A nomogram to predict metastasis of soft tissue sarcoma of the extremities

Ruo-He Li, MD, Qiang Zhou, MD, A-Bing Li, MD, Hong-Zhen Zhang, MD, Zhong-Qin Lin, MD

Abstract
Soft tissue sarcoma (STS) of the extremities are a rare tumor. Metastases develop in about 40%-50% of patients, most of whom die from their disease. We sought to identify potential risk factors associated with metastatic diseases upon presentation for patients with STS and established a reliable nomogram model to predict distant metastasis of STS at presentation. The current study retrospectively analyzed 3884 STS of the extremities or trunk patients from the Surveillance, Epidemiology, and End Results (SEER) database between 2010 and 2015. Based on patient registration, all patients were randomly allocated to training sets and validation sets (2:1). Then, univariate and binary logistic regression analysis was used to determine the significantly correlated predictors of metastasis. Finally, the nomogram model was established, using these predictors and validated it. 311 (8.21%) of the cases experienced distant metastatic disease was present at the time of presentation. The nomogram was developed from age, histology subtype, primary site, tumor size, grade and depth. Encouragingly, the nomogram showed favorable calibration with C-index 0.790 in the training set and 0.801 in validation set. The DCA showed that the novel model was clinically useful. This nomogram model had a high precision to predict the metastasis of soft tissue sarcoma of the extremities. We expect this model could be used in different clinical consultation and established risk assessment.

Abbreviations: AJCC = American Joint Committee on Cancer, C-index = the concordance Index, CIs = confidence intervals, DCA = decision curve analyses, ICD-O = International Classification of Diseases for Oncology, MPNST = malignant peripheral nerve sheath tumor, ROC = receiver operating characteristic curve, SEER = surveillance, epidemiology, and end results, STS = soft tissue sarcoma.

Keywords: extremities, metastasis, nomogram, sarcoma.

1. Introduction
Soft tissue sarcoma (STS) of the extremities or trunk is a rare tumor with over 50 histologic subtypes, morphological features, and clinical behaviors. A total of 40% to 50% of patients with STS develop metastatic disease,1–4 the most common site of which is the lungs.5 Prognosis in these patients is worse than that of patients with local disease only.6 As previous studies have shown, once distant metastasis occurs in STS patients, the main treatment is palliative chemotherapy, in which case, the median survival time is only about 12 months.7,8 Nevertheless, risk factors for metastatic STS of the extremities or trunk are not well understood. Accurate prognosis may help determine adjuvant treatment advice and provide appropriate patient counseling. A nomogram is a graphic score used to predict clinical outcomes; nomograms have been widely used in cancer research.9 To date, there have been no studies using nomograms to identify populations at risk for metastatic soft tissue sarcoma of the extremities or trunk at presentation. The surveillance, epidemiology, and end results (SEER) Program database covers approximately 28% of the US population,10 allowing detailed analysis of rare cancers. Using collected data, in this study, we identified the most relevant predictors associated with increased distant metastatic disease of STS of the extremities or trunk. We developed a nomogram to predict the probability of metastasis.

2. Methods
2.1. Data source and inclusion criteria
All data were collected from the SEER database. The inclusion criteria for STS patients of the extremities or trunk in this study were as follows:
Diagnosed with soft tissue sarcoma International Classification of Diseases for Oncology (ICD-O): sarcoma NOS (8880–8806), fibrosarcoma (8810–8815), malignant fibrohistiocytoma (8830),...
lipsarcoma (8850–8857), dedifferentiated liposarcoma (8858), leiomyosarcoma (8890, 8891, 8896), synovial sarcoma (9040–9044), and malignant peripheral nerve sheath tumor (MPNST) (9540, 9561), as primary malignancy between 2010 and 2015.

Site limited to the extremity or trunk (C47.1, C47.2, C47.3, C47.6 C49.1, C49.2, C49.6, C49.3, and C76.1.).

Patients >19 years of age.

The exclusion criteria in this study were:

- Unknown grade and stage
- Unknown tumor size
- Unknown T stage (Derived AJCC 7th)

Recorded variables included patient age, marital status, race, gender, year of diagnosis, anatomic site, tumor size, site-specific disease extent, distant metastasis at diagnosis, histologic subtype of sarcoma, histologic grade, and region.

