Development and validation of a nomogram for predicting overall survival of patients with cancer of unknown primary: a real-world data analysis

Yizi Jin[1,2#], Mingxi Lin[1,2#], Zhiguo Luo[1,2], Xichun Hu[1,2], Jian Zhang[1,2]

[1]Department of Medical Oncology, Fudan University Shanghai Cancer Center, Shanghai, China; [2]Department of Oncology, Shanghai Medical College, Fudan University, Shanghai, China

Contributions: (I) Conception and design: Y Jin, M Lin, J Zhang, X Hu; (II) Administrative support: None; (III) Provision of study materials or patients: None; (IV) Collection and assembly of data: Y Jin, M Lin; (V) Data analysis and interpretation: All authors; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

[#]These authors contributed equally to this work.

Correspondence to: Jian Zhang, PhD; Xichun Hu, PhD. Department of Medical Oncology, Fudan University Shanghai Cancer Center; Department of Oncology, Shanghai Medical College, Fudan University, No. 270, Dong'an Road, Shanghai 200032, China.
Email: zhangjian1001@csco.ac.cn; xchu2009@hotmail.com.

Background: Cancer of unknown primary (CUP) has a variable prognosis and lacks any standard staging systems. We aim to improve the prediction of survival in patients with CUP by constructing a nomogram based on a real-world, population analysis.

Methods: We performed a population analysis of patients diagnosed with CUP between 2010 and 2016 in the Surveillance, Epidemiology, and End Results (SEER) database. Patients with complete study variables were respectively assigned to training and validation cohorts by diagnostic time. A prognostic nomogram was established based on the multivariate Cox proportional hazards model and was evaluated through calculating the Harrell's C-index and plotting calibration curves.

Results: In total, 19,543 patients were identified under the selection criteria, and 3,347 cases with complete study variables were included for developing and validating the nomogram. Covariates incorporated in the final nomogram were sex, age, histological type, surgery, radiotherapy, chemotherapy, and the number of metastatic organs. The Harrell's C-index of nomogram was 0.705 (95% CI: 0.692–0.717) for the training cohort and 0.727 (95% CI: 0.703–0.752) for the validation cohort.

Conclusions: We developed and validated the first nomogram based on a large population, which showed good prediction ability for predicting overall survival of patients with CUP. The risk stratification based on this nomogram could also help clinicians in treatment planning. This nomogram requires further validation in external cohorts, since important clinical factors such as favorable/unfavorable subset, performance status, lactate dehydrogenase, blood cell counts, or metastatic patterns limited to multiple lymph nodes could not be considered due to the lack of availability of these data.

Keywords: Unknown primary tumors; nomogram; prognostic model; real-world study

Page 2 of 14

Jin et al. Nomogram for cancer of unknown primary

Introduction

Cancer of unknown primary (CUP), also known as occult primary, is a group of metastatic malignancies whose anatomical primary site cannot be detected after a complete clinical evaluation (1). The incidence of CUP has been reported to range from 2% to 7.8% of all cancers (2-4). Meanwhile, it remains third to sixth in the ranking of cancer mortality (5,6) despite the advances in diagnostic pathological tests and genomic approaches in recent years. About 20% of patients identified with some favorable clinicopathological features may benefit from locoregional treatments or specific first-line therapies, while the rest (around 80%) patients still lack standard anti-cancer regimens (4,7,8).

Due to its metastatic and highly heterogeneous nature, CUP shows a variable prognosis and lacks standard staging system for its prognosis indication and treatment planning. Several studies have investigated the prognostic models of CUP; however, they were developed based on small populations and lacked precise computation for survival probability. A more accurate and feasible prediction model is still in need for estimating individual life expectancy in the current clinical setting. In addition, the benefits of different treatments in patients with CUP have been scarcely evaluated in large populations.

In this study, we aim to identify the epidemiological characteristics of CUP, including the association of survival and different treatment regimens, and build the first prognostic nomogram for CUP based on the real-world data analysis of patients in the Surveillance, Epidemiology, and End Results (SEER) (www.seer.cancer.gov) database with internal validation. We present the following article in accordance with the TRIPOD reporting checklist (available at http://dx.doi.org/10.21037/atm-20-4826).

Methods

Data source and patient population

Patients for this study were identified in the SEER 18 Database (November 2018 submission, 1975–2016 varying) (9) with SEER*Stat Software (Version 8.3.6). Since the information for combined metastasis was only available after 2010 in SEER, only patients diagnosed between 2010 and 2016 were included in this study. The following inclusion criteria were used to determine the study population: (I) patients were histologically diagnosed with CUP site (documented as ICD-O-3 C80.9); (II) patients

aged ≥18 years; (III) patients had active follow-up data. Patients with hematological malignancies (ICD-O-3 Hist/behavior, malignant code 9590/3-9989/3) were excluded. Patients with 0 month of survival time were also excluded to eliminate cases diagnosed at autopsy alone. In total, 19,543 patients were identified under the selection criteria.

All eligible patients with available data for metastatic status at the bone, brain, liver, and/or lung (the information of metastases was only available at the above four sites in the database) were included in the further nomogram study to optimize the generalizability of the nomogram. We split the nomogram study population into training set and validation set by time, which is considered a stronger design than random division for assessing model performance when only a single data set is available (10). Ultimately, 3,347 cases with complete study variables were included for developing and validating nomogram. Patients diagnosed between 2010 and 2015 were used for model development (training set, N=2,286), and those diagnosed in 2016 were used for model evaluation (validation set, N=1,061).

