

RESEARCH ARTICLE

Integrating additional factors into the TNM staging for cutaneous melanoma by machine learning

Charles Q. Yang¹, Huan Wang², Zhenqiu Liu³, Matthew T. Hueman⁴, Aadya Bhaskaran⁵, Donald E. Henson⁶, Li Sheng⁷, Dechang Chen^{8*}

1 Department of Surgery, Walter Reed National Military Medical Center, Bethesda, MD, United States of America, **2** Department of Biostatistics, The George Washington University, Washington, DC, United States of America, **3** Department of Public Health Sciences, Penn State Cancer Institute, Hershey, PA, United States of America, **4** Department of Surgical Oncology, John P. Murtha Cancer Center, Walter Reed National Military Medical Center, Bethesda, MD, United States of America, **5** Department of Quantitative Theory and Methods, Emory University, Atlanta, GA, United States of America, **6** Deceased, was with The Department of Preventive Medicine & Biostatistics, F. Edward Hébert School of Medicine, Uniformed Services University of the Health Sciences, Bethesda, MD, United States of America, **7** Department of Mathematics, Drexel University, Philadelphia, PA, United States of America, **8** Department of Preventive Medicine & Biostatistics, F. Edward Hébert School of Medicine, Uniformed Services University of the Health Sciences, Bethesda, MD, United States of America

* dechang.chen@usuhs.edu

OPEN ACCESS

Citation: Yang CQ, Wang H, Liu Z, Hueman MT, Bhaskaran A, Henson DE, et al. (2021) Integrating additional factors into the TNM staging for cutaneous melanoma by machine learning. PLoS ONE 16(9): e0257949. <https://doi.org/10.1371/journal.pone.0257949>

Editor: Sandro Pasquali, Fondazione IRCCS Istituto Nazionale dei Tumori, ITALY

Received: January 6, 2021

Accepted: September 14, 2021

Published: September 30, 2021

Copyright: This is an open access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the [Creative Commons CC0](https://creativecommons.org/licenses/by/4.0/) public domain dedication.

Data Availability Statement: Data are available from the Surveillance, Epidemiology, and End Results (SEER) database of the National Cancer Institute to any researcher who signs a SEER Research Data Agreement form. SEER data can be accessed here: <https://seer.cancer.gov/data/>. The authors did not have any special access privileges that others would not have."

Funding: This work was partially supported by grants "Using Dendrograms to Create Prognostic Systems for Cancer" and "Creating Prognostic

Abstract

Background

Integrating additional factors into the TNM staging system is needed for more accurate risk classification and survival prediction for patients with cutaneous melanoma. In the present study, we introduce machine learning as a novel tool that incorporates additional prognostic factors to improve the current TNM staging system.

Methods and findings

Cancer-specific survival data for cutaneous melanoma with at least a 5 years follow-up were extracted from the Surveillance, Epidemiology, and End Results Program of the National Cancer Institute and split into the training set (40,781 cases) and validation set (5,390 cases). Five factors were studied: the primary tumor (T), regional lymph nodes (N), distant metastasis (M), age (A), and sex (S). The Ensemble Algorithm for Clustering Cancer Data (EACCD) was applied to the training set to generate prognostic groups. Utilizing only T, N, and M, a basic prognostic system was built where patients were stratified into 10 prognostic groups with well-separated survival curves, similar to 10 AJCC stages. These 10 groups had a significantly higher accuracy in survival prediction than 10 stages (C-index = 0.7682 vs 0.7643; increase in C-index = 0.0039, 95% CI = (0.0032, 0.0047); p-value = 7.2×10^{-23}). Nevertheless, a positive association remained between the EACCD grouping and the AJCC staging (Spearman's rank correlation coefficient = 0.8316; p-value = 4.5×10^{-13}). With additional information from A and S, a more advanced prognostic system was established using the training data that stratified patients into 10 groups and further improved the prediction

Systems for Cancer" sponsored by John P. Murtha Cancer Center Research Program and grant "Four Diamonds Fund from Penn State University" sponsored by Penn State University.

Competing interests: The authors have declared that no competing interests exist.

accuracy (C-index = 0.7865 vs 0.7643; increase in C-index = 0.0222, 95% CI = (0.0191, 0.0254); p-value = 8.8×10^{-43}). Both internal validation using the training set and temporal validation using the validation set showed good stratification and a high predictive accuracy of the prognostic systems.

Conclusions

The EACCD allows additional factors to be integrated into the TNM to create a prognostic system that improves patient stratification and survival prediction for cutaneous melanoma. This integration separates favorable from unfavorable clinical outcomes for patients and improves both cohort selection for clinical trials and treatment management.

Introduction

Cutaneous melanoma is responsible for the vast majority of skin cancer-related deaths in the US and accounts for 7 percent of new cancer cases in men and 4 percent of new cancer cases in women [1]. There are multiple prognostic factors for cutaneous melanoma that correlate with survival. These include lesion thickness, lymph node involvement, metastasis, ulceration, age, sex, anatomic location, mitotic rate, satellite/in-transit lesions, and serum lactic dehydrogenase level, etc [2]. Among these, the primary tumor feature (thickness and ulceration), nodal involvement, and distant metastasis are the three factors that constitute the AJCC staging system [3].

The initial staging system for cutaneous melanoma was developed by the American Joint Committee on Cancer (AJCC) in 1977 [4]. Through the years, the AJCC makes periodic revisions to the TNM system in an effort to reflect a better understanding of the disease. The current TNM system for melanoma classifies the disease from stage I through IV. Surgical excision is often adequate for stages I-II. Sentinel lymph node biopsy may have a role in early to intermediate cases to improve staging and identify cases who may benefit from earlier lymphadenectomy [5]. Treatment of the locally advanced or metastatic disease is complex and may require additional adjuvant treatments. Multidisciplinary tumor boards are needed to individualize treatment plans for advanced stage patients. Although melanoma has been shown historically to be a radioresistant and chemoresistant cancer, clinical trials have shown the efficacy of targeted immunotherapy, especially checkpoint blockade in metastatic cases. Additional trials are ongoing to evaluate its use in the neoadjuvant, adjuvant, and palliative settings, with or without local therapy such as surgery or radiation [6].

In the context of evolving knowledge and modern tumor registries, additional factors should be integrated to augment TNM to create prognostic systems for improved research and treatment. Cox regression modeling [7] and tree modeling [8, 9] are two approaches that could expand the AJCC staging system by integrating additional factors. Cox regression, which focuses on the optimal fitting of the data, can achieve a high accuracy in predicting survival. However, the assumptions made with Cox modeling (e.g., each factor/variable makes a linear contribution to the model) usually do not prove realistic. In addition, there is no clear rule that can guide how to use the output (e.g., the nomogram) to stratify patients into risk groups analogous to AJCC stages. As a result, studies often empirically partition patients into a small number of groups (e.g., 2 or 3 quantile groups) [10]. On the other hand, the traditional survival tree method, partitioning the space of values of factors into disjoint and non-

overlapping regions, can be used to form prognostic groups explicitly. However, the tree models in general have a low prediction accuracy [11].

In this study, we described a machine learning approach using the Ensemble Algorithm for Clustering Cancer Data (EACCD) [12–22] to expand the AJCC staging system by integrating additional factors that contribute to the prognosis of the disease. We demonstrated the approach by building 2 prognostic systems. One system, based on primary tumor (T), regional lymph nodes (N), and distant metastasis (M), was employed to examine the current 8th AJCC staging system. The second system, based on T, N, M, age (A), and sex (S), expanded the current staging system.

Materials and methods

Data

Data from patients diagnosed with cutaneous melanoma during 2004 through 2013 were obtained from 18 databases of the Surveillance, Epidemiology, and End Results Program (SEER) of the National Cancer Institute [23]. The restriction on year of diagnosis was used for reasons that 1) information of some factors' levels was not available until 2004; 2) 2013 is the latest year that ensures a minimum 5-year follow up since the current release of SEER data followed patients up to 2018. Survival time (measured in months) and SEER cause-specific death classification variable [24] were used to obtain melanoma-specific survival for analysis. In addition to restricting the years of diagnosis between 2004 and 2013, we required patients' nodes to be pathologically evaluated (CS Lymph Nodes Eval = 2,3,6,8 [25]). This study investigated 5 factors: T, N, M, A, and S. The datasets used for analysis were generated as follows.

