Development and validation of a nomogram for predicting survival of Pulmonary invasive mucinous adenocarcinoma based on Surveillance, Epidemiology, and End Results (SEER) database

Yadong Wang
Institution of Oncology, Shandong Provincial Hospital, Cheeloo College of Medicine, Shandong University; Shandong Provincial Hospital Affiliated to Shandong First Medical University.

Jichang Liu
Institution of Oncology, Shandong Provincial Hospital, Cheeloo College of Medicine, Shandong University; Shandong Provincial Hospital Affiliated to Shandong First Medical University.

Cuicui Huang
Institution of Oncology, Shandong Provincial Hospital, Cheeloo College of Medicine, Shandong University; Shandong Provincial Hospital Affiliated to Shandong First Medical University.

Yukai Zeng
Institution of Oncology, Shandong Provincial Hospital, Cheeloo College of Medicine, Shandong University; Shandong Provincial Hospital Affiliated to Shandong First Medical University.

Yong Liu
Institution of Oncology, Shandong Provincial Hospital, Cheeloo College of Medicine, Shandong University; Shandong Provincial Hospital Affiliated to Shandong First Medical University.

Jiajun Du (✉ dujiajun@sdu.edu.cn)
Institution of Oncology, Shandong Provincial Hospital Affiliated to Shandong University
https://orcid.org/0000-0003-2406-9435

Research article

Keywords: Lung cancer, pulmonary invasive mucinous adenocarcinoma, nomogram, SEER, survival

DOI: https://doi.org/10.21203/rs.3.rs-27681/v1

License: © ️ This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License
Abstract

Background: Lung cancer remains the highest tumor mortality rate. In 2015, the cancer classification guidelines of the World Health Organization were updated. The term “invasive mucinous adenocarcinoma (IMA)” arouse people's attention, while the clinicopathological factors that may influence survival were unclear.

Methods: Data of IMA patients was downloaded from SEER database. Kaplan-Meier methods and log-rank tests were used to compare the differences in OS and LCSS. The nomogram was developed based on the result of the multivariable analysis. The discrimination and accuracy were tested by Harrell's concordance index (C-index), receiver operating characteristic (ROC) curve, calibration curve and decision curve analyses (DCA). Integrated discrimination improvement (IDI) index was used to evaluate the clinical efficacy.

Results: According to multivariate analysis, the prognosis of IMAs was associated with age, differentiation grade, TNM stage and treatments. Surgery might be the only way that would improve survival. Area under the curve (AUC) of the training cohort was 0.837 and 0.840 for 3- and 5-year OS, respectively. AUC for 3- and 5-year LCSS were separately 0.839 and 0.844. The new model was then evaluated by calibration curve, DCA and IDI index.

Conclusion: Based on this study, prognosis of IMAs was systematically reviewed, and a new nomogram was developed and validated. This will help to comprehensively evaluate the prognosis of IMA patients and formulate reasonable treatment measures.

1 Background

Lung cancer remains the first cause of oncological death in recent years\textsuperscript{[1]}. Non-small cell lung cancer (NSCLC) accounts for about 85% of lung cancer, among which, lung adenocarcinoma (LUAD) is becoming the main type of NSCLC\textsuperscript{[2]}. In 2015, the cancer classification guidelines of the World Health Organization were updated, and it clarified the new classification of lung tumors. According to the new classification standard, adenocarcinoma was divided into two categories, non-mucinous adenocarcinomas and adenocarcinoma variants. Simultaneously, the term “invasive mucinous adenocarcinoma (IMA)” was proposed to replace the previously named mucinous bronchoalveolar adenocarcinoma\textsuperscript{[3]}. According to the report, IMA accounts for 0.2% of all primary lung cancer\textsuperscript{[4]}, and < 2–10% of all lung adenocarcinomas\textsuperscript{[5]}, thus it is considered as a relatively rare histologic subtype.

Due to the low incidence, the clinicopathological characteristics and prognosis of IMAs are still unclear and controversial. It was reported that compared to patients with other lung adenocarcinoma subtypes, IMA patients have poor overall survival and progression-free survival times. Meanwhile, IMAs are considered to be diagnosed at an advanced stage of inoperability\textsuperscript{[6–8]}. However, a previous study
demonstrated that patients with IMA had comparable overall survival as those with intermediate grade non-mucinous adenocarcinoma (NMA)\cite{9}. Moreover, a study by Yoshizawa et al. showed that the disease-free survival of patients with IMA were between low-grade and high-grade adenocarcinoma\cite{10}. A recent study indicated that the survival curve of IMA patients was between lepidic adenocarcinoma and other adenocarcinoma patients, and it found that ~70% of IMAs were either stage I or II at the time of diagnosis\cite{4}. Similarly, Warth et al. found that a better prognosis was available for IMA patients compared with most adenocarcinoma patients\cite{11}. The clinicopathological factors that may influence patient survival were unclear. Thus, it is vital to establish a comprehensive analytic model to accurately estimate the prognosis of each patient.

