MEDICAL SCIENCE MONITOR

CLINICAL RESEARCH

e-ISSN 1643-3750
© Med Sci Monit, 2019; 25: 4675-4690
DOI: 10.12659/MSM.915418

Received: 2019.01.28
Accepted: 2019.03.06
Published: 2019.06.24

Evaluating and Predicting the Probability of Death in Patients with Non-Metastatic Osteosarcoma: A Population-Based Study

Authors' Contribution:
Study Design A
Data Collection B
Statistical Analysis C
Data Interpretation D
Manuscript Preparation E
Literature Search F
Funds Collection G

ABCDEF 1,2 Runzhi Huang*
BCDEF 3 Shuyuan Xian*
BCDEF 4 Tingting Shi*
BCDEF 5 Penghui Yan
BCDEF 5 Peng Hu
EFG 6 Huabin Yin
ACDEFG 1,2 Tong Meng
AEFG 5 Zongqiang Huang

1 Division of Spine Surgery, Department of Orthopedics, Tongji Hospital, Tongji University School of Medicine, Shanghai, P.R. China
2 Key Laboratory of Spine and Spinal cord Injury Repair and Regeneration (Tongji University), Ministry of Education, Shanghai, P.R. China
3 Department of Orthopedics, Shanghai Tenth People's Hospital, Tongji University School of Medicine, Shanghai, P.R. China
4 Tongji University School of Medicine, Tongji University, Shanghai, P.R. China
5 Department of Orthopedics. The First Affiliated Hospital of Zhengzhou University, Zhengzhou, Henan, P.R. China
6 Department of Orthopedics, Shanghai Bone Tumor Institute, Shanghai General Hospital, School of Medicine, Shanghai Jiaotong University, Shanghai, P.R. China

* Runzhi Huang, Shuyuan Xian and Tingting Shi contributed equally to this work

Corresponding Authors: Tong Meng, e-mail: mengtong@medmail.com.cn, Zongqiang Huang, e-mail: gzhuangzq@163.com

Source of support: This study was supported in part by the National Natural Science Foundation of China (Grant No. 81702659; 81772856; 81501203); the Youth Fund of Shanghai Municipal Health Planning Commission (No.2017YQ054); and the Henan Medical Science and Technology Research Project (Grant No. 201602031)

Background: Osteosarcoma is one of the most common bone tumors, with strong local aggressiveness and early metastasis. The aim of this study was to describe the epidemiological data and evaluate the prognostic factors for overall survival (OS) and cause-specific survival (CSS) in patients with non-metastatic osteosarcoma.

Material/Methods: Patients histologically diagnosed with non-metastatic osteosarcoma between 2005 and 2014 were selected from the Surveillance, Epidemiology, and End Results (SEER) database. Survival analysis, machine learning, and Lasso regression were used to identify the prognostic factors for OS and CSS, and the accuracy of the nomograms was tested and compared with the American Joint Committee on Cancer (AJCC) staging systems.

Results: The entire cohort comprised 1000 patients with non-metastatic osteosarcoma. The multivariable analysis suggested that age, tumor size, grade, and American Joint Committee on Cancer (AJCC) T staging were independent prognostic factors for OS and CSS. Additionally, the nomograms based on these results could better predict probability of OS (Internal validation C-index, 0.7095) and CSS (0.7100) compared with the sixth (OS: 0.613; CSS: 0.628) and seventh edition AJCC staging systems (0.602, 0.613).

Conclusions: Relatively young age and low histopathological grade were favorable factors for both OS and CSS. Nomograms based on multivariable models worked well in predicting the probability of death for patients with non-metastatic osteosarcoma.

MeSH Keywords: Nomograms • Osteosarcoma • Prognosis • Survival Analysis

Full-text PDF: https://www.medscimonit.com/abstract/index/idArt/915418

3202 5 6 35

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

Background

Osteosarcoma is one of the most common categories of bone malignancies, with an incidence of 0.2% to 0.5% [1,2]. It most often occurs in teenagers and arises from bone tissue in both extremities and the spine, with the most common site being the metaphysis of long bones. With regard to osteosarcoma of the extremities, approximately 90%–95% of patients can be successfully treated with limb-sparing surgery and three-quarters can be cured with current multimodality treatment [3–6]. However, it is not easy to perform an en bloc tumor resection for spine osteosarcoma, and it exhibits strong local aggressiveness with a high rate of local recurrence. Moreover, osteosarcoma is likely to metastasize early through hematogenous spread (about 15–20%) [3,7,8]. Therefore, osteosarcoma has disastrous effects on individuals and society due to its major cohorts affected, undesirable prognosis, and early metastasis.

To improve the prognosis of non-metastatic osteosarcoma, it may be extremely important to find the truly significant prognostic factors. In previous case series, factors such as age at diagnosis, tumor size and location, pathology, presence and location of metastases, surgical strategy, surgical margin, and histologic respond to chemotherapy have been reported to affect overall survival (OS) [4,8–13]. However, criteria used for evaluation of surgical treatment vary among centers and even different surgeons, which may increase the heterogeneity of samples. Additionally, the relatively small sample size in a single center may lower the accuracy of the constructed nomograms of osteosarcoma.

Therefore, it was necessary to construct more precise prediction models and explore the most significant prognostic factors in patients with non-metastatic osteosarcoma after primary site surgery using big data. In this study, we used the SEER database and combined regression analysis methods (Kaplan-Meier method, Cox proportional hazards regression model, and Lasso regression) and a machine learning model (random forest) to explore the most significant predictors based on a relatively large sample size to identify the prognostic factors and construct nomograms for non-metastatic osteosarcoma.

Material and Methods

Patient selection

The study was approved by the Ethics Committee of the First Affiliated Hospital of Zhengzhou University (No. KEYAN-2018-LW-039). We selected patients from the SEER database, which contains data on cancer occurrences in 18 areas of the United States and covers approximately 26% of the population, and the characteristics of the SEER population are representative for the general US population. The SEER program's standard for case completeness is 98% and all patients were surveilled for 10 years after routine treatment until death or loss to follow-up [2,14]. Only patients histologically diagnosed with osteosarcoma from 2005 to 2014 were included. Exclusion criteria included tumor size code 000/888/989-999 (unknown or inaccurate) and extent codes 00 (*in situ*) or 99 (unknown extent). We also excluded patients who were not diagnosed with osteosarcoma by biopsy, who were at N0 or M0 stage, and who did not undergo primary site surgery. Patients with non-primary osteosarcoma or with missing data (survival months, race, marital status, grade, and radiotherapy data) were excluded.

Data extraction

Variables in the study were obtained from SEER database on March 6th, 2018, including baseline demographics of patients (age at diagnosis, sex, race), characteristics of tumor (tumor size, tumor extent, primary site, histologic subtype, tumor node metastasis (TNM) state), treatment (radiotherapy, chemotherapy), and socioeconomic status (marital status, education background, family income, and employment status). As researching endpoints, we retrieved osteosarcoma cause-specific survival (CSS) and overall survival (OS) from the database.