Study variables and outcomes

The demographic and clinical information of included patients were extracted, including patient ID, year of diagnosis, sex, race, age at diagnosis, insurance status, marital status at diagnosis, histology (ICD-O-3 Hist/behavior, malignant), diagnostic confirmation, surgical procedure, radiotherapy, chemotherapy, combined metastasis (metastases) at diagnosis (at the bone, brain, liver, and/or lung), survival months, and vital status (study cutoff used). We interpreted the meanings of variables according to SEER variable dictionary and coding manuals. The histological groups were classified as squamous cell carcinomas (SCC) (ICD-O-3 8050–8089), neuroendocrine carcinomas (ICD-O-3 8013/3, 8041/3, 8153/3, 824), carcinomas not otherwise specified (NOS) (ICD-O-3 801 except for 8013/3, 802, 803, 8046/3), adenocarcinomas (ICD-O-3 814, 8160/3, 819, 820, 825–855), undifferentiated malignant neoplasms (ICD-O-3 800) and other types (malignancies that not classified as carcinomas) according to the European Society for Medical Oncology (ESMO) guideline (11) and ICD-O-3 SEER Site/Histology Validation List. Although the information of metastasis was limited in the database, metastatic status at the bone, brain, liver, and/or lung was generally acknowledged as representative for patients' metastatic burden. Two new variables were set to conclude patients' metastatic status:

"metastatic status at major viscera" (brain/liver/lung) and "the number of metastatic organs" (only referred to the situation in the bone/brain/liver/lung).

Surgery regimens included non-primary surgery to CUP, other regional sites, distant lymph node(s), other distant sites, and any combination of the latter three procedures. Radiotherapy modalities included beam radiation, radioactive implants, radioisotopes, and other methods not specified. Chemotherapy regimens were not detailed for CUP patients in the database.

Patients' survival months were calculated as "FLOOR ((endpoint-date of diagnosis)/days in a month)". We chose 6- and 9-month overall survival (OS) as the endpoint respectively. The end event for OS was defined as death or last follow-up for cases alive.

Statistical analysis

The OS of patients was calculated through the Kaplan-Meier Method, with log-rank test assessing the differences between each group. We performed univariate Cox regression analyses to evaluate the association of study variables and patients' OS (12). The benefits of different treatment options were further evaluated through univariate analyses in subgroups of patients combined with or without metastasis (at the bone, brain, liver, and/or lung). Candidate predictors of the nomogram were selected based on both clinical and statistical considerations. Potential covariates for modeling were first selected according to clinical relevance based on our literature review, including sex, race, age at diagnosis, insurance status, marital status at diagnosis, histological type, surgery, radiotherapy, chemotherapy, metastatic status at major viscera, and the number of metastatic organs. Except for age (the only continuous variable) and the number of metastatic organs (the only rank variable), all the other candidates were categorical variables. A collinearity diagnosis was performed before further variable selection. The Martingale residual test was performed to confirm the linear relation of the continuous predictor and the outcome (13). Candidate predictors were selected through a backward eliminated Cox's proportional hazards analysis with the likelihood ratio tests according to the statistical recommendations (14,15). A nomogram was finally established based on the results of the multivariate analysis. In the nomogram, each factor was assigned a score, which summed up to a total score corresponding to the predicted survival probability. The predicted 6- and 9-month OS probability for individuals could be obtained

through calculation. The Cochran–Armitage test for trend was performed for the rank variable by treating it as a continuous variable in the above model. Hazard ratios (HRs) with 95% confidence intervals (CIs) were calculated for each predictor.

The discrimination performance of nomogram was evaluated through calculating Harrell's concordance-index (C-index) and drawing 6- and 9-month OS calibration plots of training and validation sets, respectively (10). A higher value of C-index (ranging from 0.5 to 1.0) indicates a better discrimination ability (16). Validations in both cohorts were performed using the bootstrap resampling method.

Based on each patient's total score in the nomogram, we stratified all patients into low-, intermediate-, and high-risk groups respectively. We also performed risk stratification separately for the training cohort and the validation cohort.

All statistical analyses were performed using IBM SPSS Statistics 26.0, and R 3.6.3 with package rms, MASS, Hmisc, formula, survival, and ggplot2. A two-tailed P value less than 0.05 was considered statistically significant in this study.

This study was performed in accordance with the Declaration of Helsinki (as was revised in 2013).

Results

Demographics, clinicopathological characteristics, and treatment options

The demographics, clinicopathological characteristics, and treatment options of patients with CUP are shown in *Table 1*. The median OS was 6.0 (95% CI: 5.8–6.2) months for all the included patients. Among all the histological types, adenocarcinoma was the most common type. Analysis of patients with available metastatic information demonstrated that liver was the most common site for combined metastasis at diagnosis among these four organs, followed by the bone, lung, and brain.

Among the 19,543 patients with CUP, only 34.2% received chemotherapy, and the proportion was even smaller for the administration of radiotherapy (24.3%) or non-primary surgery (14.4%). In the subgroup analysis (*Figure 1*), among patients either combined with or without metastasis at the bone, brain, liver, and/or lung, surgical procedures to CUP were significantly associated with better survival (P value <0.001), while patients without metastasis showed more benefits (HR =2.495, 95% CI: 2.026–3.073). The use of radiotherapy was only associated with a better

Table 1 The demographics, clinicopathological characteristics, and treatment options of patients with cancer of unknown primary (2010–2016)

