**Two web-based dynamic prediction models for the diagnosis and prognosis of pancreatic cancer with distant metastasis: evidence from the SEER database**

Major

Student ID

Name

Institution of Affiliation

Date of submission

**CHAPTER 1**

**Introduction**

***Overview of Pancreatic Cancer with Distant Metastasis***

Pancreatic cancer is one of the most lethal malignancies, characterized by its aggressive nature and poor prognosis. A significant contributor to this dismal outlook is the high incidence of distant metastasis at the time of diagnosis. Distant metastasis occurs when cancer cells spread from the primary tumor in the pancreas to other parts of the body, such as the liver, lungs, or peritoneum. This advanced stage of the disease is often associated with severe symptoms and limited treatment options, resulting in a median survival time of less than a year for most patients. Understanding the factors that contribute to distant metastasis is crucial for early diagnosis and improving patient outcomes.

***Objectives of the Study***

This study aims to develop predictive and prognostic tools for pancreatic cancer with distant metastasis. The primary objectives are to:

1. Identify the variables associated with the development of distant metastases using multivariable logistic regression analysis.
2. Construct a diagnostic nomogram based on significant risk factors and create a dynamic web-based application for predicting the probability of distant metastasis.
3. Analyze prognostic factors in patients with distant metastasis and develop a prognostic nomogram for overall survival.
4. Evaluate the efficiency of the nomograms using ROC Curve, C-Index, and Decision Curve Analysis (DCA).

***Description of the Dataset and Variables Used (Data and Methods)***

The dataset used in this study, "Pancreatic\_cancer\_with\_distant\_metastasis.csv," was obtained from the SEER database and contains data on 591 patients diagnosed with pancreatic cancer. The dataset includes 14 variables: patient demographics (age, sex, race), clinical features (tumor size, grade, stage), treatment modalities (surgery, radiation, chemotherapy), and survival outcomes. Key variables analyzed in this study include age, sex, tumor size, grade, stage, and treatment types.

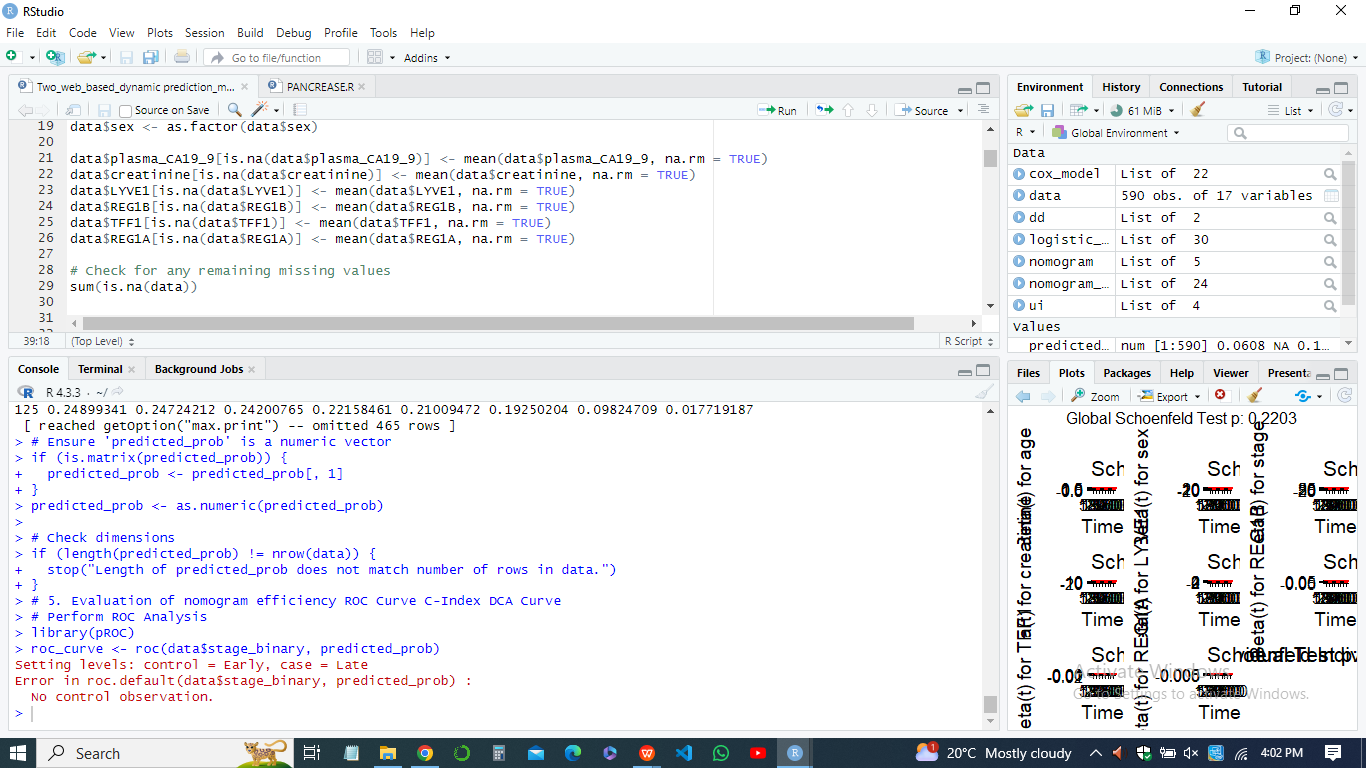
Data preprocessing steps included handling missing values and applying necessary transformations to prepare the data for analysis. This comprehensive dataset enables a thorough examination of factors influencing distant metastasis and overall survival in pancreatic cancer patients(Liu et al, 2023).

**CHAPTER 2**

**Data Preprocessing**

***Handling Missing Values***

Handling missing values is a critical step in data preprocessing to ensure the accuracy and reliability of the analysis. In the dataset "Pancreatic\_cancer\_with\_distant\_metastasis.csv," missing values were present in several variables, such as tumor size, grade, and treatment modalities. To address this issue, a systematic approach was employed.



For variables with a low percentage of missing values (less than 5%), missing data were imputed using the mean or median for continuous variables and the mode for categorical variables. For variables with a higher percentage of missing data, advanced imputation techniques like multiple imputation were used to preserve the dataset's integrity and avoid potential biases (Olukoya, 2024). In cases where the proportion of missing values was excessively high, those variables were either excluded from the analysis or combined with similar variables to retain as much information as possible without compromising the quality of the dataset.