Statistical analysis

As an initial descriptive statistic, dichotomous variables were reported as percentages, while continuous variables were reported as mean and median (range). To find the most significant predictors, we used 3 statistical methods. First, the chi-square test was used to compare outcome between the elements of each categorical variable. Continuous variables in normal distribution and homogeneity of variance were compared by using the 2-sample *t* test, otherwise, the Mann-Whitney U test was performed. In addition, the Kaplan-Meier method was applied to obtain OS and CSS between each type contained in a categorical variable. For Kaplan-Meier analysis, we divided continuous variables into new classifications following the National Comprehensive Cancer Network guidelines [15,16] (age: <15 years, 15–39 years, >39 years; tumor size: ≤80 mm, >80 mm). Histopathologically, osteosarcoma were divided into 12 subtypes according to the ICD-0-3 coding system: osteosarcoma, NOS (no other specific) (9180/3), chondroblastic osteosarcoma (9181/3), fibroblastic osteosarcoma (9182/3), telangiectatic osteosarcoma (9183/3), osteosarcoma in Paget disease (9184/3), small cell osteosarcoma (9185/3), central osteosarcoma (9186/3), intraosseous well-differentiated osteosarcoma (9187/3), parosteal osteosarcoma (9192/3), periosteal osteosarcoma (9193/3), high-grade surface osteosarcoma (9194/3), and intracortical osteosarcoma (9195/3), which contains both histological information and location [17]. Therefore, a subgroup Kaplan-Meier analysis based on histology and location information was also performed (histology group: osteosarcoma,

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

NOS, chondroblastic osteosarcoma, fibroblastic osteosarcoma, high-grade surface osteosarcoma, and others (small cell osteosarcoma, telangiectatic osteosarcoma, intraosseous well-differentiated osteosarcoma) (location: parosteal osteosarcoma, periosteal osteosarcoma and central osteosarcoma). The log-rank test was applied to compare the survival curves of each type of variable. For further analysis, the random forest (Ntree=500) was constructed for all variables. Random forest is an ensemble of unpruned decision trees, induced from bootstrap samples of the data, using random feature selection in the tree induction process. The mean decrease Gini (MDG) involved in the random forest algorithm was used to rank the influencing factors with probability of death. MDG provided ways to quantify which index contributed most to classification accuracy. Greater MDG indicated the degree of impurity arising from a category could be reduced farthest by 1 variable, thus suggesting an important associated index. Out-of-bag (OOB) error is the parameter for evaluating the classification accuracy of random forest [18].

After these procedures, we selected the best subsets of significant predictors to conduct the Cox proportional hazards model. Likelihood ratio test, Ward test, and log-rank test were used for model diagnosis. Eventually, we developed a model consisting of optimum predictors. Then, Lasso regression was performed to ensure that the multifactor models were not overfitting. The nomograms based on Cox proportional hazards model were built to predict the probability of OS and CSS. The discrimination and calibration of predictors were accessed by the C-index of internal validation and calibration curve, respectively.

Only 2-sided P value <0.05 was considered as statistical significance. All statistical analyses were conducted with R version 3.3.1 software (Institute for Statistics and Mathematics, Vienna, Austria; www.r-project.org). The R packages survival, survminer, ggplot2, pwr, and randomForest were used for modeling (including Power of Hypothesis Tests) and drawing survival curves. The nomograms were drawn by the rms package.

Results

Patient characteristics

The process of data selection is shown by the flow chart in Figure 1. The cohort consisted of 1000 patients with non-metastatic osteosarcoma from the SEER database. The characteristics of all the patients are described in Supplementary Table 1. The patients included 470 females and 530 males, with a mean age of 25.3 years (median 18.0 years, range, 3.0 to 89.0 years), similar to previous studies [10,22]. These non-metastatic osteosarcomas were dominantly localized or regional (96.5%),

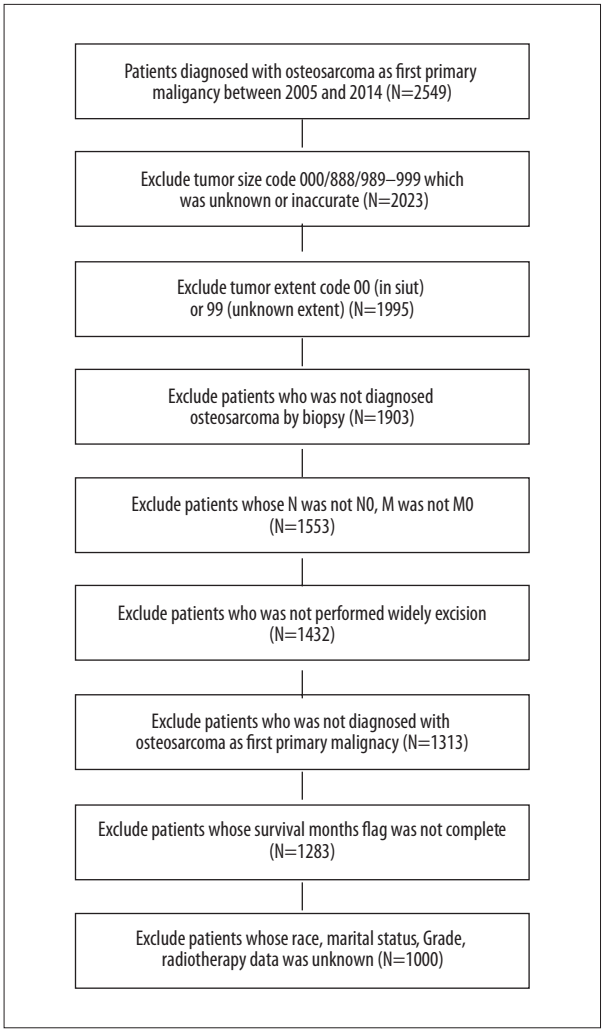


Figure 1. Flow diagram of patient selection.

grade IV (56.6%), and NOS histologically (59.9%), with a median size of 85.0 (range, 5.0 to 486.0) mm. During 10 years of follow-up, the median survival time was 46.8 (range, 0 to 119) months. The mean follow-up time was 46.77±37.90 months and all patients were active at follow-up. With the respect to the endpoint, 203 (20.3%) and 187 (18.7%) patients died of all and specific causes, respectively. Among all patients, most were unmarried (79.1%), while education levels and family incomes were distributed evenly.

Univariate analysis and random forest

The OS and CSS Kaplan-Meier curves of age and grade are shown in Supplementary Figure 1. The survival curves between age groups showed that OS and CSS were longest in patients under 15 years old and shortest in patients over 40 years old (P<0.001, Supplementary Figure 1A, P<0.001, Supplementary Figure 1B). Additionally, patients with grade I and grade II tumors had better OS (P<0.001, Supplementary Figure 1C) and

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

Table 1. Results of single-factor analysis and random forest model.

Variables	Overall survival (OS)			Cancer-specific survival (CSS)		
	P value of parameter or non-parametric test	P value of Kaplan-Meier survival analysis	MDG	P value of parameter or non-parametric test	P value of Kaplan-Meier survival analysis	MDG
Categorical age	<0.001*	<0.001*	17.9285255	<0.001*	<0.001*	16.1492625
Race	0.812	0.618	14.6161796	0.932	0.695	13.4284526
Sex	0.185	0.370	11.213705	0.093	0.215	10.166965
Marital status	<0.001*	0.046	8.571557	<0.001*	<0.001*	7.672955
Primary site	<0.001*	<0.001*	18.1616484	<0.001*	<0.001*	17.2527824
Grade	<0.001*	<0.001*	20.0879203	0.001*	0.001*	18.4483283
Histological subtype	0.002*	0.005*	26.2889922	0.004*	0.006*	23.6393802
AJCC.T staging	0.009*	0.001*	12.1364207	0.003*	<0.001*	11.6476987
Tumor size, mm	0.020*	0.110	66.4103101	0.008*	0.068	64.3342381
Radiation	0.002*	0.001*	6.437203	0.009*	0.003*	6.173988
Chemotherapy	0.116	0.124	6.346112	0.134	0.138	5.906247
9th grade education	0.100	0.152	9.34257	0.097	0.146	8.754056
High school education	0.791	0.848	7.818148	0.716	0.772	7.402536
At least bachelor's degree	0.735	0.619	7.475351	0.809	0.693	6.842956
Median family income	0.562	0.447	7.929627	0.599	0.473	7.066662
Unemployed	0.258	0.675	8.559674	0.307	0.728	7.783507
Families below poverty level	0.721	0.669	6.779123	0.991	0.634	6.658563
Families above poverty level	0.967	0.991	7.147897	0.875	0.900	7.001368

Categorical variables were compared by using the Pearson chi-square test. Continuous variables in normal distribution and homogeneity of variance were compared by using the two-sample *t* test, otherwise, the Mann-Whitney U test was performed.
OS – overall survival; CSS – cause-specific survival; MDG – Mean Decrease Gini; AJCC – American Joint Committee on Cancer. * P<0.05.

CSS (P<0.001, Supplementary Figure 1D) compared with patients with grade III and grade IV tumors. In addition, a subgroup Kaplan-Meier analysis based on histology and location information was also performed, showing no significant difference in prognosis compared with the reference group (osteosarcoma, NOS) in each subgroup (Supplementary Figure 2A–2D).

