Nomogram for predicting survival in patients with pancreatic cancer

Wei Song
Dong-Liu Miao
Lei Chen

Department of Intervention and Vascular Surgery, Affiliated Suzhou Hospital of Nanjing Medical University, Suzhou Municipal Hospital, Suzhou Cancer Medical Center, Suzhou, People’s Republic of China

Background: The purpose of this study was to develop a nomogram to predict cancer-specific survival (CSS) in pancreatic cancer (PC).

Patients and methods: We used the Surveillance, Epidemiology, and End Results (SEER) database to analyze 53,028 patients diagnosed with PC from 2004 to 2014 and randomly divided them into the training (n=26,583) cohort and validation (n=26,445) cohort. Univariate and multivariate analyses were used to select independent prognostic factors. We used significant prognostic factors for constructing a nomogram based on Cox regression analyses. Validation of the nomogram was assessed by discrimination and calibration.

Results: According to the multivariate models of training cohort, a nomogram that combined age, race, tumor location, marital status, tumor size, TNM stage, tumor grade, and surgery was constructed for predicting CSS. The internally validated and externally validated C-indexes were 0.741 and 0.734, respectively. The calibration curves showed that the nomogram was able to predict 1-, 3-, and 5-year CSS accurately.

Conclusion: A nomogram effectively predicts survival in patients with PC. This prognostic model may be considered for use in clinical practice.

Keywords: pancreatic cancer, nomogram, SEER, cancer-specific survival, prognosis

Introduction

Pancreatic cancer (PC) is known as the most intractable type of cancer and the fourth leading cause of cancer-related death worldwide. Radical resection remains the only established curative treatment for PC, but the nonspecific symptoms and high recurrence rate after curative resection are a major problem. Moreover, it is usually diagnosed at an advanced stage; most of the patients can no longer be considered candidates for curative resection. Despite significant recent developments in surgical techniques and adjuvant therapy, the overall prognosis of PC remains poor. Data from the Surveillance, Epidemiology, and End Results (SEER) program (2006–2012) demonstrated that the 5-year survival of patients with PC is 7.7%. Therefore, accurate estimates of prognosis of PC patients based on clinicopathologic factors could help clinicians implement better therapeutic strategies.

Presently, prognostic predictions and treatment strategies for PC patients are based on the American Joint Committee on Cancer (AJCC) TNM staging system. This classification system assessing PC based on the depth of invasion, number of metastasis nodes, and the status of distant metastasis has been widely used to predict the survival of patients with PC. However, survival may be different even in patients with the same AJCC stage. In fact, other patient-specific factors such as age, race, marital status, tumor size, and differentiation are associated with survival in multiple cancers.
Therefore, a more refined staging system that combines the tumor characteristics and host status is needed.

Nomograms have been accepted as a reliable alternative tool that can help clinicians make easy individual predictions.\textsuperscript{8–10} Compared to the AJCC TNM staging system, the individual survival rate can be estimated more accurately by incorporating clinically important variables.\textsuperscript{11,12} Nevertheless, nomograms for predicting the survival of patients with PC have not yet been fully developed.

In the present study, we aimed to develop a prognostic nomogram based on large population data from the SEER cancer registry program to better predict individualized survival in patients with PC.

Patients and methods

Patient population and study design

The SEER program of the US National Cancer Institute (NCI) was used as the data source for the present population-based investigation. The SEER program captures \textasciitilde97% of incident cancers, and the 17 SEER tumor registries encompass \textasciitilde28% of the US population.\textsuperscript{13} SEER program collects information on cancer incidence, prevalence, survival, and mortality of patients with cancer.