The nomogram is a commonly viable predictive model for predicting and quantifying the probability of a clinical event, which is of great value for clinical decision-making and risk stratification, especially in cancer patients\cite{12,13}. However, as far as now, no nomogram has been developed for predicting the survival outcomes of IMA patients. Thus, in the present research, we used the IMAs case data of Surveillance, Epidemiology, and End Results (SEER) database to analyze clinical characteristics and study prognostic factors of IMAs. Furthermore, a nomogram of IMA patients was developed to better predict the cancer-specific survival of patients.

2 Methods

2.1 Data source

The SEER database, a cancer incidence registry managed by the National Cancer Institute, includes about 30% of the U.S. population. The data of IMA patients was extracted from SEER database(www.seer.cancer.gov), using SEER*Stat program (version 8.3.5). From the November 2018 submission, patients were collected up to December 2016 to build our cohort following the inclusion criteria: (a) pathological diagnosis was made between 2000 and 2015, (b) the International Classification of Diseases for Oncology-3 (ICD-O-3) histology code 8253/3\footnote{8253/3: Invasive mucinous adenocarcinoma,} (c) only one malignant primary tumor. Patients with a diagnosis confirmed by autopsy and/or with incomplete survival data were excluded.

The study variables of patients we extracted and analyzed include: baseline demographics, tumor features, therapy. Baseline demographics include age (\(<59\text{y}, 60–69\text{y}, 70–79\text{y}, \geq80\)), gender (Female, Male), race (White, Black, Other), marital status (Divorced, Married, Separated, Single, Windowed), survival time (months) and vital survival status. Tumor features include grade (Well differentiated, Moderately differentiated, Poorly, differentiated/Undifferentiated), laterality (Only one side, Bilateral), SEER-stage (Localized, Regional, Distant), T stage in 8th edition AJCC system (T1, T2, T3, T4), N stage in 8th edition AJCC system (N0, N1, N2/3), M stage in 8th edition AJCC system (M0, M1). Therapy includes surgery (No, Yes), radiation (No, Yes), and chemotherapy (No/Unknown, Yes). Overall survival (OS) or lung cancer specific survival (LCSS) was set as the study endpoint. The OS was defined as the time from date of
diagnosis to date of death or last contact. The LCSS was defined as the time from date of diagnosis to date of death due to lung cancer.

2.2 Statistical analyses

To make the best use of our data for constructing the predictive model, enrolled eligible patients diagnosed in the period of 2000–2010 years were assigned to the training cohort. Enrolled eligible patients diagnosed in the period of 2011–2015 years were assigned to the validation cohort. We used training cohort to establish the predictive model and to develop the nomogram. The validation cohort was used to validate the model.

For survival analyses, age at diagnosis, gender, race, marital status, grade, laterality, SEER–stage, T stage in 8th edition AJCC system, N stage in 8th edition AJCC system, M stage in 8th edition AJCC system, surgery, radiation and chemotherapy variables were included. Kaplan-Meier methods and log-rank tests were used to compare the differences in OS and LCSS. The hazard ratio (HR) and corresponding 95% confidential interval (CI) of each potential prognostic variable were estimated by using the univariate and multivariate Cox Proportional Hazard Regression Model. SPSS 25.0 (SPSS, Chicago, IL) was used for the above analysis. Based on the results of the multivariable analysis, a nomogram was developed to provide visualized risk prediction.

Discrimination and calibration were used to evaluate the accuracy of nomogram for predicting visualized risk and survival outcomes. The Harrell's concordance index (C-index) was used as a measurement tool of discrimination. The Accuracy of calibration was represented by a calibration curve. The reliability of the model was evaluated by Decision curve analyses (DCA). Finally, integrated discrimination improvement (IDI) index was used to compare the clinical applicability between the new model and TNM staging system.

We used R version 3.6.2 (The R Foundation for Statistical Computing, Vienna, Austria) to perform analyses. R project's package 'survival', 'rms', 'foreign', 'timeROC', 'rmda' were used. Besides, “stdca.R” was downloaded from Memorial Sloan Kettering Cancer Center(www.mskcc.org) to conduct DCA. In all statistical analyses, a p value of < 0.05 was considered significant. This study followed the Declaration of Helsinki for medical research involving human subjects.

