BMJ Open

A nomogram for predicting cancer-specific survival in advanced endometrial carcinoma after surgery: A population-based analysis

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2022-070893
Article Type:	Original research
Date Submitted by the Author:	07-Dec-2022
Complete List of Authors:	Zheng, Chunqin; Shantou Central Hospital, Department of Obstetrics and Gynecology Zheng, Zhixiang; Shantou Central Hospital, Department of Obstetrics and Gynecology Liang, Xiaoling; Shantou Central Hospital, Department of Obstetrics and Gynecology Ni, Yanhong; Shantou Central Hospital Department of Obstetrics and Gynecology Fan, Fufang; Shantou Central Hospital Department of Obstetrics and Gynecology Zhang, Peili; Shantou Central Hospital Department of Obstetrics and Gynecology Fang, Danmei; Shantou Central Hospital Department of Obstetrics and Gynecology Ma, Ruijun; Shantou Central Hospital, Department of Obstetrics and Gynecology Huang, Xiaojun; Shantou Central Hospital, Department of Obstetrics and Gynecology
Keywords:	Gynaecological oncology < GYNAECOLOGY, Surgical pathology < PATHOLOGY, Radiation oncology < RADIOLOGY & IMAGING, Gynaecological oncology < ONCOLOGY, HEALTH ECONOMICS

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

A nomogram for predicting cancer-specific survival in advanced endometrial carcinoma after surgery: A population-based analysis

Chunqin Zheng[1*], Zhixiang Zheng[1], Xiaoling Liang[1], Yanhong Ni[1], Fufang Fan[1], Peili Zhang[1], Danmei Fang[1], Ruijun Ma[1], Xiaojun Huang[1]

[1]Department of Obstetrics and Gynecology, Shantou Central Hospital, Shantou 515000, China

[*] **Correspondence:**

Chunqin Zheng, Department of Obstetrics and Gynecology, Shantou Central Hospital, Shantou 515000, China. 467430374@qq.com; ORCID:0000-0003-1840-9173.

Keywords: endometrial carcinoma, nomogram, cancer-specific survival, NRI, IDI, FIGO stage

Abstract

Introduction Advanced endometrial carcinoma has a poor five-year survival rate. In this study, we developed a nomogram to predict cancer-specific survival (CSS) after surgery in patients with advanced endometrial carcinoma.

Methods From the Surveillance, Epidemiology, and End Results database, 5445 patients with advanced endometrial carcinoma were selected between 2004 and 2015. The patients were divided into training and validation cohorts. Variables for the nomogram were selected using the least absolute shrinkage and selection operator methods. The concordance index (C-index), area under the time-dependent receiver operating characteristic curve (time-dependent AUC), and calibration coefficients were used to evaluate the discrimination and calibration of the nomogram. Integrated discrimination improvement (IDI) and net reclassification index (NRI) were used to

compare the clinical utility of the nomogram with that of the International Federation of Gynecology and Obstetrics stage. Seven variables were selected to establish a

nomogram.

Results The nomogram had satisfactory discriminatory power by C-index (0.7228 for the training cohort and 0.7528 for the validation cohort) and time-dependent AUC (>0.7). Both the training and validation cohort calibration plots showed good agreement between nomogram predictions and actual observations. For the training cohort, the NRI values were 0.33, 0.308, and 0.303 for one, three, and five years, respectively, and the IDI values were 0.052, 0.089, and 0.103 for one, three, and five years of CSS prediction, respectively. For the validation cohort, the NRI values were 0.446, 0.41, and 0.396 for one, three, and five years, respectively, and the IDI values were 0.081, 0.126, and 0.145 for one, three, and five years of CSS prediction, respectively.

Conclusions In these studies, the nomogram significantly outperformed the International Federation of Gynecology and Obstetrics stage alone ($P < 0.05$). The development and validation of a prognostic nomogram may help clinicians assess the prognosis of patients with advanced endometrial carcinoma after surgery.

Introduction

Endometrial cancer is the sixth most common cancer among women, with 417,000 new cases diagnosed worldwide by 2020 [1]. Approximately 65,950 new cases of endometrial cancer are expected to be diagnosed in the United States by 2022, with 12,550 deaths expected (http://seer.cancer.gov/statfacts/html/corp.html. Cancer Stat Facts: Uterine Cancer.). There are two histological types of endometrial cancer [2 3]. Type I tumors include tumors with grade 1 or 2 endometrioid histological classifications and account for approximately 80% of endometrial cancers. Type II tumors account for 10%–20% of endometrial cancers and include grade 3 endometrioid tumors and tumors with non-endometrioid histology. The treatment of endometrial cancer is primarily surgical, with radiation and chemotherapy as common adjuvant modalities. For patients with endometrial cancer who will undergo surgery, adjuvant therapy determines disease recurrence for risk stratification on the basis of tumor stage, tumor histology, and other

pathologic factors. Endometrial cancer is divided into three groups: low, intermediate, and high risk [4]. High-risk endometrial cancer type includes International Federation of

Gynecology and Obstetrics (FIGO) stage III or higher endometrial cancer regardless of histology or grading [4]. The prognosis of endometrial cancer depends mainly on the stage and histological type of the disease (including grading and histological subtypes). In general, the five-year survival rates are 80%–90% and 70%–80% for stage I and II endometrial cancers, respectively, and 20%–60% for stage III and IV endometrial cancers [5 6]. Stage III and IV endometrial cancers are classified as advanced and high-risk endometrial cancer, and a wide range of fluctuations in five-year survival rates has been reported in the literature. It is necessary to explore the factors associated with tumor-specific survival, and the establishment of a prognostic prediction model for tumor-specific survival after surgery for advanced endometrial cancer could benefit clinicians because it may better guide oncologists in providing individualized treatment and predicting patient prognosis.

A nomogram is a simple visualization tool used by oncologists to predict and quantify patient survival on the basis of multiple variables. In contrast to the FIGO stage, nomograms focus on individual prognoses. Furthermore, nomograms play an important role in risk management, personalized clinical trials, and trial design. A nomogram has been used for patients with endometrial carcinoma [7], and Yang et al. [8] published a nomogram for patients with stage IIIC endometrial cancer following surgery; however, there is no specific prognostic prediction for advanced endometrial cancer patients after surgery. Consequently, we used 2004–2015 data from the Surveillance, Epidemiology, and End Results (SEER) database to construct a prognostic nomogram for patients with advanced endometrial carcinoma after surgery to compensate for this deficiency. The SEER database, which is the population-based database of the National Cancer Institute, covers all 18 US states, includes approximately 28% of the US population, and contains highly rigorous data [9].

Materials and Methods

Patient enrollment and variables

SEER *Stat version 8.4.0.1 (https://seer.cancer.gov/seerstat/) was used to extract data

from 2004 and 2015 on patients with endometrial carcinoma. The inclusion criteria were as follows: primary sites, C54.1-9 and C55.9 [10]; site and morphology, 8380/3 (based on the *International Classification of Tumor Diseases for Oncology* [ICD-O], Third Edition); histology, adenomas and adenocarcinomas; FIGO stage, III/IV; and therapy, surgical treatment. There were eight exclusion criteria: undetermined cause of death; undetermined survival time, survival time, or survival time < one month; (3) undetermined tumor size; undetermined lymphadenectomy; (5) unknown regional node status; unknown tumor grade; unknown months from diagnosis to treatment; and unknown race. A flow chart for patient selection is shown in Figure 1.

Variables including age at diagnosis, year of diagnosis, tumor size, race, marital status, histologic stage, tumor grade, FIGO stage, lymphadenectomy, regional node positivity, chemotherapy, radiation status, months from diagnosis to treatment, survival time, median household income, and cancer-specific survival (CSS) were collected from the SEER database. There were two categories of radiation status (with and without radiation) and chemotherapy status (with and without chemotherapy). Marital status was classified as unmarried (single, unmarried, or domestic partner), married, other (divorced, widowed, separated), and unknown. Grades were associated with tumors. ICD-O-2 defines grade I as well differentiated, grade II as moderately differentiated, grade III as poorly differentiated, and grade IV as undifferentiated. According to the SEER registry, income was examined as aggregate data, which are based on the US median income. Median household income in USD was reclassified into three groups: ≤54,999, 55,000–69,999, and ≥70,000. The historic stage was derived from the Collaborative Stage for 2004–2015 and was divided into in situ, localized, regional, distant, and unknown categories. In the localized stage, an invasive neoplasm is confined entirely to the organ of origin. In the regional stage, a neoplasm has extended 1) beyond the limits of the organ of origin directly into surrounding organs or tissues, 2) into regional lymph nodes by way of the lymphatic system, or 3) into regional lymph nodes by a combination of extension and regional lymph nodes. In the distant stage, a

neoplasm has spread to parts of the body remote from the primary tumor. This study categorizes lymphadenectomy into two categories: with and without regional lymph

node dissection. Failure to perform lymph node dissection included the failure to remove or aspirate local lymph nodes, local lymph node biopsy or aspiration, and sentinel lymph node biopsy only. Lymph node dissection includes the removal of an unknown number of regional lymph nodes, removal of one to three regional lymph nodes, removal of four or more regional lymph nodes, and regional lymph node dissection with anterior lymph node biopsy.

Statistical analysis

X-tile (Yale University, New Haven, Connecticut, USA) was used to determine the cutoff values for age at diagnosis, the tumor size, the positive regional nodes, and the risk stratification [11]. Statistical analyses were conducted using R 4.0.2 (R Foundation for Statistical Computing, Vienna, Austria, http://www.r-project.org) in the RStudio environment and by using Free Statistics 1.7 (Beijing FreeClinical Medical Technology Co., LTD). CSS was the primary endpoint of the study. Cases were randomly placed in the training and validation cohorts at a 7:3 ratio. Categorical variables are presented as frequencies and proportions. Chi-square tests were used to compare clinicopathological characteristics between the training and validation cohorts. $P < 0.05$ was considered statistically significant. To avoid overfitting, the least absolute shrinkage and selection operator (LASSO) method was used. In addition, significant prognostic factors were identified from the Cox proportional hazards model. A nomogram associated with CSS was then constructed and incorporated into the known prognostic factors. The nomogram performance was validated by training and validation. The nomogram model was evaluated using the area under the time-dependent receiver operating characteristic curve (time-dependent AUC) to assess their discriminative abilities, and calibration curves were plotted to compare the predicted CSS with the actual CSS after one, three, and five years. The area under the curve (AUC) values ranged from 0.5 to 1.0, with 0.5 representing random variability and 1.0 representing perfect fit. AUC values greater than 0.7 usually indicate rational estimation. The nomogram was

compared with the FIGO stage by using the net reclassification index (NRI) and integrated discrimination improvement (IDI). NRI and IDI can be used as alternatives

to AUC for assessing the effectiveness of a new risk prediction model and for determining its effectiveness [12 13]. The Kaplan–Meier method was used to compare the risk stratification of the nomogram.

Results

Characteristics of patients

A total of 5445 patients with advanced endometrial cancer after surgery were screened from the SEER database according to the inclusion and exclusion criteria. The training cohorts (n = 3812) and validation cohorts (n = 1633) were randomly divided at a 7:3 ratio. Table 1 shows the patient characteristics. There was no statistically significant difference in the included indicators between the two groups (all $P > 0.05$).

Nomogram variable screening

The cutoff values were determined using the X-tile procedure. For the training and validation cohorts, age was cutoff at 65 and 76 years, respectively; tumor size at 35 and 78 mm, respectively; regional node positivity at 0 and 3, respectively; and linear predictor at -0.1 and 0.8, respectively (Figure 2).

When selecting the variables for the regression equation, we used LASSO because the number of independent variables should be approximately 10 to 15 times the number of ending events. From all variables in the primary cohort, Figures 3A and 3B show the positive regional nodes, age at diagnosis, tumor size, median household income, FIGO stage, grade, and radiation status.

Multivariate Cox analysis (Table 2) included all seven variables eligible for analysis; variables with $P < 0.05$ were identified as independent risk factors, including FIGO stage and grade; radiation status; number of regional nodes positive; age at diagnosis; tumor size; and median household income.

Nomogram construction and performance

As shown in Figure 4, we developed a nomogram on the basis of LASSO to predict the

one-, three-, and five-year CSS rates. According to the training and validation cohort data, the concordance index (C-index) values were 0.7228 (95% confidence interval [CI] = 0.7056–0.7399) and 0.7528 (95% CI = 0.7264–0.7792), respectively. According to Figures 5A and 5B and Table 3, the AUC for the prediction of CSS within five years was >0.7 in both the training and validation cohorts, and this result indicated favorable discrimination. As shown in Figures 5C–5H, the calibration curves of the nomogram showed high concordance between the predicted and observed survival probabilities. The red dots represent the results of bootstrapping (resample: 1000) and the performance of the prediction nomogram. The broken straight line (the 45° line) represents an ideal prediction nomogram. As the solid red line approaches the broken straight line, survival prediction becomes more accurate. In summary, the nomogram developed for advanced endometrial cancer after surgical treatment showed considerable discrimination and calibration capabilities.

Comparative clinical value of the nomogram and the FIGO stage

Table 3 compares the accuracy of the nomogram and FIGO stage on the basis of the changes in the C-index, NRI, and IDI. According to these results, the nomogram predicted the prognosis more accurately than the FIGO stage.

An assessment of the risk of advanced endometrial cancer after surgery

In addition to the nomogram, we developed a risk stratification system on the basis of the linear predictor cutoff value for each patient in the training cohort. The patients were divided into three groups according to their linear predictors: low risk (<-0.1), intermediate risk (≥-0.1, <0.8), and high risk (≥0.8). There was a significant difference in CSS between the low-, medium-, and high-risk groups according to the Kaplan–Meier analysis (all $P < 0.001$, Figure 6). Furthermore, according to the nomogram, a total score < 140 indicated low risk, <230 indicated medium risk, and ≥230 indicated

high risk. These results showed that the nomogram had excellent risk stratification capabilities.

Discussion

During our research, we used actual patient information from patients with advanced endometrial cancer after surgery. We also developed a prognostic nomogram and risk stratification system by using SEER data. The nomogram produces excellent results both internally and externally, as shown by the calibration, C-index, and receiver operating characteristic curves.

Few studies have focused on predicting the CSS of patients with advanced endometrial cancer after surgery. The current study focused on CSS after surgery in patients with stage III–IV cancer because of two reasons. First, advanced endometrial cancer has high heterogeneity and a poor survival rate, with a five-year survival rate of 20%–60% (although different patients have varied prognoses). Owing to the lack of a reliable model that can predict survival in patients with advanced endometrial cancer after surgery, individualized clinical management and surveillance can be difficult. Second, patients with advanced endometrial cancer have significantly increased incidence and mortality rates after surgery, which may lead to confounding bias in prognostic indicators.

Endometrial cancer is usually treated with surgery, and postoperative treatment depends on different risk factors, such as age, tumor stage, myometrial infiltration depth, and histologic grade [14 15]. In the current study, a prognostic model after the surgical treatment of advanced endometrial cancer was constructed on the basis of seven variables (age at diagnosis, tumor grade, FIGO stage, lymph node dissection, tumor size, radiation therapy, and median household income) screened using LASSO regression. Scores were calculated for each item on the basis of the subtype of each independent prognostic factor. The total score was calculated by using the scores corresponding to the independent prognostic factors. Each subgroup variable was assigned a score from to 0–100 according to its contribution. All the enrolled variables were added to generate a total score on the bottom scale, which was then converted to predict CSS. The CSS at one, three, and five years was determined by drawing a vertical

line on the total score scale, with higher scores indicating a worse prognosis. According to the nomogram, tumor grade plays the largest role in prognosis, followed by FIGO

stage and age at diagnosis.

Cancer grade, histological subtype, tumor size, lymphovascular space invasion, lymph node status, and cervical involvement are important prognostic factors in endometrial cancer patients [16]. In the current study, the tumor grade, tumor size, and lymph node status were important prognostic factors after surgical treatment for advanced endometrial cancer. Tumor grade has also been shown to be a prognostic factor in endometrial cancer [17], and the nomogram in the current study indicated that poorly differentiated or undifferentiated tumors have poor prognoses. Lymph node metastasis further contributes to poor prognosis in patients with endometrial cancer, but there is no consensus on the value and extent of lymph node dissection [18]. In the current study, we found that positive lymph nodes could affect the prognosis of surgical treatment for advanced endometrial cancer, and this finding is in line with those of previous studies. However, whether lymph node dissection was beneficial was not reflected in this study and may also be related to the fact that the population selected in this study underwent lymph node dissection (98.7%), which was not comparable. Compared with women > 65 years of age, women < 65 years of age had a significant survival advantage, as indicated by previous studies [19].

By using advanced endometrial cancer after surgery as a dataset, this study examined the factors that could be included in prognostic nomograms. Nomograms combine multiple factors, including demographic and clinicopathological characteristics, into a quantitative model that makes better predictions than FIGO staging [20 21]. FIGO staging has traditionally been used to predict the prognosis of women with endometrial cancer. The staging of this system is closely associated with CSS. However, patients at the same stage have different prognoses. FIGO staging does not consider factors such as age, radiation status, and income, thus explaining prognostic heterogeneity. Therefore, we compared the nomograms that included more variables. Nomograms have a better predictive power than FIGO staging alone because of their positive NRI and IDI.

On the basis of their total nomogram scores, the patients were classified into low-,

intermediate-, and high-risk groups. There were significant differences in CSS between the three risk groups based on the Kaplan–Meier analysis (Figure 6). This nomogram

is highly effective in identifying high-risk groups owing to its poor prognosis. Patients with a total score greater than or equal to 230 should receive special attention.