Characteristics	All patients (N=19,543)		Patients in the nomogram (N=3,347)	
	No. of cases (%)	OS (months), median (95% CI)	No. of cases (%)	OS (months), median (95% CI)
Age (years)				
Mean ± SD	68.03±13.18		68.07±12.87	
Median	69		68	
Range	18 to 103		19 to 102	
Sex				
Male	1,0191 (52.1)	7.0 (6.7 to 7.3)	1,753 (52.4)	7.0 (6.1 to 7.9)
Female	9,352 (47.9)	6.0 (5.7 to 6.3)	1,594 (47.6)	6.0 (5.2 to 6.8)
Race				
White	15,914 (81.4)	7.0 (6.7 to 7.3)	2,713 (81.1)	7.0 (6.3 to 7.7)
Black	2,168 (11.1)	5.0 (4.5 to 5.5)	432 (12.9)	5.0 (3.7 to 6.3)
Other	1,461 (7.5)	6.0 (5.3 to 6.8)	202 (6.0)	6.0 (3.9 to 8.1)
Insurance status				
Insured[a]	17,828 (91.2)	6.0 (5.8 to 6.2)	3,126 (93.4)	7.0 (6.4 to 7.6)
No/unknown	1,715 (8.8)	7.0 (6.0 to 8.0)	221 (6.6)	7.0 (4.1 to 9.9)
Marital status at diagnosis				
Married/domestic partners	9,583 (49.0)	7.0 (6.6 to 7.4)	1,659 (49.6)	8.0 (7.1 to 8.9)
Single/widowed/divorced/separated	9,960 (51.0)	6.0 (5.7 to 6.3)	1,688 (50.4)	6.0 (5.4 to 6.6)
Histological types				
Squamous cell carcinoma	3,764 (19.3)	29.0 (25.4 to 32.6)	600 (17.9)	27.0 (17.1 to 36.9)
Neuroendocrine carcinoma	3,180 (16.3)	12.0 (11.1 to 12.9)	625 (18.7)	12.0 (9.6 to 14.4)
Neoplasm and other types[b]	1,370 (7.0)	6.0 (5.1 to 6.9)	225 (6.7)	8.0 (4.9 to 11.1)
Carcinoma, NOS	3,646 (18.7)	4.0 (3.7 to 4.3)	627 (18.7)	4.0 (3.4 to 4.6)
Adenocarcinoma	7,583 (38.8)	4.0 (3.8 to 4.2)	1,270 (37.9)	4.0 (3.6 to 4.4)
Non-primary surgery				
Yes	2,807 (14.4)	28.0 (24.7 to 31.3)	535 (16.0)	23.0 (16.8 to 29.2)
Surgery to CUP	1,846 (9.4)		336 (10.0)	
Surgery to other regional sites	76 (0.4)		21 (0.6)	
Surgery to distant lymph nodes	187 (1.0)		41 (1.2)	
Surgery to distant site	563 (2.9)		116 (3.5)	
Any combination[c]	135 (0.7)		21 (0.6)	
No/unknown	16,736 (85.6)	5.0 (4.8 to 5.2)	2,812 (84.0)	5.0 (4.5 to 5.5)
No	16,414 (84.0)		2,774 (82.9)	
Unknown	322 (1.6)		38 (1.1)	
Radiotherapy				
Yes	4,740 (24.3)	12.0 (11.2 to 12.8)	874 (26.1)	12.0 (10.1 to 14.0)
Beam radiation	4,603 (23.6)		846 (25.3)	

Table 1 (*continued*)

Table 1 (*continued*)

Characteristics	All patients (N=19,543)		Patients in the nomogram (N=3,347)	
	No. of cases (%)	OS (months), median (95% CI)	No. of cases (%)	OS (months), median (95% CI)
Radiation, NOS method or source not specified	61 (0.3)		11 (0.3)	
Radioactive implants	37 (0.2)		5 (0.1)	
Radioisotopes	34 (0.2)		10 (0.3)	
Combination of beam with implants or isotopes	5 (0.0)		2 (0.1)	
No/unknown[d]	14,803 (75.7)	5.0 (4.8 to 5.2)	2,473 (73.9)	5.0 (4.5 to 5.5)
None/Unknown	14,512 (74.3)		2,427 (72.5)	
Refused	166 (0.8)		24 (0.7)	
Recommended, unknown if administered	125 (0.6)		22 (0.7)	
Chemotherapy				
Yes	6,682 (34.2)	10.0 (9.6 to 10.4)	1,235 (36.9)	10.0 (9.2 to 10.8)
No/unknown[d]	12,861 (65.8)	4.0 (3.8 to 4.2)	2,112 (63.1)	4.0 (3.5 to 4.5)
Bone involvement				
Yes	–	–	670 (20.0)	3.0 (2.5 to 3.5)
No	–	–	2,677 (80.0)	8.0 (7.2 to 8.8)
Brain involvement				
Yes	–	–	190 (5.7)	3.0 (2.5 to 3.5)
No	–	–	3,157 (94.3)	7.0 (6.4 to 7.6)
Liver involvement				
Yes	–	–	982 (29.3)	4.0 (3.6 to 4.4)
No	–	–	2,365 (70.7)	9.0 (8.1 to 9.9)
Lung involvement				
Yes	–	–	536 (16.0)	3.0 (2.5 to 3.5)
No	–	–	2,811 (84.0)	8.0 (7.2 to 8.8)
Visceral metastasis[e]				
Yes	–	–	1,275 (38.1)	4.0 (3.6 to 4.4)
No	–	–	2,072 (61.9)	10.0 (8.9 to 11.1)
Number of metastatic organs[f]				
0			1,836 (54.9)	12.0 (10.6 to 13.4)
1	–	–	895 (26.7)	5.0 (4.3 to 5.7)
2	–	–	407 (12.2)	3.0 (2.6 to 3.4)
3	–	–	167 (5.0)	3.0 (2.4 to 3.6)
4	–	–	42 (1.3)	2.0 (1.2 to 2.8)

[a], including patients insured, insured/no specifics, and any Medicaid; [b], including undifferentiated malignant neoplasms and other histological types not classified as carcinomas; [c], any combination of surgical procedure to other regional sites, distant lymph nodes, and/or distant site; [d], as originally coded in the database, patients unknown if be treated with radiotherapy/chemotherapy were mixed with those who were not treated; [e], limited to brain, liver, and/or lung; refers to metastasis at diagnosis; [f], limited to bone, brain, liver, and/or lung; refers to metastasis at diagnosis. SD, standard deviation; OS, overall survival; CI, confidence interval; NOS, not other specified; CUP, cancer of unknown primary.
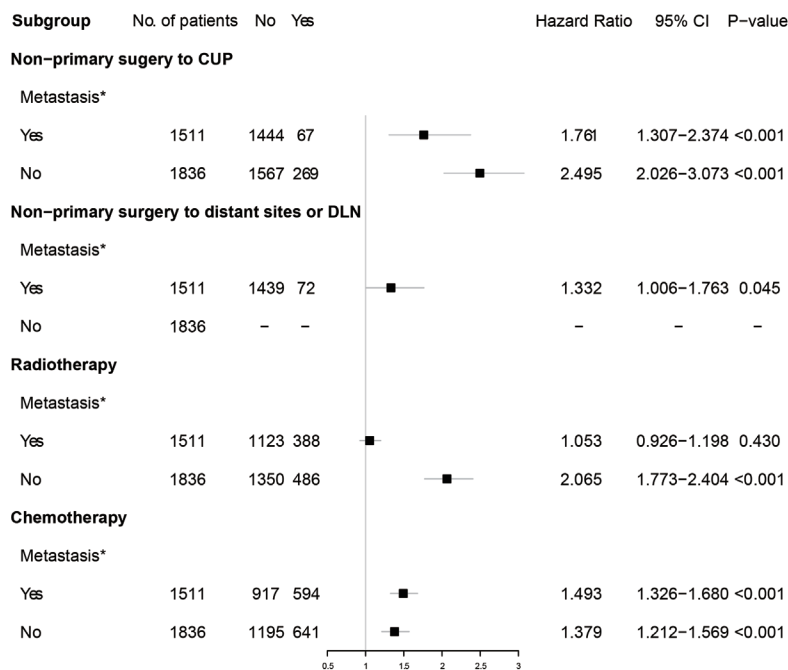
Page 6 of 14