3 Results

3.1 Patient characteristics

We identified 774 patients diagnosed as invasive mucinous adenocarcinoma (IMA) by immunohistochemistry between 2000 and 2010 (Supplementary Fig. S1). Of all the patients, 774 cases diagnosed between 2000 and 2010 were included in univariate analysis. 414 of them were included in multivariate Cox regression analysis and used as training cohort of the diagnostic nomogram. Besides, 88 cases diagnosed between 2011 and 2015 were used as validation cohort. All patients had complete
information on survival time and cause of death. The 3-year, 5-year overall survival (OS) in training cohort were 52.9% and 43.9%, respectively. The 3-year, 5-year lung cancer special survival (LCSS) in training cohort were 56.2% and 47.9%, respectively. Characteristics of the validation cohort are presented in Supplementary Table S1.

The median age was 69. The percentage of people whose age \( \leq 59, 60–69, 70–79, \) and \( \geq 80 \) years old was 25.7%, 26%, 35.4%, 12.9% respectively. Most patients are white people, about 79.6%. The proportion of female are slightly greater than male. 498 cases have detailed records of pathological grades, among which most (77.7%) are well differentiated. Of all these cases, 560 patients (72.4%) underwent surgery. While the proportion who choose chemotherapy and radiation treatment is significantly lower than surgery, only 3.0% and 30.0% of patients received radiation treatment and chemotherapy, respectively (Table 1).
Table 1
Characteristics of included IMA patients and univariate COX regression of overall survival (OS) and lung cancer specific survival (LCSS).

| Characteristics | Number | OS | LCSS |
|-----------------|--------|----|------|
|                 |        | HR | 95.0%CI | p | HR | 95.0%CI | p |
| Age             | 774    |    |         |   |    |         |   |
| ≤ 59            | 199(25.7%) | 1 | - | - | 1 | - | - |
| 60–69           | 201(26.0%) | 1.408 | 1.088–1.823 | 0.009 | 1.242 | 0.940–1.643 | 0.128 |
| 70–79           | 274(35.4%) | 2.169 | 1.716–2.741 | <0.001 | 1.711 | 1.324–2.211 | <0.001 |
| ≥ 80            | 100(12.9%) | 3.074 | 2.307–4.095 | <0.001 | 2.258 | 1.636–3.117 | <0.001 |
| Race            | 774    |    |         |   |    |         |   |
| White           | 616(79.6%) | 1 | - | - | 1 | - | - |
| Black           | 85(11.0%) | 0.968 | 0.736–1.273 | 0.815 | 1.105 | 0.821–1.487 | 0.510 |
| Other           | 73(9.4%) | 0.884 | 0.656–1.191 | 0.418 | 0.992 | 0.717–1.372 | 0.961 |
| Gender          | 774    |    |         |   |    |         |   |
| Female          | 458(59.2%) | 1 | - | - | 1 | - | - |
| Male            | 316(40.8%) | 1.313 | 1.109–1.555 | 0.002 | 1.277 | 1.054–1.546 | 0.012 |
| Grade           | 498    |    |         |   |    |         |   |
| Well differentiated | 387(77.7%) | 1 | - | - | 1 | - | - |
| Moderately differentiated | 88(17.7%) | 1.155 | 0.877–1.521 | 0.305 | 1.222 | 0.898–1.662 | 0.202 |

HR, hazard ratio; CI, confidential interval; OS, overall survival; LCSS, lung cancer specific survival.
| Characteristics                                      | Number   | OS |                 | LCSS |                 |
|----------------------------------------------------|----------|----|-----------------|------|-----------------|
|                                                    |          |    | **HR**         | **95.0%CI** | **p**     | **HR**         | **95.0%CI** | **p**     |
| Poorly differentiated/Undifferentiated              | 23(4.6%) |    | 2.175           | 1.390–3.402 | 0.001   | 2.361           | 1.471–3.790 | < 0.001  |
| HR, hazard ratio; CI, confidence interval; OS, overall survival; LCSS, lung cancer specific survival. |          |    |                 |                 |         |                 |                 |          |
| Characteristics | Number | OS | LCSS |
|-----------------|--------|----|------|
|                 |        | HR | 95.0%CI | p | HR | 95.0%CI | p |
| Mstage_8th      | 756    |    |         |   |    |         |   |
| M0              | 558(73.8%) | 1 | - | 1 | - |
| M1              | 198(26.2%) | 4.076 | 3.376–4.921 | <0.001 | 4.760 | 3.878–5.844 | <0.001 |
| Surgery         | 774    |    |         |   |    |         |   |
| No              | 214(27.6%) | 1 | - | 1 | - |
| Yes             | 560(72.4%) | 0.230 | 0.191–0.277 | <0.001 | 0.206 | 0.168–0.253 | <0.001 |
| Radiation       | 774    |    |         |   |    |         |   |
| No              | 751(97.0%) | 1 | - | 1 | - |
| Yes             | 23(3.0%) | 2.031 | 1.324–3.117 | 0.001 | 2.193 | 1.398–3.441 | 0.001 |
| Chemotherapy    | 774    |    |         |   |    |         |   |
| No/Unknown      | 542(70.0%) | 1 | - | 1 | - |
| Yes             | 232(30.0%) | 2.318 | 1.944–2.765 | <0.001 | 2.757 | 2.270–3.348 | <0.001 |
| Marital status  | 754    |    |         |   |    |         |   |
| Divorced        | 62(8.2%) | 1 | - | 1 | - |
| Married         | 477(63.3%) | 1.054 | 0.765–1.454 | 0.747 | 0.979 | 0.688–1.393 | 0.906 |
| Separated       | 13(1.7%) | 1.003 | 0.471–2.136 | 0.994 | 0.915 | 0.385–2.176 | 0.841 |
| Single          | 81(10.7%) | 1.060 | 0.711–1.582 | 0.774 | 0.856 | 0.541–1.355 | 0.508 |
| Widowed         | 121(16%) | 1.283 | 0.891–1.846 | 0.180 | 1.182 | 0.790–1.767 | 0.416 |