To investigate the potential utility of the nomogram in clinical practice, we analyzed data from the SEER database by using a large sample of data representing different regions of the population. We followed the recommendations of the Transparent Reporting of Individual Prognosis or Diagnosis Multivariate Predictive Model statement [22]. Bootstrapping and cross-validation methods were used to calculate the calibration curves, time-dependent AUCs, and C-indices. These positive results show that our nomogram may be a useful tool for assessing the prognosis of patients with advanced endometrial cancer after surgery.

Although the nomogram performed well, this study had some limitations. For example, SEER 2004-2015 did not publish data on carbohydrate antigen 125, HE4, molecular typing, surgical margins, hormonal therapy, and immunotherapy. Therefore, these variables were not assessed in this study. In addition, we did not describe the number of lymph nodes removed. Multicenter clinical validation is required to determine the external utility of our nomogram.

Conclusions

In conclusion, our nomogram is more accurate, has better clinical utility, and provides better prognostic predictions for patients with advanced endometrial cancer after surgery than FIGO staging.

Conflict of Interest

This study was conducted without any commercial or financial relationships that could be construed as potential conflicts of interest.

Author Contributions

Chunqin Zheng: study conception, data collection, data analysis, interpretation, drafting, critical revision, and final approval of the article. Ruijun Ma, Xiaojun Huang and

Danmei Fang: data collection. Yanhong Ni, Fufang Fan and Peili Zhang: data analysis. Zhixiang Zheng and Xiaoling Liang: study conception.

Funding The authors declare that the study was not funded.

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Data availability Statement This study used publicly available datasets. All data is available on https://seer.cancer.gov/data.

Ethics Statement According to local legislation and institutional requirements, no ethical review or approval was required for the study of human participants. In accordance with national legislation and institutional requirements, written informed consent was not obtained for participation in this study.

Abbreviations

SEER: Linked Surveillance, Epidemiology, and End Results. CSS: cancer-specific Survival. CI: confidence interval. FIGO: International Federation of Gynecology and Obstetrics. LASSO: least absolute shrinkage and selection operator. AUC: area under the receiver operating characteristic curve. C-index: concordance index. NRI: net reclassification index. IDI: integrated discrimination improvement.

Acknowledgments

The authors thank the SEER project for providing access to their database and gratefully thank Free Statistics software version 1.7, provided by Beijing FreeClinical Medical Technology Co., Ltd.

References

1. Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* 2021;71(3):209-49. doi: 10.3322/caac.21660 [published Online First: 20210204]
2. Bokhman JV. Two pathogenetic types of endometrial carcinoma. *Gynecol Oncol* 1983;15(1):10-7. doi: 10.1016/0090-8258(83)90111-7
3. Felix AS, Weissfeld JL, Stone RA, et al. Factors associated with Type I and Type II endometrial cancer. *Cancer Causes Control* 2010;21(11):1851-6. doi: 10.1007/s10552-010-9612-8

311 [published Online First: 20100714]
312 4. Passarello K, Kurian S, Villanueva V. Endometrial Cancer: An Overview of Pathophysiology,
313 Management, and Care. *Semin Oncol Nurs* 2019;35(2):157-65. doi:

10.1016/j.soncn.2019.02.002 [published Online First: 20190311]
5. Lewin SN, Herzog TJ, Barrena Medel NI, et al. Comparative performance of the 2009 international Federation of gynecology and obstetrics' staging system for uterine corpus cancer. *Obstet Gynecol* 2010;116(5):1141-9. doi: 10.1097/AOG.0b013e3181f39849
6. Creasman WT, Odicino F, Maisonneuve P, et al. Carcinoma of the corpus uteri. FIGO 26th Annual Report on the Results of Treatment in Gynecological Cancer. *Int J Gynaecol Obstet* 2006;95 Suppl 1:S105-43. doi: 10.1016/s0020-7292(06)60031-3
7. Zhu L, Sun X, Bai W. Nomograms for Predicting Cancer-Specific and Overall Survival Among Patients With Endometrial Carcinoma: A SEER Based Study. *Front Oncol* 2020;10:269. doi: 10.3389/fonc.2020.00269 [published Online First: 20200319]
8. Yang XL, Huang H, Kou LN, et al. Construction and validation of a prognostic model for stage IIIC endometrial cancer patients after surgery. *Eur J Surg Oncol* 2022;48(5):1173-80. doi: 10.1016/j.ejso.2021.12.462 [published Online First: 20211227]
9. Doll KM, Rademaker A, Sosa JA. Practical Guide to Surgical Data Sets: Surveillance, Epidemiology, and End Results (SEER) Database. *JAMA Surg* 2018;153(6):588-89. doi: 10.1001/jamasurg.2018.0501
10. Modh A, Ghanem AI, Burmeister C, et al. Trends in the utilization of adjuvant vaginal brachytherapy in women with early-stage endometrial carcinoma: Results of an updated period analysis of SEER data. *Brachytherapy* 2016;15(5):554-61. doi: 10.1016/j.brachy.2016.06.012 [published Online First: 20160727]
11. Camp RL, Dolled-Filhart M, Rimm DL. X-tile: a new bio-informatics tool for biomarker assessment and outcome-based cut-point optimization. *Clin Cancer Res* 2004;10(21):7252-9. doi: 10.1158/1078-0432.Ccr-04-0713
12. Uno H, Tian L, Cai T, et al. A unified inference procedure for a class of measures to assess improvement in risk prediction systems with survival data. *Stat Med* 2013;32(14):2430-42. doi: 10.1002/sim.5647 [published Online First: 20121005]
13. Pencina MJ, D'Agostino RB, Sr., Steyerberg EW. Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. *Stat Med* 2011;30(1):11-21. doi: 10.1002/sim.4085 [published Online First: 20101105]
14. Creutzberg CL, van Putten WL, Koper PC, et al. Surgery and postoperative radiotherapy versus surgery alone for patients with stage-1 endometrial carcinoma: multicentre randomised trial. PORTEC Study Group. Post Operative Radiation Therapy in Endometrial Carcinoma. *Lancet* 2000;355(9213):1404-11. doi: 10.1016/s0140-6736(00)02139-5
15. Keys HM, Roberts JA, Brunetto VL, et al. A phase III trial of surgery with or without adjunctive external pelvic radiation therapy in intermediate risk endometrial adenocarcinoma: a Gynecologic Oncology Group study. *Gynecol Oncol* 2004;92(3):744-51. doi: 10.1016/j.ygyno.2003.11.048
16. Colombo N, Creutzberg C, Amant F, et al. ESMO-ESGO-ESTRO Consensus Conference on Endometrial Cancer: diagnosis, treatment and follow-up. *Ann Oncol* 2016;27(1):16-41. doi: 10.1093/annonc/mdv484 [published Online First: 20151202]
17. Stålberg K, Kjølhede P, Bjurberg M, et al. Risk factors for lymph node metastases in women with endometrial cancer: A population-based, nation-wide register study-On behalf of

the Swedish Gynecological Cancer Group. *Int J Cancer* 2017;140(12):2693-700. doi: 10.1002/ijc.30707 [published Online First: 20170412]

18. Sari ME, Yalcin İ, Sahin H, et al. Risk factors for paraaortic lymph node metastasis in endometrial cancer. *Int J Clin Oncol* 2017;22(5):937-44. doi: 10.1007/s10147-017-1139-5 [published Online First: 20170518]
19. Chan JK, Cheung MK, Huh WK, et al. Therapeutic role of lymph node resection in endometrioid corpus cancer: a study of 12,333 patients. *Cancer* 2006;107(8):1823-30. doi: 10.1002/cncr.22185
20. Wu J, Zhang H, Li L, et al. A nomogram for predicting overall survival in patients with low-grade endometrial stromal sarcoma: A population-based analysis. *Cancer Commun (Lond)* 2020;40(7):301-12. doi: 10.1002/cac2.12067 [published Online First: 20200618]
21. Gao J, Wang S, Li F, et al. Nomograms for Predicting Overall Survival and Cancer-Specific Survival of Young Patients with Epithelial Ovarian Cancer: Analysis Based on SEER Program. *Adv Ther* 2022;39(1):257-85. doi: 10.1007/s12325-021-01955-9 [published Online First: 20211029]
22. Collins GS, Reitsma JB, Altman DG, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *Bmj* 2015;350:g7594. doi: 10.1136/bmj.g7594 [published Online First: 20150107]

Table 1 The basic characteristics of endometrial carcinoma patients in the study

Variables	Primary cohort (n=5445)	Training cohort (n = 3812)	Validation cohort (n = 1633)	*p*-value
Race, n (%)				0.927
White	4444 (81.6)	3114 (81.7)	1330 (81.4)	
Black	351 (6.4)	247 (6.5)	104 (6.4)	
Other[a]	650 (11.9)	451 (11.8)	199 (12.2)	
Chemotherapy, n (%)				0.51
No	2167 (39.8)	1528 (40.1)	639 (39.1)	
Yes	3278 (60.2)	2284 (59.9)	994 (60.9)	
Historic stage[b], n (%)				0.684
Localized	9 (0.2)	7 (0.2)	2 (0.1)	
Regional	3903 (71.7)	2719 (71.3)	1184 (72.5)	
Distant	1533 (28.2)	1086 (28.5)	447 (27.4)	
Tumor grade[c], n (%)				0.138
I	1226 (22.5)	886 (23.2)	340 (20.8)	
II	2166 (39.8)	1497 (39.3)	669 (41)	
III-IV	2053 (37.7)	1429 (37.5)	624 (38.2)	
Radiation, n (%)				0.834
No	2659 (48.8)	1858 (48.7)	801 (49.1)	
Yes	2786 (51.2)	1954 (51.3)	832 (50.9)	
Marital status, n (%)				0.791
Unmarried	1232 (22.6)	855 (22.4)	377 (23.1)	
Married	2675 (49.1)	1890 (49.6)	785 (48.1)	
Other[d]	1375 (25.3)	954 (25)	421 (25.8)	
Unknown	163 (3.0)	113 (3)	50 (3.1)	
Lymphadenectomy[e], n (%)				0.598
No	70 (1.3)	47 (1.2)	23 (1.4)	
Yes	5375 (98.7)	3765 (98.8)	1610 (98.6)	
FIGO stage, n (%)				0.326
III	4741 (87.1)	3308 (86.8)	1433 (87.8)	
IV	704 (12.9)	504 (13.2)	200 (12.2)	
Age of diagnosis, n (%)				0.653
24-65 years	3362 (61.7)	2344 (61.5)	1018 (62.3)	
66-76 years	1392 (25.6)	988 (25.9)	404 (24.7)	
77-96 years	691 (12.7)	480 (12.6)	211 (12.9)	
Regional nodes positive, n (%)				0.637
0	2415 (44.4)	1683 (44.2)	732 (44.8)	
1-3	1954 (35.9)	1383 (36.3)	571 (35)	
4-82	1076 (19.8)	746 (19.6)	330 (20.2)	
Year of diagnosis, n (%)				0.59
2004-2009	1709 (31.4)	1188 (31.2)	521 (31.9)	
2010-2015	3736 (68.6)	2624 (68.8)	1112 (68.1)	
Tumor size[f], n (%)				0.118
0-35mm	1357 (24.9)	937 (24.6)	420 (25.7)	
36-78mm	3115 (57.2)	2214 (58.1)	901 (55.2)	
79-790mm	973 (17.9)	661 (17.3)	312 (19.1)	
Income, n (%)				0.157
≤54,999$	612 (11.2)	449 (11.8)	163 (10)	
55,000-69,999$	1842 (33.8)	1281 (33.6)	561 (34.4)	
≥70,000$	2991 (54.9)	2082 (54.6)	909 (55.7)	
Diagnosis time[g], Median (IQR)	1.0 (0.0, 2.0)	1.0 (0.0, 2.0)	1.0 (0.0, 2.0)	0.777

[a], American Indian/AK Native, Asian/Pacific Islander.

[b], historic stage derived from Collaborative Stage (CS) for 2004-2015. Localized, an invasive neoplasm confined entirely to the organ of origin. Regional, a neoplasm that has extended. Distant, a neoplasm that has spread to parts of the body remote from the primary tumor.

[c], ICD-O-2 defines grade I as well differentiated, grade II as moderately differentiated, grade III as poorly differentiated, and grade IV as undifferentiated.

[d],divorced, widowed, separated.

[e], The article categorizes lymphadenectomy into two categories: those involving regional lymph node dissection and those without it. Without lymphadenectomy includes failure to remove or aspirate local lymph nodes, local lymph node biopsy or aspiration, and sentinel lymph node biopsy only. Lymphadenectomy includes removal of an unknown number of regional lymph nodes, removal of one to three regional lymph nodes, removal of four or more regional lymph nodes, and regional lymph node dissection with anterior lymph node biopsy.

[f], Based on X-tile procedure cut-offs.

[g], Months from diagnosis to treatment.

Table 2 Univariate and multivariable cox regression analysis of cancer-specific survival.

Variable	Univariate analysis		Multivariate analysis	
	HR	*P*-value	HR	*P*-value
Tumor grade*				
I	1(Ref)		1(Ref)	
II	1.51 (1.28~1.78)	<0.001	1.43 (1.21~1.68)	<0.001
III-IV	3.63 (3.12~4.23)	<0.001	2.81 (2.41~3.28)	<0.001
Radiation				
No	1(Ref)		1(Ref)	
Yes	0.67 (0.61~0.74)	<0.001	0.74 (0.67~0.82)	<0.001
FIGO stage				
III	1(Ref)		1(Ref)	
IV	3.33 (2.98~3.72)	<0.001	2.6 (2.31~2.92)	<0.001
Age of diagnosis (year)				
24-65	1(Ref)		1(Ref)	
66-76	1.47 (1.32~1.65)	<0.001	1.52 (1.36~1.7)	<0.001
77-96	2.37 (2.08~2.7)	<0.001	2.44 (2.14~2.79)	<0.001
Regional nodes positive				
Negative	1(Ref)		1(Ref)	
1-3	1.34 (1.2~1.5)	<0.001	1.34 (1.2~1.51)	<0.001
4-82	1.98 (1.76~2.24)	<0.001	1.82 (1.61~2.06)	<0.001
Tumor size (mm)				
0-35	1(Ref)		1(Ref)	
36-78	1.51 (1.32~1.71)	<0.001	1.25 (1.09~1.42)	0.001
79-790	2.36 (2.04~2.74)	<0.001	1.73 (1.49~2.02)	<0.001
Income ($)				
≤54,999	1(Ref)		1(Ref)	
55,000-69,999	0.83 (0.71~0.97)	0.018	0.83 (0.71~0.97)	0.018
≥70,000	0.72 (0.62~0.83)	<0.001	0.74 (0.64~0.86)	<0.001

*, I, well differentiated; II, moderately differentiated; III, poorly differentiated; IV, undifferentiated.

Table 3 AUC, NRI, and IDI of the nomogram and the FIGO stage in survival prediction for the advances endometrial carcinoma patients after surgical treatment.

Index	Training cohort			Validation cohort		
	Estimate	95% CI	*P*-value	Estimate	95% CI	*P*-value
NRI (vs. the FIGO stage)						
For 1-year CSS	0.33	0.269-0.398	<0.001	0.446	0.339-0.53	<0.001
For 3-year CSS	0.308	0.272-0.36	<0.001	0.41	0.34-0.465	<0.001
For 5-year CSS	0.308	0.273-0.355	<0.001	0.396	0.33-0.445	<0.001
IDI (vs. the FIGO stage)						
For 1-year CSS	0.052	0.039-0.069	<0.001	0.081	0.058-0.111	<0.001
For 3-year CSS	0.089	0.072-0.114	<0.001	0.126	0.09-0.163	<0.001
For 5-year CSS	0.103	0.083-0.125	<0.001	0.145	0.111-0.186	<0.001
The nomogram AUC						
For 1-year CSS	0.793	0.765-0.82		0.827	0.791-0.864	
For 3-year CSS	0.771	0.752-0.79		0.803	0.777-0.829	
For 5-year CSS	0.754	0.736-0.773		0.791	0.765-0.817	
The FIGO stage AUC						
For 1-year CSS	0.641	0.612-0.67		0.676	0.631-0.721	
For 3-year CSS	0.613	0.596-0.63		0.613	0.587-0.639	
For 5-year CSS	0.596	0.581-0.611		0.599	0.576-0.621	

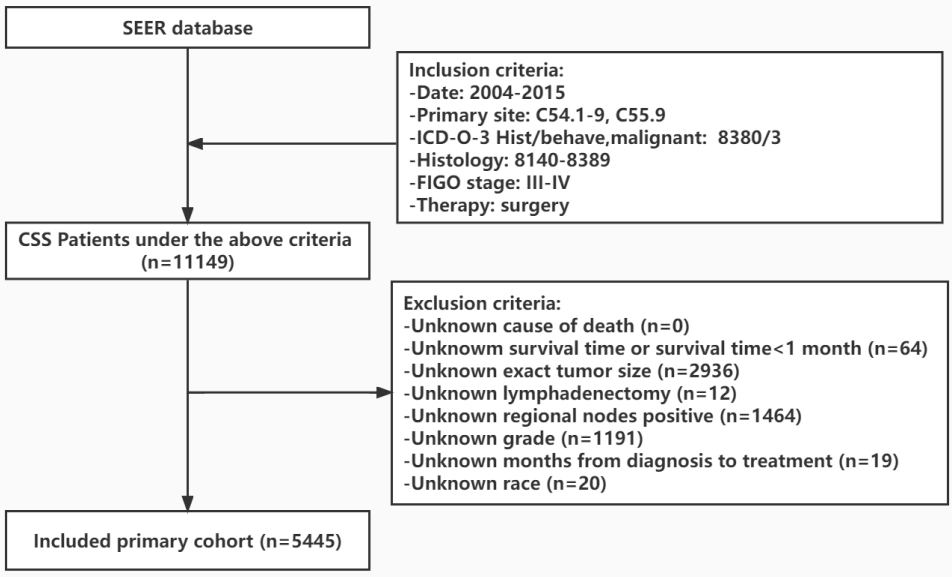


Figure 1 Flowchart of the study

Figure 1: Flowchart of the study.

