External Validation of SORG Machine Learning Algorithm to a Midwest Cohort with Spinal Metastasis

James R.L. Hall[1], Alex R. Coffman[1], Trevor R. Gulbrandsen[1], Michael Garneau[1], Olivier Groot[2], Harmen R. Kruijten[2], Cassim M. Igram[1], Andrew J. Pugely[1], Catherine R Olinger[1], Joseph H. Schwab[2]
[1]University of Iowa Hospitals and Clinics Department of Orthopedics and Rehabilitation, Iowa City, IA,
[2]Department of Orthopedic Surgery, Massachusetts General Hospital, Harvard Medical School, Boston, MA
Presenting author email: James-R-Hall@uiowa.edu

Disclosures: James R.L. Hall (N), Alex R. Coffman (4-Medtronic), Trevor R. Gulbrandsen (N), Michael Garneau (N), Olivier Groot (N), Harmen R. Kruijten (N), Cassim M. Igram (3B - Medtronic, AlloSource), Andrew J. Pugely (Medtronic, Globus Medical, United Healthcare, RBD Bioinformatics, Clinical Orthopaedics and Related Research), Catherine R Olinger (3B - Proprio, Globus Medical), Joseph H. Schwab (N)

INTRODUCTION: The spine is the most common site of bony metastasis, with autopsy studies demonstrating spinal involvement in 70% of cancer patients. Of those afflicted with spinal metastases, 2.5 – 5% present with spinal cord compression with rates varying based on primary tumor type. Treatment historically involved radiation therapy there has been a significant increase in surgical management of these patients. The increased rates of surgical management, however, have been associated with increased rates of post-operative complications. The goals of surgery remain largely palliative with preservation or restoration of neurologic function, and mechanical stabilization, but rarely oncologic control. The Skeletal Oncology Research Group machine learning algorithm (SORG – MLA) was developed in 2019 to better predict 90-day and 1-year mortality for patients with metastatic disease involving the spine. The group sought to improve accuracy in estimating immediate and long-term post-operative survival to better inform decision-making. Previous iterations of the SORG model were compared to other metastatic disease scoring systems and showed improved accuracy in short-term (30-day) and intermediate term (90-day) mortality. The current algorithm was developed based on a cohort of patients from a large institution in a Northeastern US metropolitan area. These algorithms were externally validated in an independent patient population from another large Northeastern institution. In this validation study, the algorithm showed decreased performance in patients receiving treatment after 2010. Further validation was then completed in a cohort of patients in the New York City metropolitan area who were treated between 2014 and 2016. While these results showed good performance of the algorithm in line with the initial development cohort, there has yet to be a validation performed on a patient population in the Midwestern US. Our hypothesis was that the SORG – MLA would accurately predict 90-day and 1-year mortality in a rural Midwestern cohort.

METHODS: IRB approval was obtained at our institution. Patients aged 18 years and older who underwent surgical treatment for spinal metastases between 2010 and 2022 at a single rural Midwest academic tertiary care hospital were included in this retrospective study. Baseline characteristics of the validation cohort were obtained through chart review and compared back to the developmental cohort. Discrimination (c-statistic and receiver operating curve), calibration (calibration slope, intercept, calibration plot, and observed proportions by predicted risk groups), overall performance (Brier score), and decision curve analysis were used to assess the SORG machine learning algorithm in the validation cohort.

RESULTS SECTION: Overall, there were 247 patients included in this study. Most patients were men, 146 (59%) with a median age of 62 years (interquartile range 56-70). The 90-day and 1-year mortality rates of 63 (26%) and 134 (54%) respectively. The validation cohort and the developmental cohort differed significantly regarding primary tumor histology, the presence of visceral metastasis, and pre-operative hemoglobin levels. The SORG algorithm for 90-day mortality achieved a calibration intercept of 0.05 (95 % confidence interval [CI]: -0.09 – 0.19), calibration slope of 0.81 (95% CI: 0.54 – 1.08), AUC of 0.84 (95% CI 0.78 – 0.89), Brier score of 0.14 (95% CI 0.08 – 0.20). The SORG algorithm for 1-year mortality achieved a calibration intercept of 0.05 (95% CI -0.10 – 0.20), calibration slope of 0.86 (95% CI 0.63 – 1.11), AUC of 0.82 (95% CI 0.77 – 0.87), and Brier score of 0.17 (95% CI 0.12 – 0.22).

DISCUSSION: Predicting post-operative survival is important when evaluating risks and benefits of surgical intervention in the setting of spinal metastases. The SORG-MLA has demonstrated its use as an accurate and valuable prognostic tool but has not been externally validated in a rural cohort. In accordance with the TRIPOD guidelines, external validation is a crucial step. The results show that the SORG-MLA generalized well in predicting 90-day and 1-year survival for patients with metastatic disease involving the spine in a Midwest cohort. This result supported our hypothesis. There are several limitations that are noteworthy, first being that 35% of patient were missing data regarding albumin and alkaline phosphatase. This limits the usefulness of the model as the MissForest software package has only been shown to effectively impute data to 30% missing values. Another limitation is the patient sample size, as prior studies have shown a need for 100 and as high as 200 events and nonevents for reliable calibration interpretation for external validation. This being a retrospective study is a further limitation. Finally, although this validation represents a unique geographical region, the patient population was from the United States. To address these limitations, a multi-center study in a prospective cohort would be beneficial.

SIGNIFICANCE/CLINICAL RELEVANCE: The SORG-MLA performed well in this regionally unique cohort from a rural midwestern population. The previously developed algorithms predict 90-day and 1-year survival in patients who underwent surgery for spinal metastatic disease and demonstrated good discriminative capability and overall performance, providing external validation and a tool that will inform treatment decision making.

IMAGES AND TABLES:

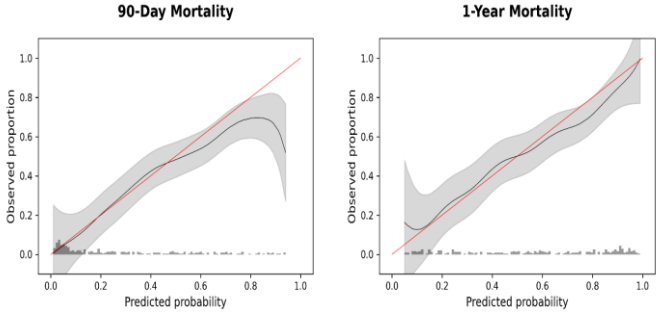

Fig 1: Calibration plots of the SORG-MLA at 90-day and 1-year mortality

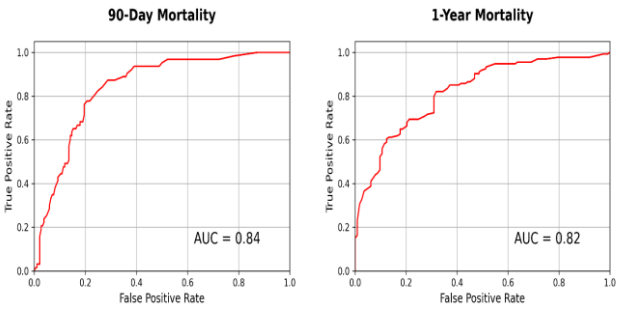

Fig 2: Receiver operating characteristic curve with area under the curve (AUC) of the SORG-MLA at 90-day and 1-year mortality