HR, hazard ratio; CI, confidential interval; OS, overall survival; LCSS, lung cancer specific survival.
3.2 Univariate and multivariate Cox proportional hazard analysis

We conducted univariable and multivariable analysis to identify the prognostic factors associated with survival of IMA patients in the training cohort.

In the univariate analysis, older age, male, poorly differentiated grade, bilateral laterality, higher TNM stage, no surgery, radiation treatment and chemotherapy predicted worse OS and LCSS (Fig. 1, Fig. 2). However, race and marital status had no significant effect on OS or LCSS (Supplementary Fig. S2, Fig. S3). Results of univariate Cox regression for OS and LCSS were stated in Table 1. All the statistically significant variables were carried into multivariate analysis.

In the multivariate analysis of both OS and LCSS, variables including age, grade, TNM stage and treatments including surgery, radiation treatment and chemotherapy were all statistically significant. However, there was a difference between OS and LCSS about gender. In multivariate analysis, gender had a statistically significant effect on the patient’s OS, but not on LCSS. In terms of OS, male patients had a higher risk than females, but there was no statistical difference for LCSS.

Results of multivariate Cox regression for OS and LCSS were stated in Table 2. According to multivariate analysis, the outcomes were improved in patients with younger age, well differentiated stage, lower TNM stage and appropriate therapies.
Table 2
Multivariate COX regression of overall survival (OS) and lung cancer specific survival (LCSS).

| Characteristics | OS          |           | LCSS          |           |
|-----------------|-------------|-----------|---------------|-----------|
|                 | HR     | 95.0%CI  | p      | HR     | 95.0%CI  | p      |
| Gender          |         |           |        |         |           |        |
| Female          | 1      | -         | -      | 1      | -         | -      |
| Male            | 1.355  | 1.053–1.743 | 0.018 | 1.314  | 0.975–1.771 | 0.072 |
| Age             |         |           |        |         |           |        |
| ≤ 59            | 1      | -         | -      | 1      | -         | -      |
| 60–69           | 1.595  | 1.098–2.316 | 0.014 | 1.340  | 0.877–2.048 | 0.176 |
| 70–79           | 2.280  | 1.623–3.204 | <0.001| 1.787  | 1.215–2.629 | 0.003 |
| ≥ 80            | 3.753  | 2.395–5.883 | <0.001| 2.603  | 1.531–4.426 | <0.001|
| Laterality      |         |           |        |         |           |        |
| Only one side   | 1      | -         | -      | 1      | -         | -      |
| Bilateral       | 0.605  | 0.205–1.786 | 0.363 | 0.597  | 0.199–1.785 | 0.355 |
| Grade           |         |           |        |         |           |        |
| Well differentiated | 1    | -         | -      |        |           |        |
| Moderately differentiated | 1.387 | 1.000–1.925 | 0.050 | 1.488  | 1.017–2.178 | 0.041 |
| Poorly differentiated/Undifferentiated | 2.111 | 1.245–3.579 | 0.006 | 2.334  | 1.313–4.152 | 0.004 |
| Tstage_8th      |         |           |        |         |           |        |
| T1              | 1      | -         | -      | 1      | -         | -      |
| T2              | 1.191  | 0.828–1.713 | 0.346 | 1.279  | 0.798–2.049 | 0.307 |