470x306mm (72 x 72 DPI)

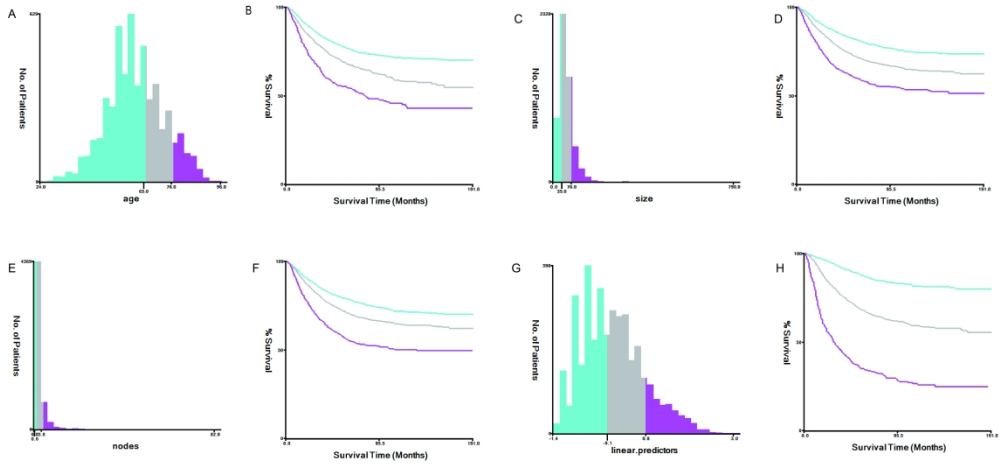


Figure 2 Cutoff values for age (A,B), tumor size (C,D), regional nodes positive (E,F), and linear predictors (G,H) that were assessed by X-tile analysis.

Figure 2 Cutoff values for age (A,B), tumor size (C,D), regional nodes positive (E,F), and linear predictors (G,H) thaFt were assessed by X-tile analysis.

585x295mm (300 x 300 DPI)

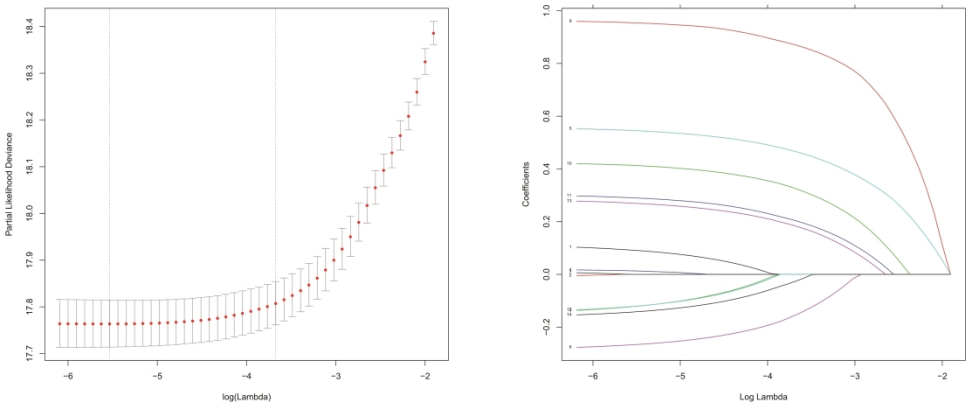


Figure 3 Selection of informative factors associated with CSS using the LASSO Cox regression model. (A) LASSO coefficient profiles of all clinical features for CSS. (B) Selection of the tuning parameter (λ. 1 se) for CSS.

Figure 3 Selection of informative factors associated with CSS using the LASSO Cox regression model. (A) LASSO coefficient profiles of all clinical features for CSS. (B) Selection of the tuning parameter (λ. 1 se) for CSS.

511x260mm (300 x 300 DPI)

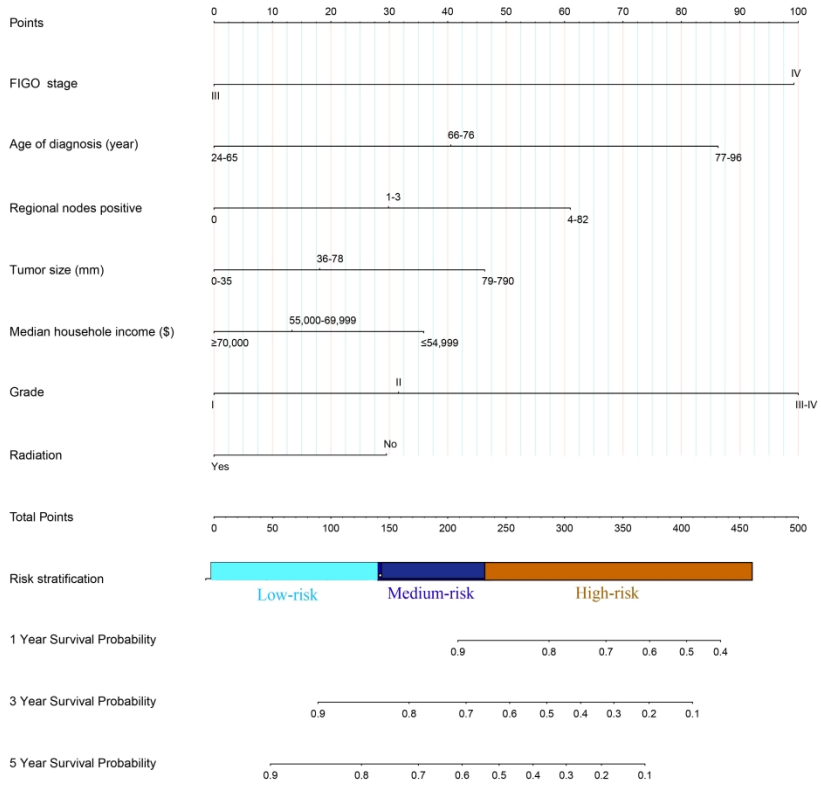


Figure 4: Nomograms for predicting 1-,3-, and 5-year cancer-specific survival.

Figure 4: Nomograms for predicting 1-,3-, and 5-year cancer-specific survival.

338x302mm (300 x 300 DPI)

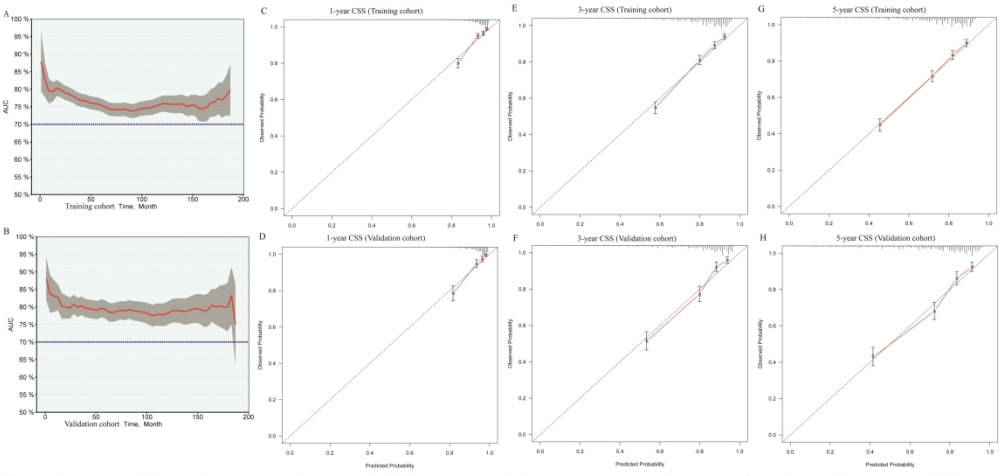
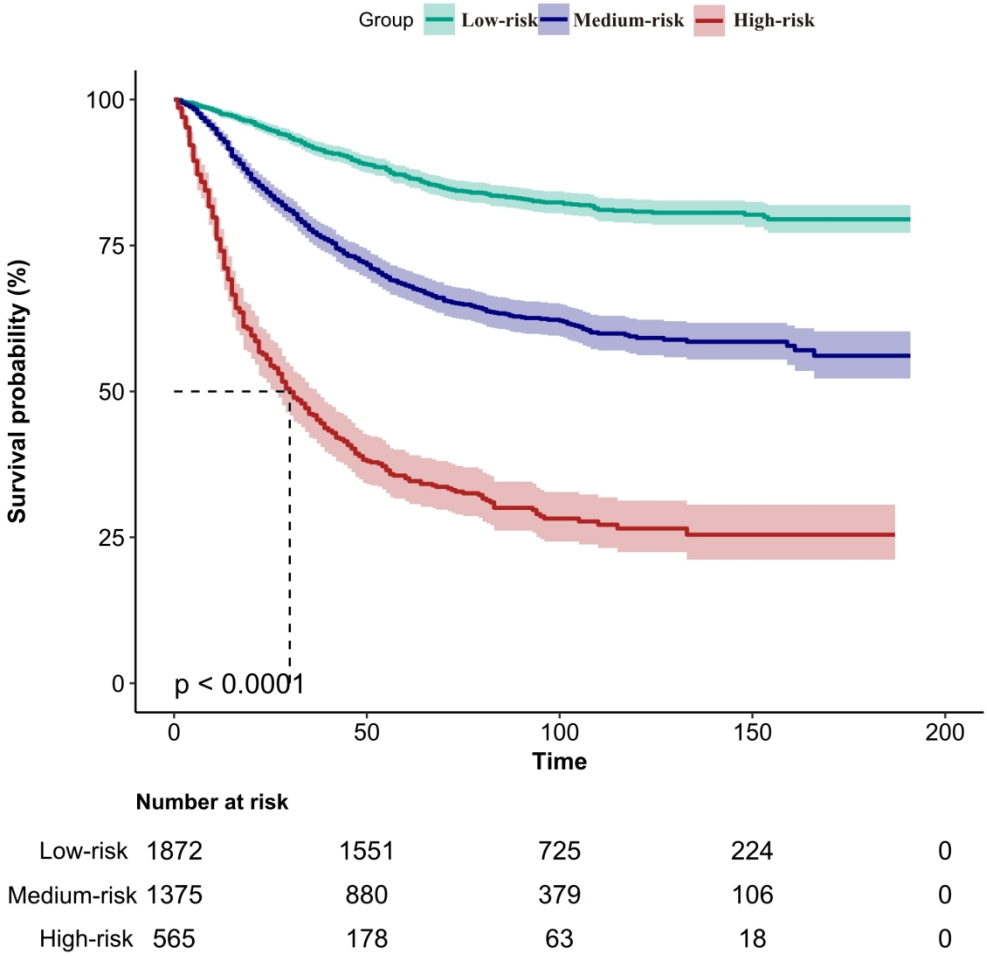


Figure 5 Time-dependent AUC and calibration curves of the nomogram. (A-B) Time-dependent AUC of using the nomogram to predict cancer-specific survival probability within 5 years in the training cohort and validation cohorts. The blue line represents AUC = 0.7, which is considered ideal. (C, E, G) Calibration curves of 1-year, 3-year and 5-year CSS for advanced endometrial carcinoma after surgery in the training cohort. (D, F, H) Calibration curves of 1-year, 3-year and 5-year CSS for advanced endometrial carcinoma after surgery in the validation cohort. The black dashed line indicates the ideal reference line where predicted probabilities would match the observed survival rates. The red dots represent the performance of the nomogram. The closer the solid red line is to the black dashed line, the more accurately the model predicts survival. AUC: area under the time-dependent receiver operating characteristic curves; CSS: cancer-specific survival.

Figure 5: Time-dependent AUC (A, B) and calibration curves (C-H) of the nomogram.

415x231mm (300 x 300 DPI)



Kaplan–Meier analysis for risk stratification.

177x177mm (300 x 300 DPI)

TRIPOD Checklist: Prediction Model Development and Validation

Section/Topic	Item		Checklist Item	Page
Title and abstract				
Title	1	D;V	Identify the study as developing and/or validating a multivariable prediction model, the target population, and the outcome to be predicted.	1
Abstract	2	D;V	Provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusions.	1-2
Introduction				
Background and objectives	3a	D;V	Explain the medical context (including whether diagnostic or prognostic) and rationale for developing or validating the multivariable prediction model, including references to existing models.	2-3
	3b	D;V	Specify the objectives, including whether the study describes the development or validation of the model or both.	3
Methods				
Source of data	4a	D;V	Describe the study design or source of data (e.g., randomized trial, cohort, or registry data), separately for the development and validation data sets, if applicable.	3-4
	4b	D;V	Specify the key study dates, including start of accrual; end of accrual; and, if applicable, end of follow-up.	3-4
Participants	5a	D;V	Specify key elements of the study setting (e.g., primary care, secondary care, general population) including number and location of centres.	4
	5b	D;V	Describe eligibility criteria for participants.	4
	5c	D;V	Give details of treatments received, if relevant.	4
Outcome	6a	D;V	Clearly define the outcome that is predicted by the prediction model, including how and when assessed.	4
	6b	D;V	Report any actions to blind assessment of the outcome to be predicted.	4
Predictors	7a	D;V	Clearly define all predictors used in developing or validating the multivariable prediction model, including how and when they were measured.	4
	7b	D;V	Report any actions to blind assessment of predictors for the outcome and other predictors.	4
Sample size	8	D;V	Explain how the study size was arrived at.	4
Missing data	9	D;V	Describe how missing data were handled (e.g., complete-case analysis, single imputation, multiple imputation) with details of any imputation method.	4
Statistical analysis methods	10a	D	Describe how predictors were handled in the analyses.	5
	10b	D	Specify type of model, all model-building procedures (including any predictor selection), and method for internal validation.	5
	10c	V	For validation, describe how the predictions were calculated.	5
	10d	D;V	Specify all measures used to assess model performance and, if relevant, to compare multiple models.	5
	10e	V	Describe any model updating (e.g., recalibration) arising from the validation, if done.	5
Risk groups	11	D;V	Provide details on how risk groups were created, if done.	6
Development vs. validation	12	V	For validation, identify any differences from the development data in setting, eligibility criteria, outcome, and predictors.	6
Results				
Participants	13a	D;V	Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful.	6
	13b	D;V	Describe the characteristics of the participants (basic demographics, clinical features, available predictors), including the number of participants with missing data for predictors and outcome.	6
	13c	V	For validation, show a comparison with the development data of the distribution of important variables (demographics, predictors and outcome).	6
Model development	14a	D	Specify the number of participants and outcome events in each analysis.	6
	14b	D	If done, report the unadjusted association between each candidate predictor and outcome.	6
Model specification	15a	D	Present the full prediction model to allow predictions for individuals (i.e., all regression coefficients, and model intercept or baseline survival at a given time point).	6
	15b	D	Explain how to the use the prediction model.	6
Model performance	16	D;V	Report performance measures (with CIs) for the prediction model.	7
Model-updating	17	V	If done, report the results from any model updating (i.e., model specification, model performance).	7
Discussion				
Limitations	18	D;V	Discuss any limitations of the study (such as nonrepresentative sample, few events per predictor, missing data).	10
Interpretation	19a	V	For validation, discuss the results with reference to performance in the development data, and any other validation data.	8-9
	19b	D;V	Give an overall interpretation of the results, considering objectives, limitations, results from similar studies, and other relevant evidence.	8-9
Implications	20	D;V	Discuss the potential clinical use of the model and implications for future research.	10
Other information				
Supplementary information	21	D;V	Provide information about the availability of supplementary resources, such as study protocol, Web calculator, and data sets.	10-11
Funding	22	D;V	Give the source of funding and the role of the funders for the present study.	11

TRIPOD Checklist: Prediction Model Development and Validation

*Items relevant only to the development of a prediction model are denoted by D, items relating solely to a validation of a prediction model are denoted by V, and items relating to both are denoted D;V. We recommend using the TRIPOD Checklist in conjunction with the TRIPOD Explanation and Elaboration document.

