the developmental cohort.

[e] Platelet count was missing for 13% (146 of 1090) of patients in the developmental cohort.

[f] The absolute lymphocyte count was missing for 2% (8 of 356) of patients in the validation cohort and 30% (326 of 1090) of patients in the developmental cohort.

[g] The absolute neutrophil count was missing for 2% (8 of 356) of patients in the validation cohort and 30% (322 of 1090) of patients in the developmental cohort.

[h] The albumin level was missing for 7% (25 of 356) of patients in the validation cohort and 29% (320 of 1090) of patients in the developmental cohort.

[i] The alkaline phosphatase level was missing for 5% (18 of 356) of patients in the validation cohort and 29% (316 of 1090) of patients in the developmental cohort.

[j] The calcium level was missing for 2% (8 of 356) of patients in the validation cohort and 18% (200 of 1090) of patients in the developmental cohort.

[k] The creatinine level was missing for 19% (66 of 356) of patients in the developmental cohort.

[l] The sodium level was missing for 0.03% (1 of 356 patient) of patients in validation cohort and 18% (199 of 1090) of patients in developmental cohort; ALP = alkaline phosphatase.

metastatic cancer; primary tumor type, classified per Katagiri et al. [25]; ECOG score; tumor location; pathologic fracture; other bone, visceral (lung and/or liver), or brain metastases; previous systemic therapy or local radiation; absolute lymphocyte and neutrophil count (in 10^3/ uL); albumin level (in g/dL); alkaline phosphatase level (in IU/L); calcium level (in mg/dL); creatinine level (in mg/dL); hemoglobin level (in g/dL); platelet count (in 10^3/ uL); sodium level (in mg/dL); and white blood cell count (in 10^3/ uL).

The primary outcomes were 90-day and 1-year mortality, which were defined as the time between the patient's first surgery for a long-bone metastasis and death of any cause. Loss to follow-up occurred in 2% (6 of 356) of patients at 90 days and in 8% (30 of 356) at 1 year. Patients whose final survival status could not be ascertained due to loss to follow-up were excluded from analyses of model performance and calculation of actual survival rates.

Missing Data

We used the missForest method [43] to impute missing values for the absolute lymphocyte count (2% [8 of 356]), absolute neutrophil count (2% [8 of 356]), albumin level (7% [25 of 356]), alkaline phosphatase level (5% [18 of 356]), calcium level (2% [8 of 356]), and sodium level (0.3% [1 of 356]). No missing data was recorded for ECOG because the hospital's electronic medical records system requires input of ECOG score every time a patient with malignancy is seen in the clinic or admitted to the hospital.

Ethical Approval

This international external validation study followed the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) guidelines

[11, 15, 16] and the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) [12] guidelines. The study was approved by our institutional review board (201912022RIND).

Assessment of Model Performance and Statistical Analysis

We manually retrieved the 90-day and 1-year SORG-MLA predictions for each patient from an internet-based application (https://sorg-apps.shinyapps.io/extremitymetssurvival/). A discrimination analysis (concordance index [c-index]), calibration analysis (intercept and slope), overall performance analysis (Brier score), and decision curve analysis were performed to validate the two set of algorithms [14, 44]. A c-index ranges from 0.5 to 1.0, with 0.5 indicating random guess and 1.0 perfect discrimination. A c-index ≥ 0.7 indicates a good model, and a c-index ≥ 0.8 an excellent model [26]. Calibration refers to the agreement between the predicted outcomes and the actual outcomes and is assessed by plotting the calibration curves and computing the calibration slope and intercept. A perfect calibration has an intercept of 0 and a slope of 1. A positive intercept suggests an underestimation of the outcome by the prediction model, and a negative intercept indicates an overestimation [48]. The Brier score refers to overall performance. It is the average mean squared difference between the model predictions and the observed outcomes, and ranges from 0 (best prediction) to 1 (worst prediction). However, the prevalence of the outcome must be considered; therefore, the Brier score of the null model was also calculated by assigning a probability equal to the prevalence of the outcome (in this case, the actual survival rate) to each patient. The net benefit of the prediction model is calculated by comparing its Brier score with that of the null model. If a prediction model's Brier score is lower than the null model's, then the prediction model is deemed as having good performance.

