Development and Validation of a Novel Nomogram for Individualized Prediction of Survival in Cancer of Unknown Primary AC

Kanwal Raghav[1], Hyunsoo Hwang[2], Alexandre A. Jácome[1], Eric Bhang[3], Anneleis Willett[1], Ryan W. Huey[1], Nishat P. Dhillon[1], Jignesh Modha[1], Brandon Smaglo[1], Aurelio Matamoros Jr[4], Jeannelyn S. Estrella[5], Justin Jao[3], Michael J. Overman[1], Xuemei Wang[2], F. Anthony Greco[6], Jonathan M. Loree[3], and Gauri R. Varadhachary[1]

ABSTRACT

Purpose: Prognostic uncertainty is a major challenge for cancer of unknown primary (CUP). Current models limit a meaningful patient-provider dialogue. We aimed to establish a nomogram for predicting overall survival (OS) in CUP based on robust clinicopathologic prognostic factors.

Experimental Design: We evaluated 521 patients with CUP at MD Anderson Cancer Center (MDACC; Houston, TX; 2012–2016). Baseline variables were analyzed using Cox regression and nomogram developed using significant predictors. Predictive accuracy and discriminatory performance were assessed by calibration curves, concordance probability estimate (CPE ± SE), and concordance statistic (C-index). The model was subjected to bootstrapping and multi-institutional external validations using two independent CUP cohorts: V1 [MDACC (2017), $N = 103$] and V2 (BC Cancer, Vancouver, Canada and Sarah Cannon Cancer Center/Tennessee Oncology, Nashville, TN; $N = 302$).

Results: Baseline characteristics of entire cohort ($N = 926$) included: median age (63 years), women (51%), Eastern Cooperative Oncology Group performance status (ECOG PS) 0–1 (64%), adenocarcinomas (52%), ≥3 sites of metastases (30%), and median follow-up duration and OS of 40.1 and 14.7 months, respectively. Five independent prognostic factors were identified: gender, ECOG PS, histology, number of metastatic sites, and neutrophil-lymphocyte ratio. The resulting model predicted OS with CPE of 0.69 [SE: ± 0.01; C-index: 0.71 (95% confidence interval: 0.68–0.74)] outperforming Culine/Seve prognostic models (CPE: 0.59 ± 0.01). CPE for external validation cohorts V1 and V2 were 0.67 (± 0.02) and 0.70 (± 0.01), respectively. Calibration curves for 1-year OS showed strong agreement between nomogram prediction and actual observations in all cohorts.

Conclusions: Our user-friendly CUP nomogram integrating commonly available baseline factors provides robust personalized prognostication which can aid clinical decision making and selection/stratification for clinical trials.

Introduction

Cancer of unknown primary (CUP) is a formidable diagnosis, one that stirs a sense of apprehension in doctors and patients alike (1–4). Although often seen as a "group of cancers" that at presentation share the common trait of being metastatic without an identifiable primary, most agree it is a very heterogeneous disease (5). This patient and tumor heterogeneity often results in wide-ranging survival outcomes amplifying the ambiguity surrounding CUP and impeding personalized care of patients.

Understanding and predicting prognosis is vital to making informed decisions for management of advanced cancers (6, 7). Candid and realistic conversations about prognosis are desired by patients and endorsed by key guidelines (8, 9). Early prognostic discussions result in better patient education about goals of care and life expectancy (10). Because baseline perceptions of prognosis can shape treatment decisions, prognostic inaccuracies can lead to overtreatment, patient and caregiver distress, poor quality of life, and adverse medical and social outcomes (7, 11–14).

Accurate and individualized prediction of survival in CUP has been a key challenge in clinical care and trials, leading to "best guess" discussions and suboptimal study designs (4, 15). Consensus reference staging systems to inform prognosis, large-scale outcome studies and prospective trials to draw approximations, which facilitate prognostication in other cancers, are all but lacking in CUP. Moreover, the relative infrequency with which oncologists encounter patients with CUP in their practice, limits experience and intuition which physicians rely on for discussing prognosis.

Despite an abundance of studies evaluating individual prognostic factors, classification models in CUP are limited (16). Culine and colleagues developed a prognostic model in 2002 with patient performance status (PS) and serum lactate dehydrogenase (LDH; or if unavailable, presence of liver metastases; ref. 17). This model assigned a good-risk and a poor-risk group with median survivals of 11.7 and 3.9 months, respectively (17). Although helpful with generalized projections, this and other prior categorical models lack individualization and are limited in enabling a meaningful dialogue between

[1]Department of Gastrointestinal Medical Oncology, The University of Texas MD Anderson Cancer Center, Houston, Texas. [2]Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, Texas. [3]BC Cancer, Vancouver, British Columbia, Canada. [4]Diagnostic Imaging, The University of Texas MD Anderson Cancer Center, Houston, Texas. [5]Pathology, The University of Texas MD Anderson Cancer Center, Houston, Texas. [6]Sarah Cannon Cancer Center/Tennessee Oncology, Nashville, Tennessee.

Note: Supplementary data for this article are available at Clinical Cancer Research Online (http://clincancerres.aacrjournals.org/).

Corresponding Author: Gauri R. Varadhachary, Gastrointestinal Medical Oncology, The University of Texas MD Anderson Cancer Center, 1515 Holcombe Blvd., Houston, TX 77030. Phone: 713-792-2828; Fax: 713-563-1701; E-mail: gvaradha@mdanderson.org

Clin Cancer Res 2021;27:3414–21

doi: 10.1158/1078-0432.CCR-20-4117

Translational Relevance

Cancer of unknown primary (CUP) is a rare and challenging clinical diagnosis, fraught with uncertainties for both patients and oncologists. Although realistic conversations about prognosis can aid informed decision making regarding patient care, models that can actually provide reliable and individualized estimates of survival for patients with CUP are lacking. Traditional risk stratification models pool patients at either good or poor risk which limit a meaningful patient-provider dialogue. Using this large multicenter cohort of 926 patients with CUP, we developed and validated a robust prognostic model and nomogram to predict overall survival with superior performance (concordance probability estimate of 0.69 and concordance index of 0.71) compared with traditional prognostic classifiers. This model uses universally available baseline factors to ensure feasibility of use in diverse clinical settings and is available as a web-based application (https://cupnomogram.shinyapps.io/Nomogram/). Personalized prognostication with this CUP nomogram can not only aid informed clinical decisions but also trial selection/stratification.

patients and providers about survival estimates (17–19). This knowledge gap of personalized prognostication puts patients with CUP and oncologists that treat CUP at a huge disadvantage (2, 3). The resulting indecision instills significant anxiety in patients and trepidation in providers toward approaching discussions of prognosis and goals of care in CUP (1–4).