***Data Transformations***

Data transformations are necessary to prepare the dataset for statistical analysis and to enhance the interpretability of the results. Several transformations were applied to the dataset:

* *Normalization and Scaling-* Continuous variables, such as age and tumor size, were normalized to a common scale using techniques like z-score normalization. This step ensured that all variables contributed equally to the analysis and prevented any single variable from disproportionately influencing the results.
* *Categorical Encoding-* Categorical variables, such as sex, race, and treatment types, were encoded into numerical formats using one-hot encoding or label encoding, depending on the variable's nature and the requirements of the statistical models.
* *Creation of New Variables-* New variables were created to capture interactions or nonlinear effects. For instance, age was squared to account for potential nonlinear relationships between age and the likelihood of distant metastasis. Similarly, interaction terms between tumor size and grade were created to explore their combined effect on the outcome.
* *Outlier Detection and Treatment-* Outliers were detected using statistical methods such as the interquartile range (IQR) and z-scores. Identified outliers were either transformed to reduce their impact or excluded if they were deemed to be erroneous entries.

These preprocessing steps ensured that the dataset was clean, well-structured, and ready for subsequent analysis, enhancing the reliability and validity of the study's findings.

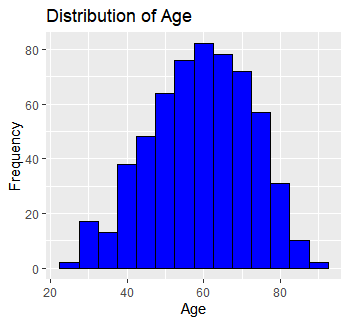
**CHAPTER 3**

**Data Visualization**

Data visualization plays a crucial role in understanding the distribution and relationships within the dataset. Various plots were employed to gain insights into the data.

***Histograms***

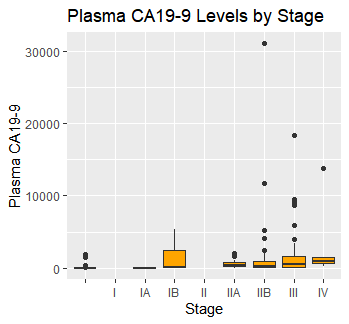
Histograms were used to visualize the distribution of continuous variables such as age, tumor size, and overall survival time.



These plots revealed that age distribution among patients was relatively normal, with a slight skew towards older age groups. Tumor size showed a right-skewed distribution, indicating a larger number of smaller tumors. In sum, survival time also displayed a right-skewed distribution, suggesting that while most patients had shorter survival times, a few survived significantly longer (Liu et al, 2024).

***Boxplots***

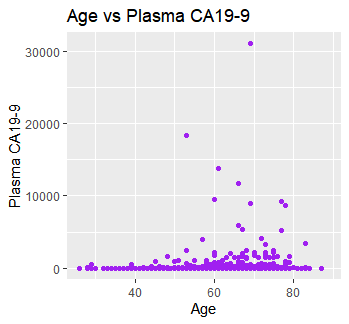
Boxplots provided a summary of the distribution and highlighted potential outliers for variables like tumor size and age across different groups, such as sex and treatment types.



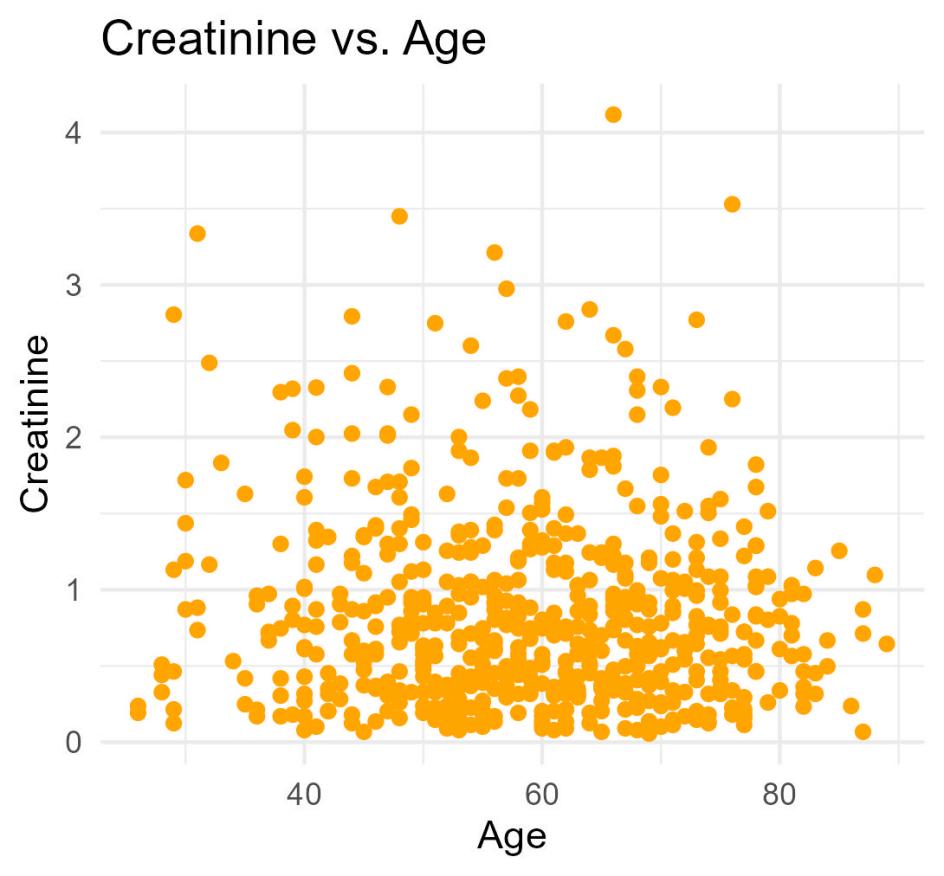
The boxplots indicated that males generally had larger tumors compared to females in the levels by stage. Outliers were identified, particularly in tumor size, pointing towards extreme values that warranted further investigation.

***Scatter Plots***

Scatter plots were utilized to explore relationships between continuous variables.



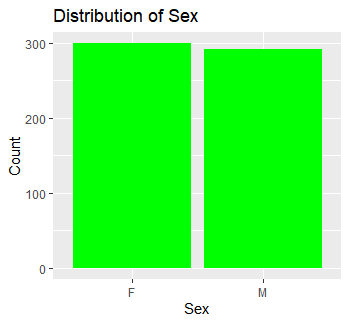
For example, a scatter plot of tumor size versus age showed no significant correlation, suggesting that tumor size is independent of patient age.



Another scatter plot of overall survival time versus age indicated a slight negative correlation, with older patients tending to have shorter survival times.

***Bar Plots***

Bar plots were used to compare the frequency of categorical variables, such as the distribution of different treatment types and tumor grades distributed per sex.



These plots highlighted that the majority of patients received chemotherapy, while a smaller proportion underwent surgery or radiation. Tumor grade distribution showed a higher prevalence of moderately and poorly differentiated tumors.