Univariate analysis and random forest for OS (OOB=20.60%) and CSS (OOB=19.50%) are shown in Table 1. All tumor characteristics, except for tumor size classified by 8 cm, showed significant associations with the survival time of patients in both parametric and non-parametric tests and in Kaplan-Meier survival analysis. In addition, they ranked in the top 7 MDG of the random forest model, the same as the age of patients. However, as a continuous variable, tumor size was significant in the parametric and non-parametric tests, ranking first in the random forest model. Multiple evidence in the guidelines showed that tumor size was associated with prognosis [19], which was also in line with our clinical experience. Therefore, we included these 6 variables (age, primary site, grade, histological subtype, AJCC T staging, and tumor size) in our further Cox modeling.

Because of the controversy over age in previous studies, we also performed univariate analysis, which suggested that older patients (>39 years) tended to have higher pathological grades (P<0.001) and higher AJCC T-stages (P=0.009) (Supplementary Tables 2, 3).

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

Table 2. Cox proportional hazards regression model for overall survival and cancer – specific survival in patients with non-metastatic osteosarcoma.

Variable	Overall survival (OS)	P	Cancer specific survival (CCS)	P
	Hazard ratio (95% CI)		Hazard ratio (95% CI)	
Categorical age				
<15	1.00 (reference)		1.00 (reference)	
15–39	1.616 (1.115 to 2.341)	0.043*	1.474 (1.012 to 2.149)	0.043*
>39	3.063 (2.020 to 4.644)	<0.001*	2.845 (1.860 to 4.351)	<0.001*
Primary site				
Bones of skull and face	1.00 (reference)		1.00 (reference)	
Bones of upper or lower limb	0.585 (0.359 to 0.953)	0.031*	0.629 (0.373 to 1.061)	0.082
Pelvic bones, sacrum, coccyx	1.486 (0.785 to 2.813)	0.224	1.759 (0.907 to 3.412)	0.095
Rib, sternum, clavicle	0.822 (0.313 to 2.157)	0.690	0.979 (0.367 to 2.607)	0.966
Vertebral column	0.842 (0.198 to 3.581)	0.816	1.018 (0.237 to 4.373)	0.981
Grade				
Grade I	1.00 (reference)		1.00 (reference)	
Grade II	1.192 (0.345 to 4.118)	0.782	1.606 (0.410 to 6.292)	0.496
Grade III	4.181 (1.461 to 11.968)	0.008*	4.702 (1.416 to 15.614)	0.011*
Grade IV	3.792 (1.347 to 10.671)	0.012*	4.394 (1.345 to 14.353)	0.014*
Histological subtype				
Osteosarcoma, NOS	1.00 (reference)		1.00 (reference)	
Central osteosarcoma	0.312 (0.099 to 0.979)	0.046*	0.337 (0.107 to 1.059)	0.063
Chondroblastic osteosarcoma	0.772 (0.517 to 1.153)	0.206	0.771 (0.509 to 1.168)	0.219
Fibroblastic osteosarcoma	0.773 (0.433 to 1.382)	0.385	0.631 (0.328 to 1.214)	0.168
High-grade surface osteosarcoma	0.832 (0.204 to 3.394)	0.797	0.906 (0.222 to 3.703)	0.891
Parosteal osteosarcoma	0.613 (0.234 to 1.602)	0.318	0.514 (0.177 to 1.493)	0.221
Periosteal osteosarcoma	0.316 (0.078 to 1.283)	0.107	0.334 (0.082 to 1.358)	0.125
Others	0.322 (0.118 to 0.881)	0.027*	0.341 (0.125 to 0.935)	0.037*
AJCC T				
T1	1.00 (reference)		1.00 (reference)	
T2	0.913 (0.605 to 1.378)	0.666	0.908 (0.593 to 1.391)	0.657
T3	2.955 (1.383 to 6.313)	0.005*	3.103 (1.441 to 6.682)	0.004*
Tumor size	1.005 (1.002 to 1.008)	0.002*	1.006 (1.002 to 1.009)	<0.001*

OS – overall survival; CSS – cause-specific survival; MDG – Mean Decrease Gini; AJCC – American Joint Committee on Cancer. * P<0.05.

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

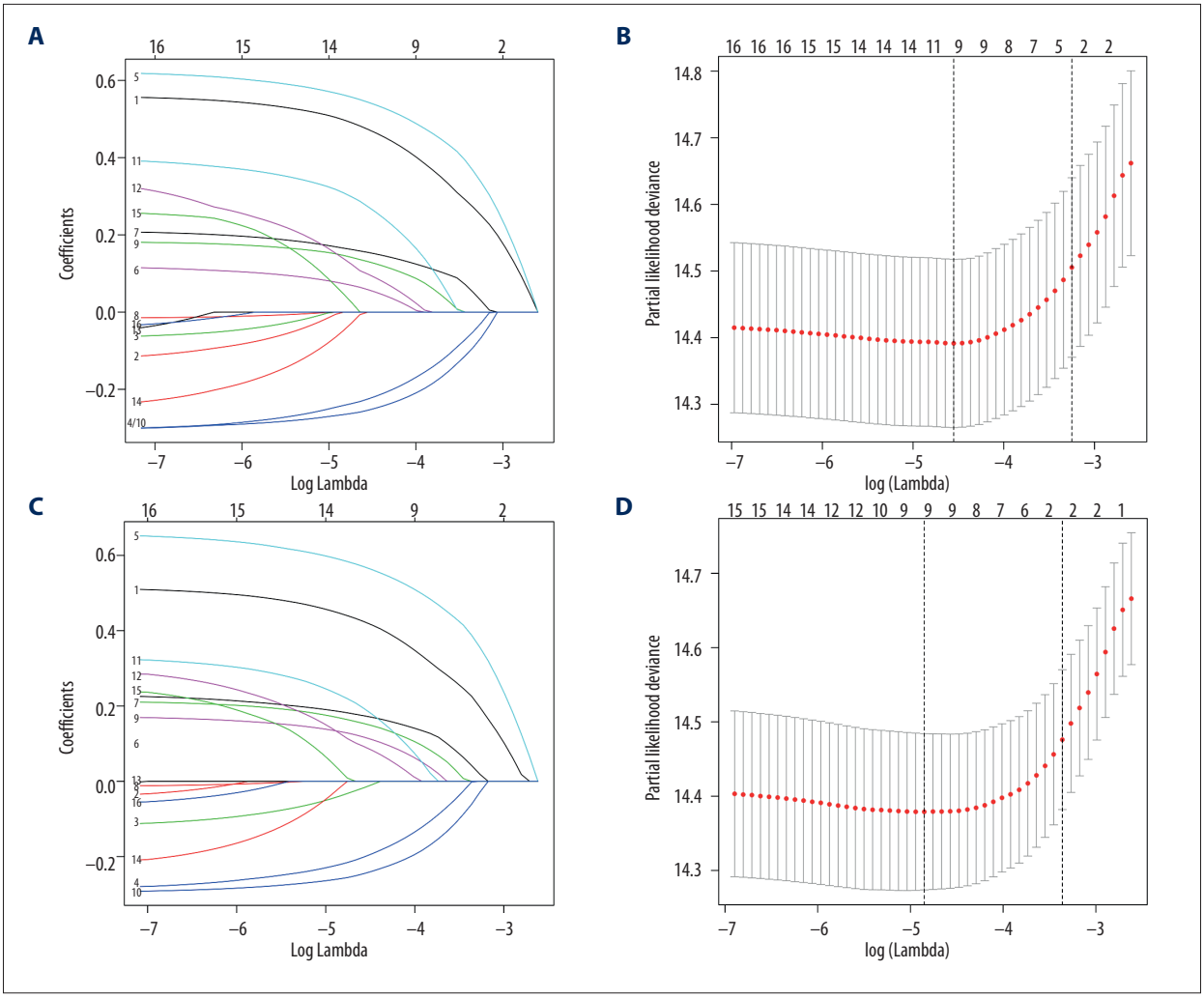


Figure 2. The Lasso regression variable-filtering process. To avoid overfitting, Lasso regression suggested including 9 and 12 variables when overall survival (OS) (**A, B**) and cause-specific survival (CSS) (**C, D**) was the endpoint, respectively.

