The Clinicopathological Features and Survival Outcomes of Patients with Different Metastatic Sites in Stage IV Endometrial Cancer: A SEER-Based Cohort Study

Mohammed Zuber

**Authors**

* Mohammed Zuber (ORCID: 0000-0000-1234-5678)

**Author affiliations**

1. College of Pharmacy, University of Georgia, Athens, GA, USA.

Corresponding author: mohammedzuber099@gmail.com

Disclaimer: The opinions expressed in this article are the author’s own and don’t reflect their employer.

# 1. Summary/Abstract

*Write a summary of your project.*

# 2. Introduction

## 2.1 General Background Information

Endometrial cancer, a neoplasm originating from the endometrial lining, is the most common gynecological cancer in the United States.(1,2) According to American Cancer Society 2024 statistics the incidence of uterine cancer is 67, 880 with 13,250 deaths. Due to its early symptomatic presentation, 70% of cases are diagnosed at stage I, where the five-year survival rate exceeds 90%. (2,3) However, around 13% of patients are diagnosed at advanced stages (III/IV).(3)

A retrospective study found that the prevalence of stage IVB endometrial cancer increased to 5.67% between 2010 and 2015.(4) Unlike stage I, which has a favorable prognosis, stage IV is associated with poor outcomes, with a five-year survival rate ranging from 17% to 30%.(5) The most frequent metastatic sites include the peritoneum, lungs, pelvic and para-aortic lymph nodes, and local pelvic recurrences.(6) Among these, lung metastasis is one of the most common distant metastases, occurring in 1.5% of patients.(7) The hematogenous spread is the primary mechanism of lung metastasis.(8) Depending on clinical features, disease stage, prior surgical history, and pathological factors, 11-13% of endometrial cancer patients experience recurrence within the first two years after initial treatment. Symptoms of lung metastasis typically include coughing up blood, shortness of breath, chest pain, persistent cough, pleural effusion, reduced appetite, and weight loss.(9)

Compared to other gynecological cancers such as cervical and ovarian cancer, endometrial carcinoma exhibits the highest rate of pulmonary metastasis, affecting up to 20-25% of patients who experience recurrence. According to Mao et al., 2020, the five-year cancer-specific survival (CSS) rate for liver and brain metastases was 21%, while bone metastases had a survival rate of 20%. In patients with lung metastases, the median overall survival was 11 months, with a five-year CSS rate of 19%. However, this data specifically pertains to stage IV lung cancer patients.(7) Patients with advanced endometrial cancer face poor survival outcomes, with metastatic disease being the leading cause of mortality. (7) Therefore, this study aimed to utilize data from the Surveillance, Epidemiology, and End Results (SEER) database to analyze survival outcomes in patients diagnosed with stage IV endometrial cancer. The primary objective was to determine whether the presence of lung/brain/bone/liver metastases in adults with stage IV endometrial cancer is associated with worse survival outcomes compared to patients without any metastasis.

## 2.2 Description of data and data source

This study will utilize the Surveillance, Epidemiology, and End Results (SEER) database, one of the most comprehensive cancer incidence and survival registries, covering over 48% of the United States. SEER collects key demographic data, cancer characteristics, treatment details, and survival outcomes. Managed by the U.S. National Cancer Institute, the database began recording information on lung, bone, liver, and brain metastases in 2010. Data for this study will be accessed through SEER\*Stat 8.4.4 software, using the SEER Research Data (2000–2019) from 17 registries.

The extracted variables include age, race, income, American Joint Committee on Cancer (AJCC) stage, year of cancer diagnosis, lung metastasis (yes/no), brain metastasis (yes/no), liver metastasis (yes/no), brain metastasis (yes/no), cause of death, survival months, vital status (dead/alive), rural/urban classification.

## 2.3 Questions/Hypotheses to be addressed

*Question* What are the clinicopathological features and survival outcomes of patients with different metastatic sites in stage IV endometrial cancer?