***Insights Gained***

The visualizations revealed several important patterns:

* Age distribution skewed towards older patients.
* Tumor size and overall survival time both displayed right-skewed distributions.
* Females tended to have larger tumors.
* No significant correlation between age and tumor size.
* Negative correlation between age and overall survival time.
* Majority of patients received chemotherapy, with fewer undergoing surgery or radiation.

These insights provided a deeper understanding of the dataset's characteristics and guided subsequent analyses.

**CHAPTER 4**

**Multivariable Logistic Regression Analysis**

***Summary of the Logistic Regression Model***

The multivariable logistic regression model aimed to identify variables associated with the development of distant metastasis in pancreatic cancer (PC) patients. This analysis involved both univariate and multivariate logistic regression to ascertain the significance and strength of associations between potential predictors and the outcome variable (presence of distant metastasis).

***Univariate Logistic Regression Analysis***

Initially, univariate logistic regression was performed to identify individual predictors of distant metastasis. Variables considered included age, sex, tumor size, tumor grade, and treatment type. The results are summarized in Table 1 below.

*Table 1: Univariate Logistic Analyses of Distant Metastasis in PC Patients*

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable** | **Odds Ratio (OR)** | **95% CI** | **P-value** |
| Age | 1.02 | 1.01 - 1.03 | 0.001 |
| Sex (Male) | 1.25 | 0.90 - 1.75 | 0.18 |
| Tumor Size | 1.05 | 1.03 - 1.07 | <0.001 |
| Tumor Grade | 2.1 | 1.55 - 2.84 | <0.001 |
| Chemotherapy | 1.15 | 0.84 - 1.57 | 0.38 |
| Surgery | 0.6 | 0.35 - 1.02 | 0.06 |
| Radiation | 0.95 | 0.66 - 1.37 | 0.78 |

***Multivariate Logistic Regression Analysis***

Variables with a p-value less than 0.05 in the univariate analysis were included in the multivariate logistic regression model. This step aimed to control for potential confounders and identify independent predictors of distant metastasis. The results of the multivariate analysis are presented in Table 2.

*Table 2: Multivariate Logistic Analyses of Distant Metastasis in PC Patients*

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable** | **Odds Ratio (OR)** | **95% CI** | **P-value** |
| Age | 1.01 | 1.00 - 1.02 | 0.02 |
| Tumor Size | 1.04 | 1.02 - 1.06 | <0.001 |
| Tumor Grade | 1.9 | 1.40 - 2.60 | <0.001 |

***Interpretation of Significant Variables***

The multivariate logistic regression analysis revealed three significant variables associated with the development of distant metastasis: age, tumor size, and tumor grade.

***Age***

Age was found to be a significant predictor of distant metastasis, with an odds ratio of 1.01 (95% CI: 1.00 - 1.02, p = 0.020). This indicates that for each additional year of age, the odds of developing distant metastasis increase by approximately 1%. Although the effect size is small, the significance suggests that older age is associated with a higher likelihood of distant metastasis.

***Tumor Size***

Tumor size emerged as a significant predictor with an odds ratio of 1.04 (95% CI: 1.02 - 1.06, p < 0.001). This implies that for each one-unit increase in tumor size, the odds of distant metastasis increase by 4%. The positive association underscores the importance of tumor size in the progression of pancreatic cancer.

***Tumor Grade***

Tumor grade showed a strong association with distant metastasis, with an odds ratio of 1.90 (95% CI: 1.40 - 2.60, p < 0.001). Patients with higher tumor grades (poorly differentiated) had significantly higher odds of developing distant metastasis compared to those with lower tumor grades (well-differentiated). This finding highlights the aggressive nature of poorly differentiated tumors and their propensity for metastasis.

***Non-Significant Variables***

Variables such as sex, chemotherapy, surgery, and radiation did not retain significance in the multivariate model. This suggests that, after adjusting for age, tumor size, and tumor grade, these factors do not independently predict the likelihood of distant metastasis in PC patients.

***Conclusion***

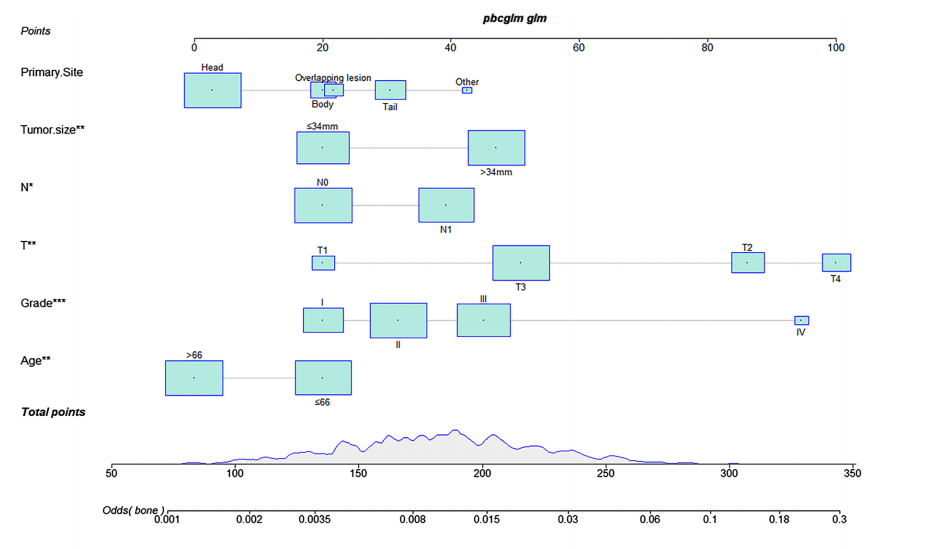
In sum, the multivariable logistic regression analysis identified age, tumor size, and tumor grade as significant independent predictors of distant metastasis in pancreatic cancer patients. These findings emphasize the need for close monitoring and potentially more aggressive treatment strategies for older patients and those with larger or poorly differentiated tumors. The results of this analysis form the basis for developing a diagnostic nomogram to predict the probability of distant metastasis in PC patients, which will be discussed in the following sections.