In this study, we aimed to establish and validate a novel prognostic model using a nomogram-based approach for predicting overall survival (OS) centered on robust and readily available baseline clinicopathologic prognostic factors in CUP. We chose this nomogram-based approach due to its ability to condense a complex statistical model embracing diverse prognostic factors into a simple graphical representation (20). A nomogram can generate a straightforward numerical probability of survival and is widely popular among oncologists due to its ability to generate individualized predictions and a user-friendly interface (20). We envision that this CUP survival nomogram will enable accurate and effective communication between patients and oncologists empowering them in making personalized decisions with the ultimate goal of improving understanding and clinical outcomes in CUP.

Materials and Methods

Patient population

Development and internal validation cohort (cohort A)

We identified a cohort of 521 consecutive patients with a diagnosis of CUP at The University of Texas MD Anderson Cancer Center (MDACC; Houston, TX) over 5 years between January 2012 and December 2016 using a retrospective-prospective CUP database (Supplementary Fig. S1). This cohort (A) served as our discovery cohort and was used for the development of the model (nomogram) with internal validation. Eligible cases were defined as those with biopsy-proven metastatic cancer without a detectable primary after an appropriate diagnostic work-up as per standard guidelines (5). To minimize diagnostic variability which can occur with CUP, only cases reviewed and confirmed by CUP pathologists and oncologists at MDACC were included. Cases were excluded if they lacked complete history and physical; CT scan of chest, abdomen, and pelvis (alternate equivalent imaging allowed if intravenous contrast contraindicated); and symptom/pathology directed endoscopy or additional imaging. Historically described "favorable subsets" such as adenocarcinoma in axillary lymph nodes in women (breast cancer), squamous cell carcinoma in neck nodes (head and neck cancer), and papillary or serous tumors in the peritoneal cavity in women (ovarian cancer) were excluded from the MDACC CUP database and this development cohort because they are treated as their putative primaries and their natural history differs from the "unfavorable subset" of CUP.

Demographic and clinicopathologic variables at baseline [age, gender, Eastern Cooperative Oncology Group performance status (ECOG PS)], tumor histology, sites of metastases (uniquely defined sites: liver, lung, peritoneum/retroperitoneum, bone, brain, lymph nodes, ovarian, adrenal, skin/subcutaneous, muscle), number of metastatic sites (NMS), laboratory values [specifically, LDH and neutrophil-lymphocyte ratio (NLR)], and survival data were retrieved from the database and electronic medical records. All patients received therapies in agreement with standard CUP guidelines as per recommendations of their treating physicians (Supplementary Table S1). The study was performed under a MDACC Institutional Review Board (IRB)-approved protocol, waiving written informed consent by patients and in accordance with the Declaration of Helsinki.

External validation cohorts (cohorts V1 and V2)

Three independent patient populations with CUP (similar eligibility criteria as cohort A) served as external validation cohorts (Supplementary Fig. S1). Cohort V1 of 103 consecutive patients with MDACC with CUP between January 2017 and December 2017 was identified using the CUP database as above. Cohort V2 included a deidentified pooled dataset of 302 consecutive patients at two centers: British Columbia Cancer (BC Cancer; Vancouver, British Columbia, Canada), between January 2014 and September 2016 ($N = 202$) and Sarah Cannon Cancer Center/Tennessee Oncology (SCCC/TO; Nashville, TN) between January 2012 and December 2015 ($N = 100$). All patients in these cohorts met the eligibility criteria stated above for the development cohorts and data regarding all variables of interest for the nomogram were collected for these patients. The study was performed under BC Cancer and SCCC/TO IRB-approved protocols, with waiver of written informed consent by patients and in accordance with the Declaration of Helsinki.

Statistical methodology

Descriptive statistics were used to summarize patient characteristics. Randomization, blinding, and power analysis was not relevant to this study. Fisher exact/χ^2 test were used for comparisons between groups. Cox proportional models were fit to assess association between variables and OS and results were expressed in HRs and 95% confidence intervals (95% CI). All tests were two sided and P values of < 0.05 were considered statistically significant. Statistical analyses were carried out with R software (version 3.6.1) and GraphPad Prism version 8.00 [GraphPad (RRID:SCR_002798) software], used for generating Kaplan–Meier curves.

Formulation of the CUP nomogram

The primary outcome was OS, defined as the time between date of diagnosis and death. Patients alive at last follow-up were censored. Baseline prognostic parameters and cutoffs for analyses were selected *a priori* based on prior research and evidence. Covariates included were age, gender, ECOG PS (0 vs. 1 vs. ≥ 2), NLR, presence of liver metastases (no vs. yes), NMS (< 3 vs. ≥ 3), and tumor histology

(non-adenocarcinoma vs. adenocarcinoma). NLR was modeled as a continuous variable using 3-knot restricted cubic spline with number of knots based on Akaike information criterion (21–23). Because of high variability in reference range and high proportion of missing values, LDH (27% missing) was not included to ensure consistency. A nomogram was constructed using significant predictors based on multivariable Cox regression analysis (backward stepwise variable selection procedure) by R software (version 3.6.1) with the survival and rms packages.

Calibration and validation of the nomogram

Predictive accuracy and discrimination performance were assessed by calibration curve (graphic representations of agreement between observed outcomes and predicted probabilities), concordance probability estimate (CPE; values range from 0.5 to 1.0, with 0.5 indicating random chance and 1.0 indicating a perfect ability to correctly discriminate the outcome with the model and are reported with their SE), and concordance index (Harrell C-index; ref. 24). The model was subjected to external validation using cohorts V1 and V2 that were not used to develop the model. Bootstrapping method (1,000 repetitions), which is based on random sampling with replacement, was used to calculate the CIs of C-index. We also compared the performance between our prognostic model and the Culine and Seve prognostic models (17).