*Hypothesis* Patients with stage IV endometrial cancer exhibit distinct clinicopathological characteristics and survival outcomes based on the site of metastasis. Specifically, metastases to different organ sites (e.g., lung, liver, bone, and brain) are associated with varying prognostic implications, with some metastatic patterns demonstrating significantly poorer survival compared to others. By analyzing data from the SEER database, we hypothesize that certain metastatic sites will be linked to more aggressive tumor behavior and worse overall survival, highlighting potential differences in disease progression and treatment response.

# 3. Methods

This study examines survival outcomes in stage IV endometrial cancer using SEER data. The primary outcomes include survival months, cause-specific death classification, and vital status at study cutoff. Key predictors include age, race, income, year of diagnosis, and presence of differnt metastasis status.

## 3.1 Data aquisition

The data was extracted from SEER\*Stat 8.4.4 software, using the SEER Research Data (2000–2019) from 17 registries. The following inclusion criteria was used to extract the data (1) primary site, C54.1–9 and C55.9 based on International Classification of Tumor Diseases for Oncology (ICD-O) codes; (2) year of diagnosis 2010-2015; (3) AJCC stage IV; (4) age more than 18 years.

The SEER database provided data on both demographic and clinical characteristics of the patients. The demographic variables included age at diagnosis, race, median household income (inflation-adjusted to 2019), year of diagnosis, and Derived AJCC Stage Group (7th edition, 2010-2015). The clinical variables comprised lung, brain, bone, and liver metastasis status, survival duration (in months), SEER cause-specific death classification, and vital status recode (used for study cutoff determination).

Age will be categorized into four groups: 18-39, 40-59, 60-79, and ≥80 years. Race is classified into Asian or Pacific Islander, American Indian or Alaska Native, Black, and White. The income variable will be divided into six brackets: < $35,000, $35,000 - $44,999, $45,000 - $54,999, $55,000 - $64,999, $65,000 - $74,999, and $75,000+. The SEER cause-specific death classification was preferred over all-cause mortality to ensure that only deaths specifically attributed to endometrial cancer will be included in the analysis.

## 3.2 Data import and cleaning

Endometrial cancer cases were extracted from the SEER database. Individuals under 18 years of age were removed from the dataset. Additionally, patients whose cause of death is unrelated to endometrial cancer or whose cause of death is missing or unknown were not included in the analysis. Cases diagnosed at AJCC stages I to III was excluded from the study. The final cohort was categorized into 4 groups: lung metastasis, brain metastasis, bone metastasis, and liver metastasis.The SEER dataset underwent a structured cleaning process to ensure data integrity and suitability for analysis. First, data types were standardized by converting categorical variables such as Year of Diagnosis, RX Summ–Surg Prim Site (1998+), and Patient ID to character format, while Survival Months was converted to numeric. Key categorical variables, including Sex, Race, Metastasis Indicators, Survival Flags, and Cause-Specific Death Classification, were transformed into factors to maintain consistency. Patients within the age categories “01-04 years” and “15-19 years” were excluded, and only those classified as Stage IVA or IVB in the Derived AJCC Stage Group, 7th ed (2010-2015) were retained. Additionally, patients with “Dead (missing/unknown COD)” or “N/A not seq 0-59” in the SEER cause-specific death classification were removed. To ensure completeness, individuals with missing or unknown survival months were excluded. The final cleaned dataset was saved for further analysis, ensuring consistency and reproducibility. The code and raw data to reproduce the data cleaning step is available in https://github.com/mohammedzuber099/ZUBER-MADA-project.