We used the SEER program to identify patients who were diagnosed with PC from 2004 to 2014. A total of 78,665 patients were initially screened. The criteria for inclusion were listed as follows: 1) no history of malignancy; 2) the diagnosis with PC as the first and only cancer diagnosis; 3) active follow-up with complete date and known outcome; and 4) reporting clinicopathological information (age, race, sex, tumor location, tumor size, marital status, TNM stage, tumor grade, and therapy). Patients were excluded if age at diagnosis was <18 years, had unknown survival time, and had multiple primary cancers. The eligible patients were randomly divided into a training (n=26,583) cohort and a validation (n=26,445) cohort.

This study was based on public data from the SEER database; we obtained permission to access research data files with the reference number 10091-Nov 2016. This study was approved by the ethics committee of the Suzhou Municipal Hospital. The data did not include the use of human subjects or personal identifying information. Thus, no informed consent was required for this part of the study.

Study variables

There were several controlled study variables, including demographics (sex, age, and race), TNM stage, tumor location (pancreatic head, body, and tail), tumor size, tumor grade (well-differentiated, moderately differentiated, poorly differentiated, undifferentiated, not differentiated/unknown), histologic type, treatment, and marital status at the time of diagnosis. The primary end point was cancer-specific survival (CSS), which was calculated from the date of diagnosis to the date of death from PC. Deaths attributed to PC were treated as events, while deaths from other causes were treated as censored observations. Widowed, single, and separated/divorced patients were classified as unmarried. Cancer stages were based on the sixth edition of the AJCC/TNM staging system. Meanwhile, as the sixth edition of the staging system was published in 2004, we limited our study to between 2004 and 2014.

Statistical analyses

Construction of the nomogram

Baseline patient demographics and disease features were compared using the Student’s t-test or the chi-square test, as appropriate. The Kaplan–Meier method was applied to assess survivor functions. The log-rank test was used to test differences between survival curves. Cox proportional hazards multivariable regression was used to assess the independent effects of the univariate prognostic factors on CSS. A graphical nomogram, derived from the multivariate logistic regression model, was constructed using the R statistical package rms (R Foundation for Statistical Computing, Vienna, Austria).

Validation of the nomogram

Validations were conducted both internally (training cohort) and externally (validation cohort) by discrimination and calibration and using bootstrap resampling (1,000 resamples). Discrimination between survival probability and actual observations was evaluated using the C-index. The value of the C-index fluctuated between 0.5 and 1.0, with 0.5 representing random chance and 1.0 representing a totally corrected discrimination.\textsuperscript{14} We constructed a calibration plot to determine whether the predicted survival and actual survival were in concordance. All statistical analyses were performed by the statistical software package SPSS for Windows, version 23 (IBM Corporation, Armonk, NY, USA) and the R software version 3.13 (http://www.r-project.org/). All P-values were two sided, and P-values \textless0.05 were considered as statistically significant.

Results

Patient characteristics

The present study identified 53,028 eligible PC patients between 2004 and 2014, including 26,583 patients in the
training cohort and 26,445 patients in the validation cohort. The flow diagram of the study selection process is shown in Figure 1. Of these patients, 27,027 (51.0%) were male and 26,001 (49.0%) were female. The most common tumor location was pancreatic head (65.2%). The majority of patients in both cohorts were elderly patients (>60 years), were married, had median tumor size (3–5 cm), and had IV stage tumor. In both cohorts, most patients did not receive surgery. Patient demographics and pathological characteristics are listed in Table 1.

**Factors associated with CSS**

Data on sex, age, race, marital status, tumor location, tumor size, tumor grade, TNM stage, and surgery were collected in the training cohort. These variables other than sex were identified as significant risk factors for poor survival in univariate analysis (Table 2). When performing multivariate analysis with Cox regression, seven factors were identified as independent prognostic factors, including age, tumor location, tumor size, marital status, TNM stage, tumor grade, and surgery. These variables were then incorporated into the nomogram.