Cox proportional hazards model and Lasso regression

The Cox proportional hazard regression model was constructed to confirm the effects of the covariates mentioned above on the OS and CSS of patients (Table 2). All variables eventually incorporated into the multivariate models were shown to be essential for modeling in the Lasso regression (Figure 2). Compared with patients younger than 15 years old, older age was associated with poorer OS (15–39 years: HR, 1.616; 95% CI, 1.115 to 2.341; P=0.043) (older than 39 years: HR, 3.063; 95% CI, 2.020 to 4.644; P<0.001) and CSS (15–39 years: HR, 1.474; 95% CI, 1.012 to 2.149; P=0.043) (older than 39 years: HR, 2.845; 95% CI, 1.860 to 4.351; P<0.001). Pelvic osteosarcoma, showing a significantly worse prognosis in the survival curve, was not a prognostic indicator in our multivariate model. Furthermore, grade III and grade IV were independently associated with worse OS and CSS (OS with grade III *vs.* grade I: HR, 4.181; 95% CI, 1.461 to 11.968; P=0.008; CSS with grade III *vs.* grade I: HR, 4.702; 95% CI, 1.416 to 15.614; P=0.011; OS with grade IV *vs.* grade I: HR, 3.792; 95% CI, 1.347 to 10.671; P=0.012; CSS with grade IV *vs.* grade I: HR, 4.394; 95% CI, 1.345 to 14.353; P=0.014). In histological subtypes, only central (HR, 0.312; 95% CI, 0.099 to 0.979; P=0.046) and other subtypes of osteosarcoma (HR, 0.322; 95% CI, 0.118 to 0.881; P=0.027) trended to better OS than osteosarcoma, NOS. As for CSS, other subtypes osteosarcoma (HR, 0.341; 95% CI, 0.125 to 0.935; P=0.037) trended to better outcome than reference. Finally, increasing tumor size was associated with worse OS (HR, 1.005; 95% CI, 1.002 to 1.008; P=0.002) and CSS (HR, 1.006; 95% CI, 1.002 to 1.009; P<0.001), while the results only showed that AJCC T3 staging was a risk factor for poor prognosis of OS (HR, 2.955; 95% CI, 1.383 to 6.313; P=0.005) and CSS (HR, 3.103; 95% CI, 1.441 to 6.682; P=0.004) compared with T1 staging, but not T2 staging, which was distinguished with T1 by 8 cm of tumor size.

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

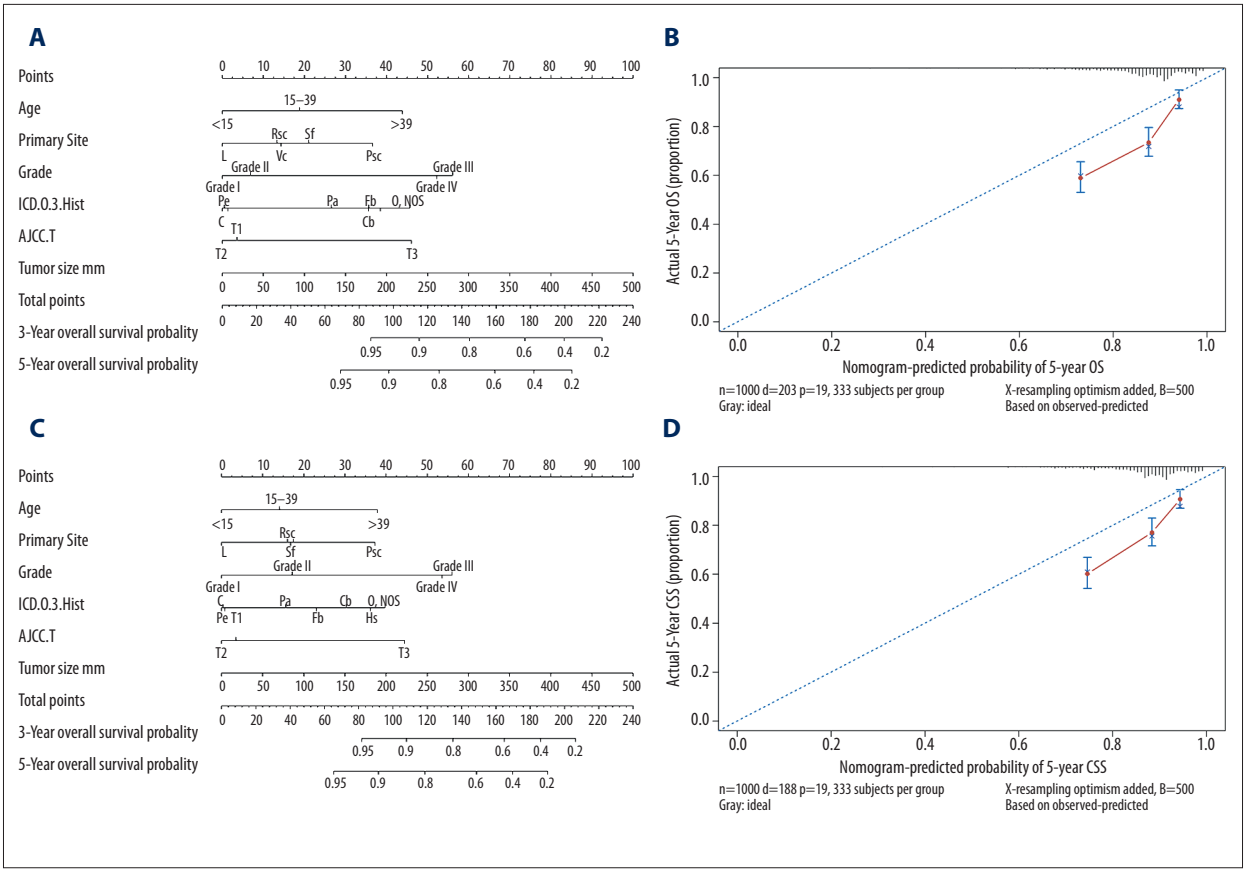


Figure 3. Nomograms and calibration curves of overall survival (OS) (**A, B**) and cause-specific survival (CSS) (**C, D**). OS – overall survival; CSS – cause-specific survival; AJCC – American Joint Committee on Cancer; ICD – International Classification of Diseases; O, NOS – osteosarcoma, No other specific; C – central osteosarcoma; Cb – chondroblastic osteosarcoma; Fb – fibroblastic osteosarcoma; Hs – high grade surface osteosarcoma; Iw – intraosseous well differentiated osteosarcoma; Pa – parosteal osteosarcoma; Pe – periosteal osteosarcoma; Sc – small cell osteosarcoma; T – telangiectatic osteosarcoma; Sf – bones of skull and face; V – vertebral column; Rsc – rib, sternum, clavicle; L – bones of upper or lower limb; Psc – pelvic bones, sacrum, coccyx.

For sensitivity analysis, we used ANOVA to compare the models reduced by each variable and the full model including all 6 variables, showing each variable was indispensable for modeling: all 6 ANOVAs showed significant statistical results (P<0.05), indicating that after removing any of the 6 variables, the multivariable model was statistically significantly different from the previous model. Therefore, each variable was essential to the modeling process. The statistical power tests also showed the sufficiency of the sample size.

Nomogram

The nomograms predicting the probability 3- and 5-year OS (Internal validation C-index, 0.7095) and CSS (Internal validation C-index, 0.7100) (Figure 3A, 3C) were constructed based on the Cox proportional hazard regression models with the total cohort as the training dataset (available in Supplementary Materials B). The calibration plot of the CIF is shown in supplementary materials Figure 3B and 3D. Even without external validation, the points slightly further from the 45-degree line indicate some inconsistencies between predictions and observations. Additionally, compared with sixth (OS: 0.613; CSS: 0.628) and seventh edition AJCC staging systems (0.602; 0.613), the nomograms predicted the probability of OS (0.7095) and CSS (0.7100) more accurately (Supplementary Figure 3).