## 3.3 Statistical analysis

The analysis aims to identify survival differences based on metastatic patterns, demographic disparities, socioeconomic influences, and temporal trends in prognosis. Kaplan-Meier survival analysis will used to compare the survival among the study cohorts. The incidence of metastasis will be calculated in stage IV endometrial cancer. In the bivariate analysis, we fitted separate Cox regression models for each variable, including metastasis status (bone, brain, liver, and lung), race, age group, year of diagnosis, and median household income, to estimate their individual hazard ratios (HRs). For the multivariate analysis, we constructed a Cox model incorporating all covariates simultaneously to adjust for potential confounders and determine independent predictors of survival. A Cox regression model was performed to compute the adjusted Hazard Ratios (aHR).

### 3.3.1 Machine leaning model

I applied a Random Survival Forest (RSF) model to evaluate the impact of clinical and demographic variables on overall survival time. RSF is a non-parametric ensemble learning method specifically designed for right-censored survival data. It extends the traditional random forest approach by using survival splitting rules and ensemble cumulative hazard estimates. The outcome of interest was overall survival time in months, with vital status (0 = alive, 1 = dead) as the event indicator. Predictor variables included metastasis status (bone, liver, lung, brain), age group, race, year of diagnosis, and median household income. All categorical variables were converted to factors, as required by the randomForestSRC package in R. The model was trained using 500 trees, and variable importance was assessed using the impurity measure. Model performance was evaluated through time-dependent ROC analysis, confusion matrix, and individualized survival curve prediction. The RSF approach enabled flexible modeling of complex interactions and non-linear effects, offering an interpretable and robust alternative to traditional Cox models.

# 4. Results

## 4.1 Exploratory/Descriptive analysis

[Table 1](#tbl-summarytable) shows a summary of the data.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 1: Data summary table.   | Variable | Type | Unique\_Values | Missing\_Values | | --- | --- | --- | --- | | age\_group | character | 14 | 0 | | sex | factor | 1 | 0 | | diagnosis\_year | character | 6 | 0 | | race | factor | 5 | 0 | | ethnicity | factor | 2 | 0 | | race\_ethnicity | factor | 6 | 0 | | ICD-O-3 Hist/behav | character | 59 | 0 | | ICD-O-3 Hist/behav, malignant | character | 59 | 0 | | Combined Summary Stage (2004+) | character | 2 | 0 | | cancer\_stage | factor | 2 | 0 | | surgery\_primary\_site | character | 30 | 0 | | RX Summ–Scope Reg LN Sur (2003+) | character | 9 | 0 | | RX Summ–Reg LN Examined (1998-2002) | character | 1 | 0 | | mets\_bone | factor | 4 | 0 | | mets\_brain | factor | 4 | 0 | | mets\_liver | factor | 4 | 0 | | mets\_lung | factor | 4 | 0 | | mets\_distant\_ln | factor | 1 | 0 | | cause\_of\_death | factor | 52 | 0 | | cause\_specific\_death | factor | 2 | 0 | | other\_cause\_of\_death | factor | 2 | 0 | | survival\_months | numeric | 120 | 0 | | survival\_flag | factor | 4 | 0 | | cod\_km | factor | 53 | 0 | | vital\_status | factor | 2 | 0 | | First malignant primary indicator | character | 2 | 0 | | patient\_id | character | 5706 | 0 | | median\_income | character | 10 | 0 | | rural\_urban\_code | character | 6 | 0 | |

[Table 2](#tbl-summarytablesurvival) gives the overview of the survival. The survival months in our dataset range from 0 to 119 months, with a mean survival of 23.1 months and a median survival of 12 months, indicating a right-skewed distribution. The standard deviation (SD) is 27.67 months, reflecting high variability in patient survival. A total of 5,839 patients were included in the analysis. This distribution suggests that while some patients survive long-term, the majority have relatively short survival durations.