Jin et al. Nomogram for cancer of unknown primary

Figure 1 Evaluating the benefits of different treatment options in the subgroups of patients combined with or without metastasis. Yes/No: refers to did/did not undergo the certain therapy. *, limited to the bone, brain, liver, and/or lung; refers to metastasis at diagnosis. CUP, cancer of unknown primary; DLN, distant lymph nodes; CI, confidence interval.

outcome in patients without metastasis at the above four sites (P value <0.001). Chemotherapy demonstrated benefits in both groups (P value <0.001).

Predictors of OS and nomogram construction

The univariate analysis of potential prognosticators is demonstrated in *Table 2*. Marital status at diagnosis was identified as a significant covariate in the univariate analysis, but was excluded from multivariate Cox model through backward elimination method. Cases combined with bone/brain/liver/lung metastasis all demonstrated significantly inferior survival compared with those without (all P values <0.001). Additionally, patients with metastasis at brain/liver/lung showed even poorer prognosis than those who without metastasis at these sites (HR =1.857, 95% CI: 1.688–2.042, P value <0.001).

Covariates incorporated in the final nomogram were sex, age, histological type, surgery, radiotherapy, chemotherapy, and the number of metastatic organs. The HR of each variable for multivariate Cox regression analysis is shown in *Table 2*. In the multivariate analysis, patients who were female and younger at diagnosis had better prognosis. SCC

showed significantly superior prognosis than carcinomas NOS and adenocarcinomas. It also indicated that more metastatic organs led to a significantly higher risk of mortality (P value for trend <0.001). The nomogram is illustrated in *Figure 2*, and its points assignment is listed in *Tables 3-5*.

Nomogram validation

The Harrell's C-index of nomogram was 0.705 (95% CI: 0.692–0.717) for the training cohort and 0.727 (95% CI: 0.703–0.752) for the validation cohort, representing a good discrimination ability. The calibration plots for 6- and 9-month OS are shown in *Figure 3*, which indicated a good agreement between predictive and observed value for both cohorts.

Performance of the nomogram in risk stratification

All patients included in the nomogram were stratified into three groups: patients with low risk (33.3%; total score <140), patients with intermediate risk (35.7%; 140≤ total score <180), and patients with high risk (31.0%; total score

Table 2 Univariate and multivariate analysis of overall survival for training cohort (N=2,286)

Characteristics	No. of patients with an outcome	Univariate analysis			Multivariate analysis		
		Hazard ratio	95% CI	P value	Hazard ratio	95% CI	P value
Sex							
Male	938	Reference	–	–	Reference	–	–
Female	870	1.066	0.972 to 1.169	0.176	0.894	0.814 to 0.983	0.020
Race							
Other	105	Reference	–	–	–	–	–
White	1,436	0.869	0.713 to 1.059	0.165	–	–	–
Black	267	1.114	0.889 to 1.397	0.347	–	–	–
Age at diagnosis	–	1.018	1.014 to 1.021	<0.001	1.015	1.011 to 1.019	<0.001
Insurance status					–	–	–
Insured	1,690	Reference	–	–			
No/unknown	118	0.93	0.771 to 1.120	0.443			
Marital status at diagnosis					–	–	–
Married/domestic partners	857	Reference	–	–			
Single/widowed/divorced/ separated	951	1.173	1.070 to 1.287	0.001			
Histological types							
Squamous cell carcinoma	235	Reference	–	–	Reference	–	–
Neuroendocrine carcinoma	284	1.611	1.355 to 1.915	<0.001	1.074	0.891 to 1.295	0.455
Neoplasm and other types[a]	100	1.455	1.151 to 1.839	0.002	1.067	0.839 to 1.357	0.598
Carcinoma, NOS	385	2.746	2.331 to 3.234	<0.001	2.084	1.758 to 2.471	<0.001
Adenocarcinoma	804	2.947	2.543 to 3.414	<0.001	2.103	1.792 to 2.469	<0.001
Non-primary surgery							
Yes	213	Reference	–	–	Reference	–	–
No/unknown	1,595	2.278	1.973 to 2.630	<0.001	1.905	1.644 to 2.207	<0.001
Radiotherapy							
Yes	414	Reference	–	–	Reference	–	–
No/unknown	1,394	1.525	1.366 to 1.702	<0.001	1.274	1.132 to 1.434	<0.001
Chemotherapy							
Yes	653	Reference	–	–	Reference	–	–
No/unknown	1,155	1.222	1.110 to 1.346	<0.001	1.367	1.235 to 1.514	<0.001
Bone involvement					–	–	–
No	1,384	Reference	–	–			
Yes	424	1.792	1.605 to 2.000	<0.001			

Table 2 (*continued*)

Page 8 of 14

Jin et al. Nomogram for cancer of unknown primary

Table 2 (*continued*)