Survival time (in months), SEER cause-specific death classification variable, T, N, M, A, and S were collected for the studied patients. The levels of T (9 levels: T0, T1a, T1b, T2a, T2b, T3a, T3b, T4a, and T4b), N (7 levels: N0, N1a, N1b, N2a, N2b, N2c, and N3), M (2 levels: M0 and M1), A (2 levels: A0, A1), and S (2 levels: S1 and S2) are defined in [S1 Table](#). Age 70 was used as the cut-off since an age greater than 70 years was associated with worse primary tumor characteristics and higher mortality compared with younger patients [26]. A combination is defined as a subset of the data corresponding to one level of each factor and is denoted by levels of factors (e.g., T1N0M0A0S1 represents a subset of patients with T = T1, N = N0, M = M0, A = A0, S = S1). We retained patients without missing values in any of the five factors (i.e., T, N, M, A, and S) so that their combinations can be formed. Then patients were assigned to the training set if they were diagnosed with melanoma before 2013, and to the validation set if diagnosed in 2013. The validation set obtained this way was used in temporal validation to examine the reproducibility of the prognostic system built on the training set [27]. For more accurate estimation in the development of prognostic systems, we removed all rare combinations each containing less than 25 patients in the training set. The final training set used for analysis includes 113 combinations giving a total number of 40,781 cases. For the validation set, we excluded patients who could not be classified into any of the 113 combinations of the training set. The final validation set contains 5,390 patients. See [Fig 1](#) and [Table 1](#) for details on data. Additional information on datasets is contained in [S1 Data](#).

EACCD

The EACCD is a machine learning algorithm that can be used to cluster patients into n^* prognostic groups so that i) patients from the same group have a similar survival experience and patients from two distinct groups have different survival experiences; and ii) n^* is as small as possible while n^* groups can achieve the maximum possible accuracy in survival prediction. In this study, we will use an expanded version of the algorithm that consists of 4 steps: 1) defining

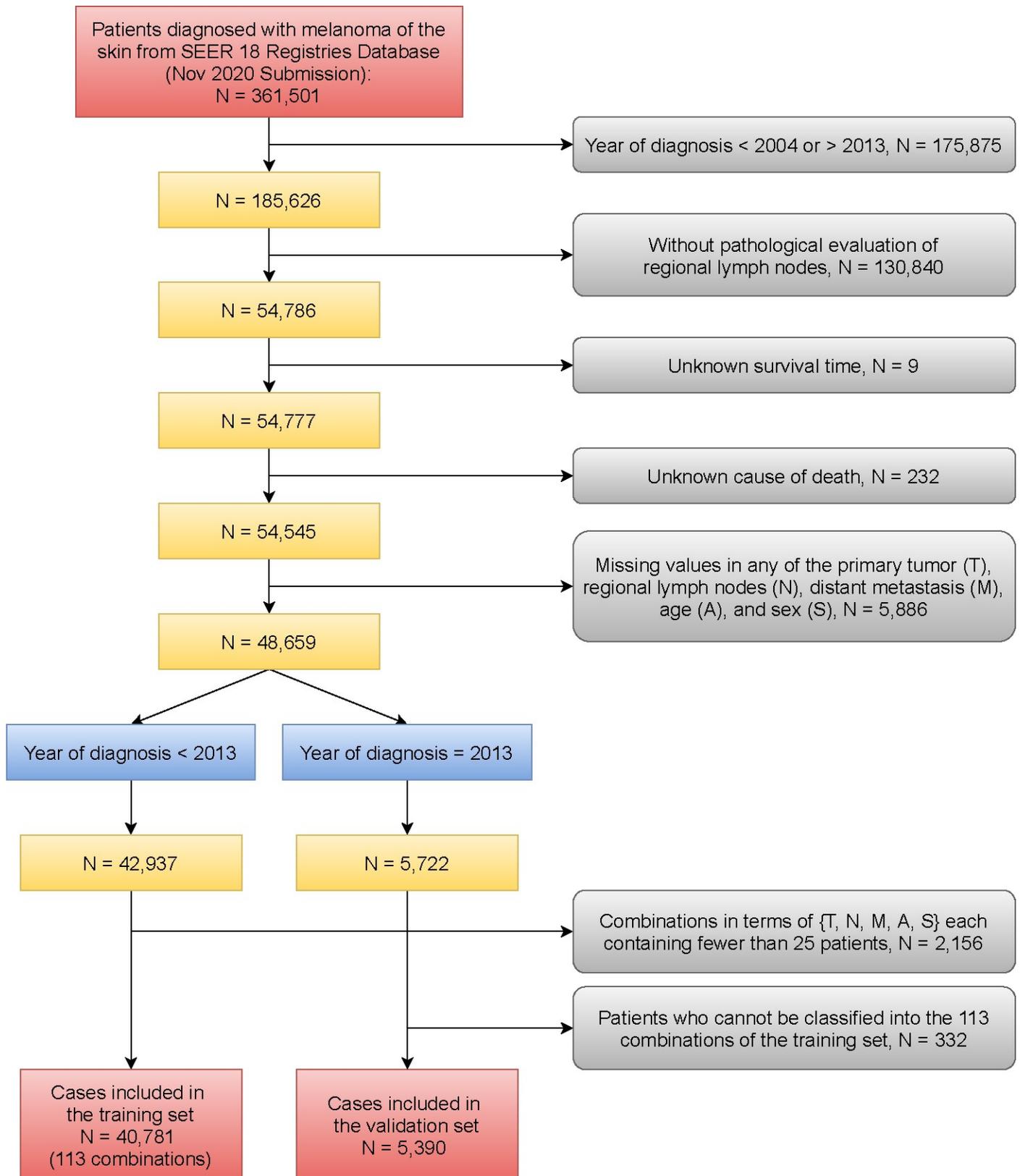


Fig 1. Flow diagram for selecting patients with skin melanoma.

<https://doi.org/10.1371/journal.pone.0257949.g001>

Table 1. Clinical and demographic characteristics of the data.

	Training Set		Validation Set	
	N	%	N	%
Primary Tumor				
T0	68	0.2	14	0.3
T1a	8,266	20.3	984	18.3
T1b	6,929	17.0	868	16.1
T2a	10,882	26.7	1,458	27.1
T2b	2,351	5.8	358	6.6
T3a	4,456	10.9	586	10.9
T3b	3,097	7.6	441	8.2
T4a	1,877	4.6	267	5.0
T4b	2,855	7.0	414	7.7
Regional Lymph Nodes				
N0	34,995	85.8	4,647	86.2
N1a	3,240	7.9	404	7.5
N1b	302	0.7	41	0.8
N2a	861	2.1	113	2.1
N2b	238	0.6	46	0.9
N2c	154	0.4	20	0.4
N3	991	2.4	119	2.2
Distant Metastasis				
M0	40,739	99.9	5,389	100.0
M1	42	0.1	1	0.0
Age				
A0	29,738	72.9	3,738	69.4
A1	11,043	27.1	1,652	30.6
Sex				
S1	24,861	61.0	3,384	62.8
S2	15,920	39.0	2,006	37.2

<https://doi.org/10.1371/journal.pone.0257949.t001>

initial dissimilarities between survival functions of any two combinations; 2) obtaining learned dissimilarities by using initial dissimilarities and an ensemble learning method; 3) applying hierarchical clustering to cluster combinations by the learned dissimilarities; and 4) generating n^* prognostic groups by exploiting the C-index. Different approaches are available for each step. In this paper, the initial dissimilarity between two combinations in step 1 was defined by Mann-Whitney based effect size $|P(T_1 > T_2) - 0.5|$, where T_1 and T_2 are the failure times of two underlying populations corresponding to the two combinations (see below). In step 2, the ensemble learning method was based on the two-phase Partitioning Around Medoids algorithm [28]. The minimax linkage method [29] was chosen for hierarchical clustering in step 3. The output of this step is a tree-structured dendrogram, which represents the relationship among survival of patients in different combinations. In step 4, a C-index curve was first produced using the dendrogram in step 3 and then the optimal number n^* of prognostic groups was found using the “knee” point of the C-index curve and then the n^* prognostic groups were generated by cutting the dendrogram [16, 18, 20].