The cutoff value of age and tumor size were determined using receiver operating characteristic curve (ROC) curves. The optimal cutoff values of tumor size were categorized as <7.6 cm and >7.6 cm (Fig. 1). The optimal age cutoffs were 36 years (Fig. 1); therefore, patients were categorized into two age groups (19–36 years and >36 years). The anatomic location of sarcoma was categorized as extremity (the upper or lower extremities) and trunk. Regarding chemotherapy and radiation, the updated SEER dataset uses “none/unknown” as a single option. Adding this information to the nomogram may introduce correlation deviations; therefore, chemotherapy and radiotherapy are not included as variables. The result of interest is the presence of metastatic disease at the time of diagnosis. The results were modeled as bivariate: “Yes” was defined as a patient with metastatic disease recorded in stage A of SEER history. Patients without distant metastases were defined as patients with only local or regional diseases. The data are anonymous, and the requirement for informed consent was therefore waived from Institutional Review Board approval at Wenzhou Hospital of Integrated Traditional Chinese and Western Medicine.

2.2. Nomogram construction and validation

All patients were randomly allocated in a 2:1 manner into training and validation cohorts. A total of 2589 patients were split into the training dataset and 1295 were in the validation dataset. The nomogram was developed based on the independent predictors included the binary logistic regression models using the training cohort, and the validation cohort was used to externally validate the nomogram. The nomogram was built as described below. All variables were involved in univariate analysis, and the $\chi^2$ test was used to determine the relationship between predictors and metastasis. Subsequently, we built binary logistic regression analysis to study the association between the presence of metastatic disease at presentation and meaningful features. $P < .05$ in the binary logistic regression analysis was considered statistically significant. $P$ values, odds ratios, and 95% confidence intervals (CIs) were used to describe the risk factors of metastasis. We developed the nomogram using significant prognostic factors from the binary logistic regression model to assess the probability of metastasis.

The validation of the nomogram was performed using the concordance index (C-index), calibration curves, and decision curve analyses (DCAs). The concordance Index (C-index) between observed and predicted outcome was calculated to evaluate the discrimination of the model. In general, C-index values over 0.7 mean a relatively accurate prediction. The predictive performance was assessed using calibration plots to compare nomogram predictions with observed outcomes. We also developed decision curve analyses to assess the potential of the nomograms for clinical application. DCA examine the clinical practical value of a predictive model by quantifying its net benefit according to the threshold probability and the relative weight between false-positive and false-negative results. The easy explanation: A good model will have a high net benefit.

All statistical analyses were performed using SPSS 25.0 software (SPSS Inc., Chicago, IL), the R software version 3.4.3.
P value of <0.05 was considered statistically significant.

3. Results

3.1. Patient characteristics

Between 2010 and 2015, all 3884 STS patients were identified from the SEER database according to the criteria, distant metastatic disease was present in 311 (8.21%) of the patients at the time of presentation. Of these, 2589 patients were split into the training dataset and 1295 were in the validation dataset. The detail clinicopathological information are listed in Table 1.

3.2. Univariate analysis and binary logistic regression analysis

For the training set, univariate analyses indicated that age, histology subtype, primary site, tumor size, grade and depth were associated with distant metastasis (Table 1). The logistic regression model (Table 2) revealed decreased odds of metastatic disease at presentation among patients with age of 36 years or more (OR = 0.496; 95% CI, 0.323 to 0.762), patients with tumor size larger than 7.6 cm (OR = 4.729; 95% CI, 3.192 to 7.006), and patients with tumors located deep to the fascia (OR = 1.713; 95% CI, 1.087 to 2.700). Patients affected by leiomyosarcoma and other histology subtypes were 2.486 and 2.450 times, respectively, more likely to have metastasis than were fibrosarcoma type. Patients whose tumor sites were in the thorax or trunk were 1.716 times more likely to have metastasis than were patients whose primary tumor site located in the upper extremity (95% CI, 1.420–3.382). Patients with advanced grade were 5.962, 5.295, and 9.066 times, respectively, more likely to metastasis than those with the grade I.