Characteristics	No. of patients with an outcome	Univariate analysis			Multivariate analysis		
		Hazard ratio	95% CI	P value	Hazard ratio	95% CI	P value
Brain involvement					–	–	–
No	1,693	Reference	–	–			
Yes	115	1.588	1.314 to 1.919	<0.001			
Liver involvement					–	–	–
No	1,224	Reference	–	–			
Yes	584	1.779	1.610 to 1.966	<0.001			
Lung involvement					–	–	–
No	1,462	Reference	–	–			
Yes	346	1.831	1.625 to 2.062	<0.001			
Visceral metastasis[b]					–	–	–
No	1,035	Reference	–	–			
Yes	773	1.857	1.688 to 2.042	<0.001			
Number of metastatic organs[c]							P trend <0.001
0	884	Reference	–	–	Reference	–	–
1	534	1.742	1.562 to 1.941	<0.001	1.649	1.473 to 1.846	<0.001
2	261	2.293	1.992 to 2.639	<0.001	2.196	1.900 to 2.538	<0.001
3	103	2.539	2.066 to 3.121	<0.001	2.471	2.001 to 3.053	<0.001
4	26	3.759	2.538 to 5.567	<0.001	5.185	3.492 to 7.699	<0.001

[a], including undifferentiated malignant neoplasms and other histological types not classified as carcinomas; [b], limited to brain, liver, and/ or lung; refers to metastasis at diagnosis; [c], limited to bone, brain, liver, and/or lung; refers to metastasis at diagnosis. CI, confidence intervals; NOS, not other specified.

≥180) (*Figure 4*). For all cohorts, the median OS of the low-, intermediate-, and high-risk groups was 28.0 (95% CI: 22.1–33.9) months, 6.0 (95% CI: 5.3–6.7) months, and 2.0 (95% CI: 1.8–2.2) months, respectively. The survival curves demonstrated that the risk stratification based on the nomogram could accurately discriminate the outcome of patients (log-rank test, P value <0.001).

Discussion

As a group of cancers with high heterogeneity and generally unsatisfactory outcome, CUP calls for a more practical and accurate prediction model for individual prognosis against the background of precision medicine. In this real-world study with a large population, we constructed the first prognostic nomogram of CUP. With good discrimination

ability, this nomogram provides a quantitative tool for survival prediction of CUP on an individual basis. Additionally, we performed a risk stratification based on this nomogram, which could distinguish between different risk groups. This may assist clinicians in identifying high-risk patients and treatment planning.

Based on the SEER database, our study showed a median OS of 6.0 (95% CI: 5.8–6.2) months for the whole study cohort with CUP. This result was within the range of survival time reported in previous studies, which varied from 3 to 14.2 months depending on the characteristics of different study populations. However, this median survival still revealed a relatively poor prognosis of our study cohort according to recent clinical guidelines (17,18).

Several studies have discussed the prognostic factors of CUP in the past decades. Parameters with prognostic value
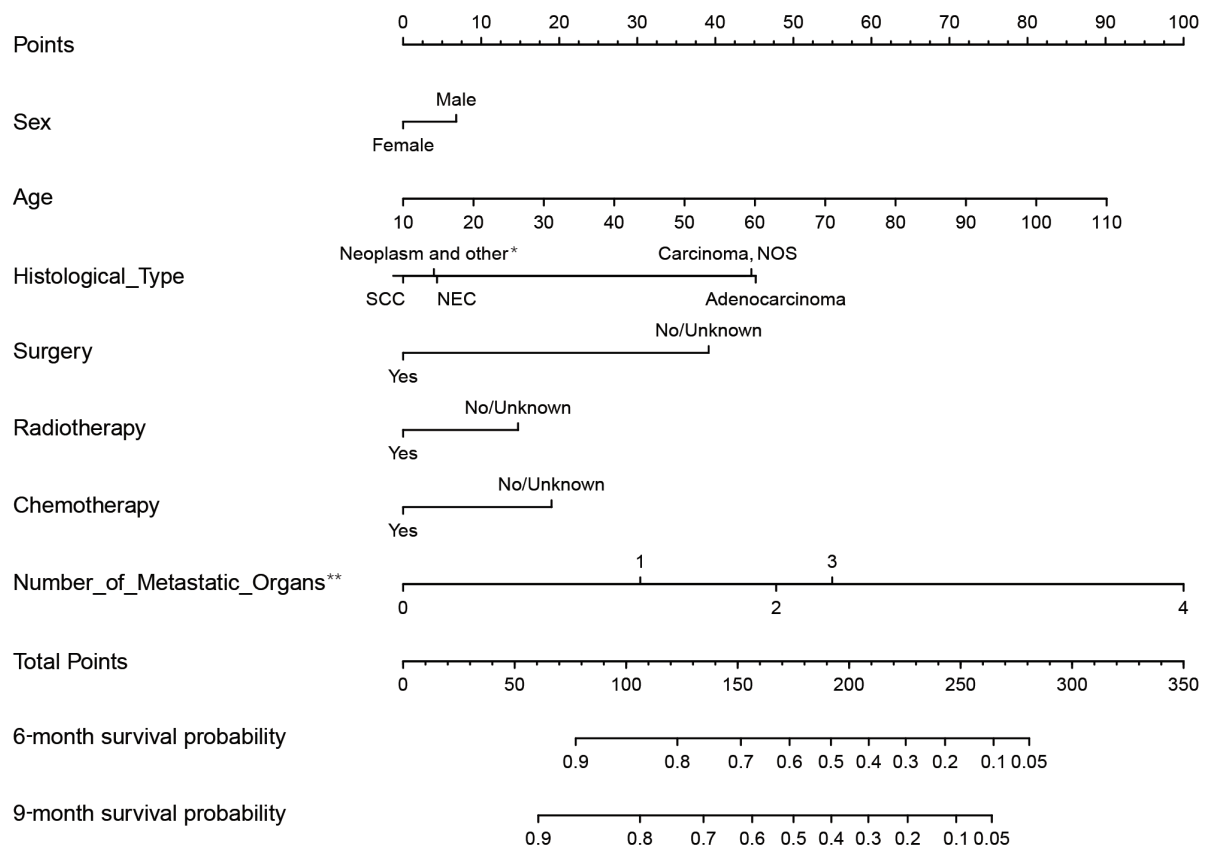
Figure 2 A nomogram for predicting 6- and 9-month overall survival of patients with cancer of unknown primary. Each variable was assigned a score, which summed up to a total score corresponding to the predicted overall survival probability. *, including undifferentiated malignant neoplasms and other histological types not classified as carcinomas; **, limited to the bone, brain, liver, and/or lung; refers to metastasis at diagnosis. SCC, squamous cell carcinoma; NEC, neuroendocrine carcinoma; NOS, not otherwise specified.