Discussion

Osteosarcoma is the most common histologic type of bone tumor, with low incidence rate but high fatality rate [14,20]. As it occurs mainly in adolescents or young adults, osteosarcoma severely harms the social workforce. Prone to early-stage metastasis, many osteosarcoma cases are diagnosed at the time when poor prognosis is inevitable, affecting patients both physically and mentally [9,21]. Thus, finding the most important

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

prognostic factors at the pre-metastasis stage enables timely management and improves prognosis to a great extent, which can be much more effective and less costly.

In our series, we divided age into childhood (<15 years), adolescent and young adult (AYA, 15-39 years), and old adult (>40 years) according to the NCCN Guidelines for Adolescent and Young Adult Oncology [16]. We confirmed that patients older than 40 years had the poorest CSS and OS, not only in univariate and multivariate analysis, but also in the random forest model. Therefore, age was regarded as an independent prognostic factor for postoperative CSS and OS of non-metastatic osteosarcoma in this study. Age at diagnosis, a controversial factor, was not considered by Harting et al. [9] to be a significant independent prognostic variable for OS and disease-free survival in extremities and torso osteosarcoma. However, Grimer et al. [23] and Kager et al. [24] did not draw the same conclusion. It is widely accepted that osteosarcoma in childhood often tends to be highly malignant, leading to poor prognosis [16]. However, in this study, we performed univariate analysis between age and other factors, showing that older patients (>39 years) tended to have higher pathological grades, higher AJCC T-stages, and larger tumor sizes. This is a new fact based on the SEER database and it can explain why increasing age was a risk factor for poor prognosis of OS and CSS.

In many previous studies, tumor size and primary site had been demonstrated to be the most significant prognostic variables for many malignant bone tumors, especially for non-metastatic osteosarcoma [9,25,26], because both of them determine the difficulty of the operation and feasibility of en bloc tumor resection [27]. For instance, either a huge osteosarcoma or a spinal one can invade and involve fateful vessels or the spinal cord, which may compromise sufficient surgical excision range. For malignant tumors, sufficient surgical excision range is the prerequisite for prolonging survival [28,29]. However, our findings suggest that the primary site and tumor size were not independent risk indicators for poor prognosis. In this study, we selected patients with non-metastatic osteosarcoma after primary site operations; therefore, all the candidates were regarded as having sufficient surgical excision range. Consequently, the primary site and tumor size did not affect the OS and CSS in this study.

The subtypes of histologic grade are defined as well differentiated (Grade I), moderately differentiated (Grade II), poorly differentiated (Grade III), and undifferentiated (Grade IV) in the SEER database, while AJCC T staging is widely used as a tumor staging method, setting 8 cm as an important cut-off value to distinguish different T stages [28,29]. Both tumor grade and AJCC T staging were suggested to be independent prognostic factors for postoperative CSS and OS in this study. Compared with AJCC T1 staging, T3 staging (but not T2 staging) was a risk factor for poor prognosis of OS and CSS.

Histopathologically, we divided osteosarcoma into 2 subgroups according to histology and position relation between tumor and bone. In line with some previous research results, in multifactor modeling, most histological subtypes were not significantly associated with patient prognosis compared with the reference group (Osteosarcoma, NOS) in our study [11,22,25,29]. Subgroup analysis obtained similar results. This might be because histological types, such as chondroblastic and fibroblastic, are only a description of the cell components of the tumor, and location only shows the positional relationship between tumor and bone, both of which had little to do with the malignancy of the tumor. However, patients with poorly differentiated and undifferentiated pathological grading showed worse prognosis in our series. Hence, for stratifying tumor pathological grading, our findings suggested that grade and AJCC T staging were better prognostic risk factors than was histological subtype.

Primary site treatment combined with preoperative and postoperative adjuvant therapy is the standard treatment strategy for osteosarcoma, aim at resecting the primary tumor completely, relieving symptoms, reconstructing function, and preventing local recurrence [16,19]. Aggressive surgery associated with systemic chemotherapy (neoadjuvant chemotherapy and adjuvant chemotherapy) and radiotherapy are essential for cure and for controlling localized and micro-metastatic disease, which was shown to be vital to postoperative prognosis in previous research [6,30]. However, our results suggested that the addition of chemotherapy or radiotherapy contributed little to the classification of the random forest model, showing that the variables had little effect on the outcome, in accordance with some previous studies [10,11]. Although patients without metastasis who received definitive surgery were regarded as good candidates for chemotherapy, the addition of chemotherapy did not have any effect on the outcome [10,11]. However, the missing data of the 2 variables in the SEER database must also be considered.

To the best of our knowledge, this is the first study combining conventional univariate and multivariate survival analysis with a machine learning model to explore the most important prognostic factors of osteosarcoma. Nevertheless, our study has some limitations. First, although the study had large sample size and multiple variables based on the SEER database, there were still some inaccurate variables in this database. Second, it has all the limitations inherent in retrospective studies. Third, although the nomogram was verified to be applicable in terms of internal validation, calibration, and clinical usefulness, it included fewer variables than other studies [31]. Last but not least, because the database contained data from multiple centers, its inter-group heterogeneity was not processed,

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

even though we used strict inclusion and exclusion criteria to minimize this heterogeneity. A more rigorous nomogram is required to compensate for the imperfection in survival prediction of this serious disease [32]. In this regard, the nomogram should contain the expression level of biomarkers (such as some newfound prognosis-related functional genetic single-nucleotide polymorphisms (SNPs) and transcription factors) most associated with these prognostic factors, which might be found by weighted correlation network analysis (WGCNA) and deep learning [33–35]. This will be our next research focus.

Conclusions

Despite its limitations, this study shows that increasing age, high histopathological grade, and T staging were the most significant prognostic factors of both OS and CSS in patients with non-metastatic osteosarcoma after primary site operations. In addition, for a tumor with such high mortality, a more accurate prediction model achieving more accuracy and higher safety should be developed, which might be constructed by adding the expression level of some hub-genes in the nomogram.

Acknowledgements

We thank the Surveillance, Epidemiology, and End Results Program team of the National Cancer Institute for allowing use of their data.

Conflicts of interest

None.

Appendix

Supplementary materials A

Supplementary Table 1. Baseline characteristics of patients with non-metastatic osteosarcoma.

Demographic or characteristic	Total patients (N=1000)		Alive cohort (N=797)		Dead cohort (N=203)	
	No.	%	No.	%	No.	%
Age, years						
Mean	25.3		23.8		31.5	
Median (range)	18.0 (3.0–89.0)		17.0 (3.0–84.0)		22.0 (3.0–89.0)	
Categorical age						
<15	316	31.6%	275	34.5%	41	20.2%
15–39	483	48.3%	385	48.3%	98	48.3%
>39	201	20.1%	137	17.2%	64	31.5%
Sex						
Female	470	47.0%	383	48.1%	87	42.9%
Male	530	53.0%	414	51.9%	116	57.1%
Race						
Black	150	15.0%	117	14.7%	33	16.3%
Other	99	9.9%	78	9.8%	21	10.3%
White	751	75.1%	602	75.5%	149	73.4%
Tumor size, mm						
Mean	96.0		93.4		106.2	
Median (range)	85.0 (5.0–486.0)		85.0 (5.0–369.0)		95.0 (14.0–486.0)	
Categorical tumor size						
≤80 mm	452	45.2%	369	46.3%	83	40.9%

