Prognostic Nomogram for pancreatic cancer with lung metastasis: a SEER database-based study

Jiachun Ding, Jiaqiang Ren, Fan Chen, Yangyang Yue, Chao Ren, Xirui Wang, Qingyong Ma, Zheng Wu, Zheng Wang

Abstract
Objectives: To establish Nomogram to predict the overall survival (OS) rate of pancreatic cancer patients with lung metastasis by utilizing the database of the Surveillance, Epidemiology, and End Results (SEER) Program.

Methods: We obtained the data of 363 pancreatic cancer patients with lung metastasis who were diagnosed between 2010 and 2016 from the SEER database. These patients were randomly divided into training (n = 255) and validation (n = 108) cohorts. The Cox proportional hazards regression model was performed to evaluate the prognostic effects of multiple clinicopathologic factors on OS. Significant prognostic factors were combined to build Nomogram. The predictive performance of Nomogram was evaluated via internal (training cohort data) and external validation (validation cohort data) by calculating index of concordance (C-index) and plotting area under curve (AUC) and calibration curves. All data from SEER database have been fully de-identified and may be used without further independent ethics committee approval.

Results: In the training cohort, the results of Cox proportional hazards regression model showed that, tumor location, surgery, chemotherapy and other organ of metastasis were significantly associated with the survival prognosis (P < .05). These factors were used to establish Nomogram. The Nomogram showed good accuracy in predicting OS rate, with C-index of 0.727 [95%CI was (0.689, 0.764)] in internal validation and C-index of 0.738 [95%CI was (0.679, 0.796)] in external validation. All calibration curves showed excellent consistency between prediction by Nomogram and actual observation.

Conclusion: Novel Nomogram for pancreatic cancer patients with lung metastasis was established to predict OS in our study. It has good prognostic significance. And it could provide the clinicians with more accurate and practical predictive tools which can quickly and accurately assess the patients’ survival prognosis individually, and make clinical suggestion for doctors in the follow-up treatment of patients.

Keywords: Nomogram, Pancreatic cancer, SEER, Survival analysis

Introduction
Pancreatic ductal adenocarcinoma is a highly malignant digestive system tumor. The prognosis of patients is very poor, and the 5-year survival rate is less than 8%.[1] Pancreatic cancer is the seventh leading cause of cancer related death worldwide.[2] Due to the lack of specific clinical symptoms in the early stage of the disease, the majority of patients with pancreatic cancer are often in the middle or late stage when they are diagnosed, which results in great difficulties for the clinical treatment and deterioration control. Although surgical resection is the only possible treatment for pancreatic cancer patients at present[4] only approximately 15% to 20% of patients with pancreatic cancer can receive surgical treatment. Some studies have found that the lung is the second most common site of pancreatic cancer metastasis. Nearly 10% of pancreatic cancer patients have lung metastasis at the time of diagnosis.[4] The impact of clinicopathologic factors on the prognosis of pancreatic cancer patients with lung metastasis is currently unclear. Accurate prediction of the survival rate of pancreatic cancer patients with lung metastasis is of great significance for clinical treatment decisions. This study retrospectively analyzed the clinical data of patients with lung metastasis when pancreatic cancer was diagnosed from 2010 to 2016 in the SEER database. The prognostic factors were also discussed, and evidence is provided to support the clinical diagnosis, treatment and prevention of lung metastasis in pancreatic cancer.

Nomograms can include many prognostic factors. The influence of different factors on survival and prognosis was quantified and the results were visualized. Nomograms are widely used to evaluate the prognosis of cancer patients.[5] The aim of this study was to build a Nomogram for predicting the survival and prognosis of pancreatic cancer patients with lung metastasis, and to establish a relatively systematic and perfect evaluation system to predict the overall survival (OS) rate of pancreatic cancer patients with lung metastasis.

Materials and methods
Patients and study design
The data for this study are from the Surveillance, Epidemiology, and End Results (SEER) Program established by the National...
Cancer Institute (NCI). The SEER database was established by the NCI in 1973. It is one of the main sources of cancer data in the United States. It is characterized by a large sample size, wide population coverage and high data accuracy. The data used in this study are composed of 2 categories: incidence rates and demographic data.