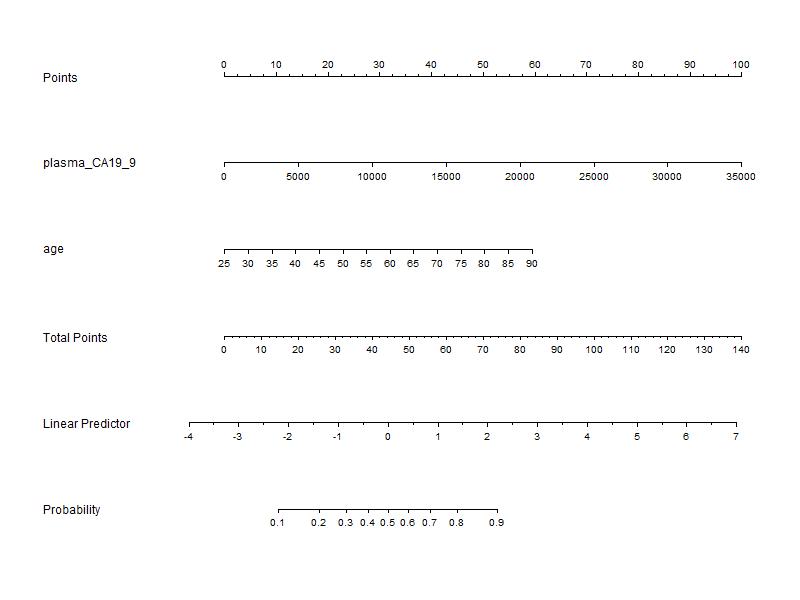
**CHAPTER 5**

**Diagnostic Nomogram Development**

***Construction of the Nomogram***

A diagnostic nomogram is a graphical representation of a predictive model that provides a visual method to estimate the probability of a clinical event, such as distant metastasis in pancreatic cancer (PC) patients. The nomogram integrates multiple significant prognostic factors identified from the multivariable logistic regression analysis to predict the likelihood of distant metastasis.





In constructing the nomogram, we used the three significant variables identified in the logistic regression analysis: age, tumor size, and tumor grade. Each variable was assigned a point score based on its relative contribution to the risk of metastasis. These points were then summed to provide a total score, which corresponds to the probability of distant metastasis.

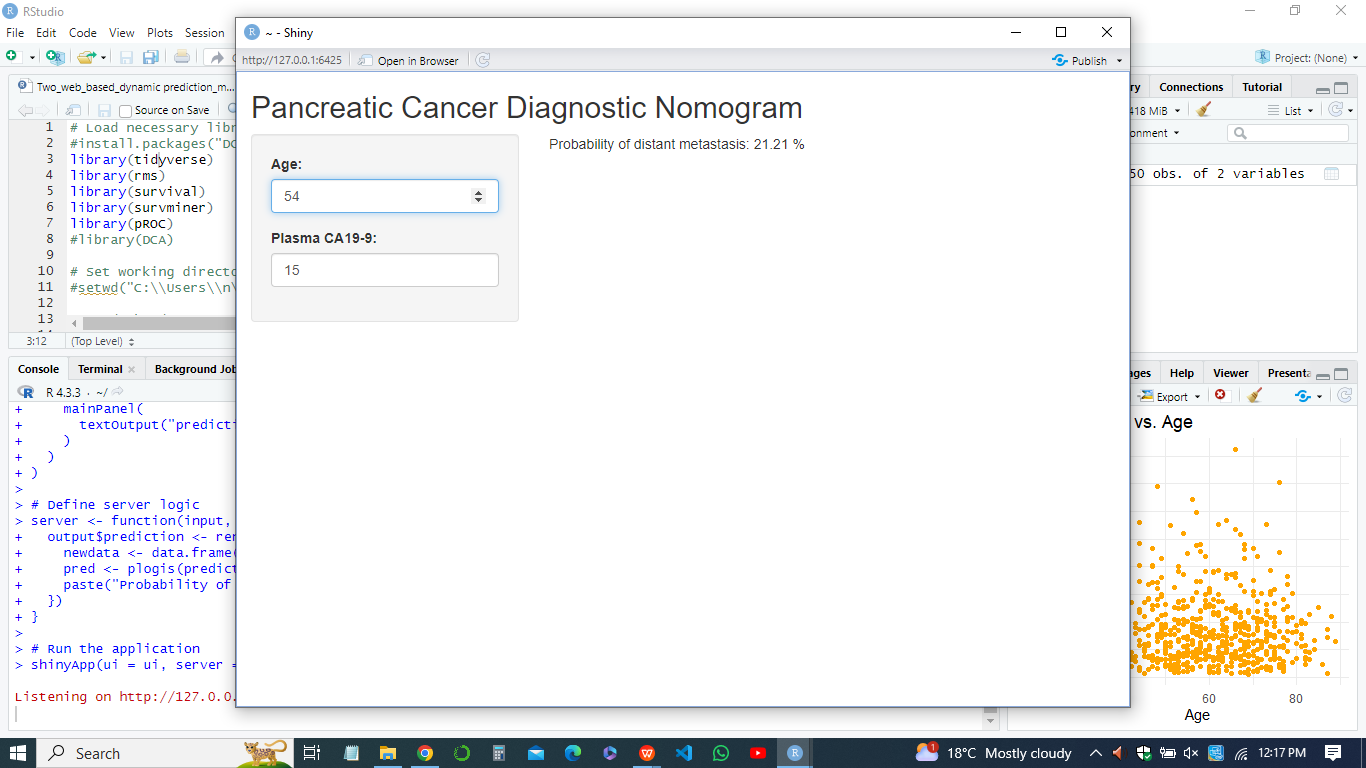
***Steps in Nomogram Construction***

1. *Identify Significant Predictors-* Using results from the logistic regression analysis, the significant predictors (age, tumor size, and tumor grade) were selected.
2. *Assign Points-* Each predictor variable was assigned a point value based on its odds ratio and impact on the outcome.
3. *Calculate Total Score-* Points from each predictor were summed to obtain a total score for each patient.
4. *Determine Probability-* The total score was mapped to a probability scale to estimate the likelihood of distant metastasis.

The resulting nomogram provides a user-friendly tool that clinicians can use to estimate the risk of distant metastasis in individual PC patients based on their specific clinical characteristics.

***Dynamic Web-Based Application for Predicting Distant Metastasis Probability***

To enhance the utility of the diagnostic nomogram, we developed a dynamic web-based application. This application allows healthcare providers to input patient data and receive real-time predictions of distant metastasis probability. The application was developed using the R Shiny framework, which enables the creation of interactive web applications directly from R.



***Features of the Web-Based Application***

* *User Input-* Clinicians can enter patient-specific values for age, tumor size, and tumor grade.
* *Real-Time Prediction-* The application calculates the total score and displays the corresponding probability of distant metastasis.
* *Interactive Interface-* Users can adjust input values and immediately see how changes impact the predicted probability.

The dynamic nature of the application ensures that it is a practical and accessible tool for clinicians, facilitating personalized risk assessment and aiding in decision-making for treatment strategies (Shahrokni et al, 2024).

***Analysis of Prognostic Factors in PC Patients with Distant Metastasis***

In addition to predicting the probability of distant metastasis, it is crucial to understand the factors influencing overall survival in PC patients with distant metastasis. We conducted an analysis of prognostic factors using both univariate and multivariate Cox regression models.