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

Demographic or characteristic	Total patients (N=1000)		Alive cohort (N=797)		Dead cohort (N=203)	
	No.	%	No.	%	No.	%
>80 mm	548	54.8%	428	53.7%	120	59.1%
Primary site						
Bones of skull and face	105	10.5%	79	9.9%	26	12.8%
Bones of upper or lower limb	823	82.3%	674	84.6%	149	73.4%
Pelvic bones, sacrum, coccyx	43	4.3%	22	2.8%	21	10.3%
Rib, sternum, clavicle	20	2.0%	15	1.9%	5	2.5%
Vertebral column	9	.9%	7	.9%	2	1.0%
Grade						
Grade I	58	5.8%	54	6.8%	4	2.0%
Grade II	92	9.2%	85	10.7%	7	3.4%
Grade III	284	28.4%	216	27.1%	68	33.5%
Grade IV	566	56.6%	442	55.5%	124	61.1%
Histological subtype						
Central osteosarcoma	45	4.5%	42	5.3%	3	1.5%
Chondroblastic osteosarcoma	151	15.1%	118	14.8%	33	16.3%
Fibroblastic osteosarcoma	59	5.9%	46	5.8%	13	6.4%
High-grade surface osteosarcoma	9	.9%	7	.9%	2	1.0%
Osteosarcoma, NOS	599	59.9%	458	57.5%	141	69.5%
Parosteal osteosarcoma	67	6.7%	62	7.8%	5	2.5%
Periosteal osteosarcoma	19	1.9%	17	2.1%	2	1.0%
Others	51	5.1%	47	5.9%	4	2.0%
AJCC T staging						
T1	445	44.5%	367	46.0%	78	38.4%
T2	534	53.4%	418	52.4%	116	57.1%
T3	21	2.1%	12	1.5%	9	4.4%
Radiation						
Beam radiation	51	5.1%	31	3.9%	20	9.9%
Combination of beam with implants or isotopes	2	0.2%	2	0.3%	0	0.0%
None/Unknown	947	94.7%	764	95.9%	183	90.1%
Chemotherapy						
No/Unknown	170	17.0%	143	17.9%	27	13.3%
Yes	830	83.0%	654	82.1%	176	86.7%
Survival time						
Mean	46.8		51.1		29.7	
Median (range)	40.0 (0–119.0)		41.0 (0–119.0)		25.0 (3.0–89.0)	
Marital status						
Married	209	20.9%	145	18.2%	64	31.5%
Single/separated/divorced/widowed	791	79.1%	652	81.8%	139	68.5%

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

Demographic or characteristic	Total patients (N=1000)		Alive cohort (N=797)		Dead cohort (N=203)	
	No.	%	No.	%	No.	%
9th grade education						
Lower 50%	480	48.0%	393	49.3%	87	42.9%
Upper 50%	520	52.0%	404	50.7%	116	57.1%
High school education						
Lower 50%	496	49.6%	397	49.8%	99	48.8%
Upper 50%	504	50.4%	400	50.2%	104	51.2%
At least bachelor's degree						
Lower 50%	423	42.3%	335	42.0%	88	43.3%
Upper 50%	577	57.7%	462	58.0%	115	56.7%
Median family income (in tens)						
Lower 50%	496	49.6%	399	50.1%	97	47.8%
Upper 50%	504	50.4%	398	49.9%	106	52.2%
Unemployed						
Lower 50%	459	45.9%	373	46.8%	86	42.4%
Upper 50%	541	54.1%	424	53.2%	117	57.6%
Families below poverty level						
Lower 50%	489	48.9%	392	49.2%	97	47.8%
Upper 50%	511	51.1%	405	50.8%	106	52.2%
Families above poverty level						
Lower 50%	420	42.0%	335	42.0%	85	41.9%
Upper 50%	580	58.0%	462	58.0%	118	58.1%

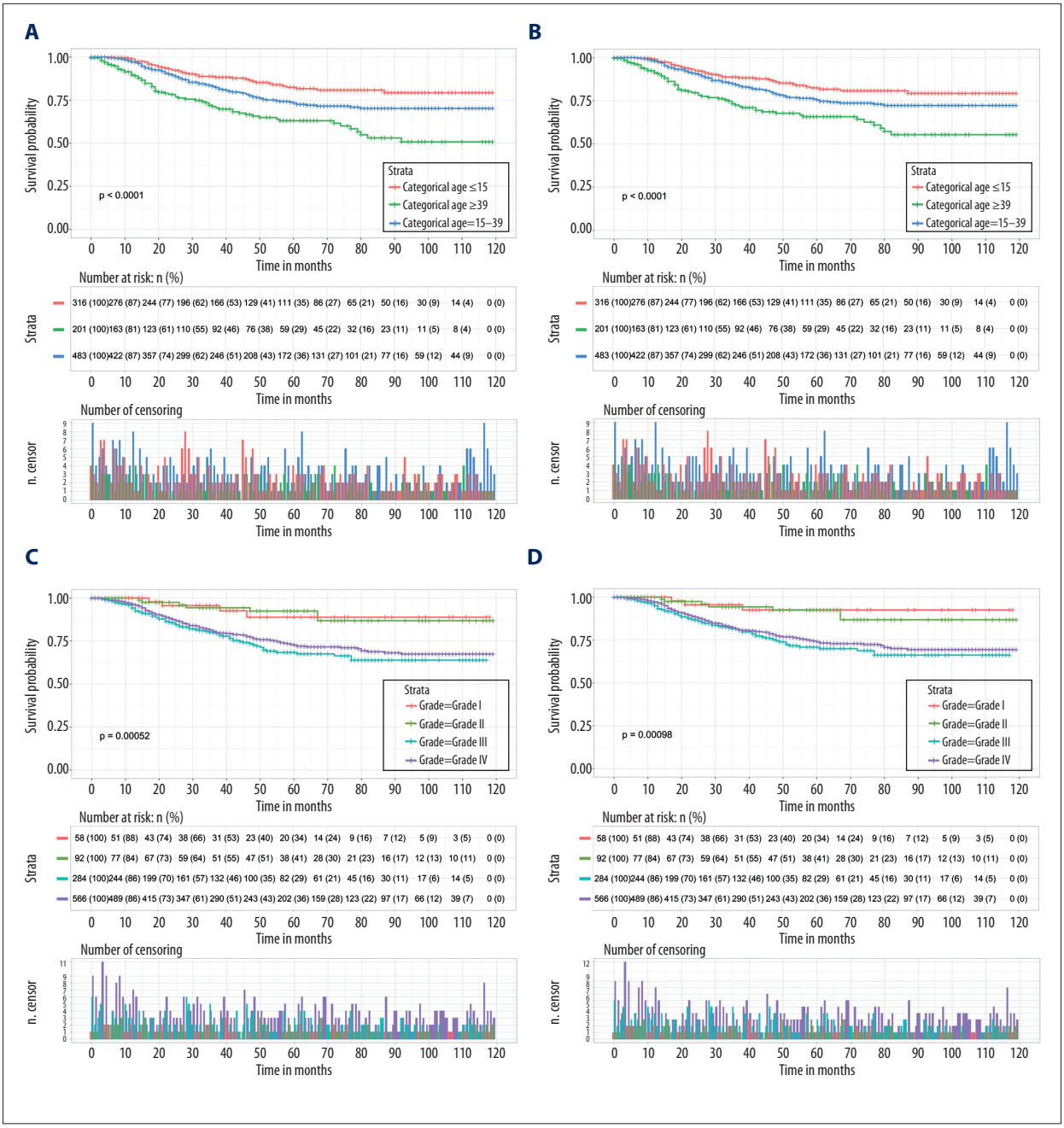
OS – overall survival; CSS – cause-specific survival; AJCC – American Joint Committee on Cancer.

Supplementary Table 2. Univariate analysis results between age and histopathological grade.

Grade	Age	OR	95% CI	P value
Grade II	0–14	1.00 (reference)		
	15–39	1.167	0.426–3.197	0.764
	>39	0.939	0.329–2.680	0.907
Grade III	0–14	1.00 (reference)		
	15–39	0.415	0.181–0.950	0.037*
	>39	0.219	0.091–0.525	0.001*
Grade IV	0–14	1.00 (reference)		
	15–39	0.380	0.170–0.851	0.019*
	>39	0.167	0.072–0.388	<0.001*