Mann-Whitney based effect size

The Mann-Whitney parameter arises from the widely used Mann-Whitney test [30] that examines whether one of two random variables is stochastically larger than the other. Let T_1 and T_2 denote the variables of survival time for patients from population 1 (with survival function $S_1(t)$) and population 2 (with survival function $S_2(t)$), respectively. The Mann-Whitney parameter is defined as $P(T_1 > T_2)$, the probability that a randomly chosen patient from population 1 has a longer survival time than a randomly chosen patient from population 2. The difference between the Mann-Whitney parameter and 0.5 suggests a difference between $S_1(t)$ and $S_2(t)$. Efron proposed to use $\hat{D} = -\int_0^\infty \hat{S}_1(t) d\hat{S}_2(t)$ to estimate the Mann-Whitney parameter for censoring data, where $\hat{S}_1(t)$ and $\hat{S}_2(t)$ are Kaplan-Meier estimates of $S_1(t)$ and $S_2(t)$, respectively [31]. However, Efron's estimator requires that both $\hat{S}_1(t)$ and $\hat{S}_2(t)$ drop to 0 at the maximum study time (the longest following-up time) and is rather unstable when censoring occurs due to incomplete follow up [32]. To overcome this problem, Wang [20] proposed to 1) use the exponential tails to complete survival functions; and 2) use the completed survival curves to approximate $P(T_1 > T_2)$. More specifically, $P(T_1 > T_2)$ is approximated by $D_e(\tau_1, \tau_2) = A+B+C$, where $A = -\int_0^{\min(\tau_1, \tau_2)} S_1(t) dS_2(t)$; $B = -\int_{\tau_1}^{\tau_2} e^{-\lambda_1 t} dS_2(t)$ if $\tau_1 < \tau_2$, 0 if $\tau_1 = \tau_2$, $\int_{\tau_2}^{\tau_1} S_1(t) \lambda_2 e^{-\lambda_2 t} dt$ if $\tau_1 > \tau_2$; $C = \frac{\lambda_2}{\lambda_1 + \lambda_2} e^{-(\lambda_1 + \lambda_2) \max(\tau_1, \tau_2)}$; and $\lambda_i = -\log(S_i(\tau_i))/\tau_i$ for $i = 1, 2$. Accordingly, $\hat{D}_e(\tau_1, \tau_2) = \hat{A} + \hat{B} + \hat{C}$, an estimate of $P(T_1 > T_2)$, can be obtained by using the Kaplan-Meier estimates and Lebesgue-Stieltjes integration. In this paper, we used $|\hat{D}_e(\tau_1, \tau_2) - 0.5|$ to compute the initial dissimilarity in survival between two combinations, with τ_1 and τ_2 set to be the maximum possible time by which the Kaplan-Meier estimates of the survival of all combinations can be calculated.

Prognostic systems

Survival curves using the Kaplan-Meier estimates [33] were plotted for the n^* prognostic groups from EACCD. These curves allow visually evaluating survival differences among the groups. The final prognostic system includes the dendrogram, group assignment, C-index, and survival curves for the prognostic groups.

Validation approach

The prognostic system was evaluated through both internal and temporal validation [27]. Internal validation, conducted on the training set, was mainly used to evaluate the validity of the system. Stratification and the predictive accuracy, two key characteristics of the system, were evaluated with internal validation. Stratification was examined by inspecting the survival curves of prognostic groups and comparing them using the univariate Cox proportional hazards regression model and the logrank test. The predictive accuracy was examined by computing the C-index of the prognostic system. The temporal validation, conducted on the validation set, was used to assess the system's reproducibility, i.e., both stratification and the predictive accuracy obtained using the training set were also valid for patients in the validation set.

Software

All statistical analyses were done in R (R version 3.6.2) using the following libraries: survival, cluster, protoclus, factoextra, and compareC.

Results

Prognostic system for T, N, M

Applying the EACCD with T, N, and M to the training set yielded the dendrogram in Fig 2A. Dissimilarity between two combinations was computed by completing survival curves of combinations with exponential tails from 149 months, i.e., $\tau_1 = \tau_2 = 149$. The C-index curve based on the dendrogram, shown in Fig 2B, was used to find the optimal number of prognostic groups n^* . The knee point of the curve corresponds to 10 groups (C-index value of 0.7682), which suggested $n^* = 10$. Cutting the dendrogram into $n^* = 10$ groups is shown by rectangles in Fig 2C. Fig 2D (and S1 Fig) shows the survival curves of these 10 groups, which are well separated. For convenience, the definition for all 10 groups is restated in the 4th column of S2 Table. The resulting prognostic system for T, N, and M includes the dendrogram in Fig 2A, the groups in S2 Table (4th column), and the survival curves in Fig 2D. Note that the risk of mortality is increased as the group number increases. For a quick comparison, Fig 3 (and S2 Fig) shows the survival curves of 10 stages of the 8th AJCC staging system.

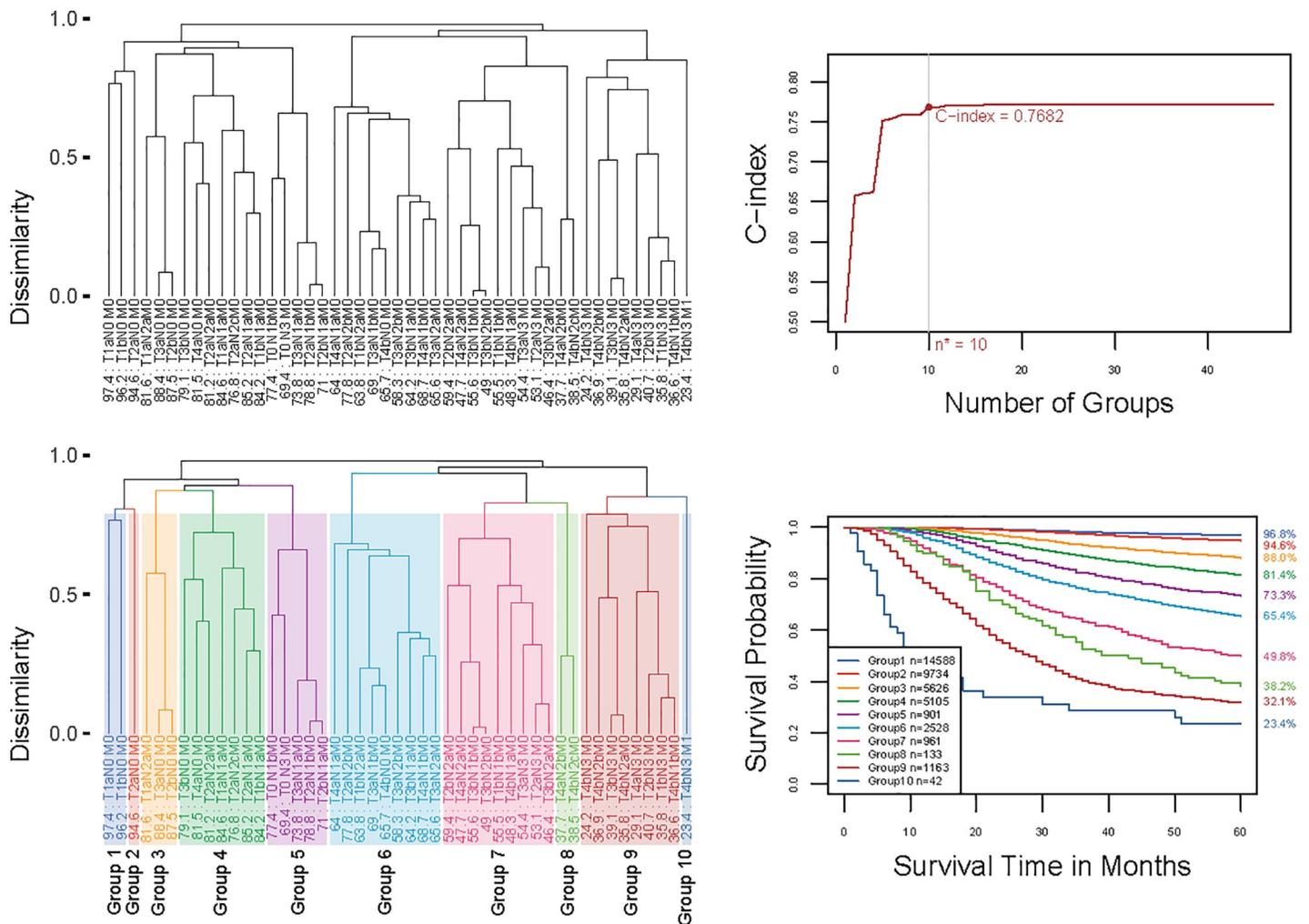


Fig 2. Creating EACCD prognostic groups on T, N, and M by the training set. A Dendrogram from running EACCD. A 5-year melanoma-specific survival rate in percentage is provided below each combination. B C-index curve based on the dendrogram in panel A. The knee point of the curve corresponds to 10 groups and a C-index value of 0.7682. C Cutting the dendrogram in panel A according to $n^* = 10$ suggested in panel B creates 10 prognostic groups, shown in rectangles. Listed on the bottom of the dendrogram are group numbers. D Melanoma-specific survival of 10 prognostic groups in panel C. The 5-year melanoma-specific survival rates for 10 groups are listed on the right side of the figure.