HR, hazard ratio; CI, confidence interval; OS, overall survival; LCSS, lung cancer specific survival.
| Characteristics | OS | LCSS |
|----------------|----|------|
|                | HR | 95.0%CI | p  | HR | 95.0%CI | p  |
| T3             | 1.671 | 1.123–2.487 | 0.011 | 2.217 | 1.363–3.605 | 0.001 |
| T4             | 3.269 | 2.284–4.680 | < 0.001 | 4.528 | 2.907–7.054 | < 0.001 |
| Nstage_8th     |     |      |     |     |      |     |
| N0             | 1 | - | 1 | - |
| N1             | 1.589 | 0.955–2.643 | 0.075 | 1.696 | 0.976–2.948 | 0.061 |
| N2/3           | 2.436 | 1.455–4.077 | 0.001 | 2.768 | 1.608–4.765 | < 0.001 |
| Mstage_8th     |     |      |     |     |      |     |
| M0             | 1 | - | 1 | - |
| M1             | 2.564 | 1.675–3.925 | < 0.001 | 2.677 | 1.682–4.258 | < 0.001 |
| Surgery        |     |      |     |     |      |     |
| No             | 1 | - | 1 | - |
| Yes            | 0.495 | 0.326–0.751 | 0.001 | 0.470 | 0.291–0.760 | 0.002 |
| Radiation      |     |      |     |     |      |     |
| No             | 1 | - | 1 | - |
| Yes            | 1.971 | 1.108–3.506 | 0.021 | 1.938 | 1.052–3.571 | 0.034 |
| Chemotherapy   |     |      |     |     |      |     |
| No/Unknown     | 1 | - | 1 | - |
| Yes            | 1.377 | 1.009–1.879 | 0.044 | 1.474 | 1.049–2.071 | 0.025 |

HR, hazard ratio; CI, confidential interval; OS, overall survival; LCSS, lung cancer specific survival.

### 3.3 Development and validation of a prognostic nomogram

These significant independent prognostic factors were used to develop a nomogram to calculate the 3- and 5-year OS or LCSS probabilities (Fig. 3). The nomogram showed that age was the most predominant
contributor to the OS followed by T stage which played a more important role for LCSS. Each subtype within these significant independent variables was assigned a score on the point scale. The total score projected to the bottom scale representing the probabilities of 3- and 5-year OS or LCSS.

In the training cohort, C-index was 0.764 (95%CI: 0.737–0.791) for OS, 0.788 (95%CI: 0.757–0.819) for LCSS. C-index was 0.780 (95%CI: 0.700–0.860) for OS, 0.807 (95%CI: 0.721–0.893) for LCSS in the validation cohort. We tested the nomogram by internal receiver operating characteristic (ROC) curves in the training cohort. The area under the curve (AUC) was 0.837(95%CI: 0.796–0.878) and 0.840(95%CI: 0.802–0.879) for 3- and 5-year OS, respectively, with 0.839(95%CI: 0.796–0.883) and 0.844(95%CI: 0.805–0.885) for 3- and 5-year LCSS (Fig. 4). Figure 5 showed the calibration plots of the nomogram. These indicated that the new prediction model had a great performance for IMAs.

DCA curve was used to compare the clinical usability and benefits between the nomogram with the 8th edition AJCC TNM staging system (Fig. 5). Discrimination improvement was confirmed by an integrated discrimination index (IDI) with 0.086(95%CI: 0.049–0.139), 0.098(95%CI: 0.066–0.143) for 3-/5- year OS and 0.062(95%CI: 0.032–0.115), 0.071(95%CI: 0.032–0.114) for 3-/5- year LCSS (Fig. 6). These results indicated that in clinical application, this new model was better than the 8th edition AJCC TNM staging system.

4 Discussion

Lung cancer remains the first cause of oncological death, and in recent years\cite{1}, LUAD is becoming more and more frequent\cite{2}. Although the proportion of IMAs in LUAD is relatively low, it was believed that people with IMA had worse prognosis.

IMAs was different from other LUAD, characterized by goblet or columnar tumor cells with abundant intracytoplasmic mucin and basally located nuclei. In some cases, IMAs showed the mixture of different pathological types\cite{14,15}. IMAs have special genetic signatures. Studies found that many genes unexpectedly enriched in mucin-producing gastrointestinal, pancreatic, and breast cancer showed significant differences in IMAs, including FOXA3, SPDEF, etc.\cite{16}. And there was evidence that B7-H4 expressed in IMAs, which was considered as a therapeutic target for immune checkpoint therapy\cite{16}. NRG1 fusion looked frequent in IMAs even without KRAS mutations\cite{17–19}. These unique pathological features may affect pathological diagnosis.

In recent years, there have been relatively few studies on systematic reviews of IMAs treatment. Therefore, we decided to constructed a nomogram to predict the prognosis for IMAs and helped to provided new sight for treatment.

In this research, patients diagnosed with IMA was included into our analysis. There are just over 1000 patients, and we included 774 patients with complete clinical information (Supplementary Fig. S1). These patients had a reasonable age distribution, and most had received surgery.
In univariate analysis, gender, age, differentiation grade, TNM stage, and treatments including surgery, radiation, chemotherapy were all related to IMAs progression (Table 1, Fig. 1, Fig. 2). As we can see, surgery treatment would decrease the HR, while radiation treatment and chemotherapy would not.