**Nomogram**

Based on the reduced multivariate models of the training cohort, a nomogram that combined all the important independent factors was constructed for predicting 1-, 3-, and 5-year CSS (Figure 2). This model demonstrated that the tumor grade contributed most to prognosis, followed by the surgery, TNM stage, age, tumor size, marital status, and tumor location. Every factor was given a score on the points scale. By summing and locating the scores on the total score scale, we could predict the possibility of 1-, 3-, and 5-year survival.

**Nomogram validation**

The internal validation demonstrated that the nomogram can accurately predict the CSS with a C-index of 0.741. Similarly, the C-index was 0.734 in the external validation. The calibration plots showed an excellent agreement between the predicted and observed values for the 1-, 3-, and 5-year CSS in not only the training cohort but also the validation cohort (Figures 3 and 4).

---

**Table 1 Patient demographics and pathological characteristics**

| Variables                           | All patients (N=53,028) | Training cohort (n=26,583) | Validation cohort (n=26,445) |
|-------------------------------------|-------------------------|---------------------------|-----------------------------|
| **Sex**                             |                         |                           |                             |
| Male                                | 27,027 (51.0)           | 13,627 (51.3)             | 13,400 (50.7)               |
| Female                              | 26,001 (49.0)           | 12,956 (48.7)             | 13,045 (49.3)               |
| **Age (years)**                     |                         |                           |                             |
| ≤60                                 | 14,925 (28.1)           | 7,537 (28.4)              | 7,388 (27.9)                |
| >60                                 | 38,103 (71.9)           | 19,046 (71.6)             | 19,057 (72.1)               |
| **Race**                            |                         |                           |                             |
| White                               | 42,072 (79.3)           | 21,171 (79.6)             | 20,901 (79.0)               |
| Black                               | 6,608 (12.5)            | 3,273 (12.3)              | 3,335 (12.6)                |
| Othera                              | 4,348 (8.2)             | 2,139 (8.1)               | 2,209 (8.4)                 |
| **Marital status**                  |                         |                           |                             |
| Married                             | 30,188 (56.9)           | 15,205 (57.2)             | 14,983 (56.7)               |
| Unmarried                           | 22,840 (43.1)           | 11,378 (42.8)             | 11,462 (43.3)               |
| **Tumor location in the pancreas**  |                         |                           |                             |
| Head                                | 34,582 (65.2)           | 17,408 (65.5)             | 17,174 (64.9)               |
| Body                                | 8,662 (16.3)            | 4,255 (16.0)              | 4,407 (16.7)                |
| Tail                                | 9,784 (18.5)            | 4,920 (18.5)              | 4,864 (18.4)                |
| **Tumor size (cm)**                 |                         |                           |                             |
| <3                                  | 12,151 (22.9)           | 6,107 (23.0)              | 6,044 (22.9)                |
| 3–5                                 | 22,467 (42.4)           | 11,306 (42.5)             | 11,161 (42.2)               |
| >5                                  | 11,345 (21.4)           | 5,697 (21.4)              | 5,648 (21.4)                |
| Unknown                             | 7,065 (13.3)            | 3,473 (13.1)              | 3,592 (13.6)                |
| **AJCC TNM stage**                  |                         |                           |                             |
| I                                   | 4,431 (8.4)             | 2,282 (8.6)               | 2,149 (8.1)                 |
| II                                  | 15,850 (29.9)           | 7,851 (29.5)              | 7,999 (30.2)                |
| III                                 | 5,031 (9.5)             | 2,590 (9.7)               | 2,441 (9.2)                 |
| IV                                  | 27,716 (52.3)           | 13,860 (52.1)             | 13,856 (52.4)               |
| **Grade**                           |                         |                           |                             |
| I                                   | 2,759 (5.2)             | 1,386 (5.2)               | 1,373 (5.2)                 |
| II                                  | 8,227 (15.5)            | 4,081 (15.4)              | 4,146 (15.7)                |
| III                                 | 7,858 (14.8)            | 3,918 (14.7)              | 3,940 (14.9)                |
| IV                                  | 355 (0.7)               | 173 (0.7)                 | 182 (0.7)                   |
| Unknown                             | 33,829 (63.8)           | 17,025 (64.0)             | 16,804 (63.5)               |
| **Therapy**                         |                         |                           |                             |
| Surgery                             | 11,338 (21.4)           | 5,645 (21.2)              | 5,693 (21.5)                |
| No surgery                          | 41,690 (78.6)           | 20,938 (78.8)             | 20,752 (78.5)               |

---

*Note:* aOther includes American Indian/Alaska native, Asian/Pacific Islander, and unknown.