BMJ Open

Development and validation of a prognostic nomogram for predicting cancer-specific survival in advanced endometrial carcinoma after surgery: A retrospective analysis of the SEER database

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2022-070893.R1
Article Type:	Original research
Date Submitted by the Author:	24-Jul-2023
Complete List of Authors:	Zheng, Chunqin; Shantou Central Hospital, Department of Obstetrics and Gynecology Zheng, Zhixiang; Shantou Central Hospital, Department of Obstetrics and Gynecology Liang, Xiaoling; Shantou Central Hospital, Department of Obstetrics and Gynecology Ni, Yanhong; Shantou Central Hospital Department of Obstetrics and Gynecology Fan, Fufang; Shantou Central Hospital Department of Obstetrics and Gynecology Zhang, Peili; Shantou Central Hospital Department of Obstetrics and Gynecology Fang, Danmei; Shantou Central Hospital Department of Obstetrics and Gynecology Ma, Ruijun; Shantou Central Hospital, Department of Obstetrics and Gynecology Huang, Xiaojun; Shantou Central Hospital, Department of Obstetrics and Gynecology
Primary Subject Heading:	Oncology
Secondary Subject Heading:	Surgery
Keywords:	Gynaecological oncology < GYNAECOLOGY, Surgical pathology < PATHOLOGY, Radiation oncology < RADIOLOGY & IMAGING, Gynaecological oncology < ONCOLOGY, HEALTH ECONOMICS

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Development and validation of a prognostic nomogram for predicting cancer-specific survival in advanced endometrial carcinoma after surgery: A retrospective analysis of the SEER database

Chunqin Zheng[1*], Zhixiang Zheng[1], Xiaoling Liang[1], Yanhong Ni[1], Fufang Fan[1], Peili Zhang[1], Danmei Fang[1], Ruijun Ma[1], Xiaojun Huang[1]

[1]Department of Obstetrics and Gynecology, Shantou Central Hospital, Shantou 515000, China

[*] **Correspondence:** Chunqin Zheng, Department of Obstetrics and Gynecology, Shantou Central Hospital, Shantou 515000, China. 467430374@qq.com; ORCID:0000-0003-1840-9173.

Abstract

Objective We aimed to construct and validate a prognostic nomogram to predict cancer-specific survival (CSS) in advanced endometrial carcinoma (EC) after surgery.

Design This study is a retrospective cohort study.

Setting and participants There were 5445 patients with advanced EC from Surveillance, Epidemiology and End Results (SEER) database diagnosed between 2004 and 2015 included in the study who were randomized 7:3 into a training cohort (n = 3812) and a validation cohort (n = 1633).

Primary and secondary outcome CSS.

Results Nomograms for CSS include ten variables (positive regional nodes, age, tumor size, FIGO stage, grade, race, income, radiation, chemotherapy and historic stage), from the result of the forward stepwise regression. Nomograms reveal the discrimination and calibration by the concordance index (C-index), area under the time-dependent receiver operating characteristic curve (time-dependent AUC), with a C-index value of 0.7324 (95% confidence interval [CI] = 0.7181–0.7468) and 0.7511 (95% CI = 0.7301–0.7722) for the training cohort and the validation cohort in the nomogram for CSS. It was also proven that there was a high degree of conformance between the predicted and observation results by calibration plots. In addition, the comparison of the nomogram and FIGO stage on the basis of the changes in the C-index, net reclassification index, and integrated discrimination improvement demonstrated that the nomogram was better in accuracy and efficacy ability.

Conclusions We successfully constructed an accurate and effective nomogram to predict CSS for patients with advanced EC, which can help clinical doctors choose individual treatment strategies for advanced EC patients. However, our findings were derived from a cohort of Americans. As a result, a larger-sample multicenter study should be conducted to determine whether our study results are more broadly applicable.

Strengths and limitations of this study

⇒SEER database is a large database with sufficiently large samples.

⇒An effective and non-invasive nomogram was constructed for the advanced EC after surgery.

⇒SEER database lacks laboratory test data, which may influence the prognosis of patients with advanced EC.

⇒Selection bias might exist, because all the cases were retrieved from the same database.

⇒Some of the classifications carried out in the SEER database were not specific enough.

Introduction

Endometrial carcinoma (EC) is the sixth most common cancer among women, with 417,000 new cases diagnosed worldwide by 2020. [1] There are two histological types of EC.[2, 3] Type I tumors include tumors with grade 1 or 2 endometrioid histological classifications and account for approximately 80% of ECs. Type II tumors account for 10%–20% of ECs and include grade 3 endometrioid tumors and tumors with non- endometrioid histology. The treatment of EC is primarily surgical, with radiation and chemotherapy as common adjuvant modalities. For patients with EC who will undergo surgery, adjuvant therapy determines disease recurrence for risk stratification on the basis of tumor stage, tumor histology, and other pathologic factors. There is overwhelming evidence that traditional pathologic features, such as histopathologic type, grade, myometrial invasion, and lymphovascular space invasion (LVSI), are important in assessing prognosis.[4] Application of the molecular classification in high-grade and/or high-risk ECs shows that POLE-mutated (POLEmut) tumors with an excellent prognosis, p53-abnormal (p53abn) tumors with a poor prognosis, and EC with mismatch repair deficient (MMRd) or non-specific molecular profile (NSMP) have an intermediate prognosis.[5] The latest European (ESGO/ESTRO/ESP 2020)/American (NCCN 2020) guidelines combining traditional pathologic and The Cancer Genome Atlas(TCGA) molecular groups proposed a novel risk stratification model：low, intermediate, high–intermediate, high, and advanced metastatic.[6] In general, the five-year survival rates are 80%–90% and 70%–80% for stage I and II ECs, respectively, and 20%–60% for stage III and IV ECs.[7, 8] Stage III and IV ECs are classified as advanced and high-risk EC. Patients with advanced and recurrent EC have a dismal prognosis with an expected 5-year survival of less than 20%.[9] Because of its high mortality, a clinical model for predicting the prognosis of advanced EC patients is necessary. Although the Federation of Gynaecology and Obstetrics (FIGO) staging system was widely used to predict the survival of EC patients,[10] a lot of limitations existed.

A nomogram is a simple visualization tool used by oncologists to predict and quantify patient survival on the basis of multiple variables. A nomogram has been used for patients with EC, [11] and Yang et al. published a nomogram for patients with stage IIIC EC following surgery; [12] however, there is no specific prognostic prediction for advanced EC patients after surgery.

Furthermore, the traditional statistical strategy only adopted the variables which were significant on univariate analysis to establish the final prediction models, which led to model overfitting and showed poor results.[13] There are some advanced statistical methodologies to minimize this limitation, such as the best subsets regression (BSR), the forward stepwise regression (FSR), and the least absolute shrinkage and selection operator (LASSO).[14-16] Therefore, our studies aimed to establish an effective and non-invasive nomogram to predict cancer-specific survival (CSS) in advanced EC after surgery, as well as adopting advanced statistical.

Methods and data

Patient and public involvement

None.

Data sources and patient selection

Case data of EC with complete follow-up records were selected from the 2004–2015 SEER database (SEER research plus data, 17 Registries, November 2021 Sub (2000–2019)) using SEER*Stat V. 8.4.0.1. The inclusion criteria were as follows: primary sites, C54.1-9 and C55.9;[17] site and morphology, 8380/3(based on the International Classification of Tumor Diseases for Oncology [ICD-O], Third Edition); histology, 8140-8389 (adenomas and adenocarcinomas); International Federation of Gynecology and Obstetrics (FIGO) stage, III/IV; and therapy, surgical treatment. There were nine exclusion criteria: (1) undetermined cause of death; (2) undetermined survival time, or survival time < one month; (3) undetermined tumor size; (4) undetermined lymphadenectomy; (5) unknown regional node status; (6) unknown tumor grade; (7) unknown months from diagnosis to treatment; (8) unknown race and (9) unknown median household income. The flow chart of the patient screening is shown in online supplemental figure 1.

Variables including age at diagnosis, year of diagnosis, tumor size, race, marital status, histologic stage, tumor grade, FIGO stage, lymphadenectomy, regional node positivity, chemotherapy, radiation status, months from diagnosis to treatment, survival time, median household income, and CSS were collected from the SEER database. There were two categories of radiation status (with and without radiation) and chemotherapy status (with and without chemotherapy). Marital status was classified as unmarried (single, unmarried, or domestic partner), married, other (divorced, widowed, separated), and unknown. Grades were associated with tumors. ICD-O-2 defines grade I as well differentiated, grade II as moderately differentiated, grade III as poorly differentiated, and grade IV as undifferentiated. According to the SEER registry, income was examined as

aggregate data, which are based on the US median income. Median household income in USD was reclassified into three groups: ≤54,999, 55,000–69,999, and ≥70,000. The historic stage was derived from the Collaborative Stage for 2004–2015 and was divided into in situ, localized, regional, distant, and unknown categories. In the localized stage, an invasive neoplasm is confined entirely to the organ of origin. In the regional stage, a neoplasm has extended 1) beyond the limits of the organ of origin directly into surrounding organs or tissues, 2) into regional lymph nodes by way of the lymphatic system, or 3) into regional lymph nodes by a combination of extension and regional lymph nodes. In the distant stage, a neoplasm has spread to parts of the body remote from the primary tumor. This study categorizes lymphadenectomy into two categories: with and without regional lymph node dissection. Failure to perform lymph node dissection included the failure to remove or aspirate local lymph nodes, local lymph node biopsy or aspiration, and sentinel lymph node biopsy only. Lymph node dissection includes the removal of an unknown number of regional lymph nodes, removal of one to three regional lymph nodes, removal of four or more regional lymph nodes, and regional lymph node dissection with anterior lymph node biopsy.

Statistical analysis

X-tile (Yale University, New Haven, Connecticut, USA) was used to determine the cutoff values for age at diagnosis, the tumor size, the positive regional nodes, and the risk stratification.[18] Statistical analyses were conducted using R 4.0.2 (R Foundation for Statistical Computing, Vienna, Austria, http://www.r-project.org) in the RStudio environment and by using Free Statistics 1.8 (Beijing FreeClinical Medical Technology Co., LTD). CSS was the primary endpoint of the study. Cases were randomly placed in the training and validation cohorts at a 7:3 ratio. Categorical variables are presented as frequencies and proportions. Chi-square tests were used to compare clinicopathological characteristics between the training and validation cohorts. $P < 0.05$ was considered statistically significant. The BSR, FSR, and LASSO was used to select variables. In addition, significant prognostic factors were identified from the Cox proportional hazards model. A nomogram associated with CSS was then constructed and incorporated into the known prognostic factors. The nomogram performance was validated by training and validation. The nomogram model was evaluated using the area under the time-dependent receiver operating characteristic curve (time-dependent AUC) to assess their discriminative abilities, and calibration curves were plotted to compare the predicted CSS with the actual CSS after one, three, and five years. The area under the curve (AUC) values ranged from 0.5 to1.0, with 0.5 representing random variability and 1.0 representing perfect fit. AUC values greater than 0.7 usually indicate rational estimation. The nomogram was compared with the FIGO stage by using the net reclassification index (NRI) and integrated discrimination improvement (IDI). NRI and IDI can be used as alternatives to AUC for assessing the effectiveness of a new risk prediction model and for

determining its effectiveness.[19, 20] The Kaplan–Meier method was used to compare the risk stratification of the nomogram.

Results

Characteristics of patients

A total of 5445 patients with advanced EC after surgery were screened from the SEER database according to the inclusion and exclusion criteria. The training cohorts (n = 3812) and validation cohorts (n = 1633) were randomly divided at a 7:3 ratio. Table 1 shows the patient characteristics. There was no statistically significant difference in the included indicators between the two groups (all P > 0.05).

Table 1 The basic characteristics of EC patients in the study.

Variables	Primary cohort (n=5445)	Training cohort (n = 3812)	Validation cohort (n = 1633)	*p*-value
Race, n (%)				0.903
White	4444 (81.6)	3107(81.5)	1337 (81.9)	
Black	351 (6.4)	245(6.4)	106 (6.5)	
Other[a]	650 (11.9)	460 (12.1)	190(11.6)	
Chemotherapy, n (%)				0.899
No	2167 (39.8)	1515(39.7)	652(39.9)	
Yes	3278 (60.2)	2297(60.3)	981(60.1)	
Historic stage[b], n (%)				0.62
Localized	9 (0.2)	5 (0.1)	4 (0.2)	
Regional	3903 (71.7)	2731(71.6)	1172 (71.8)	
Distant	1533 (28.2)	1076(28.2)	457 (28)	
Tumor grade[c], n (%)				0.631
I	1226 (22.5)	853 (22.4)	373(22.8)	
II	2166 (39.8)	1506(39.5)	660(40.4)	
III-IV	2053 (37.7)	1453(38.1)	600(36.7)	
Radiation, n (%)				0.055
No	2659 (48.8)	1894(49.7)	765(46.8)	
Yes	2786 (51.2)	1918(50.3)	868(53.2)	
Marital status, n (%)				0.438
Unmarried	1232 (22.6)	881 (23.1)	351(21.5)	
Married	2675 (49.1)	1855(48.7)	820(50.2)	
Other[d]	1375 (25.3)	967 (25.4)	408 (25)	
Unknown	163 (3.0)	109(2.9)	54(3.3)	
Lymphadenectomy[e], n (%)				0.601
No	70 (1.3)	51 (1.3)	19(1.2)	
Yes	5375 (98.7)	3761(98.7)	1614 (98.8)	
FIGO stage, n (%)				0.11
III	4741 (87.1)	3301(86.6)	1440 (88.2)	
IV	704 (12.9)	511 (13.4)	193(11.8)	
Age of diagnosis, n (%)				0.553
24-64 years	3362 (61.7)	2352(61.7)	1010 (61.8)	

65-75 years	1392 (25.6)	965 (25.3)	427(26.1)	
76-96 years	691 (12.7)	495 (13)	196 (12)	
Regional nodes positive, n (%)				0.447
0	2415 (44.4)	1694(44.4)	721(44.2)	
1-2	1954 (35.9)	1381(36.2)	573(35.1)	
3-82	1076 (19.8)	737 (19.3)	339(20.8)	
Year of diagnosis, n (%)				0.981
2004-2009	2152 (39.5)	1507(39.5)	645(39.5)	
2010-2015	3293 (60.5)	2305(60.5)	988(60.5)	
Tumor size[f], n (%)				0.319
0-35mm	1640 (30.1)	1149(30.1)	491(30.1)	
36-78mm	2847 (52.3)	1974(51.8)	873(53.5)	
79-790mm	958 (17.6)	689 (18.1)	269(16.5)	
Income, n (%)				0.701
≤54,999$	1010 (18.5)	707 (18.5)	303(18.6)	
55,000-69,999$	2168 (39.8)	1505(39.5)	663(40.6)	
≥70,000$	2267 (41.6)	1600 (42)	667(40.8)	
Diagnosis time[g], Mean±SD	1.1±1.2	1.1±1.2	1.1±1.1	0.375

a, American Indian/AK Native, Asian/Pacific Islander.
b, historic stage derived from Collaborative Stage (CS) for 2004-2015. Localized, an invasive neoplasm confined entirely to the organ of origin. Regional, a neoplasm that has extended. Distant, a neoplasm that has spread to parts of the body remote from the primary tumor.
c, ICD-O-2 defines grade I as well differentiated, grade II as moderately differentiated, grade III as poorly differentiated, and grade IV as undifferentiated.
d, divorced, widowed, separated.
e, the article categorizes lymphadenectomy into two categories: those involving regional lymph node dissection and those without it. Without lymphadenectomy includes failure to remove or aspirate local lymph nodes, local lymph node biopsy or aspiration, and sentinel lymph node biopsy only. Lymphadenectomy includes removal of an unknown number of regional lymph nodes, removal of one to three regional lymph nodes, removal of four or more regional lymph nodes, and regional lymph node dissection with anterior lymph node biopsy.
f, Based on X-tile procedure cut-offs.
g, Months from diagnosis to treatment.

Nomogram variable screening

Age, tumor size, regional node positivity and linear predictor (Linear predictor=0.448*black race+0.166* other race-0.158*chemotherapy-0.706* historic stage regional -0.702* historic stage distant +0.25* grade II+0.913* (grade III-IV) -0.261*radiation+ 0.977* FIGO stage IV + 0.471* (age of diagnosis 65-75 years)+0.881* (age of diagnosis 76-96 years)+0.263*(tumor size 36-78mm)+0.577*(tumor size 79-790mm) +0.317*(regional nodes positive 1-2)+0.619*(regional nodes positive 3-82)-0.132*(income 55,000-69,999$)-

0.195*(income≥70,000$)-0.271) were divided into three categories by X-tile software. The best cut-off age was 64 years old and 75 years old (online supplemental figure 2), the best cut-off tumor size was calculated to be 35 mm and 78 mm (online supplemental figure 2), the best cut-off regional node positivity was calculated to be 0 and 2 (online supplemental figure 2), and the best cut-off linear predictor was 0.2 and 1.2 (online supplemental figure 2).

The BSR, the LASSO, and the FSR were used to select variables. The BSR method showed great benefits on variables selection since all possible combinations of variables were calculated and the final selected combination should be optimal based on the minimum Bayesian information criterion (BIC). As shown in online supplementary figure 3 A, B, six variables (positive regional nodes, age at diagnosis, tumor size, FIGO stage, grade, and race) were selected from all variables in the training cohort. Considering that the number of independent variables included in the regression equation should be around 10 to 15 times the number of ending events, we further adopted the LASSO to select variables. As shown in online supplementary figure 3 C, D, seven variables (positive regional nodes, age at diagnosis, tumor size, FIGO stage, grade, radiation and income) were selected from all variables in the training cohort. Furthermore, the FSR selected ten variables (positive regional nodes, age at diagnosis, tumor size, FIGO stage, grade, race, chemotherapy, historic, radiation and income.) in the training cohort. As a result (online supplementary figure 4), the discrimination of the FSR was maximum in 1-, 3- and 5-year training cohort with the concordance index (C-index) 0.808(95% confidence interval (CI): 0.786–0.83), 0.787(95% CI: 0.771–0.802) and 0.771(95% CI: 0.756–0.786), respectively. Moreover, compare with the LASSO and the BSR (online supplementary table S1), the 1-, 3- and 5-year IDI of the FSR was significantly improved (FSR vs. LASSO:0.006, 0.004, 0.003, respectively. And all $P < 0.05$, FSR vs. BSR: 0.013, 0.012, 0.011, respectively. And all $P < 0.05$). Therefore, the nomogram obtained from the FSR was optimal (Figure 1). Then, these ten variables obtained from the FSR in the multivariate Cox analysis due to its optimal performance in predicting CSS for advanced EC after surgery. The results showed that race, chemotherapy, historic stage, grade, radiation, FIGO stage, age at diagnosis, tumor size, positive regional nodes and income were independent prognostic factors for CSS in advanced EC after surgery (Table 2). A nomogram for predicting 1-, 3-, and 5-year CSS was built on the basis of these ten key factors (Figure 1).