<https://doi.org/10.1371/journal.pone.0257949.g002>

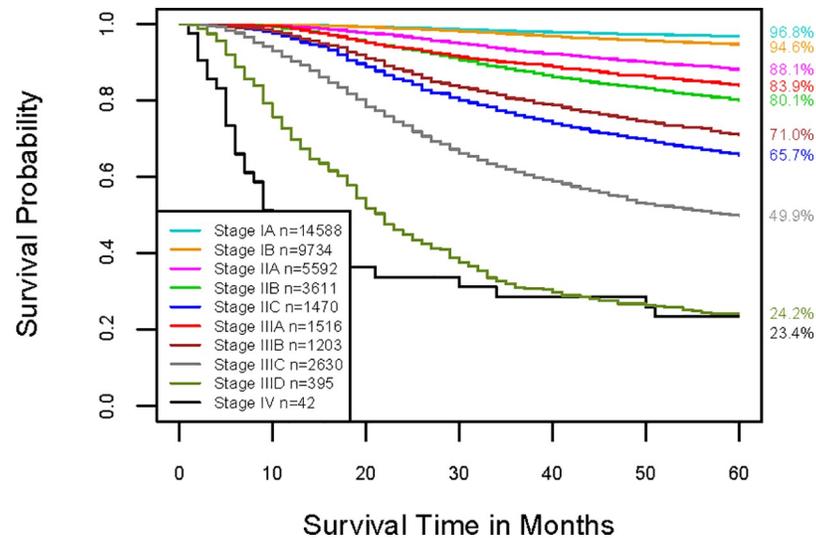


Fig 3. Melanoma-specific survival of AJCC stages of the training data. Stages are defined in the 5th column in S2 Table. The 5-year melanoma-specific survival rates for 10 stages are listed on the right side of the figure.

<https://doi.org/10.1371/journal.pone.0257949.g003>

Prognostic system for T, N, M, A, S

When T, N, M, A, and S were involved, the dissimilarity between two combinations was computed by completing survival curves of combinations with exponential tails from 115 months, i.e., $\tau_1 = \tau_2 = 115$.

Before building the EACCD prognostic system for T, N, M, A, and S by the training set, we assessed the performance in survival prediction of the models for the following three sets of factors: {T, N, M, A}, {T, N, M, S}, and {T, N, M, A, S}, as compared with the model based only on {T, N, M}. Fig 4 plots the C-index curves of the training set, each created in step 4 of EACCD, for these 4 scenarios. It is seen that the curves on {T, N, M, A} and {T, N, M, S} are higher than the curve shown on {T, N, M}. Therefore, adding A or S to {T, N, M} increases the C-index and thus improves the prediction accuracy. The curve of {T, N, M, A, S} is the highest among all 4 curves, implying that adding both A and S to {T, N, M} leads to the biggest improvement in prediction accuracy of {T, N, M}.

Exploring around the “knee” point of the red curve on {T, N, M, A, S} in Fig 5B reveals the optimal number of prognostic groups $n^* = 10$ with a corresponding C-index of 0.7865. Fig 5A shows the dendrogram and its cutting (shown in rectangles) according to $n^* = 10$ groups. A detailed definition for all 10 groups is listed in S3 Table. Fig 5C shows the survival curves for the 10 prognostic groups. The dendrogram in Fig 5A, the groups in S3 Table, and the survival curves in Fig 5C (and S3 Fig) present a prognostic system for T, N, M, A, and S.

Validation of the EACCD systems

The EACCD systems in {T, N, M} and {T, N, M, A, S} built above can be validated internally using the training data and temporally using the validation data. Below we present validation results for the EACCD system based on {T, N, M, A, S}. The same procedure can be followed for the EACCD system on {T, N, M}.

The internal validation evaluates the stratification and predictive accuracy of the EACCD system using only the training set. Fig 5 indicates that the survival curves of the prognostic groups are well separated. This good separation can be verified from the comparison between survival

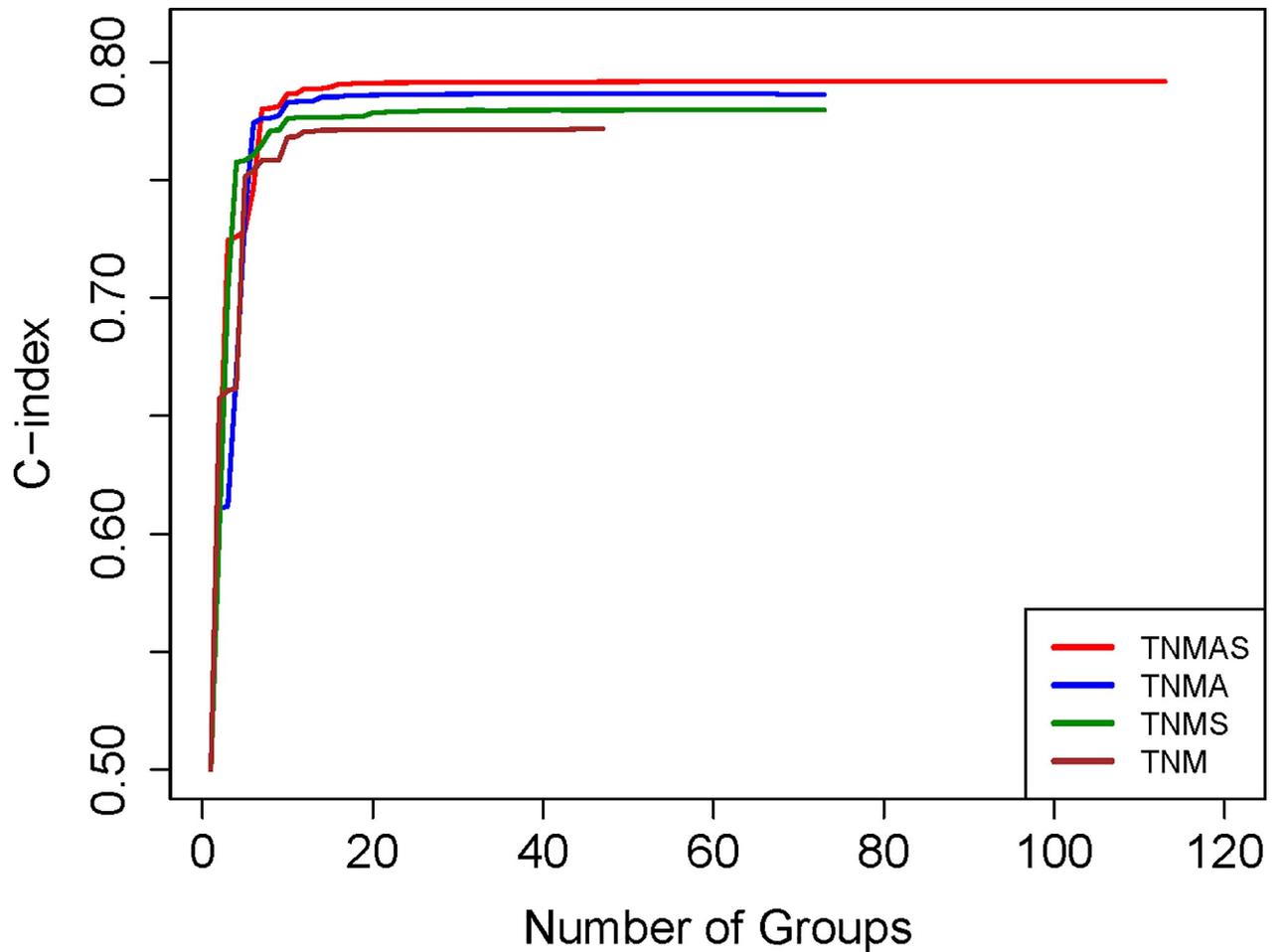


Fig 4. Training data C-index curves based on different factors.

<https://doi.org/10.1371/journal.pone.0257949.g004>

curves of adjacent groups, shown in Table 2. This table reveals that the hazard ratios between adjacent groups are uniformly significant and are mostly between 1.3 and 1.6., and that the log-rank test results between adjacent groups are also uniformly significant. The previous section showed that the C-index value of the EACCD system based on {T, N, M, A, S} has a C-index value close to 0.8, which implies a good predictive accuracy of the system [27]. Therefore, the EACCD system on {T, N, M, A, S} well stratifies and predicts the patients from the training set.

The external validation examines the stratification and predictive accuracy of the EACCD system on the validation set. Fig 6 (and S4 Fig) shows the survival curves of the 10 prognostic groups on the validation set. Although due to the small sample size these curves do not provide accurate estimates of the survival rates, they preserve the pattern from the training set (Fig 5). This is further confirmed by Table 2, which shows separation between groups, though some results are less significant due to small sample sizes. Thus a good stratification occurs in the validation set. (The large p-value between groups 6 and 7 is mainly due to their partially overlapping survival curves in early years (Fig 6 and S4 Fig).) Calculation shows that the EACCD system has a C-index value of 0.7941 on the validation set, which is high. Therefore the system has a high predictive accuracy for the patients in the validation set.