Results

Baseline characteristics

Baseline characteristics of all cohorts ($N = 926$) are shown in **Table 1**. Median age of the entire study population was 63 years (range, 18–92). Fifty-one percent were women, 64% had ECOG PS of 0 or 1, and 52% had histology consistent with adenocarcinoma. Nearly one third of patients had three or more sites of metastatic involvement (30%) and a high NLR (≥ 5; 35%). Overall cohorts A and V1 were similar, while cohort V2 was different in terms of baseline characteristics. Cohort V2 population appeared to have a higher rate of patients with poor PS, high NLR, and liver metastases. In these 926 patients, a total of 583 (63.0%) events (deaths) occurred over a median follow-up duration of 40.1 months. Median OS of entire cohort was 14.7 months (95% CI: 13.0–16.5; Supplementary Fig. S2).

Development of nomogram

Univariate and multivariable analyses were performed in cohort A and are summarized in **Table 2**. Five prognostic factors were identified to be independently associated with OS: gender, ECOG PS, histology, number of metastatic sites, and NLR. Being male, having a poor ECOG PS, an adenocarcinoma, a high number of metastatic sites, and a higher NLR were associated with worse survival in the Cox model. ECOG PS

Table 1. Baseline characteristics of development and internal validation cohort (A) and external validation cohorts (V1/V2).

	Cohort A ($N = 521$)		Cohort V1 ($N = 103$)		P^b	Cohort V2 ($N = 302$)		P^b	Overall ($N = 926$)	
Charateristic[a]	***N***	**%**	***N***	**%**	**V1 vs. A**	***N***	**%**	**V2 vs. A**	***N***	**%**
Age (years)										
Median (range)	60	18–90	65	31–88	0.022	67	22–92	<0.0001	63	18–92
<60	253	48.6	40	38.8	0.084	79	26.2	<0.0001	372	40.2
≥60	268	51.4	63	61.2		223	73.8		554	59.8
Gender										
Female	284	54.5	48	46.6	0.162	141	46.7	0.0240	473	51.1
Male	237	45.5	55	53.4		161	53.3		453	48.9
ECOG performance status										
0	127	27.3	30	31.3	0.566	50	16.7	<0.0001	207	24.0
1	202	43.4	44	45.8		99	33.0		345	40.1
2	90	19.4	13	13.5		67	22.3		170	19.8
≥3	46	9.9	9	9.4		84	28.0		139	16.1
Histology										
Adenocarcinoma	291	55.9	63	61.2	0.738	125	41.4	<0.0001	479	51.7
Non-adenocarcinoma	230	44.1	40	38.8		177	58.6		447	48.3
Carcinoma	156	29.9	24	23.3		46	17.7		226	25.6
Malignant neoplasm	45	8.6	9	8.7		16	6.2		70	7.9
Squamous cell carcinoma	20	3.8	5	4.9		48	18.5		73	8.3
Other	9	1.7	2	1.9		25	9.6		36	4.1
Number of metastatic sites										
<3	388	74.5	81	78.6	0.454	177	58.6	<0.0001	646	69.8
≥3	133	25.5	22	21.4		125	41.4		280	30.2
Liver metastasis										
Absent	363	69.7	77	74.8	0.345	158	52.3	<0.0001	598	64.6
Present	158	30.3	26	25.2		144	47.7		328	35.4
Neutrophil-lymphocyte ratio										
Median (range)	3.4	0.2–48	3.6	0.1–23	0.330	4.1	0.3–48	0.0010	3.6	0.1–48
Low (<5)	319	69.3	62	68.9	0.999	174	58.6	0.0029	555	65.5
High (≥5)	141	30.7	28	31.1		123	41.4		292	34.5

Abbreviations: ECOG, Eastern Cooperative Oncology Group; *N*, number of patients.
[a]Some variables have missing values. Proportions are derived from available data.
[b]Fisher exact test/χ^2 test as appropriate.

Table 2. Univariate and multivariable analysis of factors predicting OS in the development cohort (A) patients ($N = 521$).

	Univariate analysis			Multivariable analysis[a]		
Variable	**HR**	**95% CI**	***P***	**HR**	**95% CI**	***P***
Age	1.01	1.00–1.02	0.070	—	—	—
Gender						
Female vs. male	0.68	0.55–0.85	<0.001	0.74	0.57–0.95	0.020
ECOG performance status						
1 vs. 0	1.84	1.35–2.50	<0.001	1.65	1.19–2.28	0.002
2 vs. 0	2.57	1.80–3.67	<0.001	2.33	1.60–3.40	<0.001
>2 vs. 0	4.12	2.68–6.33	<0.001	3.43	2.15–5.47	<0.001
Histology						
Adeno vs. non-adeno	1.37	1.10–1.71	0.010	1.40	1.08–1.80	0.010
Liver metastases						
Present vs. absent	1.75	1.39–2.21	<0.001	—	—	—
Number of sites of metastases						
≥3 vs. <3	1.79	1.41–2.28	<0.001	1.47	1.12–1.94	0.010
Neutrophil-lymphocyte ratio[b]						
NLR (linear term)	1.38	1.22–1.55	<0.001	1.24	1.09–1.40	<0.001
NLR (cubic spline)	0.66	0.55–0.79	<0.001	0.83	0.68–1.01	0.060

Abbreviations: Adeno, adenocarcinoma; CI, confidence interval; ECOG, Eastern Cooperative Oncology Group; HR, hazard ratio; NLR, neutrophil-lymphocyte ratio; OS, overall survival.

[a]Multivariable analysis was performed using backward stepwise variable selection procedure.