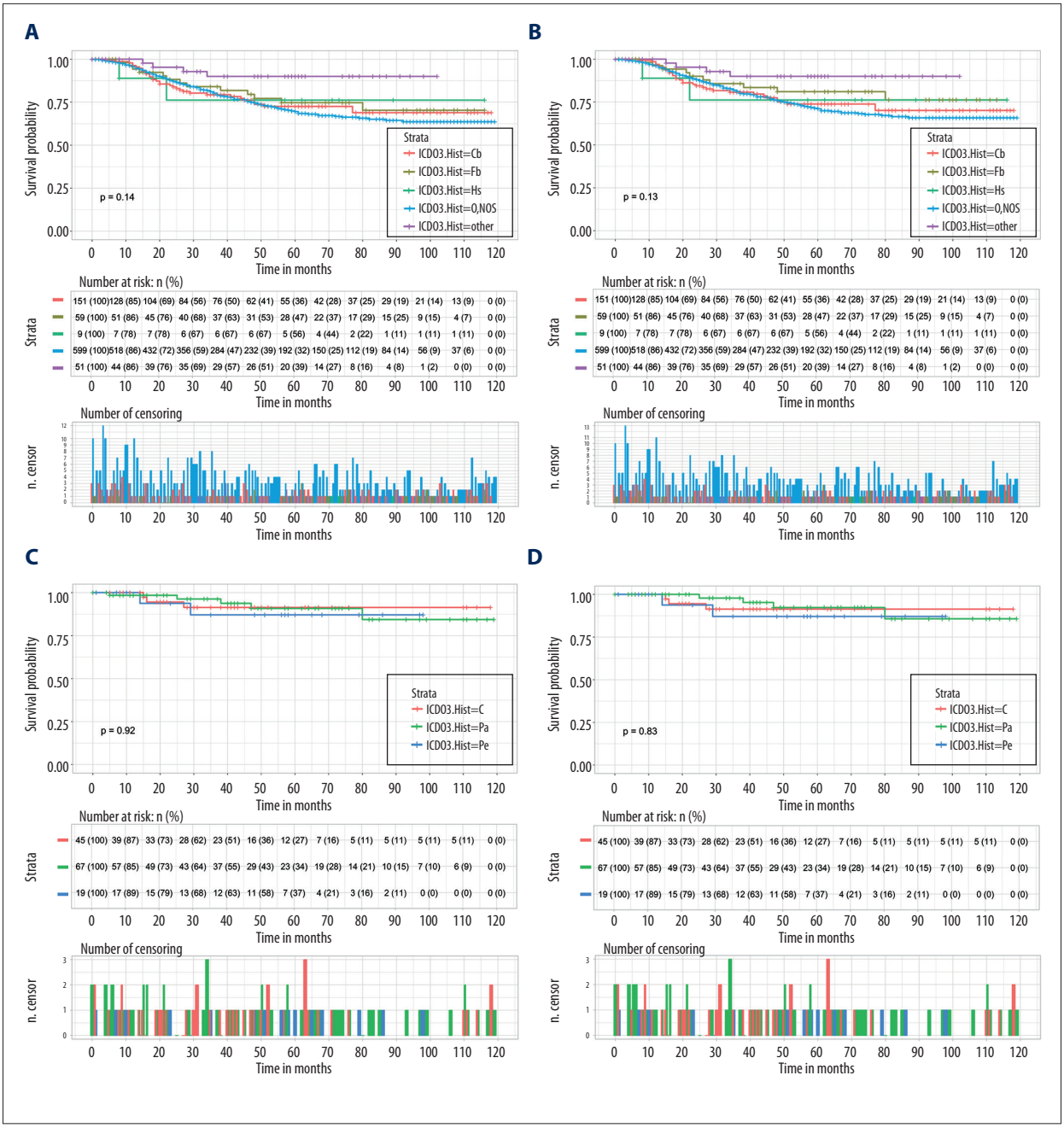
Grade I is the reference group. OR – odds ratio; CI – confidence interval. * P<0.05.

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]



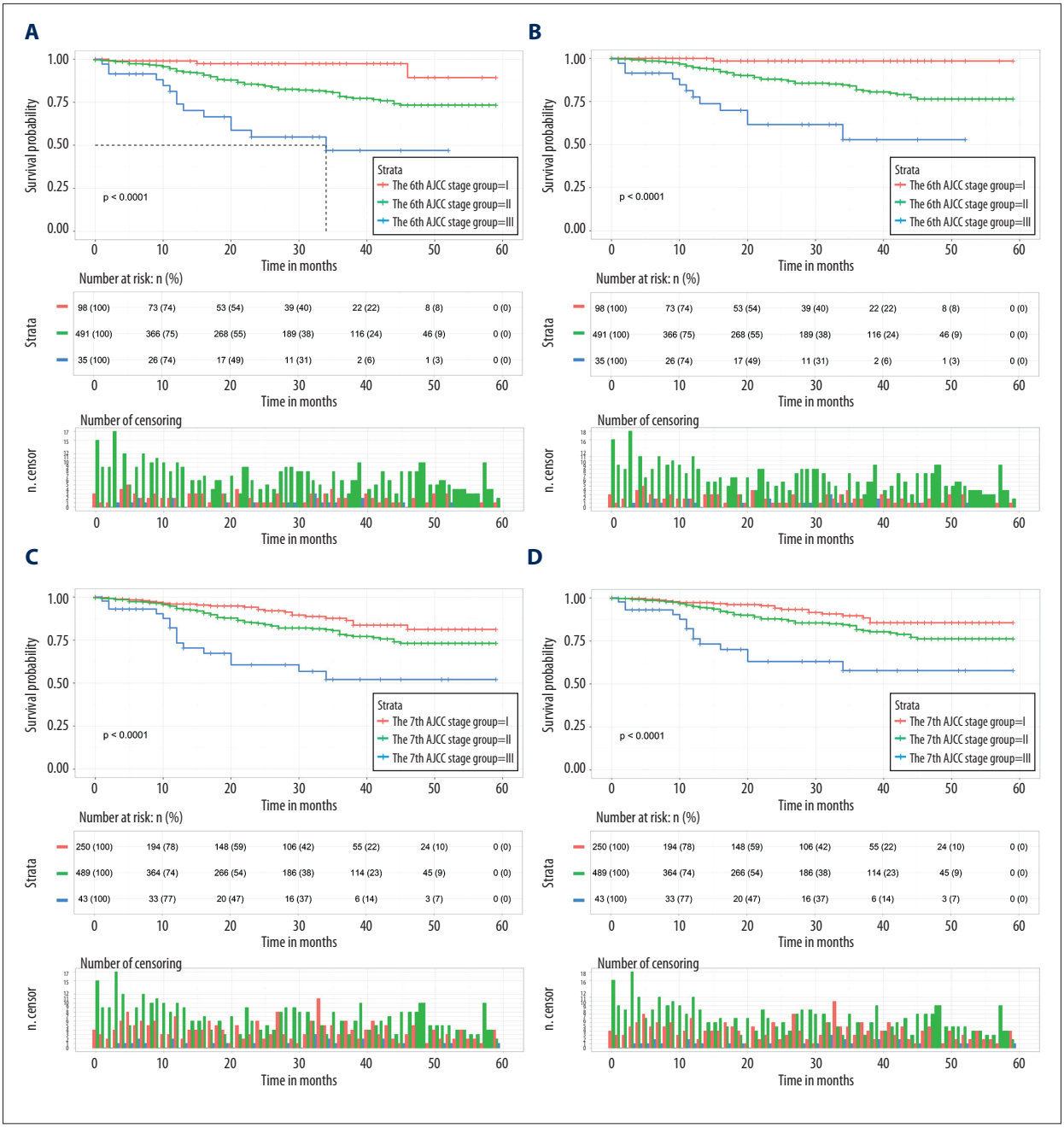
Supplementary Figure 1. Kaplan-Meier curves of overall survival (OS) (**left**) and cause-specific survival (CSS) (**right**) for age (**A, B**) and grade (C, D).

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]



Supplementary Figure 2. Kaplan-Meier curves of overall survival (OS) (**left**) and cause-specific survival (CSS) (**right**) for histological subtype (**A, B**) and location subtype (**C, D**). ICD – International Classification of Diseases; O, NOS – osteosarcoma, No other specific; C – central osteosarcoma; Cb – chondroblastic osteosarcoma; Fb – fibroblastic osteosarcoma; Hs – high-grade surface osteosarcoma; Iw – intraosseous well-differentiated osteosarcoma; Pa – parosteal osteosarcoma; Pe – periosteal osteosarcoma; Sc – small cell osteosarcoma; T – telangiectatic osteosarcoma; Other – small cell osteosarcoma, telangiectatic osteosarcoma and intraosseous well-differentiated osteosarcoma.