The decision curve analysis was designed to assess the net benefit of a model across a range of different threshold probabilities [49]. Unlike a discrimination analysis (c-index), a decision curve analysis considers the cost-to-benefit ratio. The user of the model can decide which threshold probability (such as, the ratio of potential risk to the potential benefit) of a treatment is important or applicable and determine whether the model is valuable at that threshold and see what the predicted net benefit would be. In general, if the harm of a treatment modality is relatively limited, for example, antibiotics for infection, the clinician may choose a lower threshold probability. In contrast, if the potential risks associated with a treatment are high, such as, performing extensive surgery in a fragile patient, a higher threshold possibility should be chosen for decision-making [44, 45].

We compared the baseline characteristics, 90-day mortality rate, and 1-year mortality rate of the developmental and external validation cohorts. We assessed continuous variables using one-way median tests, and we compared categorical data using chi-square tests and the Yates correction (if applicable). The actual and average predicted survival rates at 90 days and 1 year were compared with a dependent t-test. A multivariate logistic regression analysis was fitted to the ECOG performance status to estimate 90-day and 1-year mortality while adjusting for the SORG-MLA prediction outcomes. The multivariate logistic regression results are provided as odds ratios with 95% confidence intervals. A two-tailed p value ≤ 0.05 was considered significant. R for Mac (version 4.0.4, R Core Team), along with its packages of missForest, risk model decision analysis, and CalibrationCurves (downloaded through Github), was used for all statistical analyses.

Results

Is the SORG-MLA Generalizable to a Taiwanese Cohort for Predicting 90-day and 1-year Survival?

The SORG-MLA showed good discriminatory ability in predicting the postoperative 90-day and 1-year survival in the Taiwanese cohort. The c-index was 0.80 (95% CI 0.74 to 0.86) for postoperative 90-day survival prediction and 0.84 (95% CI 0.80 to 0.89) for postoperative 1-year survival prediction (Table 2). The calibration analysis provided an intercept of 0.78 (95% CI 0.46 to 1.10) and slope of 0.74 (95% CI 0.53 to 0.96) for the 90-day survival prediction, and an intercept of 0.75 (95% CI 0.49 to 1.00) and a slope of 1.22 (95% CI 0.95 to 1.49) for 1-year survival (Fig. 2). These positive calibration intercepts suggest that the SORG-MLAs tend to underestimate Taiwanese patients' survival at both 90 days postoperative and 1 year. The actual 90-day survival rate in our cohort was higher than the predicted value (82% versus 73%; dependent t-test $p < 0.01$). The actual 1-year survival rate was also higher than the predicted 1-year survival rate (49% versus 35%; dependent t-test $p < 0.01$). The Brier score of the 90-day and 1-year SORG-MLA prediction models was lower than that of their respective null model (0.12 versus 0.16 for 90-day prediction; 0.16 versus 0.25 for 1-year prediction) (Table 2), indicating good overall performance of SORG-MLAs at these two timepoints. In the decision curve analysis, the 90-day SORG-MLA was shown to provide a positive net benefit compared with a strategy of operating on either all or no patients when the threshold probabilities ranged from 0.40 to 0.95 (Fig. 3A). The 1-year SORG-MLA also provided a similar gain of positive net benefit compared with a default strategy of operating on either all or no patients when the threshold probabilities ranged from 0.15 to 1.0 (Fig. 3B). These results indicated that management changes based on the 90-day and 1-year SORG algorithms had greater net benefit than the default strategies of changing management for no patients or for all patients.