The above internal and external validation shows that the EACCD system on {T, N, M, A, S} has good stratification and a high predictive accuracy for patients used in the development

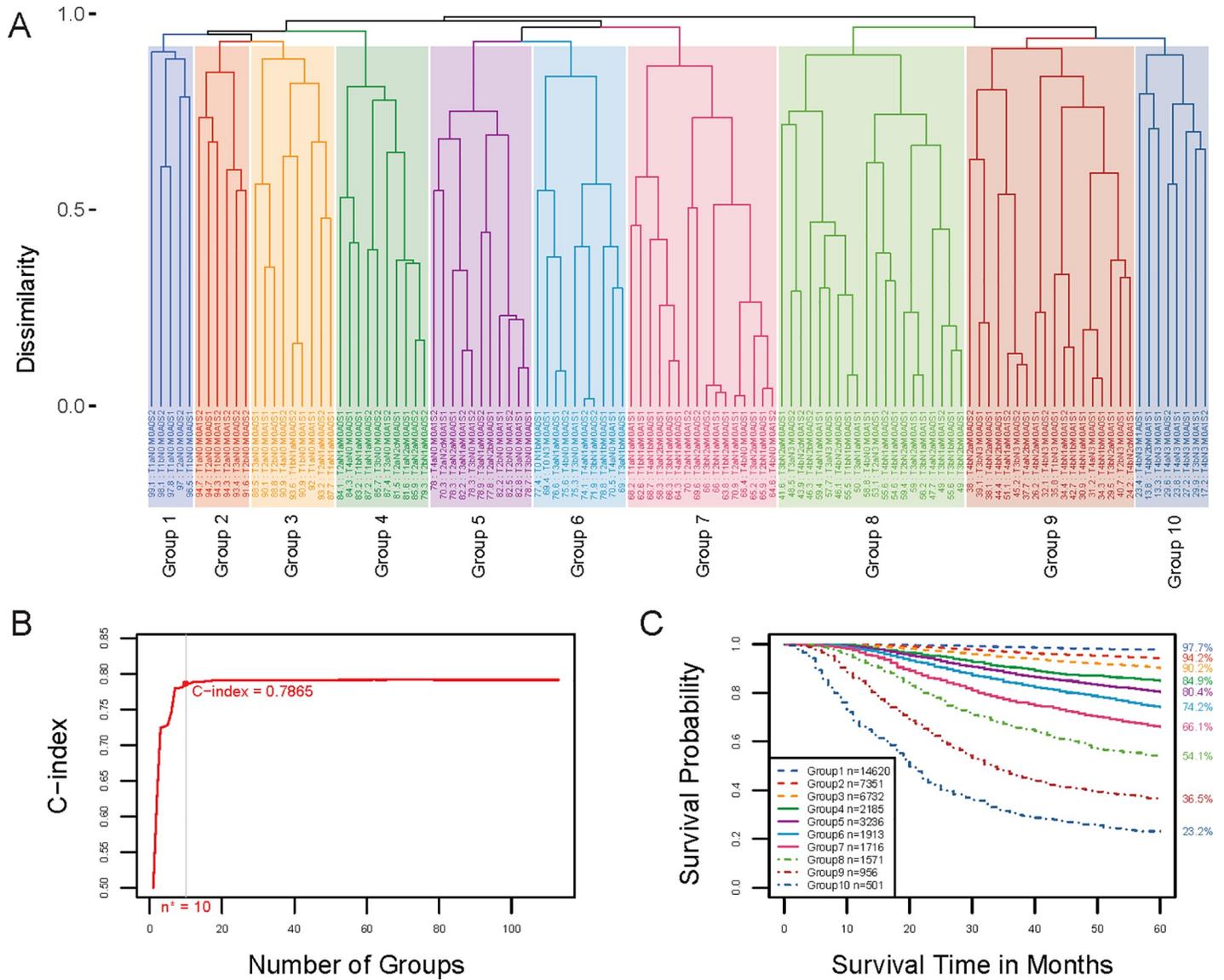


Fig 5. Creating the EACCD prognostic groups on T, N, M, A, and S by the training set. A Dendrogram and cutting the dendrogram (shown in rectangles) according to the C-index. Running the EACCD results in the tree-structured dendrogram. A 5-year melanoma-specific survival rate in percentage is provided below each combination. Cutting the dendrogram according to $n^* = 10$ in panel B creates 10 prognostic groups, shown in rectangles. Listed on the bottom of the dendrogram are group numbers. B C-index curve based on the dendrogram in panel A. The knee point of the curve corresponds to 10 groups and a C-index value of 0.7865. C Melanoma-specific survival of 10 prognostic groups in panel A. The 5-year melanoma-specific survival rates for 10 groups are listed on the right side of the figure.

<https://doi.org/10.1371/journal.pone.0257949.g005>

and patients in the future. Applying a similar validation procedure to the EACCD system on {T, N, M} led to the same conclusion for the system in three variables.

Discussion

Comparison with AJCC

With the training data, the EACCD prognostic system on {T, N, M} can be compared with the 8th ed. AJCC staging system in terms of both prediction and stratification.

Table 2. Output of the Cox proportional hazards regression model and the logrank test for EACCD grouping on the basis of T, N, M, A, and S on the training set.

Groups Compared	Training Set			Validation Set		
	Hazard Ratio (95% CI)	P-value for Hazard Ratio	P-value for Logrank Test	Hazard Ratio (95% CI)	P-value for Hazard Ratio	P-value for Logrank Test
2 vs 1	2.52 (2.26,2.81)	6.9×10^{-62}	3.0×10^{-66}	2.89 (1.87,4.47)	1.6×10^{-6}	5.0×10^{-7}
3 vs 2	1.60 (1.45,1.76)	8.5×10^{-21}	3.8×10^{-21}	1.66 (1.17,2.36)	0.0046	0.0042
4 vs 3	1.53 (1.37,1.71)	6.4×10^{-14}	4.2×10^{-14}	1.38 (0.92,2.06)	0.12	0.12
5 vs 4	1.30 (1.16,1.46)	6.4×10^{-6}	6.0×10^{-6}	1.42 (0.95,2.13)	0.089	0.088
6 vs 5	1.29 (1.16,1.44)	3.5×10^{-6}	3.3×10^{-6}	1.48 (1.07, 2.05)	0.019	0.018
7 vs 6	1.37 (1.22,1.53)	4.3×10^{-8}	3.8×10^{-8}	1.10 (0.77,1.58)	0.60	0.60
8 vs 7	1.45 (1.31,1.61)	1.8×10^{-12}	1.4×10^{-12}	2.01 (1.43,2.83)	5.9×10^{-5}	4.2×10^{-5}
9 vs 8	1.66 (1.50,1.85)	8.6×10^{-21}	3.6×10^{-21}	1.35 (0.98,1.85)	0.062	0.062
10 vs 9	1.59 (1.40,1.82)	1.9×10^{-12}	1.3×10^{-12}	1.44 (0.96,2.17)	0.080	0.079

The p-value for the hazard ratio is from the Wald test testing the null hypothesis that the hazard ratio equals 1. The p-value for the logrank test is from the logrank test testing the null hypothesis that two survival curves are equal.

<https://doi.org/10.1371/journal.pone.0257949.t002>

The 8th edition of AJCC divides the training set into 10 stages (The 5th column of S2 Table and Fig 3). Calculation shows that the AJCC staging system has a C-index of 0.7643. The p-value of the non-parametric C-index based test [34] for testing the difference (0.0039, 95% CI = (0.0032, 0.0047)) between the prediction accuracy of the above EACCD prognostic system (10 groups, C-index = 0.7682) and the AJCC staging system TNM (10 stages, C-index = 0.7643) is 7.2×10^{-23} , showing that the former has a significantly higher survival prediction accuracy than the latter.

Table 3 presents the distribution of patients for each of the 10 AJCC stages over the 10 EACCD groups. Roughly, it is seen that a higher stage in AJCC corresponds to a higher-risk group in EACCD, and vice versa. Indeed the assignment to ordered stages (from the smallest to the biggest risk) and the assignment to ordered prognostic groups have a large Spearman’s rank correlation coefficient [35] of 0.8316 with a p-value of 4.5×10^{-13} . Therefore, there is a strong positive association between AJCC staging and EACCD grouping.

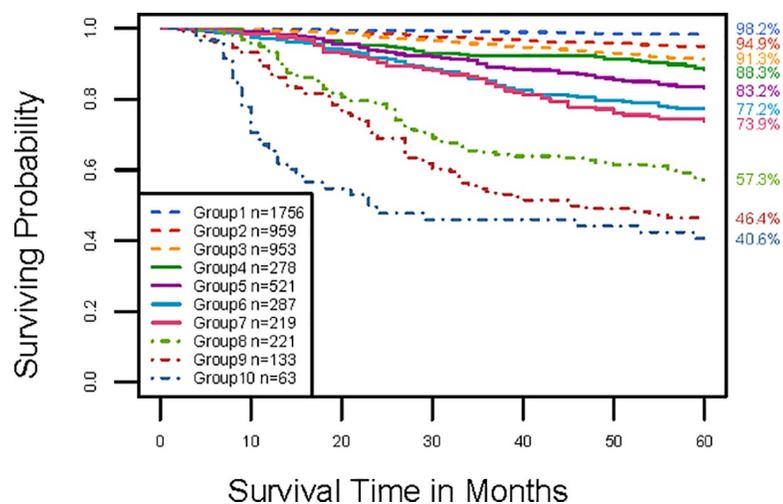


Fig 6. Melanoma-specific survival of the 10 prognostic groups from the validation set. These groups were obtained by applying to the validation set the assignment rule of the EACCD system based on the training set. The 5-year melanoma-specific survival rates for 10 groups are listed on the right side of the figure.