Table 2 Univariate and multivariable cox regression analysis of cancer-specific survival.

Variable	Univariate analysis		Multivariate analysis	
	HR	*P*-value	HR	*P*-value
Race				
White	1(Ref)		1(Ref)	
Black	1.88 (1.6~2.21)	<0.001	1.49 (1.26~1.75)	<0.001
Other[a]	1.03 (0.89~1.2)	0.697	1.15 (0.99~1.34)	0.072

Chemotherapy				
No	1(Ref)		1(Ref)	
Yes	1 (0.9~1.1)	0.958	0.84 (0.75~0.93)	0.001
Historic stage[b]				
Localized	1(Ref)		1(Ref)	
Regional	0.41 (0.17~0.98)	0.044	0.32 (0.13~0.78)	0.012
Distant	0.84 (0.35~2.03)	0.705	0.34 (0.14~0.82)	0.016
Tumor grade[c]				
I	1(Ref)		1(Ref)	
II	1.51 (1.28~1.78)	<0.001	1.43 (1.21~1.68)	<0.001
III-IV	3.63 (3.12~4.23)	<0.001	2.79 (2.39~3.26)	<0.001
Radiation				
No	1(Ref)		1(Ref)	
Yes	0.67 (0.61~0.74)	<0.001	0.76 (0.69~0.84)	<0.001
FIGO stage				
III	1(Ref)		1(Ref)	
IV	3.33 (2.98~3.72)	<0.001	2.6 (2.26~3)	<0.001
Age of diagnosis (year)				
24-64	1(Ref)		1(Ref)	
65-75	1.47 (1.32~1.65)	<0.001	1.52 (1.36~1.7)	<0.001
76-96	2.37 (2.08~2.7)	<0.001	2.38 (2.08~2.73)	<0.001
Tumor size (mm)				
0-35	1(Ref)		1(Ref)	
36-78	1.54 (1.36~1.74)	<0.001	1.25 (1.1~1.41)	<0.001
79-790	2.38 (2.07~2.74)	<0.001	1.72 (1.48~2)	<0.001
Regional nodes positive				
Negative	1(Ref)		1(Ref)	
1-2	1.34 (1.2~1.5)	<0.001	1.36 (1.21~1.53)	<0.001
3-82	1.98 (1.76~2.24)	<0.001	1.86 (1.62~2.14)	<0.001
Income				
≤54,999$	1(Ref)		1(Ref)	
55,000-69,999$	0.82 (0.72~0.94)	0.003	0.82 (0.72~0.94)	0.003
≥70,000$	0.72 (0.63~0.82)	<0.001	0.75 (0.65~0.85)	<0.001

a, American Indian/AK Native, Asian/Pacific Islander.

b, historic stage derived from Collaborative Stage (CS) for 2004-2015. Localized, an invasive neoplasm confined entirely to the organ of origin. Regional, a neoplasm that has extended. Distant, a neoplasm that has spread to parts of the body remote from the primary tumor.

c, I, well differentiated; II, moderately differentiated; III, poorly differentiated; IV, undifferentiated.

Nomogram construction and performance

As shown in figure 1, we developed a nomogram on the basis of the FSR to predict the one-, three-, and five-year CSS rates. According to the training and validation cohort data, the C-index values were 0.7324 (95% CI = 0.7181–0.7468) and 0.7511 (95% CI = 0.7301–0.7722), respectively. According to figure 2A and 2B, the AUC for the prediction of CSS within five years was >0.7 in both the training and validation cohorts, and this result indicated favorable discrimination. Figure 2C, 2E, and 2G shown the calibration curves of 1-year, 3-year and 5-year CSS for advanced EC

after surgery in the training cohort. Figure 2D, 2F, and 2H shown the calibration curves of 1-year, 3-year and 5-year CSS for advanced EC after surgery in the validation cohort. The black dashed line indicates the ideal reference line where predicted probabilities would match the observed survival rates. The red dots represent the performance of the nomogram. The closer the solid red line is to the black dashed line, the more accurately the model predicts survival. As shown in figures 2C–2H, the calibration curves of the nomogram showed high concordance between the predicted and observed survival probabilities. In summary, the nomogram developed for advanced EC after surgical treatment showed consider-able discrimination and calibration capabilities.

Comparative clinical value of the nomogram and the FIGO stage

Compares the accuracy of the nomograms and FIGO stage on the basis of the changes in the ROC curves and the time-dependent AUC (Figure 3). Furthermore, compare with the FIGO stage (online supplementary table S2), the 1-, 3- and 5-year IDI of the nomograms was significantly improved (Nomograms vs. FIGO stage:0.062, 0.099, 0.112, respectively). Besides, compare with the FIGO stage (online supplementary table S2), the 1-, 3- and 5-year NRI of the nomograms was significantly improved (Nomograms vs. FIGO stage:0.364, 0.354, 0.337, respectively). According to these results, the nomogram predicted the prognosis more accurately than the FIGO stage.

An assessment of the risk of advanced EC after surgery

In addition to the nomogram, we developed a risk stratification system on the basis of the linear predictor cutoff value for each patient in the training cohort. The patients were divided into three groups according to their linear predictors: low risk (≤0.2), intermediate risk (0.21-1.2), and high risk (>1.2). There was a significant difference in CSS between the low-, medium-, and high-risk groups according to the Kaplan–Meier analysis (all $P < 0.05$, online supplementary figure 5). Furthermore, according to the nomogram, a total score ≤ 185 indicated low risk, >185 and ≤285 indicated medium risk, and >285 indicated high risk. These results showed that the nomogram had excellent risk stratification capabilities.

Discussion

During our research, we used actual patient information from patients with advanced EC after surgery. We also developed a prognostic nomogram and risk stratification system by using SEER data. The nomogram produces excellent results both internally and externally, as shown by the calibration, C-index, and receiver operating characteristic curves.

Few studies have focused on predicting the CSS of patients with advanced EC after surgery. The current study focused on CSS after surgery in patients with stage III–IV cancer because of two reasons. First, advanced EC has high prognostic heterogeneity and a poor survival rate, with a five-year survival rate of 20%–60% (although different patients have varied prognoses). Owing to the lack of a reliable model that can predict survival in patients with advanced EC after surgery,

individualized clinical management and surveillance can be difficult. Second, patients with advanced EC have significantly increased incidence and mortality rates after surgery, which may lead to confounding bias in prognostic indicators.

EC is usually treated with surgery, and postoperative treatment depends on different risk factors, such as age, tumor stage, myometrial infiltration depth, and histologic grade.[21, 22] In the current study, a prognostic model after the surgical treatment of advanced EC was constructed on the basis of ten variables (race, chemotherapy, historic stage, tumor grade, radiation therapy, FIGO stage, age at diagnosis, tumor size, positive regional nodes and median household income) screened using the FSR. Scores were calculated for each item on the basis of the subtype of each independent prognostic factor. The total score was calculated by using the scores corresponding to the independent prognostic factors. Each subgroup variable was assigned a score from to 0–100 according to its contribution. All the enrolled variables were added to generate a total score on the bottom scale, which was then converted to predict CSS. The CSS at one, three, and five years was determined by drawing a vertical line on the total score scale, with higher scores indicating a worse prognosis. According to the nomogram, the FIGO stage plays the largest role in prognosis, followed by tumor grade and age of diagnosis.

Cancer grade, histological subtype, tumor size, LVSI, lymph node status, and cervical involvement are important prognostic factors in EC patients.[23] In the current study, the tumor grade, tumor size, and lymph node status were important prognostic factors after surgical treatment for advanced EC. Tumor grade has also been shown to be a prognostic factor in EC,[24] and the nomogram in the current study indicated that poorly differentiated or undifferentiated tumors have poor prognoses. Concerning the impact of tumor size on survival outcomes, conflicting results have been reported in the literature. A recent study shows that preoperative ultrasound tumor size does not appear as a prognostic factor for death of any cause in EC women.[25] However, Some literature showed that tumor size was an independent prognostic factor for recurrence alone[26, 27]or for recurrence and death due to EC.[28] Lymph node metastasis further contributes to poor prognosis in patients with EC, but there is no consensus on the value and extent of lymph node dissection.[29] In the current study, we found that positive lymph nodes could affect the prognosis of surgical treatment for advanced EC, and this finding is in line with those of previous studies. However, whether lymph node dissection was beneficial was not reflected in this study and may also be related to the fact that the population selected in this study underwent lymph node dissection (98.7%), which was not comparable. Compared with women ≥65 years of age, women < 65 years of age had a significant survival advantage, as indicated by previous studies.[30]

By using advanced EC after surgery as a dataset, this study examined the factors that could be

included in prognostic nomograms. Nomograms combine multiple factors, including demographic and clinicopathological characteristics, into a quantitative model that makes better predictions than FIGO staging.[31, 32] FIGO staging has traditionally been used to predict the prognosis of women with EC. The staging of this system is closely associated with CSS. However, patients at the same stage have different prognoses. FIGO staging does not consider factors such as age, radiation status, and income, thus explaining prognostic heterogeneity. Therefore, we compared the nomograms that included more variables. Nomograms have a better predictive power than FIGO staging alone because of their positive NRI and IDI.

On the basis of their total nomogram scores, the patients were classified into low-, intermediate-, and high-risk groups. There were significant differences in CSS between the three risk groups based on the Kaplan–Meier analysis (online supplementary figure 5). This nomogram is highly effective in identifying high-risk groups owing to its poor prognosis. Patients with a total score greater than 285 should receive special attention.

To investigate the potential utility of the nomogram in clinical practice, we analyzed data from the SEER database by using a large sample of data representing different regions of the population. We followed the recommendations of the Transparent Reporting of Individual Prognosis or Diagnosis Multivariate Predictive Model statement.[33] Bootstrapping and cross-validation methods were used to calculate the calibration curves, time-dependent AUCs, and C-index. These positive results show that our nomogram may be a useful tool for assessing the prognosis of patients with advanced EC after surgery.

Although the nomogram performed well, this study had some limitations. Carbohydrate antigen 125 (CA125) is a tumor marker whose levels are often elevated in patients with malignant tumors such as ovarian epithelial, fallopian tube, and EC, as well as those with lung adenocarcinoma and gastrointestinal adenocarcinoma. In the clinical diagnosis and treatment of EC, CA125 levels are often used to monitor disease changes, evaluate treatment effect, and predict the prognosis.[34] Studies have shown that CA125 is an important variable in the prognostic prediction model of EC, which can greatly improve the accuracy of the model.[35] Human epididymis protein 4 (HE4) is a whey acidic protein that was first identified in the epithelium of the distal epididymis.[36] It is expressed in the epithelium of several tissues, including the female reproductive tract, and is overexpressed in a variety of cancers. [37] HE4 is strongly associated with survival in patients with EC. [38] ECs have traditionally been classified into two subtypes according to their histopathological characteristics (type 1 and 2).[2] However, this classification system lacks reproducibility and yields heterogenous molecular groups that hamper advances and implementation of precision medicine,[39, 40] it is being replaced by a clearly-defined system based on molecular phenotypes.[41] TCGA approach results in the molecular stratification of ECs

into four distinct molecular groups: DNA Polymerase epsilon ultra-mutated classification which portends a good prognosis, microsatellite instability hypermutated (intermediate prognosis), copy number-low, and copy number-high (latter which includes p53 mutations, the worst prognosis) .[41]ESMO 2022 recommends that molecular staging testing should be performed for all EC, but in POLE testing can be omitted for low-risk patients when conditions are limited, but should still MMR and p53 testing should be performed to identify those patients who may have hereditary EC or high-risk factors.[42]In EC patients, LVSI has a prognostic value independent of TCGA signature, as well as age and adjuvant treatment, increasing the risk of death of any cause.[43] However, SEER 2004-2015 did not publish data on CA125, HE4, molecular typing, LVSI, hormonal therapy, and immunotherapy. Therefore, these variables were not assessed in this study. In addition, the chemotherapy and radiotherapy information contained in SEER database can only be obtained by signing some agreements, which can't be obtained for the time being, so we are unable to study the relationship between chemotherapy, radiotherapy, targeted therapy and the prognosis of EC. Moreover, study cases were derived from the US SEER database, which is not representative of other regions outside the USA. Multicenter clinical validation is required to determine the external utility of our nomogram.

Conclusions

In conclusion, our nomogram is more accurate, has better clinical utility, and provides better prognostic predictions for patients with advanced EC after surgery than FIGO staging. However, our findings were derived from a cohort of Americans. As a result, a larger-sample multicenter study should be conducted to determine whether our study results are more broadly applicable.

Acknowledgments

We thank the SEER project for providing access to their database and gratefully thank Free Statistics software version 1.7, provided by Beijing FreeClinical Medical Technology Co., Ltd. We gratefully thank Dr. Jie Liu of the Department of Vascular and Endovascular Surgery, Chinese PLA General Hospital, for his contribution to the statistical support, study design consultations, and comments regarding the manuscript.

Contributors

CZ study conception, data collection, data analysis, interpretation, drafting, critical revision, and final approval of the article. RM, XH and DF data collection. YN, FF and PZ data analysis. ZZ and XL study conception.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient consent for publication Not applicable.

Ethics approval Cancer is publicly reportable in every state of the United States, so informed consent was not required from patients to access the SEER database.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability Statement All data relevant to the study are included in the article or uploaded as supplementary information.

Abbreviations

EC: endometrial cancer. SEER: Linked Surveillance, Epidemiology, and End Results. CSS: cancer-specific Survival. CI: confidence interval. FIGO: International Federation of Gynecology and Obstetrics. BSR: best subsets regression. FSR: forward stepwise regression. LASSO: least absolute shrinkage and selection operator. AUC: area under the receiver operating characteristic curve. C-index: concordance index. NRI: net reclassification index. IDI: integrated discrimination improvement. LVSI: lymphovascular space invasion.

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F: **Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries**. *CA Cancer J Clin* 2021, **71**(3):209-249.

2. Bokhman JV: **Two pathogenetic types of endometrial carcinoma**. *Gynecol Oncol* 1983, **15**(1):10-17.

3. Felix AS, Weissfeld JL, Stone RA, Bowser R, Chivukula M, Edwards RP, Linkov F: **Factors associated with Type I and Type II endometrial cancer**. *Cancer Causes Control* 2010, **21**(11):1851-1856.

4. Cho KR, Cooper K, Croce S, Djordevic B, Herrington S, Howitt B, Hui P, Ip P, Koebel M, Lax S *et al*: **International Society of Gynecological Pathologists (ISGyP) Endometrial Cancer Project: Guidelines From the Special Techniques and Ancillary Studies Group**. *Int J Gynecol Pathol* 2019, **38 Suppl 1**(Iss 1 Suppl 1):S114-s122.

5. McAlpine J, Leon-Castillo A, Bosse T: **The rise of a novel classification system for endometrial carcinoma; integration of molecular subclasses**. *J Pathol* 2018, **244**(5):538-549.

6. Concin N, Matias-Guiu X, Vergote I, Cibula D, Mirza MR, Marnitz S, Ledermann J, Bosse T, Chargari C, Fagotti A *et al*: **ESGO/ESTRO/ESP guidelines for the management of patients with endometrial carcinoma**. *Int J Gynecol Cancer* 2021, **31**(1):12-39.

7. Lewin SN, Herzog TJ, Barrena Medel NI, Deutsch I, Burke WM, Sun X, Wright JD: **Comparative performance of the 2009 international Federation of gynecology and**

obstetrics' staging system for uterine corpus cancer. *Obstet Gynecol* 2010, **116**(5):1141-1149.

8. Creasman WT, Odicino F, Maisonneuve P, Quinn MA, Beller U, Benedet JL, Heintz AP, Ngan HY, Pecorelli S: **Carcinoma of the corpus uteri. FIGO 26th Annual Report on the Results of Treatment in Gynecological Cancer**. *Int J Gynaecol Obstet* 2006, **95 Suppl 1**:S105-143.

9. Costa B, Vale N: **Dostarlimab: A Review**. *Biomolecules* 2022, **12**(8).

10. Kasius JC, Pijnenborg JMA, Lindemann K, Forsse D, van Zwol J, Kristensen GB, Krakstad C, Werner HMJ, Amant F: **Risk Stratification of Endometrial Cancer Patients: FIGO Stage, Biomarkers and Molecular Classification**. *Cancers (Basel)* 2021, **13**(22).

11. Zhu L, Sun X, Bai W: **Nomograms for Predicting Cancer-Specific and Overall Survival Among Patients With Endometrial Carcinoma: A SEER Based Study**. *Front Oncol* 2020, **10**:269.

12. Yang XL, Huang H, Kou LN, Lai H, Chen XP, Wu DJ: **Construction and validation of a prognostic model for stage IIIC endometrial cancer patients after surgery**. *Eur J Surg Oncol* 2022, **48**(5):1173-1180.