[b]NLR was modeled as a continuous variable using restricted cubic spline. Relationship between OS and NLR is also depicted in the nomogram (**Fig. 1**). The probability of OS declines as value of NLR increases, and the decline is sharper as NLR value increases from 0 to 5.

and NLR were the strongest predictors of OS ($P < 0.001$). We constructed a nomogram (a graphic depiction of the model) based on these significant prognostic variables (**Fig. 1**). On the nomogram, each variable is assigned a score on a point scale based on the rank order of the effect estimates. By adding them and then assessing the total score of all variables on "total points" scale, one can draw a straight line down to derive the estimated probability of survival at either 1 or 2 years (**Fig. 1**).

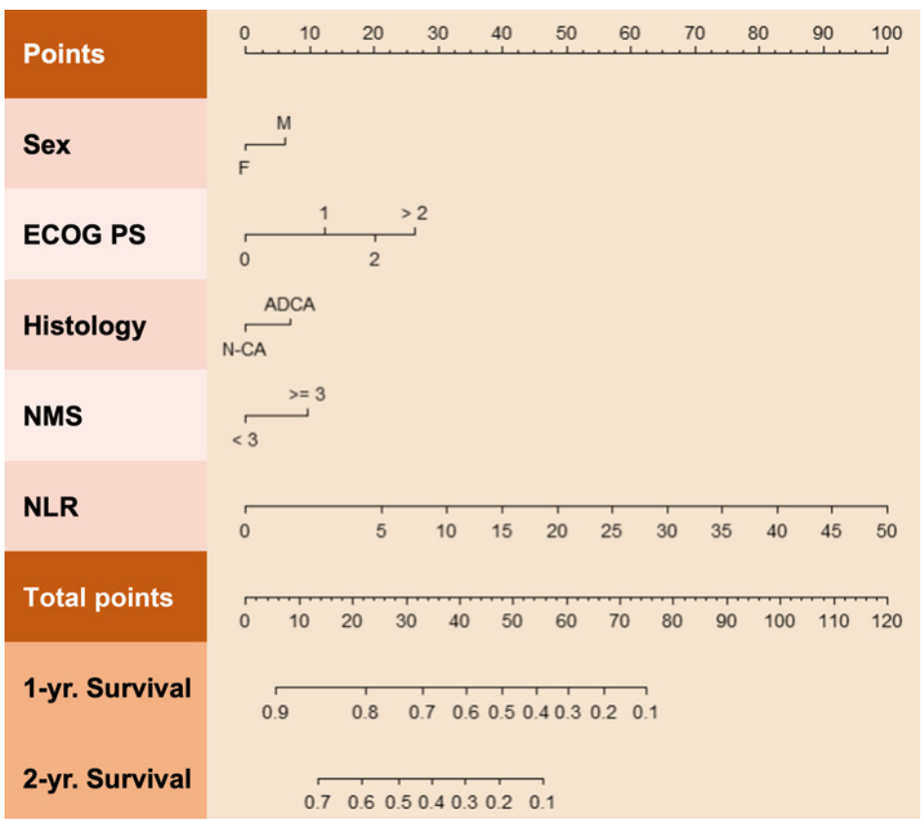

Figure 1.
Nomogram to predict the probability of 1-year or 2-year OS in patients with CUP. Predictor points are obtained from the points scale (top) according to the prognostic contribution of each variable subset. These, added together, give a total point score which can be translated into probability of survival at a specific timepoint (1 year or 2 years) by charting the score on total points scale (bottom) and projecting onto probability of survival scale. The survival estimates are given as probability of survival, for example, the 0.8 on the 1-year survival scale implies an 80% chance of survival at 1 year. ADCA, adenocarcinoma; ECOG PS, Eastern Cooperative Oncology Group performance status; F, Female; M, Male; N-CA, Non-adenocarcinoma; NLR, neutrophil-lymphocyte ratio; NMS, number of metastatic sites.

Internal and external validation and performance of nomogram

The nomogram (our prognostic model) showed good performance characteristics in both the development cohort A and the validation cohorts. Discrimination assessed with CPE to predict OS was 0.69 (SE: ± 0.01) in the internal cohort A. The CPE for the validation cohorts V1 and V2 were 0.67 (SE: ± 0.02) and 0.70 (SE: ± 0.01), respectively. The calibration curves for both 1-year and 2-year OS showed strong agreement between nomogram prediction and actual observation in both the development and validation cohorts (**Fig. 2A** and **B**). In cohort A, the CPE for nomogram predictions was greater (0.69, SE: ± 0.01) than the CPE for predictions based on Culine prognostic model (0.59, SE: ± 0.01) and Seve prognostic model (0.59, SE: ± 0.01). The corresponding Harrell C-index of the model for cohort A, V1, and V2 were 0.71 (95% CI: 0.68–0.74), 0.81 (95% CI: 0.74–0.87), and 0.76 (95% CI: 0.74–0.79), respectively. The C-index for nomogram predictions was greater than C-index for predictions based on Culine prognostic model (0.61; 95% CI: 0.58–0.64) and Seve prognostic model (0.61; 95% CI: 0.58–0.64; Supplementary Fig. S3).

Discussion

"How long do I have, doctor?" is an important question for patients with CUP that clinicians often struggle to address. Patients, caregivers, and providers all value meaningful prognostic information and its importance in making a well-informed decision regarding treatment cannot be stressed enough (6, 7, 25). However, reliable prognostication in CUP has been an inexact science and an unmet need (1, 3, 4, 26). Herein, using a large recent cohort of patients with CUP ($N = 926$) from three institutions, we have developed and validated a simple tool for reliable prediction of survival at 1 year and 2 years in CUP. This nomogram uses readily available and objective baseline clinicopathologic factors: gender, ECOG PS (0 or 1 or 2 or > 2), histology (non-adenocarcinoma or adenocarcinoma), number of sites of metastases (< 3 or ≥ 3), and NLR and provides patient-specific estimates of OS at diagnosis in CUP. On the basis of tertiles of nomogram total points, patients separated into low, intermediate, and high nomogram risk groups showed a median survival of 40.0, 15.1, and 4.1 months, respectively ($P < 0.0001$; Supplementary Fig. S4).