SEER database records include patient ID, personal information, primary lesion location, tumor size, treatment plan, cause of death, etc. The database provides high-quality data in support of clinical oncology research. Clinical cases of pancreatic ductal adenocarcinoma diagnosed pathologically during 2010 to 2016 were obtained by SEER stat 8.3.6 software. The ICD codes for pancreatic ductal adenocarcinoma are 8140 and 8500. The survival data such as age, sex, tumor location, histological grade, AJCC 6th edition stage, T stage, N stage, surgical treatment, radiotherapy, chemotherapy and other organs of metastasis were extracted. The inclusion criteria for the cases in this study were as follows: (1) pathological diagnosis of pancreatic ductal adenocarcinoma; (2) diagnosis age ≥ 18 years old; (3) diagnosis year 2010 to 2016; (4) bone, lung, liver, and brain metastases confirmed when diagnosed with pancreatic ductal adenocarcinoma; (5) known tumor type, tumor grade, and clinical stage (AJCC 6th) are known; and (6) complete follow-up information and survival time. The exclusion criteria were as follows: (1) multiple primary cancers; (2) death within one month; (3) lack of treatment information; and (4) non-tumor-related cause of death. A total of 363 patients were included in this study, of whom 255 were randomly categorized into an internal training set, and the remaining 108 were randomly categorized into an external verification set (Table 1).

All data from SEER database have been fully de-identified and may be used without further independent ethics committee approval. This study was a secondary analysis of the existing SEER dataset, and posed no more than minimal risk to patients. Therefore, the requirement to obtain informed consent from patients was waived.

### Statistical analysis

SPSS 18.0 software and X-tile software were used for statistical analysis. In the prognostic analysis, an OS curve was drawn with the Kaplan–Meier method. The difference between factors was calculated with the log-rank test. For continuous variables (such as age), the best cut-off point was obtained by X-tile software. A Cox proportional hazard model was used for multivariate analysis, and the independent factors influencing the prognosis of

### Table 1

Univariate analysis of clinicopathological factors in pancreatic cancer with lung metastases

<table>
<thead>
<tr>
<th>Clinicopathological factors</th>
<th>Case number</th>
<th>3-Month survival rate</th>
<th>6-Month survival rate</th>
<th>1-Year survival rate</th>
<th>Univariate analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (yr)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤71</td>
<td>229</td>
<td>59.3%</td>
<td>44.8%</td>
<td>19.8%</td>
<td>0.475</td>
</tr>
<tr>
<td>&gt;71</td>
<td>134</td>
<td>58.2%</td>
<td>40.3%</td>
<td>12.7%</td>
<td></td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.151</td>
</tr>
<tr>
<td>Male</td>
<td>174</td>
<td>61.5%</td>
<td>44.3%</td>
<td>18.4%</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>189</td>
<td>56.4%</td>
<td>42.1%</td>
<td>16.0%</td>
<td></td>
</tr>
<tr>
<td>Site</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Head</td>
<td>148</td>
<td>64.9%</td>
<td>50.7%</td>
<td>21.6%</td>
<td>10.589</td>
</tr>
<tr>
<td>BodyTail</td>
<td>141</td>
<td>51.8%</td>
<td>34.8%</td>
<td>12.1%</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>74</td>
<td>60.5%</td>
<td>44.0%</td>
<td>17.9%</td>
<td></td>
</tr>
<tr>
<td>Grade</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Well</td>
<td>28</td>
<td>67.9%</td>
<td>53.6%</td>
<td>21.4%</td>
<td>3.429</td>
</tr>
<tr>
<td>Moderate</td>
<td>165</td>
<td>64.7%</td>
<td>46.4%</td>
<td>18.9%</td>
<td></td>
</tr>
<tr>
<td>Poor</td>
<td>165</td>
<td>52.7%</td>
<td>38.8%</td>
<td>14.5%</td>
<td></td>
</tr>
<tr>
<td>Undiffer</td>
<td>5</td>
<td>20.0% (4/5)</td>
<td>20.0% (4/5)</td>
<td>20.0% (4/5)</td>
<td></td>
</tr>
<tr>
<td>T stage</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T0</td>
<td>2</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>2.304</td>
</tr>
<tr>
<td>T1</td>
<td>10</td>
<td>80.0%</td>
<td>50.0%</td>
<td>10.0%</td>
<td></td>
</tr>
<tr>
<td>T2</td>
<td>103</td>
<td>55.3%</td>
<td>41.7%</td>
<td>13.6%</td>
<td></td>
</tr>
<tr>
<td>T3</td>
<td>149</td>
<td>59.5%</td>
<td>44.0%</td>
<td>20.3%</td>
<td></td>
</tr>
<tr>
<td>T4</td>
<td>99</td>
<td>59.6%</td>
<td>42.4%</td>
<td>16.2%</td>
<td></td>
</tr>
<tr>
<td>N stage</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N0</td>
<td>161</td>
<td>60.2%</td>
<td>41.6%</td>
<td>13.0%</td>
<td>0.230</td>
</tr>
<tr>
<td>N1</td>
<td>202</td>
<td>57.8%</td>
<td>44.3%</td>
<td>20.4%</td>
<td></td>
</tr>
<tr>
<td>Surgery</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>36</td>
<td>80.6%</td>
<td>69.4%</td>
<td>47.2%</td>
<td>17.118</td>
</tr>
<tr>
<td>No</td>
<td>327</td>
<td>56.5%</td>
<td>40.2%</td>
<td>13.8%</td>
<td></td>
</tr>
<tr>
<td>Radiation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>33</td>
<td>75.8%</td>
<td>57.6%</td>
<td>15.2%</td>
<td>0.422</td>
</tr>
<tr>
<td>No</td>
<td>330</td>
<td>57.2%</td>
<td>41.7%</td>
<td>17.3%</td>
<td></td>
</tr>
<tr>
<td>Chemotherapy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>244</td>
<td>75.8%</td>
<td>58.2%</td>
<td>23.4%</td>
<td>86.876</td>
</tr>
<tr>
<td>No</td>
<td>119</td>
<td>23.9%</td>
<td>11.9%</td>
<td>4.3%</td>
<td></td>
</tr>
<tr>
<td>OtherMets</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>213</td>
<td>49.3%</td>
<td>33.8%</td>
<td>9.9%</td>
<td>25.502</td>
</tr>
<tr>
<td>No</td>
<td>150</td>
<td>72.5%</td>
<td>56.4%</td>
<td>27.5%</td>
<td></td>
</tr>
</tbody>
</table>
pancreatic cancer patients with lung metastasis were obtained. P < .05 was considered statistically significant.