3.3. Establishment and validation of the nomogram model

According to the results of binary logistic regression, these significant variables, including age, histological subtypes, primary location, tumor size, grade and depth, were used as

| Table 1 | Univariate analysis of risk factors in soft tissue sarcoma of the extremities or trunk. |
|---------|----------------------------------------------|
| Predictors | Non-metastasis | Metastasis | All | P value |
| Age | | | | |
| <36 | 272 (11.5%) | 37 (16.7%) | 309 (11.9%) | .023 |
| >36 | 2095 (88.5%) | 185 (83.3%) | 2280 (88.1%) | |
| Marital status | | | | |
| Married | 1417 (59.9%) | 118 (53.2%) | 1535 (59.3%) | .149 |
| Never married | 546 (23.1%) | 59 (26.6%) | 605 (23.4%) | |
| Divorced, separated, widowed | 404 (17.1%) | 45 (20.3%) | 449 (17.3%) | |
| Histology | | | | <.001 |
| Fibrosarcoma | 299 (12.6%) | 15 (6.8%) | 314 (12.1%) | |
| Leiomyosarcoma | 250 (10.6%) | 30 (13.5%) | 280 (10.8%) | |
| Liposarcoma | 713 (30.1%) | 21 (9.5%) | 734 (28.4%) | |
| Malignant fibrohistiocytoma | 177 (7.5%) | 15 (6.8%) | 192 (7.4%) | |
| MPNST | 74 (3.1%) | 8 (3.6%) | 82 (3.2%) | |
| Synovial | 138 (5.8%) | 21 (9.5%) | 159 (6.1%) | |
| Other | 716 (30.2%) | 112 (50.5%) | 828 (32.0%) | .309 |
| Gender | | | | .360 |
| Male | 1345 (56.8%) | 134 (60.4%) | 1479 (57.1%) | |
| Female | 1022 (43.2%) | 88 (39.6%) | 1110 (42.9%) | |
| Race | | | | .360 |
| White | 1906 (80.5%) | 170 (76.6%) | 2076 (80.2%) | |
| Black | 232 (9.8%) | 27 (12.2%) | 259 (10.6%) | |
| Other | 229 (9.7%) | 25 (11.3%) | 254 (9.8%) | |
| Tumor Size | | | | <.001 |
| <7.6 cm | 1076 (45.5%) | 35 (15.8%) | 1111 (42.9%) | |
| >7.6 cm | 1291 (54.5%) | 167 (42.2%) | 1458 (57.1%) | |
| Tumor Site | | | | .048 |
| Upper extremity | 410 (17.3%) | 27 (12.2%) | 437 (16.9%) | |
| Lower extremity | 1570 (66.4%) | 148 (66.7%) | 1718 (66.4%) | |
| Thoracic or trunk | 386 (16.3%) | 47 (21.2%) | 433 (16.7%) | |
| Grade | | | | <.001 |
| I | 491 (20.7%) | 4 (1.8%) | 495 (19.1%) | |
| II | 460 (19.4%) | 26 (11.7%) | 486 (18.8%) | |
| III | 534 (22.6%) | 73 (32.9%) | 607 (23.4%) | |
| IV | 882 (37.3%) | 119 (53.6%) | 1001 (38.7%) | |
| Depth | | | | <.001 |
| Superficial | 645 (27.2%) | 25 (11.3%) | 670 (25.9%) | |
| Deep | 1722 (72.8%) | 197 (88.7%) | 1919 (74.1%) | |

* malignant peripheral nerve sheath tumor.
predictors in the nomogram (Fig. 1). Based on the nomograms, we assessed the probability of metastasis of individual patients based on the personalized information. The C-index values for nomogram predictions were 0.790 in the training set and 0.801 in the validation set (Fig. 3). The internal and external calibration plots of OS and CSS nomograms showed satisfactory agreement between nomogram prediction and actual observation (Fig. 2), indicating that the logistic regression fit the data better.

### 3.4. Clinical use

Furthermore, in order to evaluate the clinical utilization, we conducted DCA (Fig. 4). The decision curve showed that if the threshold range 5% from 25% in the training set (5% from 38% in the validation set), the nomogram to predict the probability is more beneficial than using a total treatment patient regimen or no treatment regimen.