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 2: Summary of Survival.   | Min\_Survival | Max\_Survival | Mean\_Survival | Median\_Survival | SD\_Survival | Count | | --- | --- | --- | --- | --- | --- | | 0 | 119 | 23.07463 | 12 | 27.56729 | 5708 | |

[Figure 1](#fig-survival) shows the distribution of survival months. This histogram illustrates the distribution of survival months among patients in the dataset. The x-axis represents survival duration in months, while the y-axis shows the count of patients within each survival interval. The distribution is right-skewed, indicating that most patients had shorter survival times, with a gradual decline in frequency as survival duration increases. A substantial proportion of patients survived less than 12 months, while fewer cases exhibited long-term survival beyond 60 months. The presence of some patients with survival exceeding 100 months highlights potential long-term survivors.

|  |
| --- |
| Figure 1: Survival months. |

[Table 3](#tbl-summaryage) and [Figure 2](#fig-age) shows age distribution.The dataset includes patients across a wide age range, with the majority falling within the 60-69 years age group. The highest number of patients are aged 60-64 years (1,006 patients), followed closely by 65-69 years (956 patients). The number of cases gradually decreases in older age groups, with 389 patients aged 80-84 years and 280 patients aged 85+ years. The youngest age groups (20-39 years) have significantly fewer cases, with only 3 patients aged 20-24 years and 64 patients aged 35-39 years. This distribution indicates that the majority of the study population consists of older adults, which aligns with typical cancer incidence patterns that increase with age.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 3: Summary of age.   | age\_group | Count | | --- | --- | | 20-24 years | 3 | | 25-29 years | 20 | | 30-34 years | 29 | | 35-39 years | 62 | | 40-44 years | 150 | | 45-49 years | 276 | | 50-54 years | 508 | | 55-59 years | 810 | | 60-64 years | 988 | | 65-69 years | 939 | | 70-74 years | 748 | | 75-79 years | 537 | | 80-84 years | 378 | | 85+ years | 260 | |

[Figure 2](#fig-age) shows this distribution indicates that the majority of the study population consists of older adults, which aligns with typical cancer incidence patterns that increase with age.

|  |
| --- |
| Figure 2: Age |

[Table 4](#tbl-summarymetastasis) and [Figure 3](#fig-mets) shows distribution for different sit for metastasis. The table presents the distribution of metastatic sites at the time of diagnosis which is the main exposure in our study. Lung metastases were the most common, observed in 1,676 patients, followed by liver metastases (844 patients). Bone metastases were identified in 556 patients, while brain metastases were the least frequent, affecting 160 patients.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 4: Summary of metastasis.   | Metastasis\_Site | Count | | --- | --- | | Bone\_Mets | 556 | | Brain\_Mets | 160 | | Liver\_Mets | 844 | | Lung\_Mets | 1676 | |

[Figure 3](#fig-mets) shows this distribution suggests that lung and liver are the most frequent metastatic sites, which aligns with common metastatic patterns observed in many cancers. The lower occurrence of brain metastases may reflect differences in metastatic tropism or detection rates.

|  |
| --- |
| Figure 3: Metatstasis |

## 4.2 Basic statistical analysis

### 4.2.1 Bivariate Analysis: Metastasis and Survival

In the bivariate Cox proportional hazards analysis, all four metastasis types were significantly associated with an increased risk of mortality. Bone metastasis was associated with a 56% higher hazard of death (HR = 1.56), while brain metastasis showed the highest risk, nearly a 92% increased hazard of death (HR = 1.92). Liver metastasis was also strongly associated with worse survival, showing an 81% higher hazard of death (HR = 1.81), and lung metastasis increased the hazard of death by 63% (HR = 1.63). These findings indicate that the presence of metastases, particularly in the brain and liver, is a strong predictor of poor survival outcomes. [Table 5](#tbl-summaryhrmetastasis) shows the HR for different metastatis sites

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 5: Proportional hazards analysis for different metastasis.   | term | estimate | std.error | statistic | p.value | | --- | --- | --- | --- | --- | | Bone Metastasis | 1.561311 | 0.0484300 | 9.199381 | 0 | | Brain Metastasis | 1.923033 | 0.0881900 | 7.414714 | 0 | | Liver Metastasis | 1.812380 | 0.0403062 | 14.753090 | 0 | | Lung Metastasis | 1.635001 | 0.0318740 | 15.424583 | 0 | |