Using the random sampling function in R software version 4.0.0 for simple random sampling, patients were randomly divided into a training set and validation set. The log-rank test was used to screen the prognostic factors. Then, a Cox proportional risk regression model was used to analyze the data in the training set. A Nomogram was constructed with the “rms” package and the C-index was calculated. At the same time, a calibration curve was drawn. The bootstrap method (self-sampling frequency B = 50) was used for internal verification and external verification in the training set and verification set, respectively.

In an ideal calibration curve, the predicted value is equal to the observed value. The curve will be infinitely close to the ideal 45° diagonal. The C-index, is similar to the area under the receiver operating characteristic (ROC) curve. It is used to evaluate the predictive value of a Nomogram. The minimum value is 0.5 and the maximum value is 1.0. The larger the C-index is, the higher the predictive value.

**Results**

**Inclusion and exclusion of pancreatic cancer patients with lung metastasis**

We identified 57,067 eligible patients with pancreatic ductal adenocarcinoma diagnosed from 2010 to 2016 in the SEER database. Lung metastasis was present in 5921 of the 57,067 (10.3%) patients. Patients who were diagnosed with pancreatic cancer by radiography or laboratory examination markers were excluded. Some patients whose pathological grade remained unknown and who had TNM stage information deletion were also excluded. Patients who died within 30 days after the diagnosis of pancreatic cancer were not included in this study. Some patients without active follow-up were excluded. Some patients with primary tumors in other organs were excluded. Some patients were excluded due to unknown treatment information. Some patients who died of cardiovascular and cerebrovascular diseases, infections or other diseases unrelated to pancreatic cancer were excluded. Finally, 363 pancreatic cancer patients with lung metastasis were included (Fig. 1).

**Analysis of influencing factors of survival and prognosis**

A total of 363 patients were analyzed with the single-factor log-rank test. According to the results, tumor location, surgery, chemotherapy and other organs of metastasis were all related to the survival and prognosis of pancreatic cancer patients with lung metastasis (P < .05) (Table 1; Figs. 2 and 3).