We conducted multivariate analysis using these significant variables in univariate analysis (Table 2). Except for laterality, all the factors were statistically significant for OS. This result verified that older age, male, poorly differentiated grade, bilateral laterality, higher TNM stage, no surgery, radiation and chemotherapy were independent prognostic factors and improved the HR. However, for LCSS, gender was no longer statistically significant. This is different from traditional perception.

IMAs were always found in lower lobes and presented with multifocal consolidation and lung-to-lung or pleural metastasis\textsuperscript{[20]}. Many researches indicated tumor size and invasive size might be the independent factor influencing the prognosis of IMAs\textsuperscript{[9, 21]}. Apart from surgery, non-TKI chemotherapy was used in many IMA patients, while OS seemed no improvement\textsuperscript{[20]}. As we could see, more than 70% patients received surgery, and for both OS and LCSS, surgery looked like the only treatment that would improve survival. Consistent with previous reports, chemotherapy does not promote prognosis, and so is radiation therapy, suggesting that chemotherapy and radiation therapy might not bring survival benefits for IMAs. Therefore, for patients with a clear diagnosis of IMA, we still choose surgical treatment as the first choice. But the effect of surgery combined with chemoradiotherapy remains to be seen.

We plotted nomograms based on independent prognostic factors suggested in multiple factors (Fig. 3). For OS, age was the main factor that influenced prognosis, and T stage for LCSS. The accuracy of this model was measured via ROC curves and calibration plots. The training cohort AUC was 0.837(95%CI: 0.796–0.878) for 3-year OS, and 0.840(95%CI: 0.802–0.879) for 5-year OS. AUC for 3-and 5-year LCSS were 0.839(95%CI: 0.796–0.883) and 0.844(95%CI: 0.805–0.885), respectively (Fig. 4). Furthermore, we compared the model with 8th edition AJCC TNM staging system and it showed that our model had a better predictive value for IMAs than TNM staging system (Fig. 5, Fig. 6), and IDI was 0.086(95%CI: 0.049–0.139), 0.098(95%CI: 0.066–0.143) for 3-/5-year OS and 0.062(95%CI: 0.032–0.115), 0.071(95%CI: 0.032–0.114) for 3-/5-year LCSS.

To our knowledge, this is the first systematic review model for clinical characteristics and treatment of IMAs. The comprehensive clinical information of the SEER database provided great support for the study. However, there are many limitations that must be considered. IMA is difficult to diagnose till now, and many patients are classified as "adenocarcinoma" without specific pathological types. At the same time, the number of patients with a clear diagnosis of IMA in the databases around us is also very small, and we have not been able to verify the accuracy of this model in other databases. However, this model comprehensively evaluates the clinical characteristics and treatment, and provides ideas for improving the prognosis of IMA.

5 Conclusions
In conclusion, we conducted an analysis of prognosis of IMA based on a large population-based group from SEER database. Prognosis of IMAs was systematically reviewed, and a new nomogram was developed and validated. Then we elucidated the factors that affect IMA prognosis, including gender, age, TNM staging, grade of tumor differentiation and treatments. This will help us to have a deeper understanding of IMA and help to provide better treatment for patients.

**Abbreviations**

IMA, invasive mucinous adenocarcinoma; OS, overall Survival; LCSS, lung cancer specific survival; C-index, concordance index; ROC, receiver operating characteristic; AUC, area under the curve; DCA, decision curve analyses; IDI, integrated discrimination improvement; CI, confidence interval; HR, hazard ratio; SEER database, Surveillance, Epidemiology, and End Results database.

**Declarations**

**Ethics approval and consent to participate**

This study was based on publicly available data from the SEER database, and did not involve interaction with human subjects or the use of personally identifiable information. The study did not require informed consent for SEER registration cases, and the author obtained a "limited use data agreement" from SEER. No trial registration was required.

**Consent to publish**

Not applicable.

**Availability of data and materials**

All data used in this study are available at [www.seer.cancer.gov](http://www.seer.cancer.gov).

**Competing interests**

There is no competing interest regarding the publication of this paper.

**Funding**

This work is supported by National Natural Science Foundation of China (Grant No. 81672288) and Key Research and Development Plan of Shandong Province (Grant No. 2018GSF118028).

**Authors' Contributions**
JD contributed to the design and supervision of the research, and correspondence. Co-authors YW and JL collected and analyzed the data and completed the manuscript. CH, YZ, YL helped to search for references and offered guidelines of statistical methods. All authors have read and approved the final version to be published.

Acknowledgements

Not applicable.

Authors' Information

1Institute of Oncology, Shandong Provincial Hospital; Cheeloo College of Medicine, Shandong University; Shandong Provincial Hospital Affiliated to Shandong First Medical University, Jinan, PR China.