reported in the previous literature include age, sex, smoking history, performance status (PS), histology, site of metastasis, tumor location, number of metastasis, liver metastasis, symptoms, multiple comorbidities, prolonged QT interval, lactate dehydrogenase (LDH), alkaline phosphatase (ALP), albumin (ALB), leukocytosis, lymphopenia, treatment, and socioeconomic factors (3,19-24). Our study showed a consistent result of male gender and older age as unfavorable features, and also included these two factors in the final nomogram. Meanwhile, Petrakis *et al.* also pointed out that it should be biological rather than chronological age that acted as the main predictor (3). However, since PS of CUP patients was inaccessible in the database, patients' functional organ reserve could only be partially reflected through age and metastatic status in this case. For histological types, our study also confirmed that SCC and neuroendocrine carcinoma carried a superior prognosis

than adenocarcinoma, which highly agreed with previous evidence (20-22). Although the number of metastatic organs had its limitation owing to the unknown status of other sites, surveillance on these major organs largely represented the organ reserve and metastatic burden of patients.

According to the National Comprehensive Cancer Network (NCCN) clinical guidelines, chemotherapy was recommended for symptomatic patients (PS score of 1–2) or asymptomatic patients (PS score of 0) with aggressive cancer, and the regimen was largely dependent on its histological type (18). Prognostic models could offer a more accurate risk classification after a comprehensive consideration. Based on the risk stratification, low-risk patients were more likely to be treated with a curative aim, while high-risk patients should be provided with options such as low-toxicity chemotherapy and palliative care. Several prognostic models have been developed in previous

Page 10 of 14

Jin et al. Nomogram for cancer of unknown primary

Table 3 Point assignment

Variables	Assigned score
Sex	
Male	7
Female	0
Age at diagnosis, years	
10	0
20	9
30	18
40	27
50	36
60	45
70	54
80	63
90	72
100	81
110	90
Histological types	
Squamous cell carcinoma	0
Neuroendocrine carcinoma	4
Neoplasm and other types[a]	4
Carcinoma, NOS	45
Adenocarcinoma	45
Non-primary surgery	
Yes	0
No/unknown	39
Radiotherapy	
Yes	0
No/unknown	15
Chemotherapy	
Yes	0
No/unknown	19
Number of metastatic organs[b]	
0	0
1	30
2	48
3	55
4	100

[a], including undifferentiated malignant neoplasms and other histological types not classified as carcinomas; [b], limited to bone, brain, liver, and/or lung; refers to metastasis at diagnosis. NOS, not other specified.

investigations. In 1995, van der Gaast *et al.* established a prognostic model with only PS and ALP based on a cohort of 79 patients (25). This model only included single-centered patients with poorly differentiated adenocarcinoma or undifferentiated carcinomas and was not validated in external datasets. Culine *et al.* later developed and validated a model with PS score and serum LDH levels based on a population of 150 patients, which was limited to patients with carcinoma (26). Ponce Lorenzo *et al.* developed a prognostic model based on PS and absence or presence of liver metastasis (19). This study was also conducted in a single center, with a population of 100 patients and lacked independent validation. A more recent prognostic algorithm put forwarded by Petrakis *et al.* incorporated leukocytosis, clinicopathologic CUP subgroup and PS as predictors. They used classification and regression analysis in a population of 311 cases, and tested the validity in a randomly split set of their study cohort (3). This validation was relatively weak according to the current statement (10); and one of its predictors, clinicopathological subgroup, had its limitations in clinical application. All the above prognostic models were established on small populations, and lacked a quantitative method for survival probability computation and model performance evaluation. In addition, since they all served as tools for risk assessment before therapies, the impact of treatment was not considered in these prognostic models. Our nomogram incorporated surgery, radiotherapy, and chemotherapy as predictors, and all of them indicated significant improvements on OS in the multivariate analysis. However, subgroup analysis indicated that the use of radiotherapy may not bring survival benefit to patients with metastasis. In contrast, subgroup analysis suggested the value of non-primary surgery even in the metastatic context. Meanwhile, this real-world data analysis showed that the use of non-primary surgery was only considered in a rather small population. Therefore, this study suggested that non-primary surgery to CUP should also be considered as a potential option for highly selected patients, while other factors such as patients' PS, organ reserve, personal preference, tolerance to different therapies, and cost-effectiveness should also be taken into account.

This study also had several limitations. Due to the lack of data availability, the prognostic role of important clinical factors such as patients' PS, metastatic status in other organs, differentiation status, biochemical and hematological indicators such as LDH, ALP, and ALB levels could not be assessed in this study. It also limited the analysis of treatment efficacy because of the unknown

Table 4 Total prognostic score of 6-month survival probability

Total prognostic score	Estimated 6-month survival probability
77	0.9
123	0.8
151	0.7
173	0.6
192	0.5
209	0.4
225	0.3
243	0.2
265	0.1
281	0.05