According to the results of single factor analysis and professional conclusions, Cox proportional risk regression analysis was further conducted. The analyzed factors included age, sex, tumor site, histological grade, T stage, N stage, surgical
treatment, radiotherapy, chemotherapy, and other organs of metastasis. As Table 2 shows, surgery, chemotherapy and other organs of metastasis were all significant associated with the prognosis. If did not conduct surgery or chemotherapy and confirmed other organ metastasis, the HR was higher, and the prognosis was worse. The prognosis of tumors in other parts of the pancreas (uncinate process of the pancreas, intraductal pancreatic duct, etc) was better than that of pancreatic head and body tail tumors. In addition, the survival and prognosis of patients with different age, sex, histological grade, T stage, N stage and radiotherapy were not significantly different, which was consistent with the results of univariate analysis.

Nomogram for predicting overall survival

The Nomogram included all statistically significant prognostic factors in the Cox proportional hazards regression model, including tumor location, surgical treatment, chemotherapy, and other organs of metastasis. The prediction results of the 1-year and 2-year OS rates are shown in Figure 4. According to the different classifications of each feature, the score of each item was obtained by projecting upward to the small points. The survival rate of the patients was obtained by projecting downward from the total points. The Nomogram was used according to the different conditions of different patients. Because the survival rate was predicted individually, the Nomogram improved the efficiency and accuracy of prediction.

Verification of the Nomogram

The bootstrap method was used to verify the Nomogram internally and externally. The self sampling frequency was 50. The verification results showed that, the C-index of the training set was 0.727 [95% CI: (0.689, 0.764)], and the C-index of the validation set was 0.738 [95% CI: (0.679, 0.796)]. All of the results had a fine predictive value. The AUC and calibration curves of the 1-year survival rate in the training set and validation set are shown in Figure 5. The calibration curves of the training set and verification set were close to the ideal 45° dotted line. This indicates that there is good consistency between the predicted value and the actual observation value.

Discussion

Distant metastasis of pancreatic cancer indicates that the disease is advanced. The liver is the most frequent organ of metastasis in patients with pancreatic cancer. At present, a large number of studies have been conducted on liver metastasis of pancreatic cancer,[8–11] and lung metastasis is gradually becoming a hot topic.

A systematic prognostic evaluation is of great importance for the treatment and follow-up of patients with pancreatic cancer. For pancreatic cancer, clinicians often make empirical judgments based on the patient’s age, AJCC stage, pathological results, etc. There is no basis for a more perfect scoring system. Compared with the traditional method, Nomograms can make predictions more quickly, conveniently and accurately. Their predictive value has been reported to be better than other evaluation systems,[12] which is of great significance for clinical decision-making. Nomograms based on the SEER database have been widely used in the study of lymphoma, breast cancer and colorectal cancer.[13–15] The SEER database collects demographic, tumor, and survival data from 17 regions across the United States. This provides good data support for the establishment of Nomograms, which are impossible to achieve in general single-center research.

In this study, the prognosis of primary tumors in the body and tail of the pancreas was relatively poorer. However, the prognosis of primary tumors in the uncinate process of the pancreas and pancreatic duct was relatively better. Previous studies[16,17] have shown that tumors originating from the body and tail of the
pancreas are more likely to have distant metastasis than those from the head of the pancreas. The reason for these results may be that when tumors are located in the body and tail of the pancreas, symptoms will appear later. On the other hand, tumors in the head and duct of the pancreas can be detected earlier because of the appearance of obstructive jaundice. Therefore, different sites of primary pancreatic tumors may lead to different characteristics of distant metastasis.

Surgery is the only possible way to cure pancreatic cancer patients. However, most patients have no chance of surgery because tumors are in the advanced stage when patients are diagnosed with pancreatic cancer. For the application value of surgical treatment for local cancer sites in solid tumor patients with metastasis, there have been literature reports on metastatic renal cell carcinoma, metastatic nonfunctional pancreatic neuroendocrine tumors and metastatic hepatocellular carcinoma. However, for pancreatic cancer patients with metastasis, there is not enough evidence to support the consensus that primary tumor surgery should be performed. Early retrospective studies have suggested that simultaneous resection of primary pancreatic cancer and liver metastasis has no benefit for advanced patients. Recently, some studies have shown that, among highly selective patients with stage IV pancreatic cancer, patients who respond to systemic chemotherapy after primary resection may benefit. However, the studies described above did not have a sufficient sample size. At present, the Chinese pancreatic cancer research group is conducting a multicenter, prospective, randomized, phase III controlled clinical trial on the comparison of synchronous resection for primary tumors with liver metastasis after conversion chemo-