### 4. Discussion

Analysis of the SEER database from 2010 to 2015 revealed that 8.21% patients with STS of the extremities or trunk presented with distant metastatic disease at diagnosis. Younger age, location in the trunk, larger tumor size, advancing grade, deep location, and special histologic subtypes were all associated with a greater odds ratio of metastatic disease at diagnosis. There may be a causal link between marital status, race, gender, and distant metastasis; but statistical significance was lost in the univariate analysis. A convenient and comprehensive prognostic nomogram was then built to predict the probability of metastasis. The
nomogram performed well in the general population and showed superior predictive ability.

In this report, we found that deep tumor location, advanced grade, and larger tumor size were independent risk factors for metastasis at presentation. Koea et al[13] conducted a retrospective study in STS of the extremity STS patients admitted at Memorial Sloan-Kettering Cancer Center. They showed that larger tumor size (>5cm) was a poor prognostic factor for distant metastasis. Deep tumors (RR, 2.4; 95% CI, 1.4–4.0) and high tumor grade (RR, 4.2; 95% CI, 2.6–6.7) were also adverse prognostic factors for distant metastasis.

We found that deep location was a highly significant risk factor, as such patients were 1.919 times more likely to experience subsequent metastasis (95% CI: 1.323–2.783, \( P = .001 \)). We drew a similar conclusion that grade IV was the greatest predictor for metastasis in the nomogram (100 points), followed by size >7.6 cm (75 points) and age <44 (30 points).

Histologic subtype has been reported to be a significant predictive factor for metastatic disease of STS patients.[14,15] Leiomyosarcoma was an independent predictor of distant disease, suggesting that extremity leiomyosarcomas have greater metastatic potential than do other histotypes.[5,13] Consistent with the findings of this study, high-grade sarcomas and leiomyosarcoma were most common in patients with metastases. Similarly, synovial sarcomas were often viewed as high-grade sarcomas with high metastatic risk.[16] These observations are concordant with our experience in which we observed differences among histological subtypes and similar risks in truncal and extremity sarcomas.

We found that patients with tumors of the trunk were at higher risk for metastasis. It was previously predicted that the development of metastatic disease with STS was significantly associated with the site of primary tumor location.[4,17] The variety in risk rates for metastasis between trunk and extremities could be explained by anatomic characteristics, with no clear

---

**Figure 3.** Receiver operating characteristic (ROC) curve analysis for distant metastasis with Soft tissue sarcoma of the extremities in the training set (a) and the testing set (b). The AUC was calculated, and its 95% CI was estimated by bootstrapping. The \( P \) values were two-sided. ROC = receiver operating characteristic; 95% CI = 95% confidence interval.

**Figure 4.** Decision curve analysis (DCA) for the new model in the training set (a) and the testing set (b). The horizontal solid black line represents the assumption that no patients will experience the event, and the solid gray line represents the assumption that all patients will relapse. On decision curve analysis, the nomogram showed more net benefit compared with that of treat-all or treat-none across a range of threshold probabilities.
anatomic boundaries and compartments making it increasingly difficult to resect a large tumor with adequate margins in the thoracic wall compared to tumors in the extremities.\textsuperscript{[13]}

The surprising observation was that the risk of distant metastasis was significantly lower in patients over 44 years of age. In multivariate analysis, the risk of distant metastasis decreased with age, and the probability of metastatic disease increased in patients aged 44 years or younger. A previous study\textsuperscript{[19]} reported a significant reduction in the risk of distant metastasis to bone or viscera in all age groups of breast cancer patients over the age of 40. There are several possible reasons for lower distant metastasis incidence with age: lack of angiogenic factors,\textsuperscript{[20]} host inhibitors or systemic humoral factors\textsuperscript{[21]} and local factor environments\textsuperscript{[22]}; transplantable murine tumors were difficult to grow and metastasize in older mice. Studies showed that CD4+ T cells are an important part of the metastasis process.\textsuperscript{[23]} Tsuda et al\textsuperscript{[20]} examined the relationship between immunobiology and aging and found a decreased capacity of T cells from old mice to generate angiogenic factors.