13. Harrell FE, Jr., Lee KL, Matchar DB, Reichert TA: **Regression models for prognostic prediction: advantages, problems, and suggested solutions**. *Cancer Treat Rep* 1985, **69**(10):1071-1077.

14. Collins GS, Reitsma JB, Altman DG, Moons KG: **Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. The TRIPOD Group**. *Circulation* 2015, **131**(2):211-219.

15. Debray TP, Damen JA, Snell KI, Ensor J, Hooft L, Reitsma JB, Riley RD, Moons KG: **A guide to systematic review and meta-analysis of prediction model performance**. *Bmj* 2017, **356**:i6460.

16. Walter S, Tiemeier H: **Variable selection: current practice in epidemiological studies**. *Eur J Epidemiol* 2009, **24**(12):733-736.

17. Modh A, Ghanem AI, Burmeister C, Rasool N, Elshaikh MA: **Trends in the utilization of adjuvant vaginal brachytherapy in women with early-stage endometrial carcinoma: Results of an updated period analysis of SEER data**. *Brachytherapy* 2016, **15**(5):554-561.

18. Camp RL, Dolled-Filhart M, Rimm DL: **X-tile: a new bio-informatics tool for biomarker assessment and outcome-based cut-point optimization**. *Clin Cancer Res* 2004, **10**(21):7252-7259.

19. Uno H, Tian L, Cai T, Kohane IS, Wei LJ: **A unified inference procedure for a class of measures to assess improvement in risk prediction systems with survival data**. *Stat Med* 2013, **32**(14):2430-2442.

20. Pencina MJ, D'Agostino RB, Sr., Steyerberg EW: **Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers**. *Stat Med* 2011, **30**(1):11-21.

21. Creutzberg CL, van Putten WL, Koper PC, Lybeert ML, Jobsen JJ, Wárlám-Rodenhuis CC, De Winter KA, Lutgens LC, van den Bergh AC, van de Steen-Banasik E *et al*: **Surgery and postoperative radiotherapy versus surgery alone for patients with stage-1 endometrial carcinoma: multicentre randomised trial. PORTEC Study Group. Post**

Operative Radiation Therapy in Endometrial Carcinoma. *Lancet* 2000, **355**(9213):1404-1411.

22. Keys HM, Roberts JA, Brunetto VL, Zaino RJ, Spirtos NM, Bloss JD, Pearlman A, Maiman MA, Bell JG: **A phase III trial of surgery with or without adjunctive external pelvic radiation therapy in intermediate risk endometrial adenocarcinoma: a Gynecologic Oncology Group study**. *Gynecol Oncol* 2004, **92**(3):744-751.

23. Colombo N, Creutzberg C, Amant F, Bosse T, González-Martín A, Ledermann J, Marth C, Nout R, Querleu D, Mirza MR *et al*: **ESMO-ESGO-ESTRO Consensus Conference on Endometrial Cancer: diagnosis, treatment and follow-up**. *Ann Oncol* 2016, **27**(1):16-41.

24. Stålberg K, Kjølhede P, Bjurberg M, Borgfeldt C, Dahm-Kähler P, Falconer H, Holmberg E, Staf C, Tholander B, Åvall-Lundqvist E *et al*: **Risk factors for lymph node metastases in women with endometrial cancer: A population-based, nation-wide register study-On behalf of the Swedish Gynecological Cancer Group**. *Int J Cancer* 2017, **140**(12):2693-2700.

25. Ambrosio M, Raffone A, Alletto A, Cini C, Filipponi F, Neola D, Fabbri M, Arena A, Raimondo D, Salucci P *et al*: **Is preoperative ultrasound tumor size a prognostic factor in endometrial carcinoma patients?** *Front Oncol* 2022, **12**:993629.

26. Senol T, Polat M, Ozkaya E, Karateke A: **Tumor Diameter for Prediction of Recurrence, Disease Free and Overall Survival in Endometrial Cancer Cases**. *Asian Pac J Cancer Prev* 2015, **16**(17):7463-7466.

27. Sozzi G, Uccella S, Berretta R, Petrillo M, Fanfani F, Monterossi G, Ghizzoni V, Frusca

T, Ghezzi F, Chiantera V *et al*: **Tumor Size, an Additional Risk Factor of Local Recurrence in Low-Risk Endometrial Cancer: A Large Multicentric Retrospective Study**. *Int J Gynecol Cancer* 2018, **28**(4):684-691.

28. Chattopadhyay S, Cross P, Nayar A, Galaal K, Naik R: **Tumor size: a better independent predictor of distant failure and death than depth of myometrial invasion in International Federation of Gynecology and Obstetrics stage I endometrioid endometrial cancer**. *Int J Gynecol Cancer* 2013, **23**(4):690-697.

29. Sari ME, Yalcin İ, Sahin H, Meydanli MM, Gungor T: **Risk factors for paraaortic lymph node metastasis in endometrial cancer**. *Int J Clin Oncol* 2017, **22**(5):937-944.

30. Chan JK, Cheung MK, Huh WK, Osann K, Husain A, Teng NN, Kapp DS: **Therapeutic role of lymph node resection in endometrioid corpus cancer: a study of 12,333 patients**. *Cancer* 2006, **107**(8):1823-1830.

31. Wu J, Zhang H, Li L, Hu M, Chen L, Xu B, Song Q: **A nomogram for predicting overall survival in patients with low-grade endometrial stromal sarcoma: A population-based analysis**. *Cancer Commun (Lond)* 2020, **40**(7):301-312.

32. Gao J, Wang S, Li F, Xu H, Li X, Yan L, Zhu L: **Nomograms for Predicting Overall Survival and Cancer-Specific Survival of Young Patients with Epithelial Ovarian Cancer: Analysis Based on SEER Program**. *Adv Ther* 2022, **39**(1):257-285.

33. Collins GS, Reitsma JB, Altman DG, Moons KG: **Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement**. *Bmj* 2015, **350**:g7594.

34. Jiang T, Huang L, Zhang S: **Preoperative serum CA125: a useful marker for surgical**

management of endometrial cancer. *BMC Cancer* 2015, **15**:396.

35. Li Q, Kong F, Ma J, Wang Y, Wang C, Yang H, Li Y, Ma X: **Nomograms Based on Fibrinogen, Albumin, Neutrophil-Lymphocyte Ratio, and Carbohydrate Antigen 125 for Predicting Endometrial Cancer Prognosis**. *Cancers (Basel)* 2022, **14**(22).

36. Kirchhoff C, Habben I, Ivell R, Krull N: **A major human epididymis-specific cDNA encodes a protein with sequence homology to extracellular proteinase inhibitors**. *Biol Reprod* 1991, **45**(2):350-357.

37. Galgano MT, Hampton GM, Frierson HF, Jr.: **Comprehensive analysis of HE4 expression in normal and malignant human tissues**. *Mod Pathol* 2006, **19**(6):847-853.

38. Behrouzi R, Barr CE, Crosbie EJ: **HE4 as a Biomarker for Endometrial Cancer**. *Cancers (Basel)* 2021, **13**(19).

39. Gilks CB, Oliva E, Soslow RA: **Poor interobserver reproducibility in the diagnosis of high-grade endometrial carcinoma**. *Am J Surg Pathol* 2013, **37**(6):874-881.

40. de Boer SM, Wortman BG, Bosse T, Powell ME, Singh N, Hollema H, Wilson G, Chowdhury MN, Mileshkin L, Pyman J *et al*: **Clinical consequences of upfront pathology review in the randomised PORTEC-3 trial for high-risk endometrial cancer**. *Ann Oncol* 2018, **29**(2):424-430.

41. Kandoth C, Schultz N, Cherniack AD, Akbani R, Liu Y, Shen H, Robertson AG, Pashtan I, Shen R, Benz CC *et al*: **Integrated genomic characterization of endometrial carcinoma**. *Nature* 2013, **497**(7447):67-73.

42. Oaknin A, Bosse TJ, Creutzberg CL, Giornelli G, Harter P, Joly F, Lorusso D, Marth C, Makker V, Mirza MR *et al*: **Endometrial cancer: ESMO Clinical Practice Guideline for**

diagnosis, treatment and follow-up. *Ann Oncol* 2022, **33**(9):860-877.

43. Raffone A, Travaglino A, Raimondo D, Neola D, Maletta M, Santoro A, Insabato L, Casadio P, Fanfani F, Zannoni GF *et al*: **Lymphovascular space invasion in endometrial carcinoma: A prognostic factor independent from molecular signature**. *Gynecol Oncol* 2022, **165**(1):192-197.

Figure Legend:

Figure 1 - Nomograms for predicting 1-,3-, and 5-year cancer-specific survival.
*, American Indian/AK Native, Asian/Pacific Islander.

Figure 2 - Time-dependent AUC and calibration curves of the nomogram. (A-B) Time-dependent AUC of using the nomogram to predict cancer-specific survival probability within 5 years in the training cohort and validation cohorts. The red line represents AUC = 0.7, which is considered ideal. (C, E, G) Calibration curves of 1-year, 3-year and 5-year CSS for advanced EC after surgery in the training cohort. (D, F, H) Calibration curves of 1-year, 3-year and 5-year CSS for advanced EC after surgery in the validation cohort. The black dashed line indicates the ideal reference line where predicted probabilities would match the observed survival rates. The red dots represent the performance of the nomogram. The closer the solid red line is to the black dashed line, the more accurately the model predicts survival. AUC: area under the time-dependent receiver operating characteristic curves; CSS: cancer-specific survival. EC: endometrial carcinoma.

Figure 3 - Compares the accuracy of the nomograms and FIGO stage on the basis of the changes in the ROC curves and the time-dependent AUC.

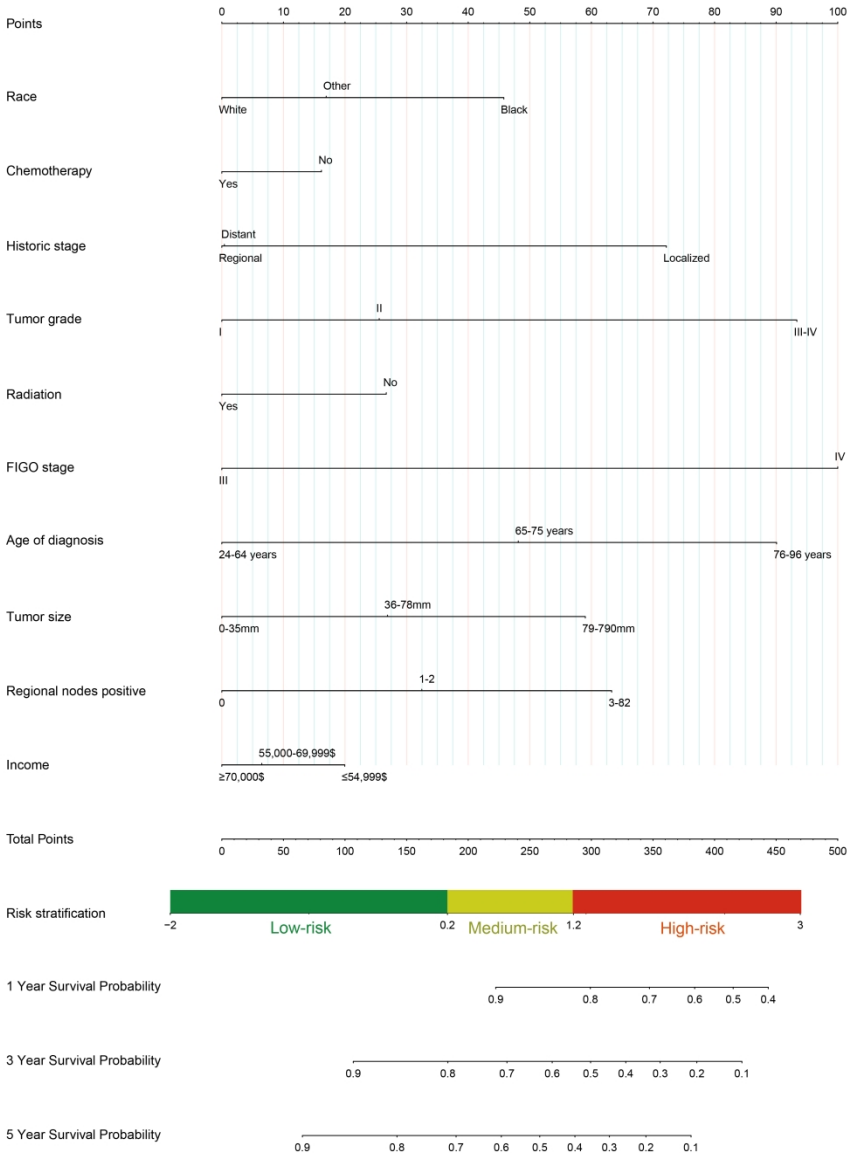


Figure 1 Nomograms for predicting 1-,3-, and 5-year cancer-specific survival.
*, American Indian/AK Native, Asian/Pacific Islander.

288x385mm (300 x 300 DPI)

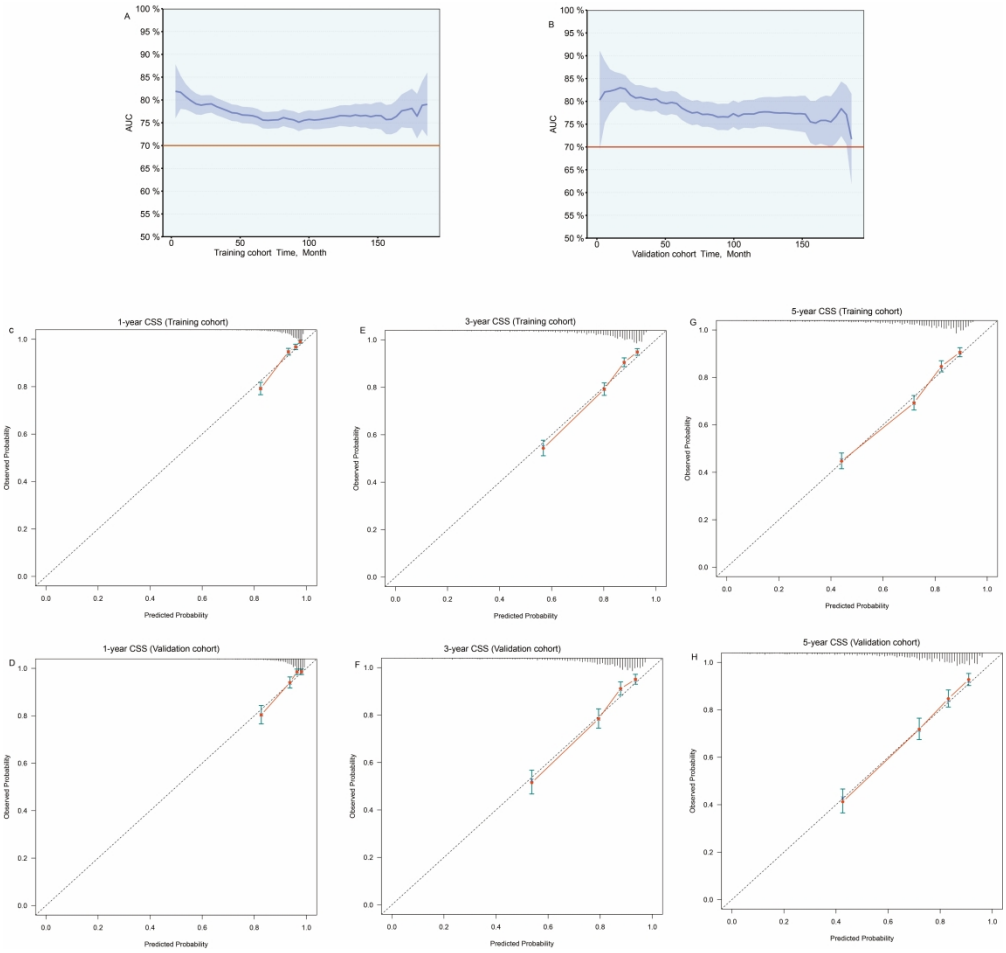


Figure 2 Time-dependent AUC and calibration curves of the nomogram. (A-B) Time-dependent AUC of using the nomogram to predict cancer-specific survival probability within 5 years in the training cohort and validation cohorts. The red line represents AUC = 0.7, which is considered ideal. (C, E, G) Calibration curves of 1-year, 3-year and 5-year CSS for advanced EC after surgery in the training cohort. (D, F, H) Calibration curves of 1-year, 3-year and 5-year CSS for advanced EC after surgery in the validation cohort. The black dashed line indicates the ideal reference line where predicted probabilities would match the observed survival rates. The red dots represent the performance of the nomogram. The closer the solid red line is to the black dashed line, the more accurately the model predicts survival. AUC: area under the time-dependent receiver operating characteristic curves; CSS: cancer-specific survival. EC: endometrial carcinoma.