Figure 2. The cut-off value of age. Shown are the results calculated and analyzed by X-tile software. The best cut-off value of the age of pancreatic cancer patients with lung metastasis is 71 years old.
Figure 3. The survival curves of all the clinicopathological factors. Kaplan–Meier survival curves for age, sex, site, grade, T stage, N stage, surgery, radiation, chemotherapy, other organ with metastasis in the SEER database for pancreatic cancer with lung metastasis.
therapy and standard therapy in pancreatic cancer with liver oligometastasis.\textsuperscript{[26]}

At present, the treatment of pancreatic cancer is mainly based on chemotherapy. Chemotherapy plays an irreplaceable role in the treatment of pancreatic cancer. The results showed that the HR of patients receiving chemotherapy was only 0.322 compared with patients without chemotherapy. It is suggested that chemotherapy is a protective factor for pancreatic cancer patients, providing a strong basis for pancreatic cancer patients with lung metastasis to receive chemotherapy to prolong their survival time.

In this study, all the above mentioned prognostic factors were included, and the different classifications of each index were quantified to construct a more systematic and perfect evaluation system. The C-index values of the Nomogram in the training set and verification set were 0.727 and 0.738, respectively. The Nomogram had high predictive value and the calibration curve also showed good consistency.

However, there are some limitations in this study. First, lung metastasis in pancreatic cancer is a very complex pathological process, that involves many mechanisms and influencing factors. Because the SEER database is not sufficiently comprehensive, this study was limited to some clinicopathological factors. In this study, some important clinical data such as tumor markers, specific chemotherapy drugs, dose regimen, peritoneal dissemination, the number and size of the metastatic sites in other organs and the treatment of metastatic sites in other organs were not evaluated in detail, which easily led to partial bias. A small proportion of the patients were enrolled for analysis, and there could be obvious selective bias. In addition, the SEER database, as the main cancer clinical database in the United States, involves a variety of ethnic groups, but most of the patients are white and black people. The clinical data of Asian races are less frequently recorded. Therefore, the application of this Nomogram in China has certain limitations. In this study, all cases were confirmed as pancreatic cancer by pathological examination, while the existence of lung metastasis was only confirmed by imaging examination, not by pathological examination. Nevertheless, most of the main clinical features affecting the survival and prognosis of pancreatic cancer patients were included in this study. The constructed Nomogram can provide clinicians with more accurate and practical prediction tools, and can make rapid and accurate evaluations of the survival and prognosis of patients individually, to better guide the follow-up treatment of pancreatic cancer patients.

\begin{figure}
\centering
\includegraphics[width=\textwidth]{nomogram.png}
\caption{Nomogram for the pancreatic cancer patient with lung metastasis. Predictive model consisted of significant clinical and pathological factors on the prognosis of pancreatic cancer with lung metastasis. OS = overall survival, OtherMets = other organ with metastasis.}
\end{figure}
Conclusion
In conclusion, we established a Nomogram for predicting the risk of lung metastasis in patients with pancreatic ductal adenocarcinoma in this study. Physicians can assess a diverse range of parameters in patients with more objectives and precision for predicting lung metastasis using this Nomogram, resulting in reduced health care costs, limited radiation exposure and few unnecessary diagnostic investigations.

Acknowledgments
We thank the National Cancer Institute’s SEER Program for collection of the SEER data.

Author contributions
Jiachun Ding, QingYong Ma, Zheng Wu and Zheng Wang designed and supervised the project, Jiachun Ding, Jiaqiang Ren, Fan Chen, YangYang Yue, Chao Ren and Xirui Wang analyzed the data, Jiachun Ding, QingYong Ma and Zheng Wang wrote and revised the manuscript.

Financial support
This study was financially supported by the National Natural Science Foundation of China (NSFC 81872008 to ZW and NSFC 8207102675 to ZW) and the Science and Technology Innovation as a Whole Plan Projects of Shaanxi Province, China (No. 2016KJZDSF01-05-01).

Conflicts of interest
The authors declare no conflicts of interest.

Ethics approval
All data from SEER database have been fully de-identified and may be used without further independent ethics committee approval. This study was a secondary analysis of the existing SEER dataset, and posed no more than minimal risk to patients. Therefore, the requirement to obtain informed consent from patients was waived.

REFERENCES