Several groups have reported on CUP prognostic algorithms in the past using a few baseline variables. Culine and colleagues published a CUP prognostic model almost two decades ago to select patients for clinical trials (17). It separated patient populations into good risk (1-year OS: 45%) and poor risk (1-year OS: 11%), using a classification scheme built on PS and LDH (or liver metastases) from a dataset of 150 patients with CUP and a validation set of 130 patients. Similarly, Seve and colleagues published a model based on liver metastases and serum albumin, again separating patients into good risk (1-year OS: 39%) and poor risk (1-year OS: 12%; ref. 18). Another CUP prognostic algorithm developed by Petrakis and colleagues classified patients in low-risk, intermediate-risk, and high-risk groups with median survival of 36, 11–14, and 5–8 months, respectively (19). Our nomogram, using a large sample size and diverse characteristics, offers major technical and functional improvements over these prior prognostic CUP models (17, 18). First, integrating multiple clinical and pathologic factors, as opposed to one or two factors in prior models, increases the accuracy and robustness of the nomogram and accommodates relative contribution of multiple prognostic factors (16). Second, the data elements required at baseline are objective and universally available during routine work-up of patients with CUP. We purposely excluded variables such as LDH or serum albumin, used in prior models, because they are often not available, their cutoffs are prone to interlaboratory variations, and they can be nonspecifically influenced by cancer-unrelated factors such as liver functions, age, and hydration. In fact, LDH and albumin were missing in 38%–46% and 27% cases in the reference populations used for prior studies (17, 18). Likewise, in our development cohort, LDH was missing in 27% cases and albumin in 15% cases. Finally, the categorical nature of prior systems forces the transformation of continuous variables into uncompromising bracketed outcomes, good or bad, limiting predictive accuracy and extent which is overcome by a continuous nature of prediction by the nomogram.

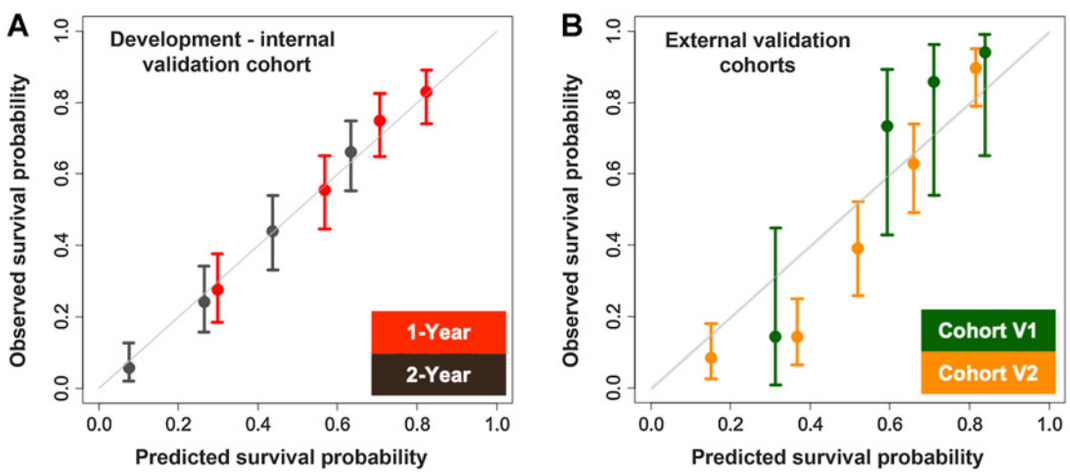

Figure 2.

Calibration plot of overall survival at 1 and 2 years for development and internal validation cohort (**A**) and external validation cohort (**B**). Observed and predicted (as per our prognostic nomogram) survival is plotted on *y*-axis and *x*-axis, respectively. Means of survival predicted by our model was compared with the means of observed survival assessed by Kaplan–Meier estimates after grouping of equal sample sizes. The 45-degree line through the origin point represents perfect calibration model with identical actual and predicted survival outcome probabilities. Vertical arrows represent 95% CIs for observed survival.

Use of pretreatment NLR is a unique and important attribute of this model. NLR acts as a potential surrogate biomarker of tumor microenvironment, immune milieu, and systemic inflammatory state of malignancy and has been recognized as a prognostic factor for several tumors (27–29). NLR has also been shown to have impact on prognosis in patients with CUP (30, 31). It can be easily derived from a complete blood count with differential done as a part of routine initial CUP management. Furthermore, it is a ratio and therefore not affected by testing site or methodology. This makes it an ideal and objective prognostic parameter that does not add expense or resource utilization. In fact, all elements on the nomogram can be easily gathered at diagnosis without any added effort.

The fact that the source population for creating the nomogram was derived from a single referral institution may be seen as a limitation. Notably, the median OS of our cohort is higher when compared with published population-based reports of CUP. However, we believe that this selection ensured development based on a large and vigorous dataset which has consistent availability of detailed data elements (20). This is necessary in light of the complexity of CUP diagnosis in general. In the development cohort, CUP diagnosis was rigorously evaluated by an experienced team and the primary outcome and prognostic variables to be included in the nomogram were defined *a priori* adding support to its performance. Importantly, we validated the model in multicenter cohorts, adding to the generalizability of the nomogram.