To our knowledge, we report here the first prognostic nomogram for predicting distant metastasis of soft tissue sarcoma. Our comprehensive prognostic nomogram used all significant clinicopathologic variables to accurately predict the probability of distant metastasis with C-index was 0.790 and 0.801, for internal and external validation, indicating that the nomogram model was an effective prediction system. This nomogram could be used as a supportive graphic tool in soft tissue sarcoma to help clinicians to distinguish, assess and evaluate the risk of distant involvement with clinicopathologic factors and determine whether more attention should be paid to distant organs during the course of disease. To use the nomogram for a patient, we first identify the point corresponding to the value of each predictor using Figure 1, and then sum these points together. The total point is associated with a probability that the patient will develop metastatic diseases. For example, a 65-year-old patient was diagnosed with grade II deep synovial STS with a primary tumor of 8.0 cm in left leg. Totaling the points for this patient, we see that she had 133 points in risk of metastasis of nomogram. This results in estimated distant metastasis rates of 11% according to the nomogram. Finally, it provides the accurate and deserved answer to our patient’s question: Doctor, what chance do I have distant metastasis?

The study has several limitations. First, the SEER database is a retrospective cohort, it is inevitably missing data that leads to reduced sample size. Prospective study should be performed to further confirm our conclusion. Second, as previously noted, SEER does not provide information on whether chemotherapy, exact type of surgical procedure, final margin status, as well as some causal and informative factors, such as living habits were unavailable in the SEER dataset. These variables may be an effective complement to this study, which will be an important section of our future research. Despite these shortcomings, the SEER program database served as an unparalleled resource when studying rare cancers. Finally, we only evaluated distant metastasis as the main endpoint. Nevertheless, local recurrence can be regarded as one of the endpoints of the corresponding nomogram. Despite these limitations, we identified that independent prognostic factors were associated with metastatic disease upon presentation for patients with STS and established a reliable nomogram model to predict synchronous distant metastasis of STS at presentation.

5. Conclusion

Based on the SEER cohort, younger age, location in the trunk, larger tumor size, advancing grade, deep location, and special histologic subtypes were associated with metastatic disease upon presentation for patients with STS. Our study was the first to develop and validate a prognostic nomogram based on these variables. The C-index was good in both internal and external validation. Utilizing our nomogram as a useful tool, it could meaningfully assist clinicians improve individual treatment, make clinical decisions and guide follow-up management strategies.

Acknowledgments

The authors thank the National Cancer Institute for providing the SEER data set.

Author contributions

Guarantor: Qiang Zhou.
Conception and design: Qiang Zhou, Ruo-he Li, A-bing Li, Zhong-qin Lin, Hong-zhen Zhang.
Administrative support: Qiang Zhou, Ruo-he Li, A-bing Li.
Collection and assembly of data: Qiang Zhou, Ruo-he Li, A-bing Li, Zhong-qin Lin.
Data analysis and interpretation: Qiang Zhou, A-bing Li.
Manuscript writing: Qiang Zhou, Ruo-he Li, A-bing Li, Zhong-qin Lin, Hong-zhen Zhang.
Final approval of manuscript: Qiang Zhou, Ruo-he Li, A-bing Li, Zhong-qin Lin, Hong-zhen Zhang.