2Department of Thoracic Surgery, Shandong Provincial Hospital, Cheeloo College of Medicine, Shandong University; Shandong Provincial Hospital Affiliated to Shandong First Medical University, Jinan, PR China

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. CA Cancer J Clin 2020, 70(1): 7-30.
2. Huang C, Qu X, Du J. Proportion of lung adenocarcinoma in female never-smokers has increased dramatically over the past 28 years. Journal of thoracic disease 2019, 11(7): 2685-2688.
3. Travis WD, Brambilla E, Nicholson AG, Yatabe Y, Austin JHM, Beasley MB, Chirieac LR, Dacic S, Duhig E, Flieder DB, Geisinger K, Hirsch FR, Ishikawa Y, Kerr KM, Noguchi M, Pelosi G, Powell CA, Tsao MS, Wistuba I, Panel WHO. The 2015 World Health Organization Classification of Lung Tumors: Impact of Genetic, Clinical and Radiologic Advances Since the 2004 Classification. J Thorac Oncol 2015, 10(9): 1243-1260.
4. Moon SW, Choi SY, Moon MH. Effect of invasive mucinous adenocarcinoma on lung cancer-specific survival after surgical resection: a population-based study. J Thorac Dis 2018, 10(6): 3595-3608.
5. Travis WD, Brambilla E, Noguchi M, Nicholson AG, Geisinger K, Yatabe Y, Powell CA, Beer D, Riely G, Garg K, Austin JH, Rusch VW, Hirsch FR, Jett J, Yang PC, Gould M, American Thoracic S. International Association for the Study of Lung Cancer/American Thoracic Society/European Respiratory Society: international multidisciplinary classification of lung adenocarcinoma: executive summary. Proc Am Thorac Soc 2011, 8(5): 381-385.
6. Dacic S. Pros: the present classification of mucinous adenocarcinomas of the lung. Transl Lung Cancer Res 2017, 6(2): 230-233.
7. Yoshizawa A, Motoi N, Riely GJ, Sima CS, Gerald WL, Kris MG, Park BJ, Rusch VW, Travis WD. Impact of proposed IASLC/ATS/ERS classification of lung adenocarcinoma: prognostic subgroups and implications for further revision of staging based on analysis of 514 stage I cases. Mod Pathol 2011, 24(5): 653-664.
8. Wislez M, Antoine M, Baudrin L, Poulot V, Neuville A, Pradere M, Longchampt E, Isaac-Sibille S, Lebitasy MP, Cadranel J. Non-mucinous and mucinous subtypes of adenocarcinoma with bronchioloalveolar carcinoma features differ by biomarker expression and in the response to gefitinib. Lung Cancer 2010, 68(2): 185-191.

9. Lee HY, Cha MJ, Lee KS, Lee HY, Kwon OJ, Choi JY, Kim HK, Choi YS, Kim J, Shim YM. Prognosis in Resected Invasive Mucinous Adenocarcinomas of the Lung: Related Factors and Comparison with Resected Nonmucinous Adenocarcinomas. J Thorac Oncol 2016, 11(7): 1064-1073.

10. Yoshizawa A, Sumiyoshi S, Sonobe M, Kobayashi M, Fujimoto M, Kawakami F, Tsuruyama T, Travis WD, Date H, Haga H. Validation of the IASLC/ATS/ERS lung adenocarcinoma classification for prognosis and association with EGFR and KRAS gene mutations: analysis of 440 Japanese patients. J Thorac Oncol 2013, 8(1): 52-61.

11. Warth A, Muley T, Meister M, Stenzinger A, Thomas M, Schirmacher P, Schnabel PA, Budczies J, Hoffmann H, Weichert W. The novel histologic International Association for the Study of Lung Cancer/American Thoracic Society/European Respiratory Society classification system of lung adenocarcinoma is a stage-independent predictor of survival. J Clin Oncol 2012, 30(13): 1438-1446.

12. Balachandran VP, Gonen M, Smith JJ, DeMatteo RP. Nomograms in oncology: more than meets the eye. The Lancet Oncology 2015, 16(4): e173-e180.

13. Fakhry C, Zhang Q, Nguyen-Tân PF, Rosenthal DI, Weber RS, Lambert L, Trotti AM, Barrett WL, Thorstad WL, Jones CU, Yom SS, Wong SJ, Ridge JA, Rao SSD, Bonner JA, Vigneault E, Raben D, Kudrimoti MR, Harris J, Le Q-T, Gillison ML. Development and Validation of Nomograms Predictive of Overall and Progression-Free Survival in Patients With Oropharyngeal Cancer. Journal of clinical oncology : official journal of the American Society of Clinical Oncology 2017, 35(36): 4057-4065.