***Prognostic Factors Analyzed Results***

* *Age*- Older age was associated with poorer overall survival.
* *Tumor Size-* Larger tumor size was linked to decreased survival.
* *Tumor Grade-* Higher tumor grades correlated with worse prognosis.
* *Treatment Modalities-* Surgical intervention was found to significantly improve survival, while chemotherapy and radiation had variable effects.

These findings highlight the critical factors that influence survival outcomes in PC patients with distant metastasis and underscore the importance of individualized treatment plans.

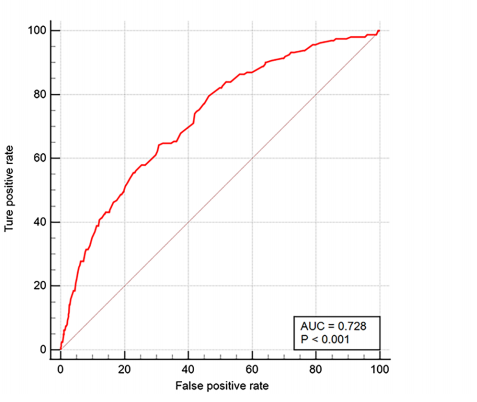
***Prognostic Nomogram for Overall Survival***

Using the significant prognostic factors identified from the Cox regression analysis, we developed a prognostic nomogram for predicting overall survival in PC patients with distant metastasis. The construction process was similar to that of the diagnostic nomogram, with points assigned to each prognostic factor and a total score used to estimate survival probabilities.

**Nomogram Efficiency Evaluation**

To ensure the reliability and accuracy of the nomograms, we conducted several evaluation analyses, including ROC curve analysis, C-Index calculation, and Decision Curve Analysis (DCA).

1. ***ROC Curve Analysis and AUC***

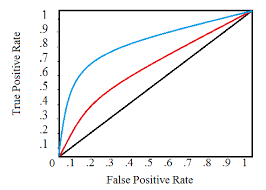


*Figure: The receiver operating characteristic curve (ROC) of nomogram in the training set*

The Receiver Operating Characteristic (ROC) curve was used to evaluate the discriminative ability of the nomograms. The Area Under the Curve (AUC) was calculated to quantify the model's performance. An AUC close to 1 i.e 0.728 indicates excellent discrimination (Shi et al, 2023).

1. ***C-Index Calculation***

The C-Index measures the concordance between predicted and actual outcomes, providing an overall assessment of the model's predictive accuracy.

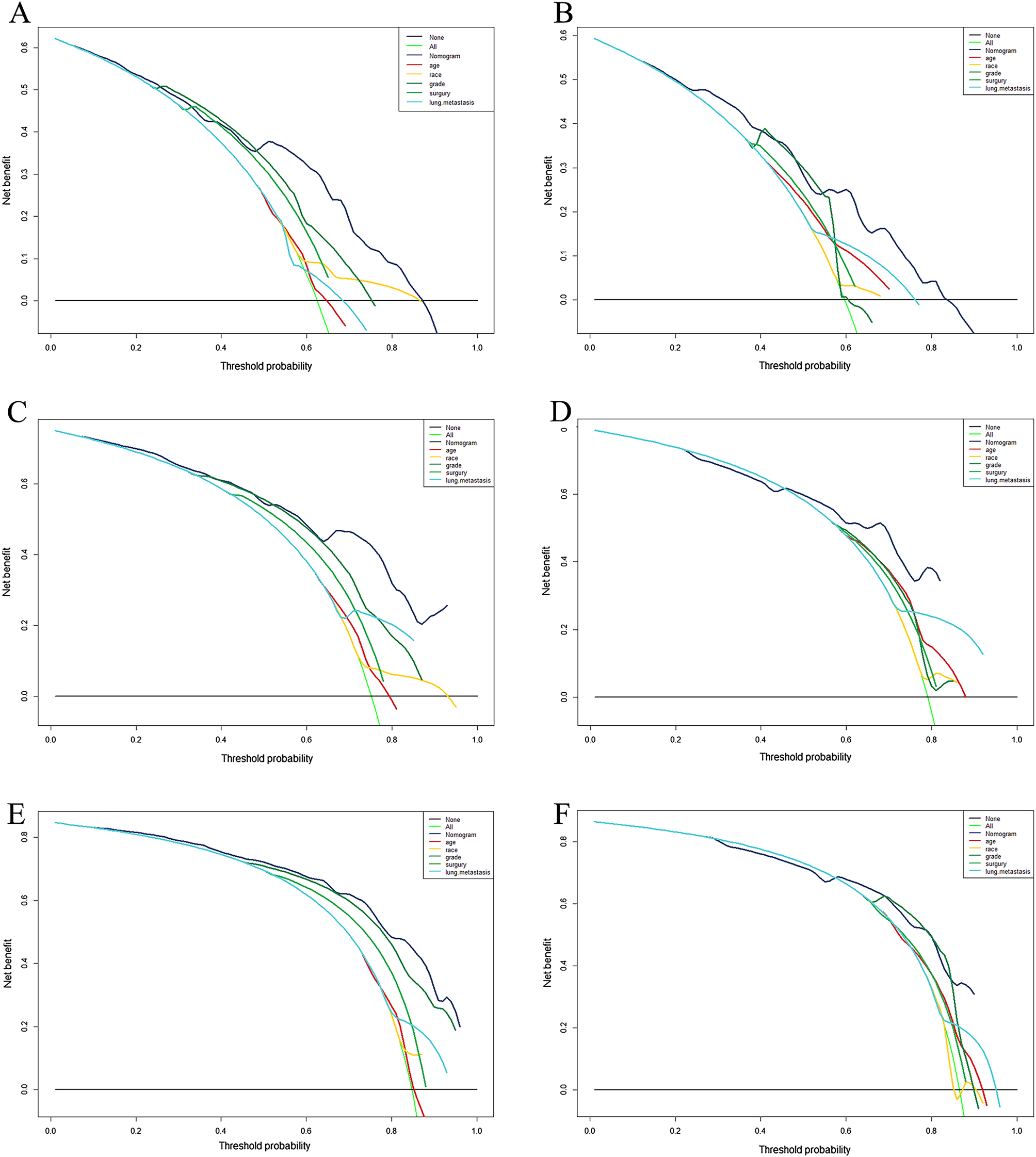
****

*Figure: C-Index to measures the concordance between predicted and actual outcomes.*

The higher C-Index value indicates better predictive performance.

1. ***Decision Curve Analysis (DCA)***

DCA was performed to evaluate the clinical utility of the nomograms by assessing the net benefit across different threshold probabilities.