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]



Supplementary Figure 3. Kaplan-Meier curves of overall survival (OS) (**left**) and cause-specific survival (CSS) (**right**) for sixth (**A, B**) and seventh (**C, D**) edition American Joint Committee on Cancer (AJCC) staging systems.

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

Supplementary materials B

Supplementary Table 3. Univariate analysis results between age and American Joint Committee on Cancer T staging.

AJCC T	Age	OR	95% CI	P value
T2	0–14	1.00 (reference)		
	15–39	0.984	0.736–1.316	0.913
	>39	0.651	0.454–0.933	0.019*
T3	0–14	1.00 (reference)		
	15–39	0.316	0.116–0.864	0.025*
	>39	0.306	0.084–1.113	0.072

T1 is the reference group. OR – odds ratio; CI – confidence interval; AJCC – American Joint Committee on Cancer. * P<0.05.

References:

1. Noone AM, Howlader N, Krapcho M et al: SEER cancer statistics review, 1975–2015, National Cancer Institute. Bethesda, MD. Based on November 2017 SEER data submission, posted to the SEER web site, April 2018. Available from: URL: *https://seer.cancer.gov/csr/1975_2015/*
2. Zambo I, Vesely K: [WHO classification of tumours of soft tissue and bone 2013: The main changes compared to the 3rd edition]. Cesk Patol, 2014; 50: 64–70 [in Czech]
3. Chou AJ, Geller DS, Gorlick R: Therapy for osteosarcoma: Where do we go from here? Paediatr Drugs, 2008; 10: 315–27
4. Chen Y, Gokavarapu S, Shen Q et al: Chemotherapy in head and neck osteosarcoma: Adjuvant chemotherapy improves overall survival. Oral Oncol, 2017; 73: 124–31
5. Aksnes LH, Bauer HC, Jebsen NL et al: Limb-sparing surgery preserves more function than amputation: A Scandinavian sarcoma group study of 118 patients. J Bone Joint Surg Br, 2008; 90: 786–94
6. Federman N, Bernthal N, Eilber FC, Tap WD: The multidisciplinary management of osteosarcoma. Curr Treat Options Oncol, 2009; 10: 82–93
7. Gelderblom H, Jinks RC, Sydes M et al: Survival after recurrent osteosarcoma: Data from 3 European Osteosarcoma Intergroup (EOI) randomized controlled trials. Eur J Cancer, 2011; 47: 895–902
8. Kim W, Han I, Lee JS et al: Post-metastasis survival in high-grade extremity osteosarcoma: A retrospective analysis of prognostic factors in 126 patients. J Surg Oncol, 2018; 117(6): 1223–31
9. Harting MT, Lally KP, Andrassy RJ et al: Age as a prognostic factor for patients with osteosarcoma: an analysis of 438 patients. J Cancer Res Clin Oncol, 2010; 136: 561–70
10. Pakos EE, Nearchou AD, Grimer RJ et al: Prognostic factors and outcomes for osteosarcoma: an international collaboration. Eur J Cancer, 2009; 45: 2367–75
11. Iwata S, Ishii T, Kawai A et al: Prognostic factors in elderly osteosarcoma patients: A multi-institutional retrospective study of 86 cases. Ann Surg Oncol, 2014; 21: 263–68
12. Colding-Rasmussen T, Thorn AP, Horstmann P et al: Survival and prognostic factors at time of diagnosis in high-grade appendicular osteosarcoma: A 21-year single-institution evaluation from east Denmark. Acta Oncol, 2018; 57: 420–25
13. Hegyi M, Semsei AF, Jakab Z et al: Good prognosis of localized osteosarcoma in young patients treated with limb-salvage surgery and chemotherapy. Pediatr Blood Cancer, 2011; 57: 415–22
14. Howlader NNA, Krapcho M, Miller D et al: SEER cancer statistics review, 1975–2014, National Cancer Institute. Bethesda, MD, 2017. Based on November 2016 SEER data submission, posted to the SEER web site, April 2017. Available from: URL: *https://seer.cancer.gov/csr/1975_2014/*
15. Biermann JS, Chow W, Reed DR et al: NCCN guidelines insights: Bone cancer, Version 2.2017. J Natl Compr Canc Netw, 2017; 15: 155–67
16. Coccia PF, Pappo AS, Beaupin L et al: Adolescent and young adult oncology, Version 2.2018, NCCN clinical practice guidelines in oncology. J Natl Compr Canc Netw, 2018; 16: 66–97
17. Quan HD, Sundararajan V, Halfon P et al: Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. Med Care, 2005; 43: 1130–39
18. James G, Witten D, Hastie T, Tibshirani R: An introduction to statistical learning with applications in R introduction. Springer Texts Stat, 2013; 103: 1–14
19. Biermann JS, Adkins DR, Agulnik M et al: Bone cancer. J Natl Compr Canc Netw, 2013; 11(6): 688–723
20. Ottaviani G, Jaffe N: The etiology of osteosarcoma. Cancer Treat Res, 2009; 152: 15–32
21. Lagmay JP, Krailo MD, Dang H et al: Outcome of patients with recurrent osteosarcoma enrolled in seven phase II trials through children's cancer group, pediatric oncology group, and children's oncology group: Learning from the past to move forward. J Clin Oncol, 2016; 34: 3031–38
22. Miller BJ, Cram P, Lynch CF, Buckwalter JA: Risk factors for metastatic disease at presentation with osteosarcoma: An analysis of the SEER database. J Bone Joint Surg Am, 2013; 95: e89
23. Grimer RJ, Cannon SR, Taminiau AM et al: Osteosarcoma over the age of forty. Eur J Cancer, 2003; 39: 157–63
24. Kager L, Zoubek A, Potschger U et al: Primary metastatic osteosarcoma: Presentation and outcome of patients treated on neoadjuvant Cooperative Osteosarcoma Study Group protocols. J Clin Oncol, 2003; 21: 2011–18
25. Joo MW, Shin SH, Kang YK et al: Osteosarcoma in Asian populations over the age of 40 years: A multicenter study. Ann Surg Oncol, 2015; 22: 3557–64
26. Whelan JS, Jinks RC, McTiernan A et al: Survival from high-grade localised extremity osteosarcoma: Combined results and prognostic factors from three European Osteosarcoma Intergroup randomised controlled trials. Ann Oncol, 2012; 23: 1607–16
27. Biermann JS Chow W, Reed DR et al: Bone cancer, version 2.2017 featured updates to the NCCN guidelines. J Natl Compr Canc Netw, 2017; 15(2): 155–67
28. Bacci G, Forni C, Longhi A et al: Local recurrence and local control of nonmetastatic osteosarcoma of the extremities: A 27-year experience in a single institution. J Surg Oncol, 2007; 96: 118–23
29. Cates JMM: Modeling continuous prognostic factors in survival analysis: implications for tumor staging and assessing chemotherapy effect in osteosarcoma. Am J Surg Pathol, 2018; 42: 485–91
30. Harrison DJ, Geller DS, Gill JD et al: Current and future therapeutic approaches for osteosarcoma. Expert Rev Anticancer Ther, 2018; 18: 39–50
31. Ogura K, Fujiwara T, Yasunaga H et al: Development and external validation of nomograms predicting distant metastases and overall survival after neoadjuvant chemotherapy and surgery for patients with nonmetastatic osteosarcoma: A multi-institutional study. Cancer-Am Cancer Soc, 2015; 121: 3844–52

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

32. Balachandran VP, Gonen M, Smith JJ, DeMatteo RP: Nomograms in oncology: More than meets the eye. Lancet Oncol, 2015; 16: E173–80
33. Langfelder P, Horvath S: WGCNA: An R package for weighted correlation network analysis. BMC Bioinformatics, 2008; 9: 559
34. Zhang H, Mao JS, Hu WF: Functional genetic single-nucleotide polymorphisms (SNPs) in cyclin-dependent kinase inhibitor 2A/B (CDKN2A/B) locus are associated with risk and prognosis of osteosarcoma in Chinese populations. Med Sci Monit, 2019; 25: 1307–13
35. Jiang Z, Zhang W, Chen Z et al: Transcription factor 21 (TCF21) rs12190287 polymorphism is associated with osteosarcoma risk and outcomes in east Chinese population. Med Sci Monit, 2017; 23: 3185–91

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]