Table 2. C-indices and Brier scores of the SORG-MLA by primary tumor histology in the validation cohort (n = 356)

	90-day prediction			1-year prediction		
Validation cohort	c-index (95% CI)[a]	Brier score[b]	Actual vs predicted survival rate	c-index (95% CI)[a]	Brier score[b]	Actual vs predicted survival rate
Overall (n = 356)	0.80 (0.74-0.86)	0.12 (0.16)	82% vs 73%	0.84 (0.80-0.89)	0.16 (0.25)	49% vs 35%
Solid-organ (n = 333)	0.79 (0.73-0.86)	0.13 (0.16)	81% vs 72%	0.84 (0.80-0.89)	0.16 (0.25)	47% vs 34%
Lung (n = 116)	0.87 (0.77-0.97)	0.10 (0.16)	82% vs 73%	0.89 (0.83-0.95)	0.13 (0.25)	44% vs 34%
Breast (n = 58)	0.58 (0.16-1.00)	0.07 (0.07)	93% vs 83%	0.75 (0.58-0.91)	0.15 (0.17)	78% vs 43%
Liver (n = 37)	0.72 (0.53-0.91)	0.13 (0.14)	85% vs 58%	0.76 (0.58-0.94)	0.18 (0.25)	47% vs 26%
Hematologic malignancies (n = 23)	0.95[c]	0.04 (0.05)	96% vs 83%	0.82 (0.58-1.00)	0.15 (0.20)	71% vs 49%
Kidney (n = 21)	0.65[c]	0.05 (0.05)	95% vs 80%	0.80 (0.53-1.00)	0.17 (0.24)	39% vs 44%
Prostate (n = 19)	0.69 (0.43-0.94)	0.21 (0.22)	68% vs 77%	0.98 (0.92-1.00)	0.08 (0.24)	42% vs 39%

[a] A c-index of 0.5 indicates random guess and 1.0 indicates perfect discriminatory ability. A c-index of 0.8 is typically considered to denote great discriminatory ability.
[b] The Brier score of the prediction model should be compared with that of the null model. The Brier score of the null model is presented in parentheses. A lower Brier score of the prediction model indicates good overall model performance. Solid-organ malignancies include all kind of malignancies except for hematopoietic malignancies.
[c] 95% CI could not be calculated because only one patient died within 90 days of surgery.

Is the ECOG Score an Independent Factor Associated with 90-day and 1-year Survival While Controlling for SORG-MLA Predictions?

The ECOG score was an independent factor associated with 90-day survival but not 1-year survival while controlling for SORG-MLA predictions. In the multivariate analysis that adjusted for SORG-MLA's 90-day survival prediction, patients with an ECOG score of 3 or 4 had higher 90-day mortality (OR 1.94 [95% CI 1.01 to 3.73]; p = 0.04) but not 1-year mortality (OR 1.07 [95% CI 0.53 to 2.17]; p = 0.85) than those with a score of 0 to 2.

Discussion

Patients with metastatic bone disease in the extremities should ideally be managed with a personalized strategy that takes their life expectancy into consideration to avoid under- or overtreatment. The SORG-MLAs incorporate state-of-the-art machine learning techniques and have demonstrated excellent performance on internal validation. However, SORG-MLAs have not been externally validated outside the United States, especially in the Han Chinese population, who represent nearly one-fifth of the global population. In this study, we found that SORG-MLAs retained excellent discriminatory ability and provided net benefits to surgical decision-making when used to estimate both 90-day and 1-year survival probabilities in Taiwanese patients with extremity metastasis. However, the calibration analysis and a comparison of the actual and the predicted survival rates indicated that SORG-MLA tended to underestimate patient survival in our Taiwanese validation cohort. Clinicians should keep this underestimation in mind when they use SORG-MLAs for survival prediction in patients of Han Chinese descent. The SORG-MLAs can be accessed online at https://sorg-apps.shinyapps.io/extremitymetssurvival/.