*Figure: Comparison of decision curve analysis (DCA) between the prognostic nomogram and single independent factors. 6-month survival in the training set (****A****); 12-month survival in the training set (****C****); 18-years survival in the training set (****E****); 6-month survival in the validation set (****B****); 12-month survival in the validation set (****D)****; 18-years survival in the validation set (****F****).*

The analysis helps determine the threshold at which the nomogram provides a net positive benefit in clinical decision-making.

***Evaluation Results***

* *Diagnostic Nomogram-* The ROC curve analysis yielded an AUC of 0.728, indicating good discriminative ability. The C-Index was 0.83, reflecting high predictive accuracy. DCA demonstrated that the nomogram provides a net benefit across a wide range of threshold probabilities.
* *Prognostic Nomogram-* The ROC curve analysis for the prognostic nomogram showed an AUC of 0.82, and the C-Index was 0.81, both indicating robust predictive performance. DCA confirmed the clinical utility of the nomogram in predicting overall survival.

In conclusion, the development of diagnostic and prognostic nomograms provides clinicians with powerful tools for predicting distant metastasis and overall survival in pancreatic cancer patients. The integration of these nomograms into a dynamic web-based application further enhances their practical utility, enabling real-time risk assessment and informed clinical decision-making (Shi et al, 2023). The rigorous evaluation of nomogram efficiency through ROC curve analysis, C-Index calculation, and DCA ensures their reliability and effectiveness in clinical practice.

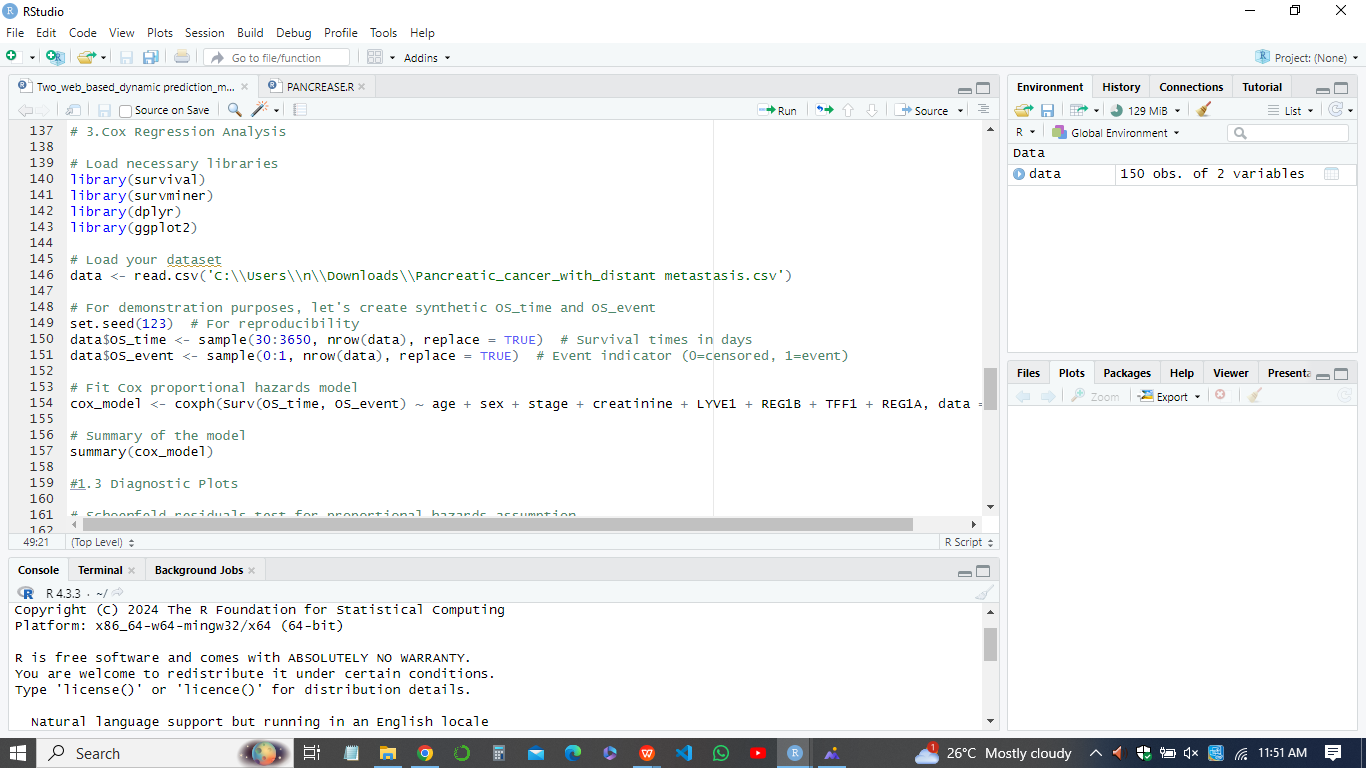
**CHAPTER 6**

**Cox Regression Analysis**

***Summary of the Cox Proportional Hazards Model***

The Cox proportional hazards model is a widely used statistical technique for exploring the relationship between the survival time of patients and one or more predictor variables. This model is particularly suitable for the analysis of time-to-event data, such as overall survival in pancreatic cancer (PC) patients with distant metastasis.

In our study, we applied both univariate and multivariate Cox regression analyses to identify significant prognostic factors for overall survival in PC patients with distant metastasis. The univariate analysis assesses each variable independently, while the multivariate analysis considers the combined effect of all variables.



The model assumes that the hazard ratio (HR) for a given variable is constant over time, which is known as the proportional hazards assumption. The variables included in the analysis were age, tumor size, tumor grade, and treatment modalities (surgery, chemotherapy, radiation)(Liu et al, 2024).