508x481mm (300 x 300 DPI)

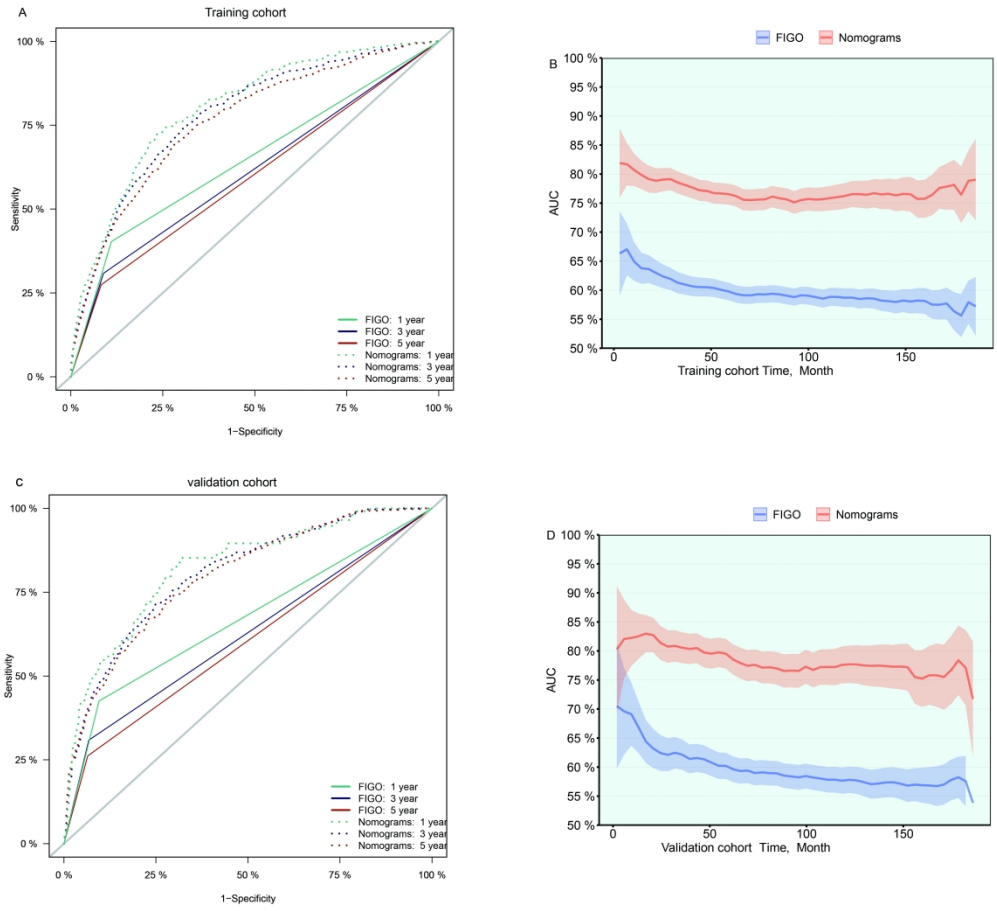
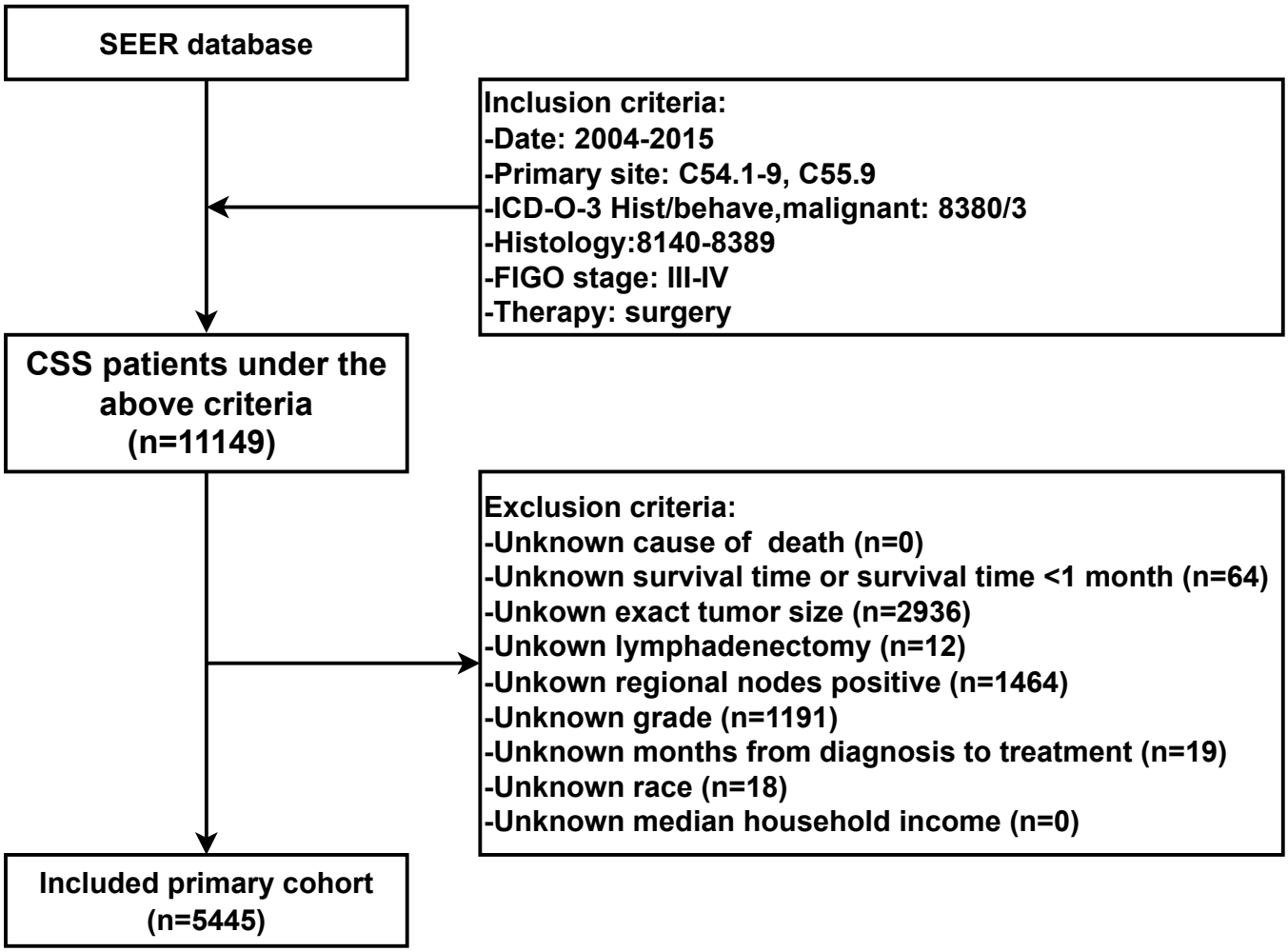
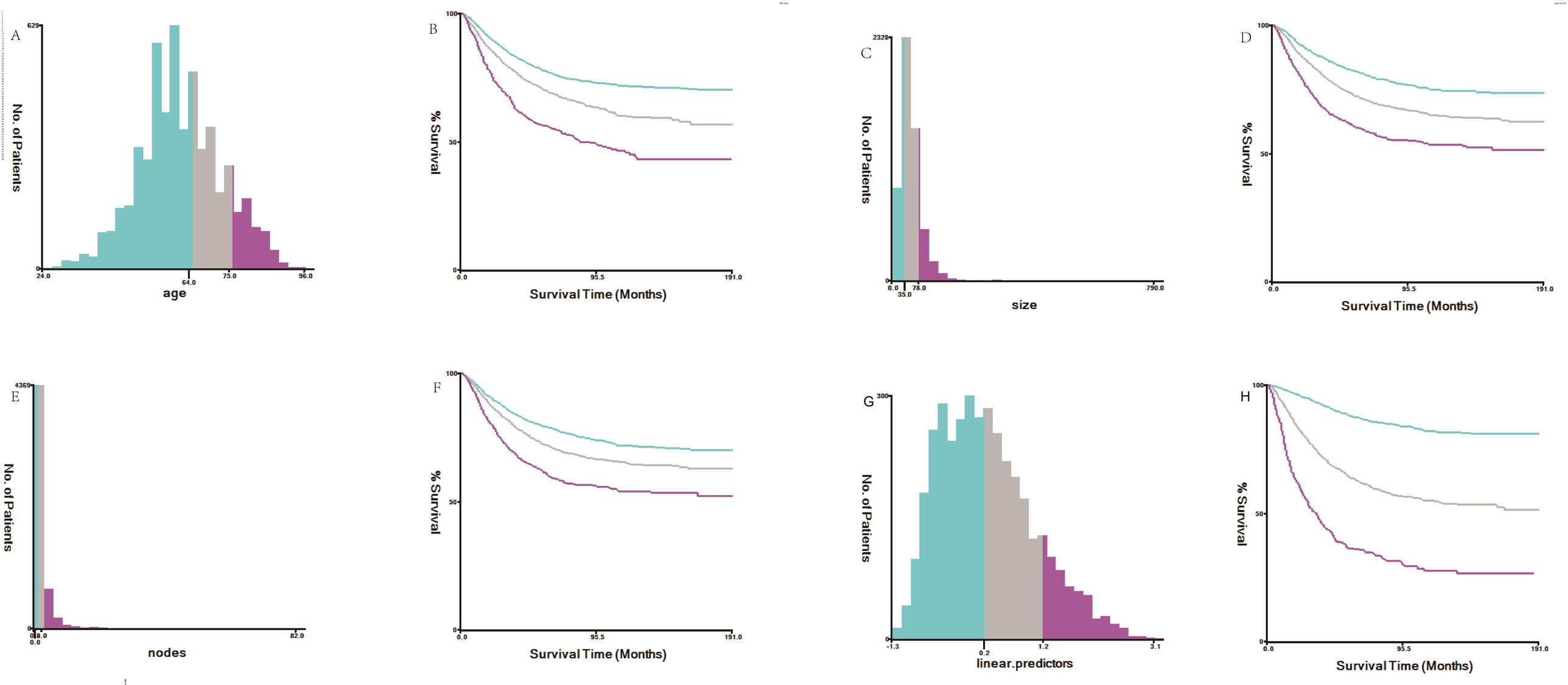


Figure 3 Compares the accuracy of the nomograms and FIGO stage on the basis of the changes in the ROC curves and the time-dependent AUC.

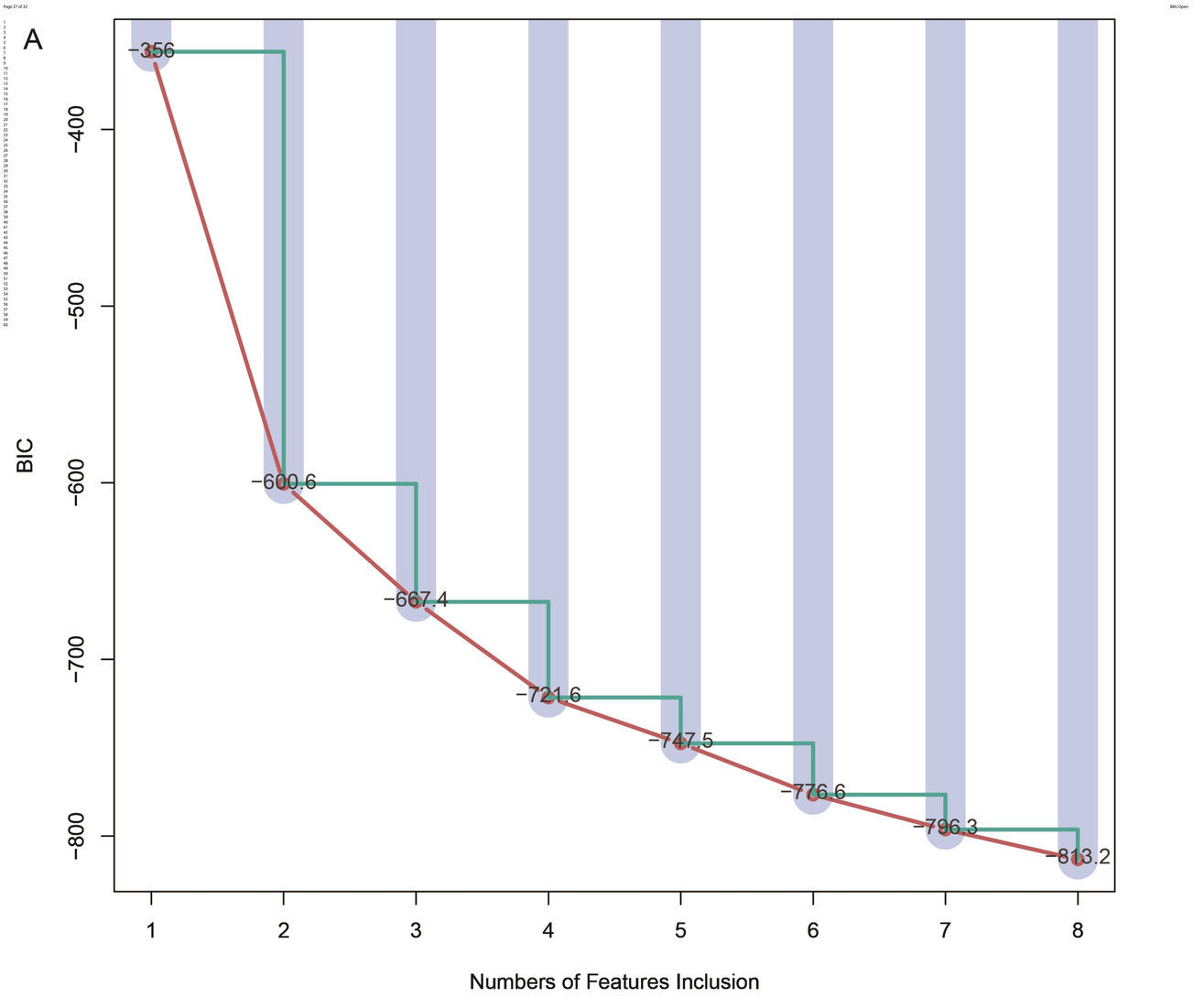
431x388mm (300 x 300 DPI)