Limitations

This study has several limitations. First, this was a single-institution study, and more than 98% of patients in our cohort were of Han Chinese descent. This might limit the reference value of the current study for physicians treating patients from other racially distinct regions. In addition, this cohort is unique because the Taiwanese healthcare system consists of the government-run National Health Insurance program, which covers every citizen and legal foreign resident, rendering molecularly targeted treatment and radiotherapy readily accessible and relatively

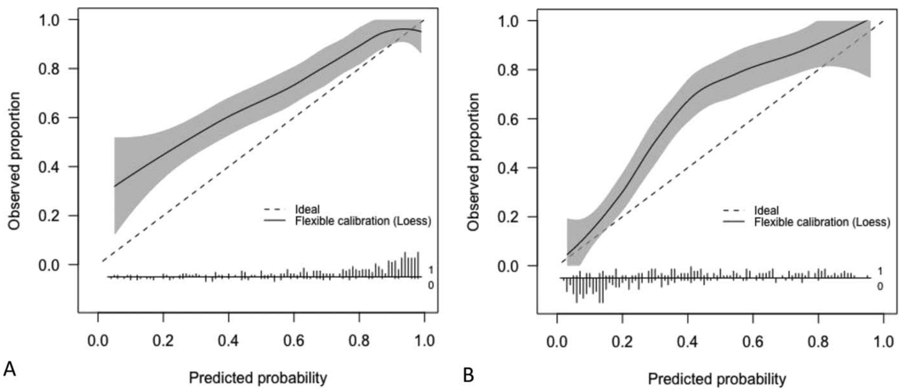

Fig. 2 A-B Calibration plots representing the predictions of the SORG-MLAs are shown for (**A**) 90-day and (**B**) 1-year survival. The calibration plot visualizes how accurate the predictions are for different probabilities. The diagonal dashed line represents the perfect prediction (predicted probabilities = observed probabilities); the closer the model curve is to the diagonal line, the more accurate the prediction.

affordable. For example, the price of gefitinib, an effective targeted agent for lung cancer, is 10 times more expensive in the United States (USD 270 per tablet) than it is in Taiwan (USD 26 per tablet) [1, 18]. As a result, patients in Taiwan might be less financially constrained with use of newer medical therapies such as targeted agents and immunotherapy. Second, although we accounted for most known prognostic variables, additional factors—in particular, tumor-specific variables such as response to systemic therapy, use of oral targeted therapies or bone-modifying agents, administration of immunotherapy, and tumor molecular profiling—may predict survival, but we did not include them. Lack of consideration of these granular details could have contributed to the underestimation of patient survival in our validation cohort. We believe current predictive models can be improved not only by considering incremental factors such as the ECOG score identified in this study, but also by investigating the added value of these aforementioned variables. Third, this study is retrospective. The data used for input into the SORG-MLAs, such as results of laboratory tests and variables based on imaging studies or clinical evaluation, were not acquired in a standardized fashion and not all at the same time before surgery. Validation of the SORG-MLAs based on data from a prospectively enrolled cohort evaluated with a standardized preoperative protocol is an avenue for future research. Fourth, survival is only one aspect to consider when deciding on surgical treatment. For example, some patients with femoral pathologic fractures might benefit from surgical fixation even though their expected survival is short because in this situation acceptable pain control and quality of life is seldom achieved with nonsurgical treatment. Future studies should attempt to develop predictive models for outcomes such as postoperative ambulatory status, hospitalization, reoperations, systemic complications, level of pain, and quality of life, the latter of which is often considered the most important aspect in the care of patients with incurable cancer.