[Figure 4](#fig-km-curve-bone-mets) illustrates the survival probability of patients with and without bone metastasis over time. The red curve represents patients without bone metastasis, while the blue curve represents those with bone metastasis. Patients with bone metastasis have significantly worse survival outcomes, as indicated by the lower survival probability throughout the follow-up period.

|  |
| --- |
| Figure 4: Kaplan-Meier survival curves for Bone Metastasis |

[Figure 5](#fig-km-curve-brain-mets) shows that patients with brain metastasis have significantly lower survival probabilities than those without. The log-rank test (p < 0.0001) confirms a statistically significant difference, indicating that brain metastasis is associated with worse survival outcomes.

|  |
| --- |
| Figure 5: Kaplan-Meier survival curves for Brain Metastasis |

[Figure 6](#fig-km-curve-liver-mets) indicates that patients with liver metastasis (blue) have significantly lower survival probabilities than those without (red). The log-rank test (p < 0.0001) confirms a statistically significant difference, suggesting that liver metastasis is associated with poorer survival outcomes.

|  |
| --- |
| Figure 6: Kaplan-Meier survival curves for Liver Metastasis |

[Figure 7](#fig-km-curve-lung-mets) shows that patients with lung metastasis (blue) have significantly lower survival probabilities than those without (red). The log-rank test (p < 0.0001) confirms a statistically significant difference, indicating that lung metastasis is associated with worse survival outcomes.

|  |
| --- |
| Figure 7: Kaplan-Meier survival curves for Lung Metastasis |

### 4.2.2 Bivariate Analysis: Race and Survival

Patients identified as had significantly worse survival outcomes than White patients, with a 37% higher risk of death (HR = 1.37, p < 0.0001). In contrast, Asian patients had a lower risk of death compared to White patients (HR = 0.86, p = 0.0076), suggesting a potential survival advantage. The survival difference between American Indian and White patients was not statistically significant (HR = 1.21, p = 0.29), indicating no clear evidence of an increased or decreased risk. [Table 6](#tbl-summaryhrrace) shows the HR for patients with different race.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 6: Proportional hazards analysis for different Race.   | term | estimate | std.error | statistic | p.value | | --- | --- | --- | --- | --- | | Black vs. White | 1.3748971 | 0.0375889 | 8.470012 | 0.0000000 | | Asian vs. White | 0.8648224 | 0.0544274 | -2.668346 | 0.0076226 | | American Indian vs. White | 1.2069971 | 0.1776483 | 1.059034 | 0.2895843 | |

### 4.2.3 Bivariate Analysis: Age and Survival

The hazard ratio (HR) estimates for different age groups indicate a general trend of increasing risk with age, though statistical significance varies. Compared to the reference group (20-24 years), younger age groups (25-74 years) show HRs between 3.5 and 6.2, but their p-values are above 0.05, suggesting no statistically significant difference in survival risk. However, individuals aged 80-84 years (HR = 7.99, p = 0.038) and 85+ years (HR = 11.01, p = 0.017) have significantly higher mortality risk, indicating poorer survival outcomes in older patients. These findings suggest that age is an important predictor of survival, with the highest risk observed in the oldest age group. [Table 7](#tbl-summaryhrage) shows the HR for diferrent age groups.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 7: Proportional hazards analysis for different age group.   | term | estimate | std.error | statistic | p.value | | --- | --- | --- | --- | --- | | 25-29 | 3.932239 | 1.035126 | 1.322746 | 0.1859200 | | 30-34 | 3.585583 | 1.025991 | 1.244573 | 0.2132888 | | 35-39 | 3.572297 | 1.012148 | 1.257927 | 0.2084181 | | 40-44 | 3.536501 | 1.004861 | 1.257027 | 0.2087438 | | 45-49 | 4.035494 | 1.002636 | 1.391461 | 0.1640858 | | 50-54 | 4.762009 | 1.001340 | 1.558582 | 0.1190954 | | 55-59 | 4.913059 | 1.000828 | 1.590579 | 0.1117043 | | 60-64 | 5.207184 | 1.000657 | 1.648955 | 0.0991568 | | 65-69 | 5.110758 | 1.000690 | 1.630223 | 0.1030545 | | 70-74 | 6.178850 | 1.000836 | 1.819611 | 0.0688182 | | 75-79 | 6.935713 | 1.001099 | 1.934557 | 0.0530447 | | 80-84 | 7.993028 | 1.001564 | 2.075323 | 0.0379566 | | age\_group85+ years | 11.008004 | 1.002149 | 2.393480 | 0.0166894 | |