<https://doi.org/10.1371/journal.pone.0257949.g006>

Table 3. Contingency table between EACCD grouping and AJCC staging on the basis of T, N, M on the training set.

AJCC \ EACCD	1	2	3	4	5	6	7	8	9	10	Total
IA	14,588	0	0	0	0	0	0	0	0	0	14,588
IB	0	9,734	0	0	0	0	0	0	0	0	9,734
IIA	0	0	5,592	0	0	0	0	0	0	0	5,592
IIB	0	0	0	3,611	0	0	0	0	0	0	3,611
IIC	0	0	0	0	0	1,470	0	0	0	0	1,470
IIIA	0	0	34	1,439	0	43	0	0	0	0	1,516
IIIB	0	0	0	0	864	274	65	0	0	0	1,203
IIIC	0	0	0	55	37	741	896	133	768	0	2,630
IIID	0	0	0	0	0	0	0	0	395	0	395
IV	0	0	0	0	0	0	0	0	0	42	42
Total	14,588	9,734	5,626	5,105	901	2,528	961	133	1,163	42	40,781

<https://doi.org/10.1371/journal.pone.0257949.t003>

However, survival curves of different AJCC stages often overlap, which can cause confusion in treatment planning. For instance, stages IIID and IV have overlapping survival curves (Fig 3 and S2 Fig). In addition, due to heterogeneity, higher stage groups can have higher survival rates, which is counterintuitive. For example, stage IIIA has a more favorable survival curve than stage IIB, and stage IIIB has a more favorable survival curve than stage IIC. These drawbacks were also reported by Abdel-Rahman et al. [36], though they applied a different data management process to the SEER data. We must emphasize that the EACCD prognostic groups rarely have these problems—the survival curves are well separated and the index of prognostic groups preserves the risk order. Table 3 details how this happens.

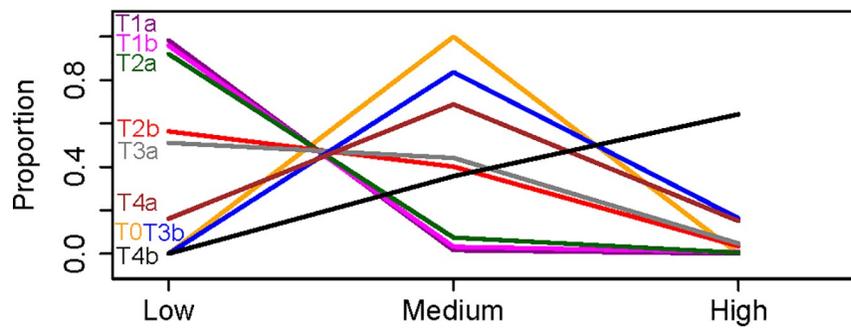
In fact, Table 3 shows whether cases of an AJCC stage should be divided according to prognostic groups included in the EACCD. Horizontally, we can examine all stages line by line. First, we observe that no splitting for stage IA and all 14,588 cases are assigned to group 1. Then we observe all cases included in stage IB are assigned to prognostic group 2. We continue until reaching the last line where stage IV equals prognostic group 10. The key observation is that each of stages IIIA, IIIB, and IIIC is split into cases of multiple prognostic groups. For instance, cases of stage IIIC are decomposed into those in prognostic groups 4–9. We note that based on its design, EACCD grouping represents an assignment of cases so that cases within a group are more homogeneous in survival and cases from different groups differ in survival. Therefore, the above observation may indicate that each of stages IIIA, IIIB, and IIIC contains more heterogeneous cases than other stages. This is consistent with the AJCC definition of stages IIIA, IIIB, and IIIC [3].

In summary, in predicting survival, the EACCD prognostic system on {T, N, M} has a significantly higher accuracy than the AJCC staging system TNM. In stratifying patients, the EACCD grouping produces well-separated survival curves while AJCC staging does not, but EACCD grouping and AJCC staging are strongly positively associated.

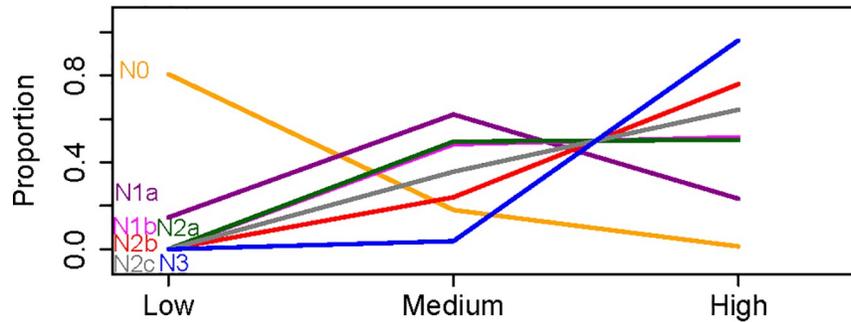
Effect of factor levels on survival

The EACCD prognostic system on {T, N, M, A, S} represents an expansion to TNM. Compared with the EACCD system on {T, N, M}, this expanded system further improved the prediction accuracy of the AJCC staging system (C-index = 0.7865 vs 0.7643; increase in C-index = 0.0222, 95% CI = 0.0191, 0.0254); p-value = 8.8×10^{-43}).

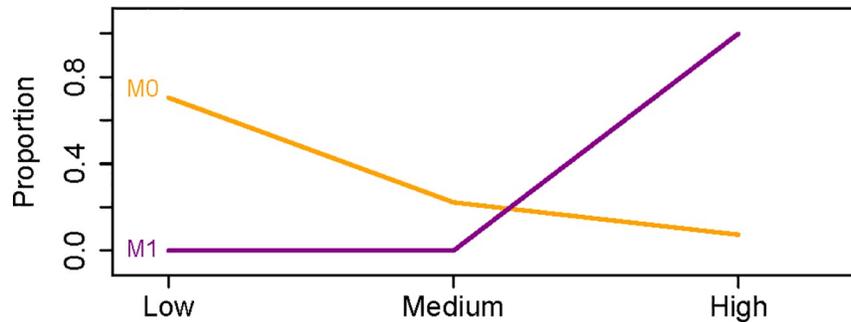
Below we applied this expanded system to examine the effect of levels of individual factors on survival. To simplify the analysis, we consider the following three risk categories by combining 10 prognostic groups from the system: low risk (groups 1–3), medium risk (groups



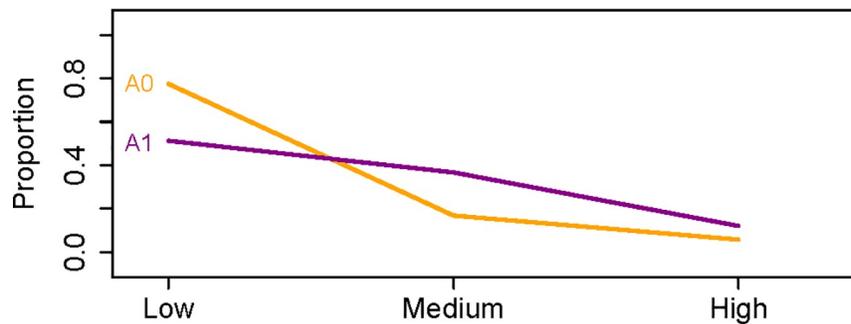
	Low	Medium	High
<i>Primary Tumor (T)</i>			
T0	0.0%	100.0%	0.0%
T1a	98.4%	1.6%	0.0%
T1b	95.9%	3.2%	0.9%
T2a	92.1%	7.5%	0.5%
T2b	56.4%	40.1%	3.5%
T3a	51.1%	44.2%	4.7%
T3b	0.0%	83.7%	16.3%
T4a	16.1%	68.8%	15.1%
T4b	0.0%	35.7%	64.3%



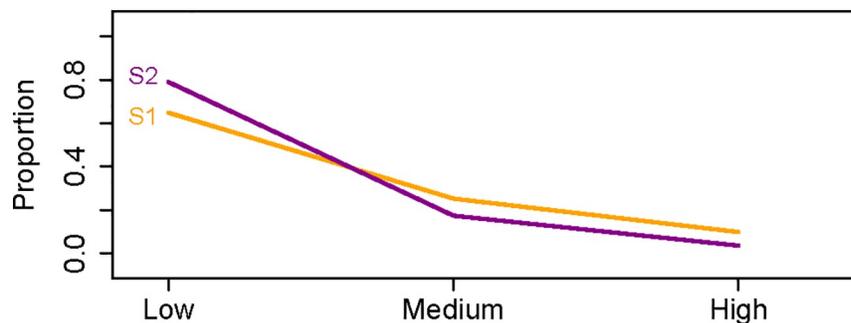
	Low	Medium	High
<i>Regional Lymph Nodes (N)</i>			
N0	80.7%	18.1%	1.3%
N1a	14.7%	62.1%	23.3%
N1b	0.0%	48.3%	51.7%
N2a	0.0%	49.6%	50.4%
N2b	0.0%	23.9%	76.1%
N2c	0.0%	35.7%	64.3%
N3	0.0%	3.7%	96.3%



	Low	Medium	High
<i>Distant Metastasis (M)</i>			
M0	70.5%	22.2%	7.3%
M1	0.0%	0.0%	100.0%



	Low	Medium	High
<i>Age (A)</i>			
A0	77.5%	16.8%	5.7%
A1	51.3%	36.7%	12.0%



	Low	Medium	High
<i>Sex (S)</i>			
S1	64.9%	25.3%	9.9%
S2	79.0%	17.4%	3.6%

Fig 7. Distributions of patients over risk categories on the training set. In each panel, one factor is concerned, and for each level of the factor, the distribution of patients (3 proportions at 3 risk categories) is presented in two ways: plot on the left and tabulation on the right.