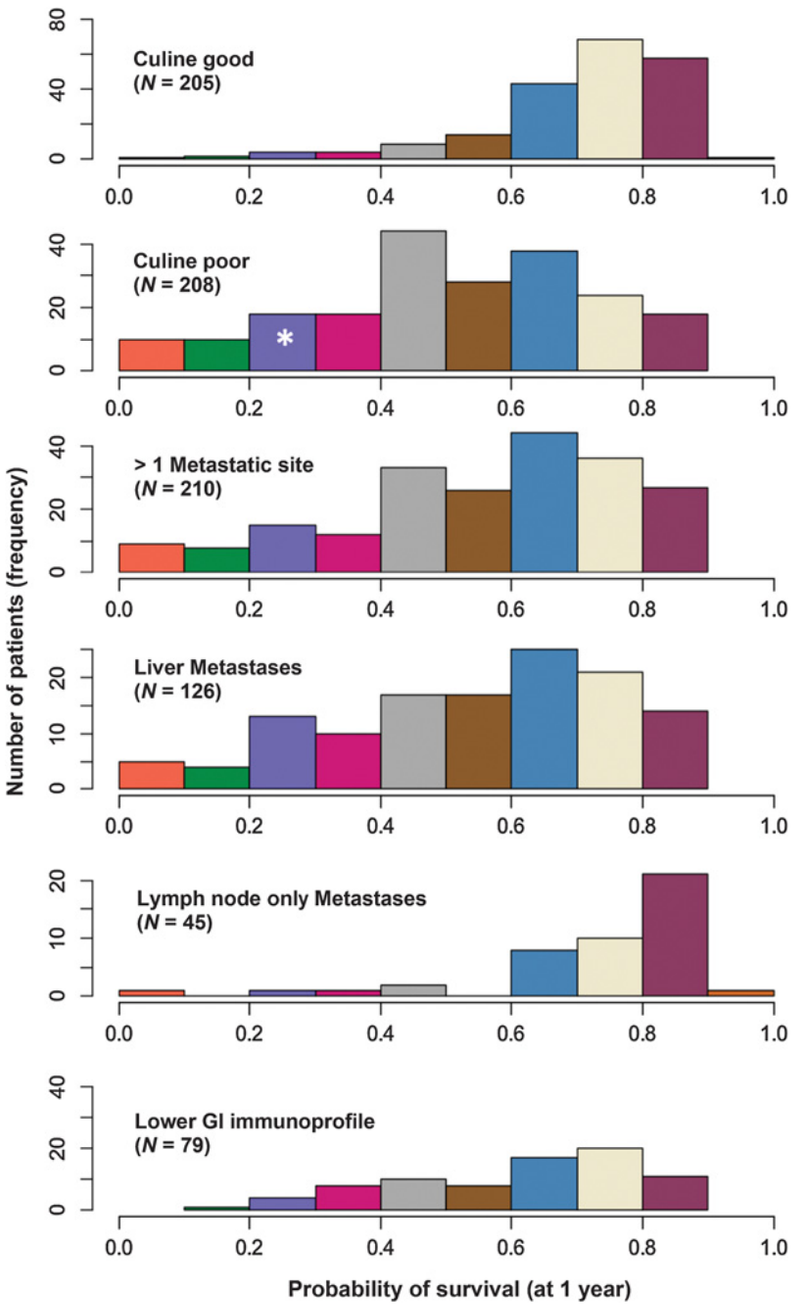

Figure 3.

Histogram of nomogram-predicted OS in various clinicopathologic settings: Culine prognostic groups, > 1 site of metastases, presence of liver metastases, lymph node only disease, and IHC consistent with a lower GI profile ($CK20^+$, $CDX2^+$). The heterogeneity within each clinicopathologic setting is apparent by the range of predicted probabilities of OS as calculated using the nomogram. The *y*-axis stands for number of patients (frequency) and *x*-axis stands for OS probability at 1 year, which ranges from 0 to 1 and was divided by 0.1 (10%). For example, the third bin (*) in Culine poor group represents that there were 18 patients with OS probability from 20% to 30%.

Notably, even with differing baseline patient characteristics and outcomes between MDACC and external (BC Cancer and SCCC/TO) cohorts, which may reflect referral bias and varying treatment patterns, the nomogram performed exceptionally well in these validation cohorts. In a population-based cohort like the BC Cancer cohort, which comprised of consecutively diagnosed patients from an entire province in Canada, we had optimal prognostic discrimination (CPE: 0.70 $\pm$ 0.02). However, further efforts are required to investigate the performance of the nomogram in a more diverse population (e.g., outside of North American populations).

The merits discussed above not only make the nomogram distinctive but also enhance its performance. Our analysis using the asymptotically unbiased CPE showed good discriminatory power of the model in all cohorts (CPE of 0.69, 0.67, and 0.70 in cohorts A, V1, and V2, respectively) and performance superior to predictions based on the currently validated prognostic Culine (CPE: 0.59) and Seve (CPE: 0.59) prognostic models in the discovery cohort (24). The commonly used Harrel C-indices of the nomogram were correspondingly 0.71, 0.81, and 0.76 in cohorts A, V1, and V2, respectively. This performance estimate was also superior to predictions based on Culine (0.61) and Seve (0.61) prognostic models in the discovery cohort (Supplementary Fig. S3). We also investigated the ability of the nomogram to dissect the heterogeneity of outcome within the Culine prognostic groups and other clinical and pathologic subsets in CUP (**Fig. 3**). **Figure 3** shows distribution of nomogram-predicted probabilities within each of these groups [Culine good- and poor-risk group, patients with > 1 site of metastases, presence of liver metastases, lymph node only presentation and immunophenotype ($CK20^+$, $CDX2^+$) consistent with a lower gastrointestinal (GI) profile] and clearly the nomogram demonstrates the variations in predicted outcomes within these categories of CUP. For example, about 8% of all patients that would be classified as poor risk in the Culine model and deemed to have a 1-year OS of 45%, can have a 1-year OS-predicted probability of 80%–90%, distinguishing patients that may do significantly better than others plausibly due to favorable response to therapy despite poor PS, high LDH, or liver metastases.

A large body of evidence has established that frank discussions regarding prognosis in advanced cancer results in realistic patient expectations and enhances the emotional well-being of patients and the doctor-patient relationship (32–34). However, studies have shown that clinical prediction of survival by health care professionals is inaccurate with at best "moderate" agreement between predicted and actual patient survival. In most cases, no group accurately predicted the length of patient survival more than 50% of the time (26, 35). Prognostic uncertainty in CUP is greater compared with other cancers and causes apprehension in patients and providers (2–4). This nomogram can lessen this uncertainty and improve patients and provider understanding regarding CUP prognosis. However, it should be recognized that prognosis in any advanced cancer is a dynamic phenomenon. In this era, increasing use of molecular diagnostics and targeted/immune therapies can alter the course of any disease and make prognostic accuracy a moving target and a challenge (36–39). Any prognostic model, including this nomogram, has to evolve over time to account for these changes. Early evidence suggests that certain genomic alterations (such as *KRAS* mutations and *CDKN2A* deletions) may be prognostic and others are druggable using targeted agents that are highly effective (such as *NTRK* fusions). However, the lack of universal availability of molecular profiling for CUP and validation of these biomarkers, limits current inclusion in a clinic ready nomogram (37, 40). In addition, due to low prevalence of these biomarkers and limited access to testing and therapy in CUP, we believe that the nomogram will play a critical role in management of a substantial subset of patients with CUP. Ongoing efforts will be needed to study continued performance and refinement as understanding of molecular biology improves and as more therapies become available for patients with CUP. Integration into prospective trials for risk stratification will be key to understanding this impact.