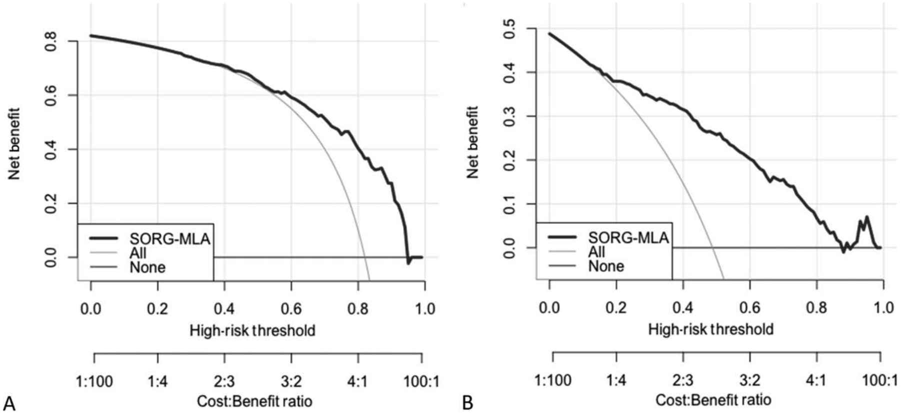

Fig. 3 A-B Decision curve analysis plots of SORG-MLA predictions of (**A**) 90-day and (**B**) 1-year survival. A color image accompanies the online version of this article.

Physicians should be aware of these potential pitfalls when using SORG-MLAs in the clinical setting.

Is the SORG-MLA Generalizable to a Taiwanese Cohort for Predicting 90-day and 1-year Survival?

In this study, we found that SORG-MLAs performed well in a cohort comprised mostly of Han Chinese, who represent a substantial portion of the world's population and may be more frequently seen in many clinicians' practices in this age of globalization. This tool can help physicians and their Han Chinese patients in the shared decision-making process, but users should be aware that SORG-MLAs might underestimate survival rates in this patient population. In a study comparing six state-of-the-art preoperative scoring systems for patients undergoing surgical treatment for long-bone metastases, Meares et al. [28] reported that the PathFx model had the best performance for 90-day survival prediction (c-index 0.70 [95% CI 0.69 to 0.70]) and the OPTIModel was the best for predicting 1-year survival (c-index 0.79 [95% CI 0.78 to 0.79]). Compared with these two benchmarks (the PathFx model and OPTIModel), the SORG-MLAs had better discriminatory ability at both timepoints (c-index 0.80 [95% CI 0.74 to 0.86] for 90-day survival prediction and c-index 0.84 [95% CI 0.80 to 0.89] for 1-year survival prediction). However, PathFx was recently updated and has now been externally validated not only in patients treated with surgery but also in patients treated nonoperatively with external beam radiation therapy [2]. In addition, PathFx provides postoperative survival predictions at six timepoints: 1 month, 3 months, 6 months, 12 months, 18 months, and 24 months. By contrast, the SORG-MLAs currently offer only 90-day and 1-year survival predictions and remain to be validated in nonoperatively treated patients. The SORG-MLAs should ideally be retrained to make up for these shortcomings. Furthermore, cancer therapeutics have evolved, and there have been rapid advances in recent years. More emphasis is now placed on tumor-specific characteristics such as the histologic subtype, mutation status, hormone receptor expression profile, and response to novel treatment strategies. We believe future studies should focus on collecting granular tumor-specific data of individual cancer types to enhance the SORG-MLA's performance.

Is the ECOG Score an Independent Factor Associated with 90-day and 1-year Mortality While Controlling for SORG-MLA Predictions?

In our Taiwanese cohort, the ECOG performance scale was an independent factor associated with 90-day mortality but not with 1-year mortality after controlling for SORG-MLA predictions in multivariate analysis. This finding was consistent with results from several previous studies, in which investigators found that 90-day survival depended more on the patient's general condition (for example, the ECOG performance status or albumin level) and 1-year survival was influenced more by the primary tumor type [23, 24, 27, 33, 39, 47]. One study specifically assigned quantified importance to various survival prognosticators for patients with spinal metastases [23, 39]. On a scale of 0 to 100, where 100 indicated the most important prognosticators and 0 indicated the least important ones, the primary tumor type scored 100, the albumin level scored 90, and ECOG performance status scored less than 20 in 1-year survival prediction. On the other hand, these three factors scored 60, 100, and 40, respectively, in 90-day survival prediction. We propose that developers of survival prediction algorithms should consider incorporating the ECOG score into their (machine learning) algorithms for predicting survival in patients with long-bone metastases. We believe that current predictive models can be improved by considering incremental factors such as the ECOG. Future studies should investigate the benefit of additional predictive factors such as tumor mutation profiles, novel systemic therapies, or body composition measurements based on imaging [21].