References

[1] Carneiro A, Bendahl PO, Engellau J, et al. A prognostic model for soft tissue sarcoma of the extremities and trunk wall based on size, vascular invasion, necrosis, and growth pattern. Cancer 2011;117:1279–87.
[2] Kleihues P, Cavenee WJP. Genetics of tumours of the nervous system. IARC L: World Health Organization classification of tumours; 2000.
[3] Comnre JM, Terner P, Guillou L, et al. Predictive value of grade for metastasis development in the main histologic types of adult soft tissue sarcomas: a study of 1240 patients from the French Federation of Cancer Centers Sarcoma Group. Cancer; Interdisciplinary International Journal of the American Cancer Society 2001;91:1914–26.
[4] Italiano A, Mathoulin-Pelissier S, Cesne AL, et al. Trends in survival for patients with metastatic soft-tissue sarcoma. Cancer 2011;117:1049–54.
[5] Gadd MA, Casper ES, Woodruff JM, et al. Development and treatment of pulmonary metastases in adult patients with extremity soft tissue sarcoma. Annals of Surgery 1993;218:705.
[6] Gutierrez JC, Perez EA, Franceschi D, et al. Outcomes for soft-tissue sarcoma in 8249 cases from a large state cancer registry. Journal of Surgical Research 2007;141:105–14.
[7] Van Glabbeke M, Van Oosterom A, Oosterhuis J, et al. Prognostic factors for the outcome of chemotherapy in advanced soft tissue sarcoma: an analysis of 2,185 patients treated with anthracycline-containing first-line regimens—a European Organization for Research and Treatment of Cancer Soft Tissue and Bone Sarcoma Group Study. Journal of Clinical Oncology 1999;17:150–7.
[8] Karavasilis V, Seddon BM, Ashley S, et al. Significant clinical benefit of first-line palliative chemotherapy in advanced soft-tissue sarcoma: retrospective analysis and identification of prognostic factors in 488 patients. Cancer 2008;112:1585–91.
[9] Bianco JF. Nomograms and medicine. European Urology 2006.
[10] Cates, Justin MM. The AJCC 8th edition staging system for soft tissue sarcoma of the extremities or trunk: a cohort study of the SEER database. Journal of the National Comprehensive Cancer Network 16.2 (2018): 144–152.
[11] Yan C, Yang M-F, Huang Y-wJWn. A reliable nomogram model to predict the recurrence of chronic subdural hematoma after burr hole surgery. World Neurosurgery 2018;118:e336–66.
[12] Balachandran VP, Gonen M, Smith JJ, et al. Nomograms in oncology: more than meets the eye. The Lancet Oncology 2015;16:e173–80.
[13] Torosian MH, Friedrich M, Godbold J, et al. Soft-tissue sarcoma: initial characteristics and prognostic factors in patients with and without metastatic disease. Seminars in surgical oncology. New York: John Wiley & Sons, Inc., 1988, 4: 13-19.
[14] Fletcher CD, Mertens F, Unni K. World Health Organization classification of tumours: pathology and genetics of tumours of soft tissue and bone. 2002; IARC Publications,
[15] Koea JB, Leung D, Lewis JJ, et al. Histopathologic type: an independent prognostic factor in primary soft tissue sarcoma of the extremity? Cancer 2003;10:432–40.
[16] Von Mehren M, Randall RL, Benjamin RS, et al. Soft tissue sarcoma, version 2.2016, NCCN clinical practice guidelines in oncology. Journal of the National Comprehensive Cancer Network 2016;14:758–86.
[17] Salah S, Yaser S, Salem A, et al. Factors influencing survival in metastatic synovial sarcoma: importance of patterns of metastases and the first-line chemotherapy regimen. Medical Oncology 2013;30:639.
[18] Hajdu SI, Shiu MH, Fornier JGJC. Tendosynovial sarcoma. A clinicopathological study of 136 cases. Cancer 1977;39:1201–17.
[19] Stefanovski P, Bidoli E, De Paoli A, et al. Prognostic factors in soft tissue sarcomas: a study of 395 patients. European Journal of Surgical Oncology 2002;28:153–64.
[20] Soerensen TR, Raedkjaer M, Jørgensen PH, et al. Soft Tissue Sarcomas of the Thoracic Wall, More Prone to Higher Mortality, and Local Recurrence: a Single Institution Long-Term Follow-up Study. International Journal of Surgical Oncology 2019;2019:
[21] Purushotham A, Shamil E, Cariati M, Agbaje O, Muhidin A, Gillett C, et al. Age at diagnosis and distant metastasis in breast cancer—a surprising inverse relationship. European Journal of Cancer 2014;50:1697–705.
[22] Weksler M, Tsuda T, Kim Y, et al. Immunobiology of aging and cancer. Cancer Detection and Prevention 1990;14:609–11.
[23] Pili R, Guo Y, Chang J, et al. Altered angiogenesis underlying age-dependent changes in tumor growth. Journal of the National Cancer Institute 1994;86:1303–14.