Table 5 Total prognostic score of 9-month survival probability

Total prognostic score	Estimated 9-month survival probability
61	0.9
106	0.8
135	0.7
157	0.6
175	0.5
192	0.4
209	0.3
226	0.2
248	0.1
264	0.05
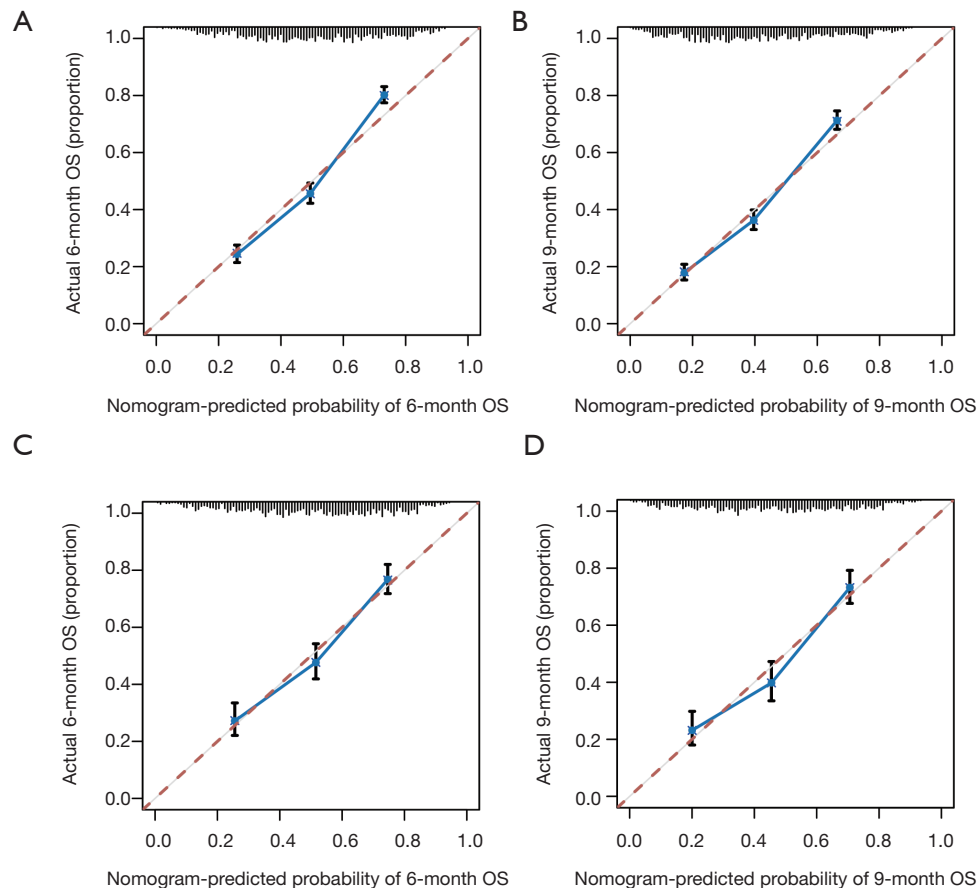
Figure 3 The calibration plots of nomograms. (A,B) The calibration plots of 6-month (A) and 9-month (B) overall survival in the training cohort (patients diagnosed in 2010–2015); (C,D) the calibration plots of 6-month (C) and 9-month (D) overall survival in the validation cohort (patients diagnosed in 2016). The x-axis and y-axis represent the overall survival probability predicted by nomogram and the actual overall survival probability, respectively. A plot along the 45-degree line reveals a perfect agreement between predicted and actual survival. OS, overall survival.

Page 12 of 14

Jin et al. Nomogram for cancer of unknown primary
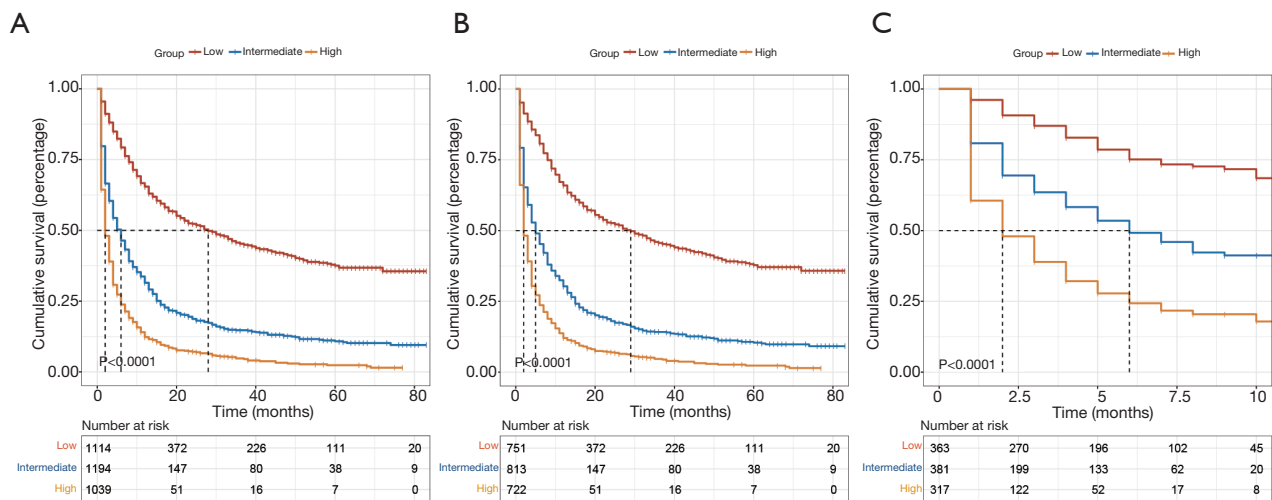
Figure 4 The risk stratification of all cohorts, training set, and validation set. (A) The survival curves of patients in all cohorts. The median OS of the low-, intermediate-, and high-risk groups was 28.0 (95% CI: 22.1–33.9), 6.0 (95% CI: 5.3–6.7), and 2.0 (95% CI: 1.8–2.2) months, respectively. (B) The survival curves of patients in the training set. The median OS of the low-, intermediate-, and high-risk groups was 29.0 (95% CI: 22.5–35.5), 5.0 (95% CI: 4.2–5.8), and 2.0 (95% CI: 1.7–2.3) months, respectively. (C) The survival curves of patients in the validation set. The median OS was not reached for the low-risk group. The median OS of the intermediate- and the high-risk groups was 6.0 (95% CI: 4.5–7.5) and 2.0 (95% CI: 1.5–2.5) months respectively.

sequence of being treated and disease progressing, unknown regimen details, and unknown response to a certain therapy. In addition, our nomogram was established on the U.S. population and has not been validated in any external datasets. Its predicting and discriminating abilities need to be further assessed in current patients from other registries or institutions. All of the above limitations need to be considered before widespread application, and more investigations are yet to be done for the improvement of the current prognostic nomogram for CUP.