Conclusion

SORG-MLAs performed well in this Taiwanese cohort in terms of both discrimination and decision curve analysis. However, they tended to underestimate the patient's actual survival. The ECOG performance status may provide additional prognostic value for survival predictions, with further research warranted regarding this possibility. More international, larger, and preferably prospective studies in search of additional prognosticators that add incremental value to the current model are needed to confirm and refine the findings of this study. The SORG-MLAs for extremity metastases can be accessed freely as an internet application at https://sorg-apps.shinyapps.io/extremitymetssurvival/.

Acknowledgment We thank all healthcare professionals from the various departments at the National Taiwan University Hospital for providing multidisciplinary care for our patients.

References

1. Aguiar PN Jr, Haaland B, Park W, et al. Cost-effectiveness of osimertinib in the first-line treatment of patients with EGFR-mutated advanced non-small cell lung cancer. *JAMA Oncol.* 2018;4:1080-1084.

2. 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 Relat Res.* 2020;478:808-818.
3. Bauer HC, Wedin R. Survival after surgery for spinal and extremity metastases. Prognostication in 241 patients. *Acta Orthop Scand.* 1995;66:143-146.
4. Bickels J, Dadia S, Lidar Z. Surgical management of metastatic bone disease. *J Bone Joint Surg Am.* 2009;91:1503-1516.
5. Bongers MER, Karhade AV, Villavieja J, et al. Does the SORG algorithm generalize to a contemporary cohort of patients with spinal metastases on external validation? *Spine J.* 2020;20: 1646-1652.
6. Chang CW, Tai HC, Cheng NC, Li WT, Lai HS, Chien HF. Risk factors for complications following immediate tissue expander based breast reconstruction in Taiwanese population. *J Formos Med Assoc.* 2017;116:57-63.
7. Chen CH, Lu YS, Cheng AL, et al. Disparity in tumor immune microenvironment of breast cancer and prognostic impact: Asian versus western populations. *Oncologist.* 2020;25:e16-e23.
8. Chen CH, Tzai TS, Huang SP, et al. Clinical outcome of Taiwanese men with metastatic prostate cancer compared with other ethnic groups. *Urology.* 2008;72:1287-1292.
9. Chiang CJ, Lo WC, Yang YW, You SL, Chen CJ, Lai MS. Incidence and survival of adult cancer patients in Taiwan, 2002-2012. *J Formos Med Assoc.* 2016;115:1076-1088.
10. Coleman RE. Metastatic bone disease: clinical features, pathophysiology and treatment strategies. *Cancer Treat Rev.* 2001;27: 165-176.
11. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *Ann Intern Med.* 2015;162:55-63.
12. Cuschieri S. The STROBE guidelines. *Saudi J Anaesth.* 2019;13: S31-S34.
13. 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:e19956.
14. Fredon A, Radchenko AK, Cuppen HM. Quantification of the role of chemical desorption in molecular clouds. *Acc Chem Res.* 2021;16:745-753.
15. Groot OQ, BJJ Bindels, Ogink PT, et al. Availability and reporting quality of external validations of machine-learning prediction models with orthopedic surgical outcomes: a systematic review. *Acta Orthop.* 2021:1-9.
16. Groot OQ, Ogink PT, Lans A, et al. Machine learning prediction models in orthopedic surgery: a systematic review in transparent reporting. *J Orthop Res*. Published online March 18, 2021. DOI: 10.1002/jor.25036.
17. Harvey N, Ahlmann ER, Allison DC, Wang L, Menendez LR. Endoprostheses last longer than intramedullary devices in proximal femur metastases. *Clin Orthop Relat Res.* 2012;470: 684-691.
18. Hsu JC, Wei CF, Yang SC. Effects of removing reimbursement restrictions on targeted therapy accessibility for non-small cell lung cancer treatment in Taiwan: an interrupted time series study. *BMJ Open.* 2019;9:e022293.
19. Janssen SJ, van der Heijden AS, van Dijke M, et al. 2015 Marshall Urist young investigator award: prognostication in patients with long bone metastases: does a boosting algorithm improve survival estimates? *Clin Orthop Relat Res.* 2015;473: 3112-3121.