<https://doi.org/10.1371/journal.pone.0257949.g007>

4–7), and high risk (groups 8–10). Fig 7 shows how patients associated with a factor level are distributed across these three risk categories.

The 1st panel of Fig 7 shows the distribution of patients associated with different T levels. T1a, T1b, and T2a curves peak at low risk with a near 100% proportion, indicating that patients associated with T1a, T1b, and T2a are almost exclusively distributed in groups with low risk. The T2b and T3a curves also peak at low risk. Patients with these two levels are more likely to appear in low-risk groups, but there is a reasonably big proportion (about 40%) of them classified into middle-risk groups. Patients with T0, T3b, and T4a are mostly distributed in middle-risk groups. We note that T0 is not a highly optimistic level as what we would typically imagine. The reason is that patients with T0 are always associated with N levels higher than N0. The indication of a medium risk is also reflected in the AJCC staging system where T0 patients are assigned to either stage IIIB or IIIC. T4b is the level with the worst prognosis and patients with T4b are mostly distributed in high-risk groups. The above indicates an increased severity of the disease in terms of the order {T1a, T1b, T2a}, {T2b, T3a}, {T0, T3b, T4a}, T4b.

The 2nd panel shows that: 1) N0 patients are most likely to have a low risk; 2) N1a patients are more likely to assume a medium risk; 3) N1b and N2a patients are most likely at either a medium or high risk; and 4) N2b, N2c, and N3 patients are most likely to have a high risk. These indicate the increased severity of the disease in terms of the order N0, N1a, {N1b, N2a}, {N2b, N2c, N3}.

The 3rd panel reveals that M0 patients have a decreasing distribution from low to high-risk groups, whereas the entire M1 patients are distributed in the high-risk group only. This clearly suggests that patients with M1 have a poor prognosis.

The 4th and 5th panels clearly show that A0 and S2 have a better prognosis than A1 and S1, respectively.

The above analysis shows how a factor level is associated with risk. Although these findings are known from the literature, this is the first time these factor levels are integrated together and described explicitly in ordered risk groups of the prognostic system TNMAS created in this paper.

Relationship to stratification and prediction

It is seen that an EACCD prognostic system is derived by partitioning, according to survival, patients into a minimum number of groups that yields approximately the maximum value of the predictive accuracy, i.e., C-index. A prognostic system is neither a model only targeting at the “best” stratification nor a model only aiming at the “highest” predictive accuracy. A prognostic system is a model that shows a tradeoff between stratification and prediction. It is this balance that possesses unique utilizations in clinical settings such as staging patients, offering treatment, and designing trials, etc.

Limitations of analyses

Melanoma-specific survival data were used in this study. Although the SEER cause-specific death classification is determined according to various features (e.g., tumor sequence, site of the original cancer diagnosis, and comorbidities), death certificate errors could introduce a bias to the estimation of cause-specific survival. Another limitation is that the EACCD requires

a relatively large data set in order to obtain robust estimates of survival. This report includes combinations with a minimum of 25 cases, which may exclude some “rare” but interesting combinations. The impact of this requirement on the size of combinations will be minimized as more data become available. In addition, the absence of N1c subcategory in our study could affect, to some extent, the comparison between our system and the AJCC system ([S1 Data](#)).

Conclusions

We used SEER data and the EACCD to build two prognostic systems to demonstrate how machine learning techniques can be used to improve prognostic assessment and risk stratification for cutaneous melanoma. The basic system, on the basis of T, N, and M, classifies patients into prognostic groups that are strongly positively correlated with the AJCC TNM stages but have a higher prediction accuracy in survival. Survival curves of the prognostic system from EACCD do not overlap and are reasonably ordered, which is in contrast to the AJCC TNM system where different stages can present overlapping survival. The second system, incorporating age and sex as additional prognostic factors, further improves prediction accuracy. The prognostic systems from EACCD serve the same role as the TNM staging system and are created through optimizing both prediction and stratification, which can impact clinical and treatment decisions.

As our understanding of melanoma and other cancers improves, more factors will come to the forefront of clinical significance. These additional factors will be very difficult to manually incorporate into conventional TNM staging systems. Our approach to staging allows incorporation of unlimited additional factors to create nuanced staging systems that will provide real-time prognosis information. When sufficient data become available, other variables/factors, such as treatment and immune checkpoint inhibitors, could be readily integrated into known systems by the EACCD to generate prognostic systems for refinements in stratification and outcome prediction that are necessary for patient care, such as monitoring large therapeutic trials.

Supporting information

S1 Data. Supplementary data.

(DOCX)

S1 Fig. [Fig 2D](#) with the number of patients at risk at different times.

(TIF)

S2 Fig. [Fig 3](#) with the number of patients at risk at different times.

(TIF)

S3 Fig. [Fig 5C](#) with the number of patients at risk at different times.

(TIF)

S4 Fig. [Fig 6](#) with the number of patients at risk at different times.

(TIF)

S1 Table. Definitions of levels of T, N, M, A, and S for SEER melanoma of the skin. Refer to AJCC Cancer Staging Manual [1] and SEER Research Data Record Description [2] for specifics of the 3rd column.

(DOCX)

S2 Table. EACCD and AJCC grouping of melanoma of the skin patients according to T, N, and M.

(DOCX)

S3 Table. EACCD grouping of melanoma of the skin patients according to T, N, M, A, and S.

(DOCX)

Acknowledgments

Dr. Donald E. Henson passed away before the submission of the final version of this manuscript. Dr. Dechang Chen accepts responsibility for the integrity and validity of the data collected and analyzed.

Disclaimer The contents, views or opinions expressed in this publication or presentation are those of the authors and do not necessarily reflect official policy or position of Uniformed Services University of the Health Sciences, the Department of Defense (DoD), or Departments of the Army, Navy, or Air Force. Mention of trade names, commercial products, or organizations does not imply endorsement by the U.S. Government.

Author Contributions

Conceptualization: Charles Q. Yang, Matthew T. Hueman, Dechang Chen.

Data curation: Charles Q. Yang, Huan Wang, Aadya Bhaskaran, Donald E. Henson.

Formal analysis: Huan Wang, Dechang Chen.

Funding acquisition: Zhenqiu Liu, Matthew T. Hueman, Dechang Chen.

Methodology: Huan Wang, Zhenqiu Liu, Dechang Chen.

Validation: Zhenqiu Liu, Li Sheng.

Writing – original draft: Charles Q. Yang, Huan Wang, Aadya Bhaskaran, Dechang Chen.

Writing – review & editing: Charles Q. Yang, Huan Wang, Zhenqiu Liu, Matthew T. Hueman, Aadya Bhaskaran, Donald E. Henson, Li Sheng, Dechang Chen.