**Abbreviation:** AJCC, American Joint Committee on Cancer.

---

Figure 1 Flow diagram of the study selection process.
Table 2 Univariate and multivariate analyses of CSS in the training cohort

| Variable                        | Univariate analysis | Multivariate analysis |
|---------------------------------|---------------------|-----------------------|
|                                 | P-value             | HR (95% CI)           | P-value   |
| Sex                             | 0.586               |                       |           |
| Male                            |                     |                       |           |
| Female                          |                     |                       |           |
| Age (years)                     | <0.001              | 1.430 (1.356–1.509)   | <0.001    |
| ≤60                             | Reference           |                       |           |
| >60                             |                     |                       |           |
| Race                            | <0.001              | 1.040 (0.964–1.123)   | 0.313     |
| White                           |                     |                       |           |
| Black                           |                     |                       |           |
| Other                           | <0.001              | 0.981 (0.897–1.072)   | 0.671     |
| Marital status                  | <0.001              | 1.191 (1.132–1.252)   | <0.001    |
| Married                         | Reference           |                       |           |
| Unmarried                       |                     |                       |           |
| Tumor location in the pancreas  | <0.001              |                       |           |
| Head                            | Reference           |                       |           |
| Body                            | 0.834 (0.772–0.900) | <0.001                |           |
| Tail                            | 0.886 (0.823–0.954) | 0.001                 |           |
| Tumor size (cm)                 | <0.001              | 1.144 (1.078–1.214)   | <0.001    |
| <3                              | Reference           |                       |           |
| 3–5                             | 1.275 (1.187–1.369) | <0.001                |           |
| >5                              |                     |                       |           |
| AJCC TNM stage                  | <0.001              |                       |           |
| I                               | Reference           |                       |           |
| II                              | 1.438 (1.306–1.583) | <0.001                |           |
| III                             | 1.560 (1.378–1.767) | <0.001                |           |
| IV                              | 2.226 (2.004–2.472) | <0.001                |           |
| Grade                           | <0.001              |                       |           |
| I                               | Reference           |                       |           |
| II                              | 1.766 (1.624–1.921) | <0.001                |           |
| III                             | 2.322 (2.135–2.526) | <0.001                |           |
| IV                              | 2.039 (1.695–2.454) | <0.001                |           |
| Surgery                         | <0.001              |                       |           |
| Yes                             | Reference           |                       |           |
| No                              | 2.792 (2.613–2.982) | <0.001                |           |

Note: *Other includes American Indian/Alaska native, Asian/Pacific Islander, and unknown.
Abbreviations: HR, hazard ratio; CSS, cancer-specific survival; NL, not included in the multivariate survival analysis; AJCC, American Joint Committee on Cancer.

Discussion

The nomogram, as a statistical tool, can provide the most accurate predictions by a simple graphical presentation. The nomogram is simple, easy to understand, and easy to apply to clinical practice. Moreover, the model can achieve individualized predictions, and thus, clinicians can use the tool to assess patients for their participation in clinical trials. To date, several nomograms have been constructed for predicting prognosis of PC patients. Vernerey et al identified five parameters predicting OS before chemotherapy in 442 locally advanced PC patients. Hamada et al evaluated six parameters predicting survival of nonresectable PC patients. Nevertheless, all of these nomograms were based on very limited cases and variables. Therefore, a nomogram for PC with universal applicability still needs to be further validated. In fact, inclusion of sufficient data can improve the accuracy of the nomogram. Therefore, we developed the nomogram predicting the 1-, 3-, and 5-year CSS for PC based on a larger population in the SEER database.