20. Kantarjian H, O'Brien S, Cortes J, et al. Results of intensive chemotherapy in 998 patients age 65 years or older with acute myeloid leukemia or high-risk myelodysplastic syndrome: predictive prognostic models for outcome. *Cancer.* 2006;106:1090-1098.
21. Kapoor ND, Twining PK, Groot OQ, et al. Adipose tissue density on CT as a prognostic factor in patients with cancer: a systematic review. *Acta Oncol.* 2020;59:1488-1495.
22. Karhade AV, Ahmed AK, Pennington Z, et al. External validation of the SORG 90-day and 1-year machine learning algorithms for survival in spinal metastatic disease. *Spine J.* 2020;20:14-21.
23. Karhade AV, Thio Q, Ogink PT, et al. Predicting 90-day and 1-year mortality in spinal metastatic disease: development and internal validation. *Neurosurgery*. 2019;85:E671-E681.
24. Karhade AV, Thio Q, Ogink PT, et al. Development of machine learning algorithms for prediction of 30-day mortality after surgery for spinal metastasis. *Neurosurgery*. 2019;85:E83-E91.
25. Katagiri H, Okada R, Takagi T, et al. New prognostic factors and scoring system for patients with skeletal metastasis. *Cancer Med.* 2014;3:1359-1367.
26. Mandrekar JN. Receiver operating characteristic curve in diagnostic test assessment. *J Thorac Oncol.* 2010;5:1315-1316.
27. Massaad E, Shin JH. Commentary: Sarcopenia as a prognostic factor for 90-day and overall mortality in patients undergoing spine surgery for metastatic tumors: a multi-center retrospective cohort study. *Neurosurgery.* 2020;87:E550-E551.
28. Meares C, Badran A, Dewar D. Prediction of survival after surgical management of femoral metastatic bone disease - a comparison of prognostic models. *J Bone Oncol.* 2019;15:100225.
29. Mirels H. Metastatic disease in long bones. A proposed scoring system for diagnosing impending pathologic fractures. *Clin Orthop Relat Res.* 1989;249:256-264.
30. Nathan SS, Healey JH, Mellano D, et al. Survival in patients operated on for pathologic fracture: implications for end-of-life orthopedic care. *J Clin Oncol*. 2005;23:6072-6082.
31. Nystrom LM, Reimer NB, Reith JD, et al. Multidisciplinary management of soft tissue sarcoma. *ScientificWorldJournal.* 2013;2013:852462.
32. Park JW, Chen M, Colombo M, et al. Global patterns of hepatocellular carcinoma management from diagnosis to death: the BRIDGE Study. *Liver Int.* 2015;35:2155-2166.
33. Pielkenrood BJ, van Urk PR, van der Velden JM, et al. Impact of body fat distribution and sarcopenia on the overall survival in patients with spinal metastases receiving radiotherapy treatment: a prospective cohort study. *Acta Oncol.* 2020;59:291-297.
34. Ratasvuori M, Wedin R, Keller J, et al. Insight opinion to surgically treated metastatic bone disease: Scandinavian sarcoma group skeletal metastasis registry report of 1195 operated skeletal metastasis. *Surg Oncol.* 2013;22:132-138.
35. Roodman GD. Mechanisms of bone metastasis. *N Engl J Med.* 2004;350:1655-1664.
36. Ruggieri P, Mavrogenis AF, Casadei R, et al. Protocol of surgical treatment of long bone pathological fractures. *Injury.* 2010;41: 1161-1167.
37. San-Julian M, Diaz-de-Rada P, Noain E, Sierrasesumaga L. Bone metastases from osteosarcoma. *Int Orthop.* 2003;27:117-120.
38. Scotti C, Camnasio F, Peretti GM, Fontana F, Fraschini G. Modular prostheses in the treatment of proximal humerus metastases: review of 40 cases. *J Orthop Traumatol.* 2008;9:5-10.
39. Shah AA, Karhade AV, Park HY, et al. Updated external validation of the SORG machine learning algorithms for prediction of ninety-day and one-year mortality after surgery for spinal metastasis. *Spine J.* Published online March 31, 2021. DOI: 10.1016/j.spinee.2021.03.026.
40. Shieh SH, Hsieh VC, Liu SH, Chien CR, Lin CC, Wu TN. Delayed time from first medical visit to diagnosis for breast cancer patients in Taiwan. *J Formos Med Assoc*. 2014;113:696-703.