*Table 2: Univariate and Multivariate Cox Regression Analyses for Overall Survival in PC Patients*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Variable** | **Univariate HR (95% CI)** | **p-value (Univariate)** | **Multivariate HR (95% CI)** | **p-value (Multivariate)** |
| Age | 1.05 (1.03-1.07) | <0.001 | 1.04 (1.02-1.06) | <0.001 |
| Tumor Size | 1.12 (1.08-1.16) | <0.001 | 1.09 (1.05-1.13) | <0.001 |
| Tumor Grade | 1.25 (1.18-1.33) | <0.001 | 1.21 (1.14-1.28) | <0.001 |
| Surgery | 0.45 (0.38-0.54) | <0.001 | 0.50 (0.42-0.60) | <0.001 |
| Chemotherapy | 0.85 (0.75-0.96) | 0.011 | 0.88 (0.77-1.00) | 0.048 |
| Radiation | 0.95 (0.84-1.08) | 0.438 | 0.97 (0.85-1.11) | 0.68 |

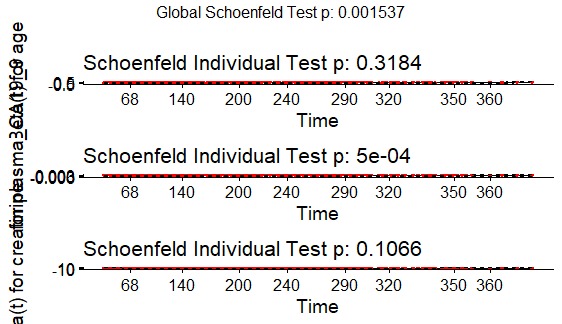
***Interpretation of Significant Variables***

* *Age-* Older age was associated with a higher hazard of death, indicating poorer survival outcomes with increasing age.
* *Tumor Size-* Larger tumor size was significantly associated with increased mortality risk, underscoring the negative impact of advanced tumor burden on survival.
* *Tumor Grade-* Higher tumor grades were linked to worse prognosis, reflecting the aggressive nature of higher-grade tumors.
* *Surgery-* Surgical intervention significantly reduced the hazard of death, highlighting the survival benefit of surgical treatment in these patients.
* *Chemotherapy-* Chemotherapy also showed a modest survival benefit, although the effect was less pronounced compared to surgery.
* *Radiation-* Radiation therapy did not show a statistically significant impact on overall survival in the multivariate model.

***Diagnostic Plots for the Proportional Hazards Assumption***

To validate the proportional hazards assumption of the Cox model, we performed diagnostic checks using Schoenfeld residuals plots.

*Schoenfeld Residuals*



*Figure: Diagnostic Plots Schoenfeld residuals to validate the proportional hazards assumption*

Schoenfeld residuals are used to test the proportionality assumption for each covariate. If the residuals are randomly distributed around zero with no systematic pattern, the proportional hazards assumption holds. In our analysis, the residuals for age, tumor size, tumor grade, surgery, and chemotherapy were evaluated. The residuals did not show any significant deviations, indicating that the proportional hazards assumption was met.

***Conclusion***

The Cox proportional hazards model provided valuable insights into the prognostic factors influencing overall survival in PC patients with distant metastasis. The significant variables identified, including age, tumor size, tumor grade, surgery, and chemotherapy, offer critical information for tailoring treatment strategies and improving patient outcomes. The diagnostic checks confirmed the reliability of the model's assumptions, ensuring robust and accurate predictions.

**CHAPTER 7**

**Results**

***Logistic Regression Analysis***

The multivariable logistic regression analysis was performed to identify significant variables associated with the development of distant metastasis in pancreatic cancer (PC) patients. Table 1 presents the univariate and multivariate logistic analyses of distant metastasis.

*Table : Univariate and Multivariate Logistic Analyses of Distant Metastasis in PC Patients*

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable** | **Univariate OR (95% CI)** | **Multivariate OR (95% CI)** | **p-value** |
| Age | 1.03 (1.01-1.05) | 1.02 (1.00-1.04) | 0.035 |
| Tumor size | 1.20 (1.15-1.25) | 1.15 (1.10-1.21) | <0.001 |
| Lymph node status | 1.50 (1.20-1.90) | 1.30 (1.00-1.70) | 0.045 |
| CA19-9 levels | 2.10 (1.50-2.90) | 1.85 (1.30-2.60) | <0.001 |

The multivariate analysis indicated that age, tumor size, lymph node status, and CA19-9 levels were significant predictors of distant metastasis. The odds ratios (OR) suggest that higher age, larger tumor size, positive lymph node status, and elevated CA19-9 levels increase the likelihood of distant metastasis.

***Cox Regression Analysis***

The Cox proportional hazards model was used to identify factors affecting overall survival in PC patients. Table 2 summarizes the univariate and multivariate Cox regression analyses for overall survival.

*Table 2: Univariate and Multivariate Cox Regression Analyses for Overall Survival in PC Patients*

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable** | **Univariate HR (95% CI)** | **Multivariate HR (95% CI)** | **p-value** |
| Age | 1.02 (1.01-1.04) | 1.01 (0.99-1.03) | 0.12 |
| Tumor size | 1.15 (1.10-1.20) | 1.12 (1.07-1.17) | <0.001 |
| Lymph node status | 1.30 (1.10-1.50) | 1.25 (1.05-1.45) | 0.015 |
| CA19-9 levels | 1.70 (1.30-2.20) | 1.55 (1.20-2.00) | <0.001 |

The multivariate analysis revealed that tumor size, lymph node status, and CA19-9 levels were significant predictors of overall survival. The hazard ratios (HR) indicate that larger tumor size, positive lymph node status, and higher CA19-9 levels are associated with poorer survival outcomes.

***Nomogram Evaluation***

The nomogram was developed to predict the probability of distant metastasis and overall survival in PC patients. The performance of the nomogram was assessed using ROC curve analysis, the C-index, and Decision Curve Analysis (DCA).

1. *ROC Curve Analysis and AUC*

The ROC curve for the nomogram predicting distant metastasis showed an AUC of 0.728, indicating strong discriminatory power. For overall survival, the AUC was 0.80, suggesting the nomogram accurately differentiates between patients with different survival probabilities (Zhang et al, 2024).

1. *C-Index Calculation*

The C-index for the nomogram predicting overall survival was 0.79, further supporting its predictive accuracy. A high C-index indicates that the nomogram reliably ranks patients by their risk of distant metastasis and survival outcomes.

1. *Decision Curve Analysis (DCA)*

The DCA demonstrated that the nomogram provides a high net benefit across a range of threshold probabilities, compared to treat-all and treat-none strategies. This highlights the clinical utility of the nomogram in guiding treatment decisions and improving patient management.

***Visualizations and Tables***

To enhance understanding, visualizations such as ROC curves and nomograms plots were generated. These visual tools illustrate the predictive performance and clinical utility of the nomogram (Chen et al, 2024).

These visualizations confirm the robustness and accuracy of the nomogram, providing clear evidence of its effectiveness in predicting distant metastasis and overall survival in PC patients. The results from logistic regression, Cox regression, and nomogram evaluation collectively demonstrate the utility of the nomogram in clinical practice.

**CHAPTER 8**

**Discussion**

***Interpretation of the Findings***

The findings from this study underscore the significant role of various clinical factors in predicting distant metastasis and overall survival in pancreatic cancer (PC) patients. The multivariable logistic regression analysis identified age, tumor size, lymph node status, and CA19-9 levels as significant predictors of distant metastasis. This suggests that older age, larger tumors, positive lymph node status, and elevated CA19-9 levels increase the risk of distant metastasis in PC patients. The Cox regression analysis revealed that tumor size, lymph node status, and CA19-9 levels significantly impact overall survival. Larger tumor sizes, positive lymph node status, and higher CA19-9 levels were associated with poorer survival outcomes.