### 4.2.4 Multivariable analysis

The multivariable Cox proportional hazards model results indicate that metastasis to the bone (HR = 1.30, p < 0.001), brain (HR = 1.72, p < 0.001), liver (HR = 1.66, p < 0.001), and lung (HR = 1.51, p < 0.001) are all significantly associated with poorer survival outcomes. Among racial groups, Black patients (HR = 1.39, p < 0.001) had significantly worse survival compared to White patients, while Asian or Pacific Islander and American Indian/Alaska Native groups did not show a statistically significant difference.

Regarding age, survival risk generally increased with older age, with statistically significant increases observed in the 70-74 years (HR = 7.26, p = 0.048), 75-79 years (HR = 7.91, p = 0.039), 80-84 years (HR = 9.53, p = 0.025), and 85+ years (HR = 12.75, p = 0.011) groups, indicating poorer survival outcomes in older patients.

Year of diagnosis did not show a consistent or significant trend, suggesting no clear improvement or decline in survival based on the year of diagnosis. Household income was associated with survival, with patients in the $50,000 - $54,999 (HR = 1.16, p = 0.011) and $65,000 - $69,999 (HR = 1.12, p = 0.030) categories showing significantly higher risk compared to those earning $75,000+. However, other income groups did not show significant differences.

Overall, metastasis, age, and race appear to be key determinants of survival, with Black race and older age groups associated with poorer outcomes.

### 4.2.5 Random Survival Forest

A Random Survival Forest (RSF) model was used to evaluate the contribution of clinical and demographic predictors to overall survival. The RSF model demonstrated strong performance in predicting survival at 12 months, achieving an AUC of 0.74 ([Figure 8](#fig-rsf-auc)). The confusion matrix revealed a sensitivity of 86.5% and a specificity of 47.9%, with an overall accuracy of 53.8% and balanced accuracy of 67.2%, indicating that the model was effective in identifying patients at high risk of death within the first year.

Variable importance analysis revealed that age group, median household income, and year of diagnosis were the most influential predictors of survival ([Figure 9](#fig-rsf-variable-importance)). Other important variables included race and metastasis to the bone, lung, liver, and brain. The individualized survival prediction for a representative patient showed a steady decline in survival probability over time, reflecting the nature of long-term risk in this cohort ([Figure 10](#fig-rsf-ind-sur-pre)). [Figure 11](#fig-rsf-con-matrix). Confusion matrix showing prediction results of the Random Survival Forest model at 12 months. The model correctly classified 709 survivors and 2,182 non-survivors, with moderate misclassification of 2,376 survivors and 111 non-survivors.

|  |
| --- |
| Figure 8: Random Survival Forest AUC |

|  |
| --- |
| Figure 9: Random Survival Forest Variable Importance |

|  |
| --- |
| Figure 10: Random Survival Forest Individualized Survival Prediction |

|  |
| --- |
| Figure 11: RSF Confusion Matrix (12 Months) |

# 5. Discussion

## 5.1 Summary and Interpretation

*Summarize what you did, what you found and what it means.*

## 5.2 Strengths and Limitations

*Discuss what you perceive as strengths and limitations of your analysis.*

## 5.3 Conclusions

*What are the main take-home messages?*

*Include citations in your Rmd file using bibtex, the list of references will automatically be placed at the end*

This paper (10) discusses types of analyses.