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. *CA Cancer J Clin.* 2019; Jan; 69(1):7–34. <https://doi.org/10.3322/caac.21551> PMID: 30620402
2. Homs J, Kashani-Sabet M, Messina JL, Daud A. Cutaneous melanoma: prognostic factors. *Cancer Control.* 2005 Oct; 12(4):223–9. <https://doi.org/10.1177/107327480501200403> PMID: 16258493
3. Amin MB, Edge S, Greene F, Byrd DR, Brookland RK, Washington, MK, et al. *AJCC Cancer Staging Manual*, 8th ed. Springer International Publishing. 2017.
4. Beahrs O, Carr D, Rubin P. *Manual for staging of cancer.* Chicago, IL: American Joint Committee. 1977. <https://cancerstaging.org/references-tools/deskreferences/Documents/AJCC1stEdCancerStagingManual.pdf> Accessed 20 June 2021.
5. Gershenwald JE, Thompson W, Mansfield PF, Lee JE, Colome MI, Tseng CH, et al. Multi-institutional melanoma lymphatic mapping experience: the prognostic value of sentinel lymph node status in 612 stage I or II melanoma patients. *J Clin Oncol.* 1999 Mar; 17(3):976–83. <https://doi.org/10.1200/JCO.1999.17.3.976> PMID: 10071292
6. Albittar AA, Alhalabi O, Glitza Oliva IC. Immunotherapy for Melanoma. *Adv Exp Med Biol.* 2020; 1244:51–68. https://doi.org/10.1007/978-3-030-41008-7_3 PMID: 32301010.
7. Maurichi A, Miceli R, Camerini T, Mariani L, Patuzzo R, Ruggeri R, et al. Prediction of survival in patients with thin melanoma: results from a multi-institution study. *J Clin Oncol.* 2014; 32(23):2479–85. <https://doi.org/10.1200/JCO.2013.54.2340> PMID: 25002727

8. Garbe C, Büttner P, Bertz J, Burg G, d'Hoedt B, Drepper H, et al. Primary cutaneous melanoma. Identification of prognostic groups and estimation of individual prognosis for 5093 patients. *Cancer*. 1995 May 15; 75(10):2484–91. [https://doi.org/10.1002/1097-0142\(19950515\)75:10<2484::aid-cncr2820751014>3.0.co;2-u](https://doi.org/10.1002/1097-0142(19950515)75:10<2484::aid-cncr2820751014>3.0.co;2-u) PMID: 7736392
9. Averbook BJ, Fu P, Rao JS, Mansour EG. A long-term analysis of 1018 patients with melanoma by classic Cox regression and tree-structured survival analysis at a major referral center: implications on the future of cancer staging. *Surgery*. 2002 Oct 1; 132(4):589–604. <https://doi.org/10.1067/msy.2002.127546> PMID: 12407342
10. Liu H, Xu YB, Guo CC, Li MX, Ji JL, Dong RR, et al. Predictive value of a nomogram for melanomas with brain metastases at initial diagnosis. *Cancer Med*. 2019 Dec; 8(18):7577–85. <https://doi.org/10.1002/cam4.2644> PMID: 31657530
11. Wang H, Li G. A selective review on random survival forests for high dimensional data. *Quant Biosci*. 2017; 36(2):85. <https://doi.org/10.22283/qbs.2017.36.2.85> PMID: 30740388
12. Chen D, Xing K, Henson D, Sheng L, Schwartz AM, Cheng X. Developing prognostic systems of cancer patients by ensemble clustering. *Biomed Res Int*. 2009 Jun 23; 2009. <https://doi.org/10.1155/2009/632786> PMID: 19584918
13. Qi R, Wu D, Sheng L, Henson D, Schwartz A, Xu E, et al. On an ensemble algorithm for clustering cancer patient data. *BMC Syst Biol*. 2013 Oct; 7(4):S9.
14. Chen D, Hueman MT, Henson DE, Schwartz AM. An algorithm for expanding the TNM staging system. *Future Oncol*. 2016 Apr; 12(8):1015–24. <https://doi.org/10.2217/fon.16.5> PMID: 26904925
15. Wang H, Chen D, Hueman MT, Sheng L, Henson DE. Clustering big cancer data by effect sizes. In 2017 IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE). 17 July 2017; pp. 58–63. <https://ieeexplore.ieee.org/abstract/document/8010615> Accessed 20 June 2021.
16. Hueman MT, Wang H, Yang CQ, Sheng L, Henson DE, Schwartz AM, et al. Creating prognostic systems for cancer patients: A demonstration using breast cancer. *Cancer Med*. 2018 Aug; 7(8):3611–21. <https://doi.org/10.1002/cam4.1629> PMID: 29968970
17. Wang H, Hueman M, Pan Q, et al. Creating Prognostic Systems by the Mann-Whitney Parameter. In 2018 IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE) 26 September 2018: 33–39. <https://ieeexplore.ieee.org/abstract/document/8648686> Accessed 20 June 2021.
18. Yang C, Gardiner L, Wang H, Hueman M, Chen D. Creating prognostic systems for well differentiated thyroid cancer using machine learning. *Front Endocrinol*. 2019; 10:288. <https://doi.org/10.3389/fendo.2019.00288> PMID: 31139148
19. Hueman M, Wang H, Henson D, Chen D. Expanding the TNM for cancers of the colon and rectum using machine learning: a demonstration. *ESMO Open*. 2019 Jun 1; 4(3):e000518. <https://doi.org/10.1136/esmoopen-2019-000518> PMID: 31275615
20. Wang H. Development of Prognostic Systems for Cancer Patients. The George Washington University; 2020. Doctoral dissertation. <https://search.proquest.com/openview/3addd97db64a77ddab4b1ceca94b44ec/1?pq-origsite=gscholar&cbl=18750&diss=y> Accessed 20 June 2021, 2020.
21. Grimley PM, Liu Z, Darcy KM, Hueman MT, Wang H, Sheng L, et al. A prognostic system for epithelial ovarian carcinomas using machine learning. *Acta Obstet Gynecol Scand*. 2021 Mar 4. (Accepted) <https://doi.org/10.1111/aogs.14137> PMID: 33665831
22. Hueman M, Wang H, Liu Z, Henson D, Nguyen C, Park D, et al. Expanding TNM for lung cancer through machine learning. *Thoracic cancer*. 2021 May; 12(9):1423–30. <https://doi.org/10.1111/1759-7714.13926> PMID: 33713568
23. Surveillance, Epidemiology, and End Results (SEER) Program (www.seer.cancer.gov) Research Data (2000–2018), National Cancer Institute, DCCPS, Surveillance Research Program, released April 2021, based on the November 2020 submission. Available from: <https://seer.cancer.gov/> (cited 20 June 2021).
24. SEER Cause-specific Death Classification. Available from URL: <https://seer.cancer.gov/causespecific/> (accessed 20 June 2021).
25. CS Lymph Nodes Eval. Available from URL: [https://staging.seer.cancer.gov/cs/input/02.05.50/melanoma_skin/nodes_eval/?breadcrumbs=\(~schema_list-\),\(~view_schema-,~melanoma_skin-\)](https://staging.seer.cancer.gov/cs/input/02.05.50/melanoma_skin/nodes_eval/?breadcrumbs=(~schema_list-),(~view_schema-,~melanoma_skin-)) (accessed 20 June 2021).
26. Balch CM, Thompson JF, Gershenwald JE, Soong SJ, Ding S, McMasters KM, et al. Age as a predictor of sentinel node metastasis among patients with localized melanoma: an inverse correlation of melanoma mortality and incidence of sentinel node metastasis among young and old patients. *Ann Surg Oncol*. 2014 Apr; 21(4):1075–81. <https://doi.org/10.1245/s10434-013-3464-x> PMID: 24531700

27. Ramspek CL, Jager KJ, Dekker FW, Zoccali C, van Diepen M. External validation of prognostic models: what, why, how, when and where?. *Clinical Kidney Journal*. 2021 Jan; 14(1):49–58. <https://doi.org/10.1093/ckj/sfaa188> PMID: 33564405
28. Kaufman L, Rousseeuw PJ. Finding groups in data: an introduction to cluster analysis. Hoboken, New Jersey: John Wiley & Sons; 1990.
29. Bien J, Tibshirani R. Hierarchical clustering with prototypes via minimax linkage. *J Am Stat Assoc*. 2011 Sep 1; 106(495):1075–84. <https://doi.org/10.1198/jasa.2011.tm10183> PMID: 26257451
30. Mann HB, Whitney DR. On a test of whether one of two random variables is stochastically larger than the other. *Ann Math Stat*. 1947; 18:50–60.
31. Efron B. The two sample problem with censored data. In Lucien M., Cam Le and Jerzy N (eds) Fifth Berkeley Symposium on Mathematical Statistics and Probability, Statistical Laboratory of the University of California, Berkeley, June 21–18 July 1965 and 27 December 1965–7 January 1966, p.666. Berkeley, Calif: University of California Press 1967.
32. Brentnall AR, Cuzick J. Use of the concordance index for predictors of censored survival data. *Stat Methods Med Res*. 2018; 27(8): 2359–73. <https://doi.org/10.1177/0962280216680245> PMID: 27920368
33. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *J Am Stat Assoc*. 1958 Jun 1; 53(282):457–81.
34. Kang L, Chen W, Petrick NA, Gallas BD. Comparing two correlated C indices with right-censored survival outcome: a one-shot nonparametric approach. *Stat Med*. 2015 Feb 20; 34(4):685–703. <https://doi.org/10.1002/sim.6370> PMID: 25399736
35. Daniel WW. Biostatistics: a foundation for analysis in the health sciences. 7th ed. New York: John Wiley & Sons; 1999.
36. Abdel-Rahman O. Evaluation of the eighth American Joint Committee on Cancer staging system for malignant melanoma of the skin. *Future Oncol*. 2018 Feb; 14(5):471–81. <https://doi.org/10.2217/fon-2017-0500> PMID: 29322826