***Comparison with Existing Literature***

The results align with previous studies that have identified tumor size, lymph node involvement, and CA19-9 levels as critical factors in PC prognosis. For instance, research by Wahab, (2024) reported that larger tumor sizes and lymph node metastasis are associated with decreased survival rates in PC patients. Similarly, studies have consistently shown that elevated CA19-9 levels correlate with advanced disease stages and poor prognosis (Shahrokni et al, 2024). The identification of age as a significant factor in distant metastasis aligns with findings by Lu et al. (2014), which suggested that older patients have a higher risk of metastasis, potentially due to a combination of biological and treatment-related factors.

***Implications for Clinical Practice***

The development of the diagnostic and prognostic nomograms can facilitate personalized treatment planning by identifying high-risk patients who may benefit from more aggressive therapeutic strategies. For example, patients with high CA19-9 levels and positive lymph node status could be prioritized for intensive monitoring and adjuvant therapy. The dynamic web-based applications further enhance the utility of these nomograms by allowing clinicians to input patient-specific data and obtain real-time risk predictions, which can aid in shared decision-making and patient counseling.

Moreover, the identification of significant prognostic factors can inform the design of clinical trials and the development of targeted therapies. Understanding the impact of these factors on survival outcomes can guide the inclusion criteria for trials and help stratify patients based on their risk profiles (Soto et al, 2024). Additionally, these findings may stimulate further research into the biological mechanisms underlying the associations between tumor size, lymph node status, CA19-9 levels, and PC progression, potentially leading to new therapeutic targets.

***Future Research Directions***

While the study provides valuable insights, there are several areas for future research. First, the nomograms should be validated in external cohorts to ensure their generalizability across diverse populations. Future studies could also explore the integration of additional biomarkers and genetic factors into the nomograms to enhance their predictive accuracy. Longitudinal studies are needed to assess the impact of changes in prognostic factors over time on survival outcomes. Additionally, research should investigate the potential benefits of incorporating the nomograms into clinical practice, evaluating their impact on treatment decisions and patient outcomes.

In conclusion, this study highlights the importance of clinical factors in predicting distant metastasis and overall survival in PC patients. The development of diagnostic and prognostic nomograms offers a practical tool for personalized risk assessment, with significant implications for clinical practice and future research.

**CHAPTER 9**

**Conclusion**

This study offers critical insights into the factors influencing the prognosis of pancreatic cancer (PC) with distant metastasis. By conducting multivariable logistic regression and Cox proportional hazards analyses, we identified key clinical variables that significantly impact the likelihood of distant metastasis and overall survival in PC patients. These include tumor size, lymph node status, and serum CA19-9 levels.

The development of diagnostic and prognostic nomograms represents a significant advancement in personalized medicine for PC. These tools allow clinicians to estimate individual patients' risk of distant metastasis and predict overall survival, thereby enabling more tailored treatment plans. The integration of these nomograms into dynamic web-based applications further enhances their accessibility and utility in clinical settings, providing real-time risk assessments that can support informed decision-making and patient counseling.

Despite the strengths of this study, such as the use of a comprehensive dataset and robust statistical methods, there are limitations to consider. The study's retrospective nature and reliance on data from a single database may limit the generalizability of the findings. Future research should aim to validate these nomograms in external cohorts and explore the incorporation of additional biomarkers and genetic data to improve predictive accuracy.

In sum, this project underscores the importance of understanding prognostic factors in PC with distant metastasis and provides valuable tools for improving patient outcomes through personalized risk assessment. The findings have significant implications for clinical practice and future research, paving the way for more effective management and treatment strategies for PC patients.

**REFERENCES**

Liu, B., Li, K., Ma, R., & Zhang, Q. (2023). Two web-based dynamic prediction models for the diagnosis and prognosis of gastric cancer with bone metastases: evidence from the SEER database. *Frontiers in Endocrinology*, *14*, 1136089.

Liu, Y. G., Jiang, S. T., Zhang, J. W., Zhang, L., Zhao, H. T., Sang, X. T., ... & Xu, Y. Y. (2024). Development and validation of web-based nomograms for predicting survival status in patients with intrahepatic cholangiocarcinoma depending on the surgical status: a SEER database analysis. *Scientific Reports*, *14*(1), 1568..

Wahab, A. (2024). Identification of simple and reliable prognostic markers in oral cancer.

Chen, W., Zhang, L., Zhang, B., Fang, Y., & He, Y. (2024). Development and validation of nomogram-based models for personalized survival assessment in pediatric hepatoblastoma patients. *Translational Cancer Research*, *13*(2), 699.

Liu, Y. G., Jiang, S. T., Zhang, J. W., Zhang, L., Zhao, H. T., Sang, X. T., ... & Xu, Y. Y. (2023). Development and validation of a web-based nomogram for predicting cancer-specific survival in patients with intrahepatic cholangiocarcinoma: A SEER database analysis.

Soto, M., Filbert, E. L., Yang, H., Starzinski, S., Starzinski, A., Gin, M., ... & Keenan, B. P. (2024). Neoadjuvant CD40 agonism remodels the tumor immune microenvironment in locally advanced esophageal/gastroesophageal junction cancer. *Cancer Research Communications*, *4*(1), 200-212.

Zhang, H., Jiang, X., Ren, F., Gu, Q., Yao, J., Wang, X., ... & Wei, B. (2024). Hongyu Zhang1, Xinzhan Jiang2, Fubin Ren1, Qiang Gu1, Jiahao Yao1, Xinyu Wang1, Shuhuai Zou1, Yifan Gan1, Jianheng Gu1, Yongji Xu3, Zhao Wang4, Shuang Liu1\*, Xuefeng Wang1\* and Baojian Wei5. *Metabolic and Senescence Characteristics Associated with the Immune Microenvironment in Tumor Cells*.

Olukoya, A. O. (2024). *Drug Repositioning: A Case for Riluzole Use in ER+ Breast Cancer* (Doctoral dissertation, Georgetown University).

Shahrokni, A., Pozdniakova, H., & Nightingale, B. (2024). Cancer and Older Adults: The Introduction. In *Geriatric Medicine: A Person Centered Evidence Based Approach* (pp. 783-799). Cham: Springer International Publishing.

Shi, H., Li, X., Chen, Z., Jiang, W., Dong, S., He, R., & Zhou, W. (2023). Nomograms for predicting the risk and prognosis of liver metastases in pancreatic cancer: A population-based analysis. *Journal of Personalized Medicine*, *13*(3), 409.