These papers (11,12) are good examples of papers published using a fully reproducible setup similar to the one shown in this template.

Note that this cited reference will show up at the end of the document, the reference formatting is determined by the CSL file specified in the YAML header. Many more style files for almost any journal [are available](https://www.zotero.org/styles). You also specify the location of your bibtex reference file in the YAML. You can call your reference file anything you like.

# 6. References

1. Monk BJ, Smith G, Lima J, et al. Real-world outcomes in patients with advanced endometrial cancer: A retrospective cohort study of US electronic health records. *Gynecologic Oncology* [electronic article]. 2022;164(2):325–332. (<http://dx.doi.org/10.1016/j.ygyno.2021.12.008>)

2. Morice P, Leary A, Creutzberg C, et al. Endometrial cancer. *The Lancet* [electronic article]. 2016;387(10023):1094–1108. (<https://www.sciencedirect.com/science/article/pii/S0140673615001300>)

3. Galaal AM K, Lawrie T. Adjuvant chemotherapy for advanced endometrial cancer. *Cochrane Database of Systematic Reviews* [electronic article]. 2014;(5). (<https://doi.org//10.1002/14651858.CD010681.pub2>)

4. Guo J, Cui X, Zhang X, et al. The clinical characteristics of endometrial cancer with extraperitoneal metastasis and the value of surgery in treatment. *Technology in Cancer Research & Treatment* [electronic article]. 2020;19:1533033820945784. ( [https://doi.org/10.1177/1533033820945784]( https://doi.org/10.1177/1533033820945784))

5. Carvalho JP, Del Giglio A, Achatz MI, et al. Complete clinical response in stage IVB endometrioid endometrial carcinoma after first-line pembrolizumab therapy: Report of a case with isolated loss of PMS2 protein. *Case Reports in Oncology* [electronic article]. 2020;13(3):1067–1074. (<https://doi.org/10.1159/000510000>)

6. De Nardo C, Landini N, Massaro M, et al. “Cheerios in the lung” as first metastases from endometrial endometrioid adenocarcinoma with adequate response to immunotherapy. *Egyptian Journal of Radiology and Nuclear Medicine* [electronic article]. 2023;54(1). (<http://dx.doi.org/10.1186/s43055-023-01079-w>)

7. Mao W, Wei S, Yang H, et al. Clinicopathological study of organ metastasis in endometrial cancer. *Future Oncology* [electronic article]. 2020;16(10):525–540. (<http://dx.doi.org/10.2217/fon-2020-0017>)

8. Ballon SC, Berman ML, Donaldson RC, et al. Pulmonary metastases of endometrial carcinoma. *Gynecologic Oncology* [electronic article]. 1979;7(1):56–65. (<http://dx.doi.org/10.1016/0090-8258(79)90081-7>)

9. Jamil A, Kasi A. Lung metastasis. In: *StatPearls*. Treasure Island (FL): StatPearls Publishing; 2025

10. Leek JT, Peng RD. [Statistics. What is the question?](https://doi.org/10.1126/science.aaa6146) *Science (New York, N.Y.)*. 2015;347(6228):1314–1315.

11. McKay B, Ebell M, Billings WZ, et al. [Associations Between Relative Viral Load at Diagnosis and Influenza A Symptoms and Recovery.](https://doi.org/10.1093/ofid/ofaa494) *Open forum infectious diseases*. 2020;7(11):ofaa494.

12. McKay B, Ebell M, Dale AP, et al. [Virulence-mediated infectiousness and activity trade-offs and their impact on transmission potential of influenza patients.](https://doi.org/10.1098/rspb.2020.0496) *Proceedings. Biological sciences*. 2020;287(1927):20200496.