In order to ensure the predictive accuracy of the nomogram, the Kaplan–Meier method and the Cox proportional hazards regression were used to select the factors used in developing the CSS nomogram. Moreover, C-index and calibration plots were also used to assess the predictive accuracy of the model. All nomogram C-indexes were >0.7, indicating excellent agreement between predicted and actual survival.

Our model is easy to use in comparison with the widely used TNM staging and has the capability to provide quantitative prognosis to individual patients. First, according to the contribution degree of each factor in the regression model, the influence score of each factor is given, and then, the total score of an individual is calculated. Consider, for example, two stage III PC patients: the first patient who is 65-years old is diagnosed with a grade III tumor of 4 cm and the other patient is 55-years old and is diagnosed with a grade II tumor of 2 cm. Using nomograms, the two patients have 1-year CSS probabilities of 52% and 76%, respectively (Table 3). However, according to TNM staging, both of the two patients are categorized with stage III tumor, which shows the same result.

We identified seven clinicopathological characteristics that can predict CSS for patients with PC, including age, tumor location, marital status, tumor size, TNM stage, tumor grade, and surgery, which are in line with previous studies. Several studies demonstrated that age is an important prognostic factor, although the exact mechanism remains unclear. Accumulated evidence showed that Black PC patients were at significantly greater risk of mortality and that Black ethnicity is associated with decreased survival. Our result confirmed that Black patients have the lowest CSS than other patients. Carbohydrate antigen 19-9 (CA 19-9) is widely recognized as a prognostic factor for PC. However, patients with negative red blood cell phenotypes on both Lewis A and B antigens cannot secrete CA 19-9 into their serum. In order to ensure use of the nomogram for the general population, we excluded this variable in the model but included the tumor size, which was positively associated with CA 19-9.

Our study has several merits. Compared with previous PC nomograms, our model was constructed based on a large
Figure 2: Nomogram for predicting 1-, 3-, and 5-year pancreatic CSS.
Abbreviation: CSS, cancer-specific survival.

Figure 3: Internal calibration plot.
Note: (A) 1-year, (B) 3-year, and (C) 5-year CSS nomogram calibration curves.
Abbreviation: CSS, cancer-specific survival.
population-based cohort which could improve the accuracy of the nomogram. Moreover, discrimination and calibration emphasized satisfaction in the presentation and validity of the model.

Furthermore, our nomogram identified nine variables that can be easily obtained. These variables reflect the common status of the patients and disease activity, thus providing clinically relevant information in PC. They also enhance the relevance of the tools developed.

Nevertheless, our study has several limitations. First, the SEER database did not provide data on radiotherapy and chemotherapy, which might result in bias. Moreover, data on several important clinicopathological parameters were not complete, decreasing the number of eligible cases. Second, as this nomogram was based on the SEER database, some potential predictive variables such as pain, albumin, C-reactive protein, neutrophil-to-lymphocyte ratio, and platelet-to-lymphocyte ratio were not included. Third, TNM classification data were not available until 2004. Therefore, we were not able to predict a longer survival time.

**Conclusion**

We developed and validated a prognostic nomogram based on a population-based database predicting survival for patients with PC. This nomogram could help clinicians to calculate an individualized survival prediction and provide more individualized treatment.

**Acknowledgment**

The authors would like to thank the SEER program for providing open access to the database.

**Author contributions**

WS and LC conceived and designed the study. WS and D-LM searched databases and collected the data. WS, D-LM, and LC wrote the manuscript. All authors contributed toward data analysis, drafting and critically revising the paper and agree to be accountable for all aspects of the work.

**Disclosure**

The authors report no conflicts of interest in this work.