In summary, this novel CUP survival nomogram is a user-friendly tool comprised of readily available baseline objective data elements that allows robust estimates for survival in patients with CUP, overcoming the epistemic uncertainty of prognostication in this disease. The nomogram is publicly accessible for use in a user-friendly web-based application at (https://cupnomogram.shinyapps.io/Nomogram/; Supplementary Fig. S5). Besides facilitating a meaningful dialogue for optimizing routine clinical management of patients with CUP, the ability to generate individualized predictions enables its use in the identification and stratification of patients with CUP for clinical trials. While communicating prognosis will always remain a multifaceted and arduous undertaking, we believe this nomogram will substantially lessen the fear and ambiguity that accompanies a diagnosis of CUP for both our patients and our providers.

Authors' Disclosures

B. Smaglo reports personal fees from Taiho Oncology and Sirtex outside the submitted work. F.A. Greco reports speakers' bureau and serves as a medical advisor to Biotheranostics Inc. J.M. Loree reports grants and personal fees from Amgen and Ipsen, as well as personal fees from Eisai, Novartis, and Pfizer outside the submitted work. No disclosures were reported by the other authors.

Authors' Contributions

K. Raghav: Conceptualization, resources, data curation, supervision, validation, investigation, visualization, methodology, writing–original draft, project administration, writing–review and editing. **H. Hwang:** Data curation, software, formal analysis, validation, methodology, writing–original draft, writing–review and editing. **A.A. Jácome:** Data curation, writing–review and editing. **E. Bhang:** Data curation, writing–review and editing. **A. Willett:** Data curation, writing–review and editing. **R.W. Huey:** Writing–review and editing. **N.P. Dhillon:** Data curation, writing–review and editing. **J. Modha:** Data curation, writing–review and editing. **B. Smaglo:** Writing–review and editing. **A. Matamoros Jr:** Writing–review and editing. **J.S. Estrella:** Writing–review and editing. **J. Jao:** Software, writing–review and editing. **M.J. Overman:** Visualization, writing–original draft, writing–review and editing. **X. Wang:** Data curation, software, formal analysis, validation, visualization, methodology, writing–original draft, writing–review and editing. **F.A. Greco:** Resources, data curation, writing–original draft, writing–review and editing. **J.M. Loree:** Resources, data curation, software, formal analysis, validation, investigation, writing–original draft, writing–review and editing. **G.R. Varadhachary:** Conceptualization, resources, data curation, formal analysis, supervision, funding acquisition, validation, investigation, visualization, methodology, writing–original draft, project administration, writing–review and editing.

Acknowledgments

We thank our patients for entrusting us with their care and for giving us the opportunity to learn and understand their cancer and hopefully help others who may suffer from this orphan disease in the future. This work was supported in part by Painter Research Funds (to G.R. Varadhachary) and Cancer Center Support Grant (CCSG) - PA30 CA016672 (to K. Raghav and G.R. Varadhachary).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received October 19, 2020; revised December 23, 2020; accepted April 9, 2021; published first April 15, 2021.