Conclusions

We developed and validated the first nomogram for predicting individual survival of patients with CUP based on a real-world study with a large population. Covariates included in the nomogram were sex, age, marital status, histological type, surgery, radiotherapy, chemotherapy, and the number of metastatic organs. It demonstrated good performance on prediction and discrimination, with a C-index of 0.705 (95% CI: 0.692–0.717) for the training cohort and 0.727 (95% CI: 0.703–0.752) for the validation cohort. Further validations of this nomogram in independent cohorts are needed before its widespread application.

Footnote

of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. This study was performed in accordance with the Declaration of Helsinki (as was revised in 2013). Access was permitted by SEER to the database, which is available for public use. Therefore, consent and ethics approval were not required.

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: https://creativecommons.org/licenses/by-nc-nd/4.0/.

References

1. Varadhachary GR, Raber MN. Cancer of unknown primary site. N Engl J Med 2014;371:757-65.
2. Pavlidis N, Briasoulis E, Hainsworth J, et al. Diagnostic and therapeutic management of cancer of an unknown primary. Eur J Cancer 2003;39:1990-2005.
3. Petrakis D, Pentheroudakis G, Voulgaris E, et al. Prognostication in cancer of unknown primary (CUP): development of a prognostic algorithm in 311 cases and review of the literature. Cancer Treat Rev 2013;39:701-8.
4. Pavlidis N, Pentheroudakis G. Cancer of unknown primary site. Lancet 2012;379:1428-35.
5. Hemminki K, Bevier M, Hemminki A, et al. Survival in cancer of unknown primary site: population-based analysis by site and histology. Ann Oncol 2012;23:1854-63.
6. Jones W, Allardice G, Scott I, et al. Cancers of unknown primary diagnosed during hospitalization: a population-based study. BMC Cancer 2017;17:85.
7. Greco FA. Cancer of unknown primary site: still an entity, a biological mystery and a metastatic model. Nat Rev Cancer 2014;14:3-4.
8. Economopoulou P, Pentheroudakis G. Cancer of unknown primary: time to put the pieces of the puzzle together? Lancet Oncol 2016;17:1339-40.
9. Surveillance, Epidemiology, and End Results (SEER) Program (www.seer.cancer.gov) SEER*Stat Database: Incidence - SEER 18 Regs Custom Data (with additional treatment fields), Nov 2018 Sub (1975-2016 varying) - Linked To County Attributes - Total U.S., 1969-2017 Counties, National Cancer Institute, DCCPS, Surveillance Research Program, released April 2019, based on the November 2018 submission. Available online: https://seer.cancer.gov/data-software/documentation/seerstat/nov2018/
10. Collins GS, Reitsma JB, Altman DG, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. Ann Intern Med 2015;162:55-63. Erratum in: Ann Intern Med. 2015 Apr 21;162(8):600. doi: 10.7326/L15-0078-4.
11. Fizazi K, Greco FA, Pavlidis N, et al. Cancers of unknown primary site: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol 2015;26 Suppl 5:v133-8.
12. Iasonos A, Schrag D, Raj GV, et al. How to build and interpret a nomogram for cancer prognosis. J Clin Oncol 2008;26:1364-70.
13. Lin DY, Wei LJ, Ying Z. Checking the Cox Model with Cumulative Sums of Martingale-Based Residuals. Biometrika 1993;80:557-72.
14. Heinze G, Wallisch C, Dunkler D. Variable selection - A review and recommendations for the practicing statistician. Biom J 2018;60:431-49.
15. Mantel N. Why Stepdown Procedures in Variable Selection. Technometrics 1970;12:621-5.
16. Harrell FE Jr, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med 1996;15:361-87.
17. Losa F, Soler G, Casado A, et al. SEOM clinical guideline on unknown primary cancer (2017). Clin Transl Oncol 2018;20:89-96.
18. NCCN. Clinical Practice Guidelines in Oncology. Occult Primary (Cancer of Unknown Primary [CUP]), Version 1.2020. 2019. Available online: https://www.nccn.org/professionals/physician_gls/pdf/occult.pdf
19. Ponce Lorenzo J, Segura Huerta A, Díaz Beveridge R, et al. Carcinoma of unknown primary site: development in a single institution of a prognostic model based on clinical and serum variables. Clin Transl Oncol 2007;9:452-8.
20. Abbruzzese JL, Abbruzzese MC, Hess KR, et al. Unknown primary carcinoma: natural history and prognostic factors in 657 consecutive patients. J Clin Oncol 1994;12:1272-80.
21. Fernandez-Cotarelo MJ, Guerra-Vales JM, Colina F, et

Page 14 of 14

Jin et al. Nomogram for cancer of unknown primary

al. Prognostic factors in cancer of unknown primary site. Tumori 2010;96:111-6.

22. Hess KR, Abbruzzese MC, Lenzi R, et al. Classification and regression tree analysis of 1000 consecutive patients with unknown primary carcinoma. Clin Cancer Res 1999;5:3403-10.

23. Lu HJ, Chen KW, Tzeng CH, et al. Evaluation of prognosis for carcinoma of unknown origin in elderly patients. Oncology 2012;83:24-30.

24. Grajales-Álvarez R, Martin-Aguilar A, Silva JA, et al. ECOG is as independent predictor of the response to

chemotherapy, overall survival and progression-free survival in carcinoma of unknown primary site. Mol Clin Oncol 2017;6:643-50.

25. van der Gaast A, Verweij J, Planting AS, et al. Simple prognostic model to predict survival in patients with undifferentiated carcinoma of unknown primary site. J Clin Oncol 1995;13:1720-5.

26. Culine S, Kramar A, Saghatchian M, et al. Development and validation of a prognostic model to predict the length of survival in patients with carcinomas of an unknown primary site. J Clin Oncol 2002;20:4679-83.