14. Cha YJ, Shim HS. Biology of invasive mucinous adenocarcinoma of the lung. Transl Lung Cancer Res 2017, 6(5): 508-512.

15. Shimizu K, Okita R, Saisho S, Maeda A, Nojima Y, Nakata M. Clinicopathological and immunohistochemical features of lung invasive mucinous adenocarcinoma based on computed tomography findings. Onco Targets Ther 2017, 10: 153-163.

16. Guo M, Tomoshige K, Meister M, Muley T, Fukazawa T, Tsuchiya T, Karns R, Warth A, Fink-Baldauf IM, Nagayasu T, Naomoto Y, Xu Y, Mall MA, Maeda Y. Gene signature driving invasive mucinous adenocarcinoma of the lung. EMBO Mol Med 2017, 9(4): 462-481.

17. Nakaoku T, Tsuta K, Ichikawa H, Shiraishi K, Sakamoto H, Enari M, Furuta K, Shimada Y, Ogiwara H, Watanabe S, Nokihara H, Yasuda K, Hiramoto M, Nammo T, Ishigame T, Schetter AJ, Okayama H, Harris CC, Kim YH, Mishima M, Yokota J, Yoshida T, Kohno T. Druggable oncogene fusions in invasive mucinous lung adenocarcinoma. Clin Cancer Res 2014, 20(12): 3087-3093.

18. Kim HS, Han JY, Shin DH, Lim KY, Lee GK, Kim JY, Jacob W, Ceppi M, Weisser M, James I. EGFR and HER3 signaling blockade in invasive mucinous lung adenocarcinoma harboring an NRG1 fusion. Lung Cancer 2018, 124: 71-75.
19. Drilon A, Somwar R, Mangatt BP, Edgren H, Desmeules P, Ruusulehto A, Smith RS, Delasos L, Vojnic M, Plodkowski AJ, Sabari J, Ng K, Montecalvo J, Chang J, Tai H, Lockwood WW, Martinez V, Riely GJ, Rudin CM, Kris MG, Arcila ME, Matheny C, Benayed R, Rekhtman N, Ladanyi M, Ganji G. Response to ERBB3-Directed Targeted Therapy in NRG1-Rearranged Cancers. Cancer Discov 2018, 8(6): 686-695.

20. Cha YJ, Kim HR, Lee HJ, Cho BC, Shim HS. Clinical course of stage IV invasive mucinous adenocarcinoma of the lung. Lung Cancer 2016, 102: 82-88.

21. Oki T, Aokage K, Nomura S, Tane K, Miyoshi T, Shiiya N, Funai K, Tsuboi M, Ishii G. Optimal method for measuring invasive size that predicts survival in invasive mucinous adenocarcinoma of the lung. J Cancer Res Clin Oncol 2020.

Figures
OS for IMA patients stratified by (A) Age, p<=0.001; (B) Gender, p=0.001; (C) Grade, p=0.002; (D) Laterality, p<=0.001; (E) SEER stage, p<=0.001; (F) T-stage, p<=0.001; (G) N-stage, p<=0.001; (H) M-stage, p<=0.001; (I) Surgery, p<=0.001; (J) Radiation, p=0.001; (K) Chemotherapy p<=0.001.
Figure 2

LCSS for IMA patients stratified by (A) Age, \( p \leq 0.001 \); (B) Gender, \( p = 0.012 \); (C) Grade, \( p = 0.001 \); (D) Laterality, \( p \leq 0.001 \); (E) SEER stage, \( p \leq 0.001 \); (F) T-stage, \( p \leq 0.001 \); (G) N-stage, \( p \leq 0.001 \); (H) M-stage, \( p \leq 0.001 \); (I) Surgery, \( p \leq 0.001 \); (J) Radiation, \( p \leq 0.001 \); (K) Chemotherapy \( p \leq 0.001 \).
Figure 3

Prognostic nomogram predicting the probability of 3- and 5-year (A) overall survival (OS) and (B) lung cancer specific survival (LCSS). Each subtype within these significant independent variables was assigned a score on the point scale. The total score projected to the bottom scale.
Figure 4

(A, B) ROC curves for 3- and 5-year OS based on the nomogram. The AUC was 0.837 and 0.840, respectively; (C, D) ROC curves for 3- and 5-year LCSS. The AUC was 0.839 and 0.845, respectively.
Figure 5

(A, B) Calibration plots for 3- and 5-year OS in the training cohort; (C, D) Calibration plots for 3- and 5-year LCSS in the training cohort.
Figure 6

(A. B) DCA curves for 3- and 5-year OS comparing this new model with 8th TNM staging system; (C. D) DCA curves for 3- and 5-year LCSS comparing this new model with 8th TNM staging system.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryTableS1.docx
• Supplementarymaterials.docx