References

1. Richardson A, Wagland R, Foster R, Symons J, Davis C, Boyland L, et al. Uncertainty and anxiety in the cancer of unknown primary patient journey: a multiperspective qualitative study. BMJ Support Palliat Care 2015;5:366–72.
2. Wagland R, Bracher M, Drosdowsky A, Richardson A, Symons J, Mileshkin L, et al. Differences in experiences of care between patients diagnosed with metastatic cancer of known and unknown primaries: mixed-method findings from the 2013 cancer patient experience survey in England. BMJ Open 2017;7: e017881.
3. Boyland L, Davis C. Patients' experiences of carcinoma of unknown primary site: dealing with uncertainty. Palliat Med 2008;22:177–83.
4. Karapetis CS, Guccione L, Tattersall MH, Gooden H, Vajdic CM, Lambert S, et al. Perceptions of cancer of unknown primary site: a national survey of Australian medical oncologists. Intern Med J 2017;47:408–14.
5. Varadhachary GR, Karanth S, Qiao W, Carlson HR, Raber MN, Hainsworth JD, et al. Carcinoma of unknown primary with gastrointestinal profile: immunohistochemistry and survival data for this favorable subset. Int J Clin Oncol 2014; 19:479–84.
6. Temel JS, Greer JA, Admane S, Gallagher ER, Jackson VA, Lynch TJ, et al. Longitudinal perceptions of prognosis and goals of therapy in patients with metastatic non-small-cell lung cancer: results of a randomized study of early palliative care. J Clin Oncol 2011;29:2319–26.
7. Weeks JC, Cook EF, O'Day SJ, Peterson LM, Wenger N, Reding D, et al. Relationship between cancer patients' predictions of prognosis and their treatment preferences. JAMA 1998;279:1709–14.
8. Hagerty RG, Butow PN, Ellis PA, Lobb EA, Pendlebury S, Leighl N, et al. Cancer patient preferences for communication of prognosis in the metastatic setting. J Clin Oncol 2004;22:1721–30.
9. Peppercorn JM, Smith TJ, Helft PR, Debono DJ, Berry SR, Wollins DS, et al. American Society of Clinical Oncology statement: toward individualized care for patients with advanced cancer. J Clin Oncol 2011;29:755–60.
10. Liu PH, Landrum MB, Weeks JC, Huskamp HA, Kahn KL, He Y, et al. Physicians' propensity to discuss prognosis is associated with patients' awareness of prognosis for metastatic cancers. J Palliat Med 2014;17:673–82.
11. Goodlin SJ, Zhong Z, Lynn J, Teno JM, Fago JP, Desbiens N, et al. Factors associated with use of cardiopulmonary resuscitation in seriously ill hospitalized adults. JAMA 1999;282:2333–9.
12. Wright AA, Zhang B, Ray A, Mack JW, Trice E, Balboni T, et al. Associations between end-of-life discussions, patient mental health, medical care near death, and caregiver bereavement adjustment. JAMA 2008;300:1665–73.
13. Gramling R, Stanek S, Han PKJ, Duberstein P, Quill TE, Temel JS, et al. Distress due to prognostic uncertainty in palliative care: frequency, distribution, and outcomes among hospitalized patients with advanced cancer. J Palliat Med 2018; 21:315–21.
14. Obermeyer Z, Makar M, Abujaber S, Dominici F, Block S, Cutler DM. Association between the Medicare hospice benefit and health care utilization and costs for patients with poor-prognosis cancer. JAMA 2014;312:1888–96.
15. Hayashi H, Kurata T, Takiguchi Y, Arai M, Takeda K, Akiyoshi K, et al. Randomized phase II trial comparing site-specific treatment based on gene expression profiling with carboplatin and paclitaxel for patients with cancer of unknown primary site. J Clin Oncol 2019;37:570–9.
16. Culine S. Prognostic factors in unknown primary cancer. Semin Oncol 2009;36: 60–4.
17. Culine S, Kramar A, Saghatchian M, Bugat R, Lesimple T, Lortholary A, 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.
18. Seve P, Ray-Coquard I, Trillet-Lenoir V, Sawyer M, Hanson J, Broussolle C, et al. Low serum albumin levels and liver metastasis are powerful prognostic markers for survival in patients with carcinomas of unknown primary site. Cancer 2006; 107:2698–705.
19. Petrakis D, Pentheroudakis G, Voulgaris E, Pavlidis N. 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.
20. Iasonos A, Schrag D, Raj GV, Panageas KS. How to build and interpret a nomogram for cancer prognosis. J Clin Oncol 2008;26:1364–70.
21. Atkinson AC. A note on the generalized information criterion for choice of a model. Biometrika 1980;67:413–8.
22. Van Houwelingen JC, Le Cessie S. Predictive value of statistical models. Stat Med 1990;9:1303–25.
23. Harrell FE Jr. Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis. New York: Springer; 2015.
24. Mithat Gönen GH. Concordance probability and discriminatory power in proportional hazards regression. Biometrika 2005;92:965–70.
25. Kirk P, Kirk I, Kristjanson LJ. What do patients receiving palliative care for cancer and their families want to be told? A Canadian and Australian qualitative study. BMJ 2004;328:1343.
26. Twomey F, O'Leary N, O'Brien T. Prediction of patient survival by healthcare professionals in a specialist palliative care inpatient unit: a prospective study. Am J Hosp Palliat Care 2008;25:139–45.
27. Templeton AJ, McNamara MG, Seruga B, Vera-Badillo FE, Aneja P, Ocana A, et al. Prognostic role of neutrophil-to-lymphocyte ratio in solid tumors: a systematic review and meta-analysis. J Natl Cancer Inst 2014;106:dju124.
28. Howard R, Kanetsky PA, Egan KM. Exploring the prognostic value of the neutrophil-to-lymphocyte ratio in cancer. Sci Rep 2019;9:19673.
29. Chen ZY, Raghav K, Lieu CH, Jiang ZQ, Eng C, Vauthey JN, et al. Cytokine profile and prognostic significance of high neutrophil-lymphocyte ratio in colorectal cancer. Br J Cancer 2015;112:1088–97.
30. Mohamed Z, Pinato DJ, Mauri FA, Chen KW, Chang PM, Sharma R. Inflammation as a validated prognostic determinant in carcinoma of unknown primary site. Br J Cancer 2014;110:208–13.
31. Huey RW, Makawita S, Xiao L, Matamoros A, Estrella JS, Overman MJ, et al. Sarcomatoid carcinoma presenting as cancers of unknown primary: a clinicopathological portrait. BMC Cancer 2019;19:965.
32. Enzinger AC, Zhang B, Schrag D, Prigerson HG. Outcomes of prognostic disclosure: associations with prognostic understanding, distress, and relationship with physician among patients with advanced cancer. J Clin Oncol 2015;33: 3809–16.
33. Mack JW, Fasciano KM, Block SD. Communication about prognosis with adolescent and young adult patients with cancer: information needs, prognostic awareness, and outcomes of disclosure. J Clin Oncol 2018;36:1861–7.
34. Gordon EJ, Daugherty CK. 'Hitting you over the head': oncologists' disclosure of prognosis to advanced cancer patients. Bioethics 2003;17:142–68.
35. Jenkins V, Solis-Trapala I, Langridge C, Catt S, Talbot DC, Fallowfield LJ. What oncologists believe they said and what patients believe they heard: an analysis of phase I trial discussions. J Clin Oncol 2011;29:61–8.
36. Temel JS, Shaw AT, Greer JA. Challenge of prognostic uncertainty in the modern era of cancer therapeutics. J Clin Oncol 2016;34:3605–8.
37. Varghese AM, Arora A, Capanu M, Camacho N, Won HH, Zehir A, et al. Clinical and molecular characterization of patients with cancer of unknown primary in the modern era. Ann Oncol 2017;28:3015–21.
38. Ross JS, Wang K, Gay L, Otto GA, White E, Iwanik K, et al. Comprehensive genomic profiling of carcinoma of unknown primary site: new routes to targeted therapies. JAMA Oncol 2015;1:40–9.
39. Naing A, Meric-Bernstam F, Stephen B, Karp DD, Hajjar J, Rodon Ahnert J, et al. Phase 2 study of pembrolizumab in patients with advanced rare cancers. J Immunother Cancer 2020;8:e000347.
40. Bochtler T, Reiling A, Endris V, Hielscher T, Volckmar AL, Neumann O, et al. Integrated clinicomolecular characterization identifies RAS activation and CDKN2A deletion as independent adverse prognostic factors in cancer of unknown primary. Int J Cancer 2020;146:3053–64.