41. Sorensen MS, Gerds TA, Hindso K, Petersen MM. External validation and optimization of the SPRING model for prediction of survival after surgical treatment of bone metastases of the extremities. *Clin Orthop Relat Res.* 2018;476:1591-1599.
42. Steensma M, Boland PJ, Morris CD, Athanasian E, Healey JH. Endoprosthetic treatment is more durable for pathologic proximal femur fractures. *Clin Orthop Relat Res*. 2012;470: 920-926.
43. Stekhoven DJ, Buhlmann P. MissForest–non-parametric missing value imputation for mixed-type data. *Bioinformatics.* 2012;28: 112-118.
44. 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.
45. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology*. 2010;21:128-138.
46. Tattersall MH, Thomas H. Recent advances: oncology. *BMJ*. 1999;318:445-448.
47. Thio Q, Karhade AV, Bindels BJJ, et al. Development and internal validation of machine learning algorithms for preoperative survival prediction of extremity metastatic disease. *Clin Orthop Relat Res.* 2020;478:322-333.
48. Van Calster B, McLernon DJ, van Smeden M, et al. Calibration: the Achilles heel of predictive analytics. *BMC Med*. 2019;17:230.
49. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making.* 2006;26: 565-574.
50. Wedin R. Surgical treatment for pathologic fracture. *Acta Orthop Scand Suppl*. 2001;72:1-29.
51. Wells A, Grahovac J, Wheeler S, Ma B, Lauffenburger D. Targeting tumor cell motility as a strategy against invasion and metastasis. *Trends Pharmacol Sci*. 2013;34:283-289.
52. Willeumier JJ, van der Linden YM, van der Wal C, et al. An easy-to-use prognostic model for survival estimation for patients with symptomatic long bone metastases. *J Bone Joint Surg Am.* 2018; 100:196-204.
53. Wu CE, Chen SC, Chang HK, Lo YF, Hsueh S, Lin YC. Identification of patients with hormone receptor-positive breast cancer who need adjuvant tamoxifen therapy for more than 5 years. *J Formos Med Assoc.* 2016;115:249-256.
54. Yang JJ, Chen CW, Fourman MS, et al. International external validation of the SORG machine learning algorithms for predicting 90-day and 1-year survival of patients with spine metastases using a Taiwanese cohort. *Spine J*. Published online February 2, 2021. DOI: 10.1016/j.spinee.2021.01.027.
55. Yap YS, Lu YS, Tamura K, et al. Insights into breast cancer in the east vs the west: a review. *JAMA Oncol*. 2019;5:1489-1496.
56. Zhao YB, Zhang Y, Zhang QC, et al. Ancient DNA reveals that the genetic structure of the northern Han Chinese was shaped prior to 3,000 years ago. *PLoS One.* 2015;10:e0125676.