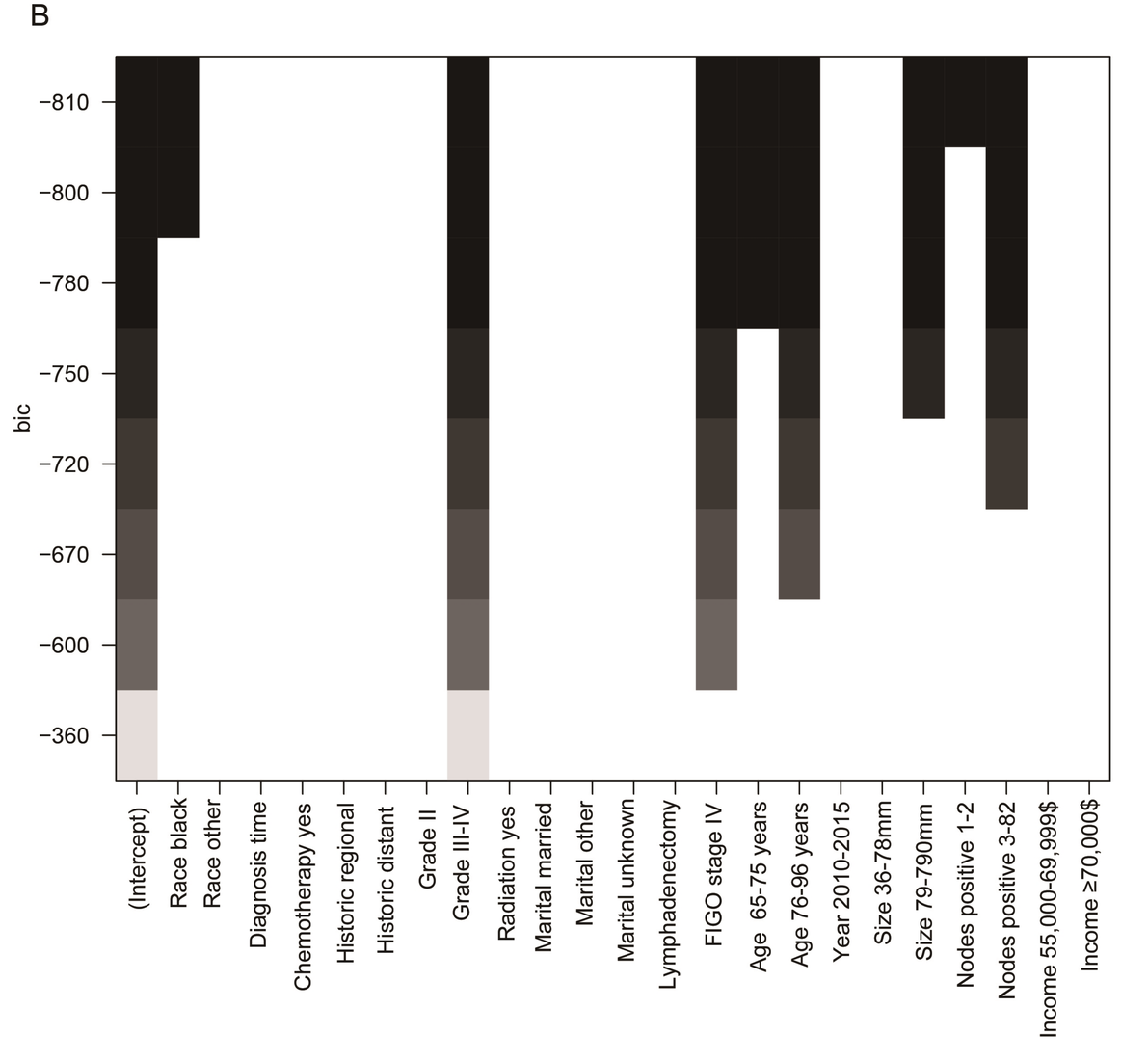


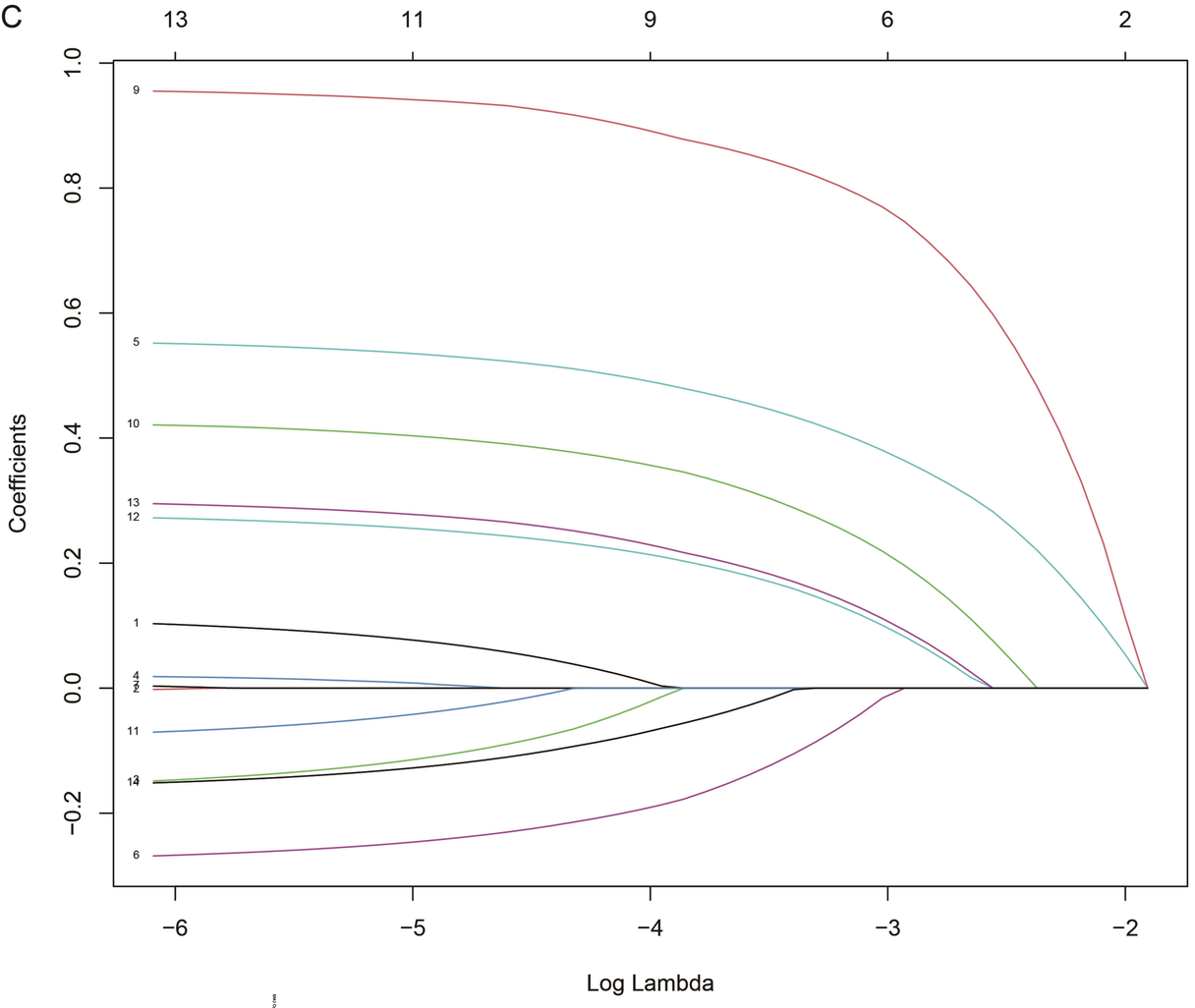
Supplementary Figure 1 Flowchart of the study

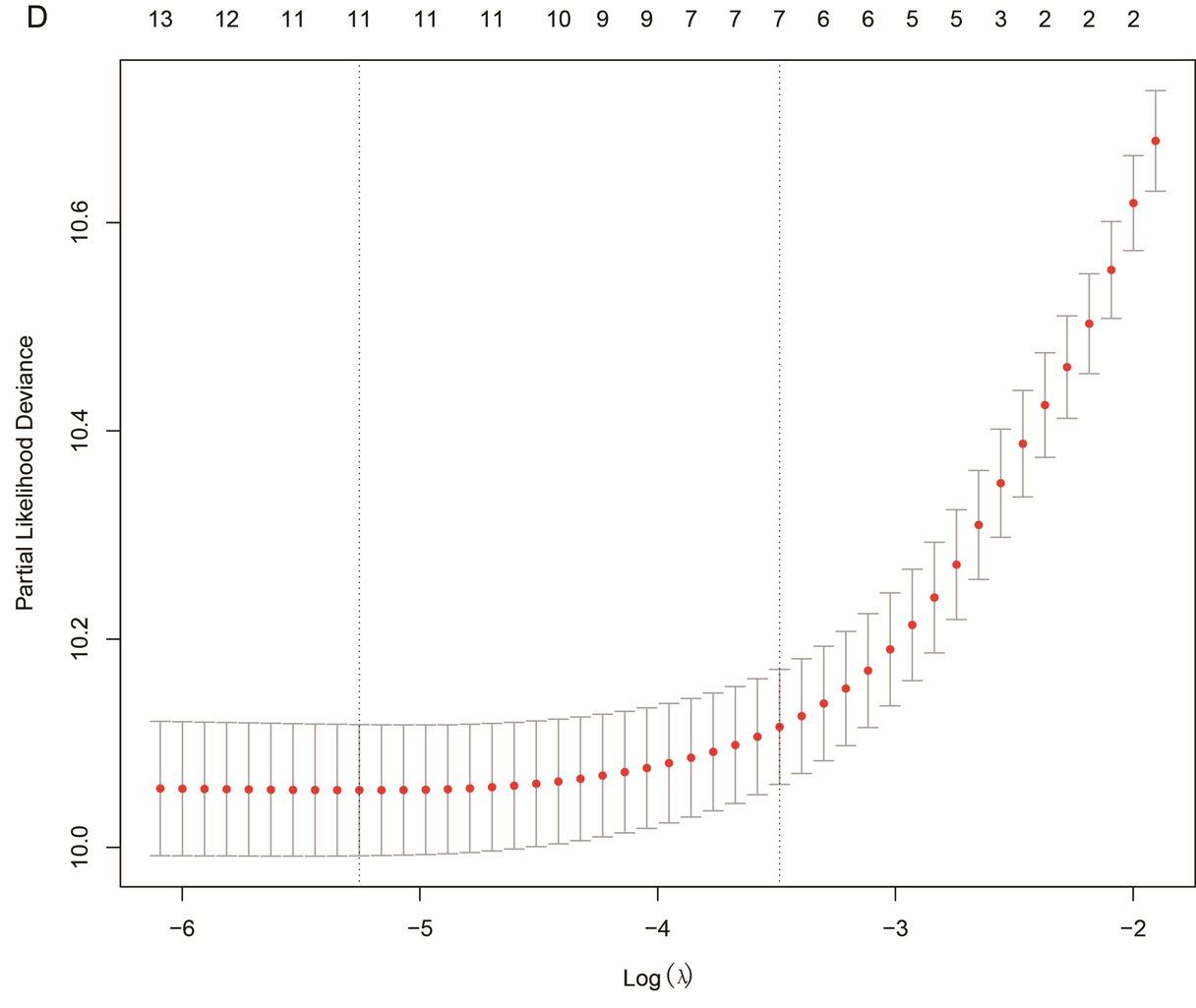


Supplementary Figure 2 Cutoff values for age (A,B), tumor size (C,D), regional nodes positive (E,F), and linear predictors (G,H) that were assessed by X-tile analysis.

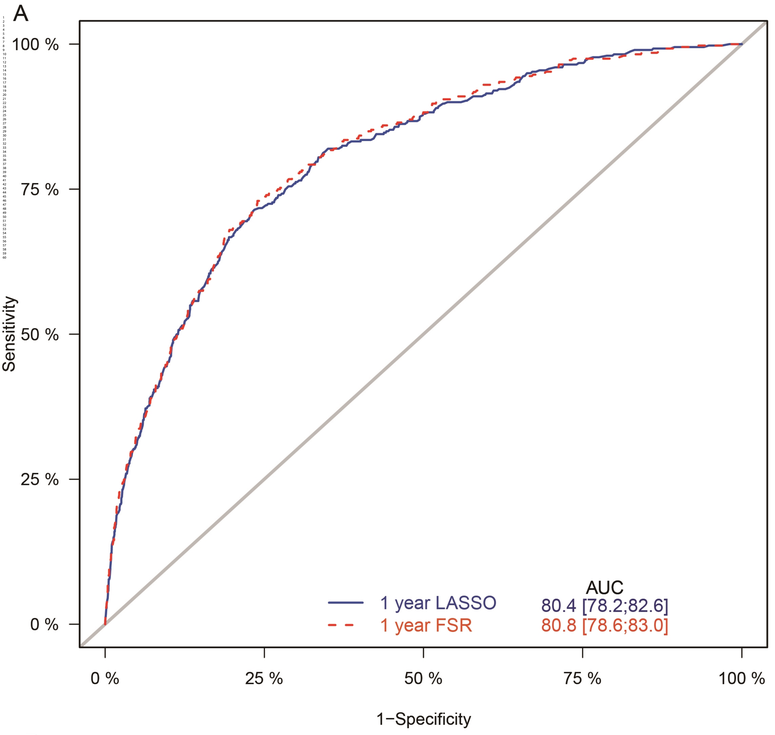


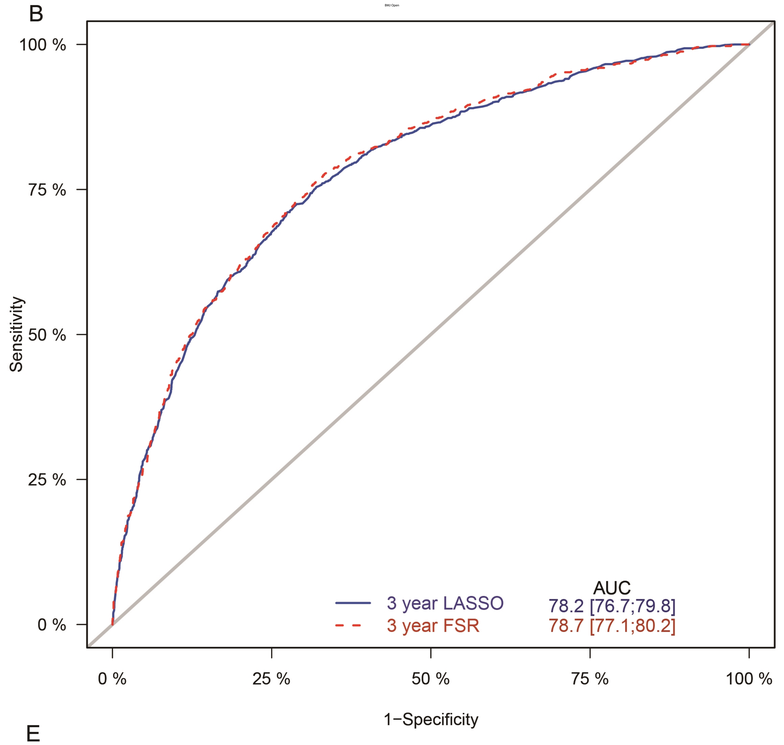


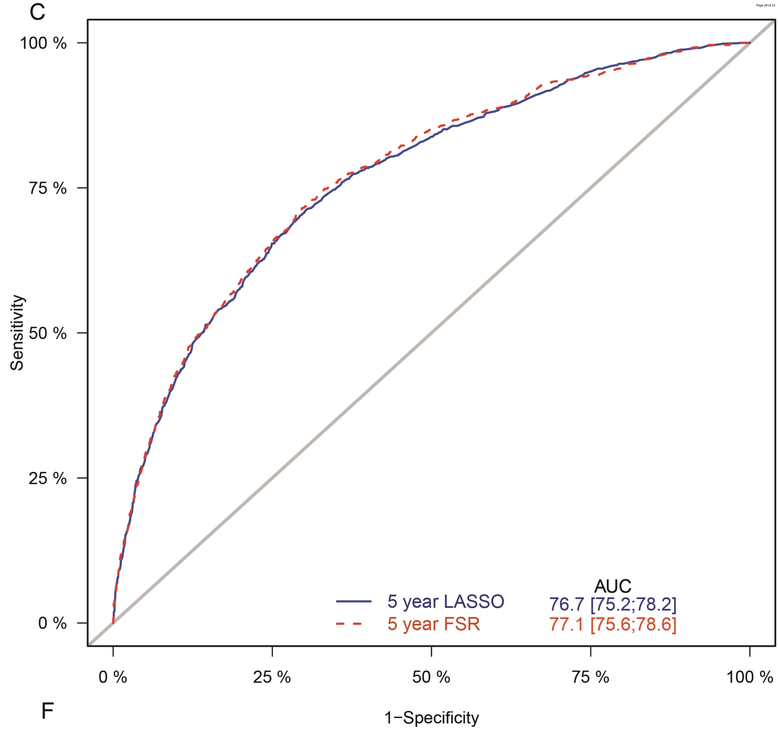


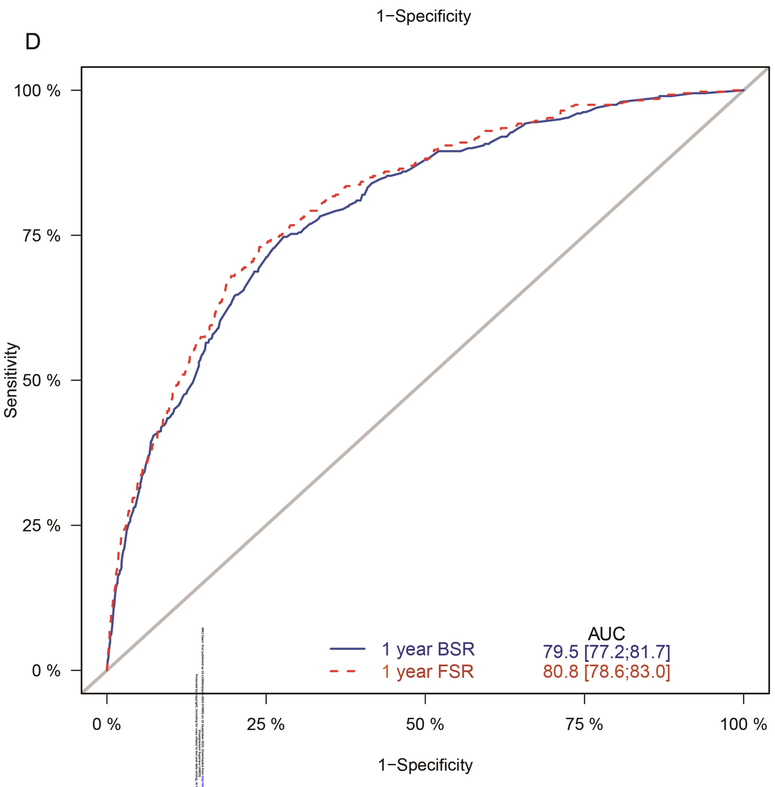


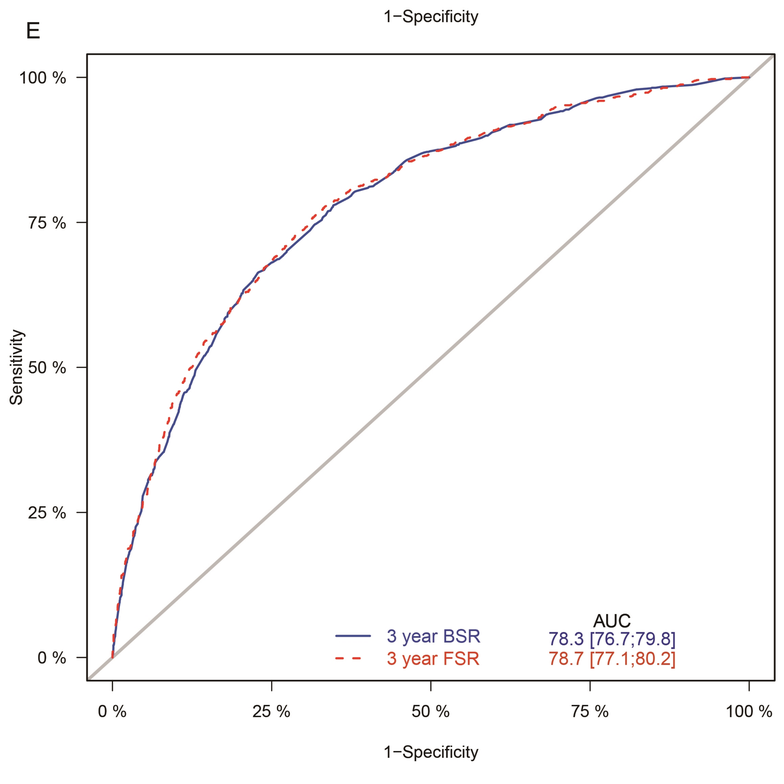
Supplementary Figure 3 Variables selection methods. (A, B) The selection of variables using the BSR method. (C) The LASSO coefficient profile of 14-related variables in primary cohort. (D) 10-fold cross-validation (CV) for tuning parameter (λ) selection.

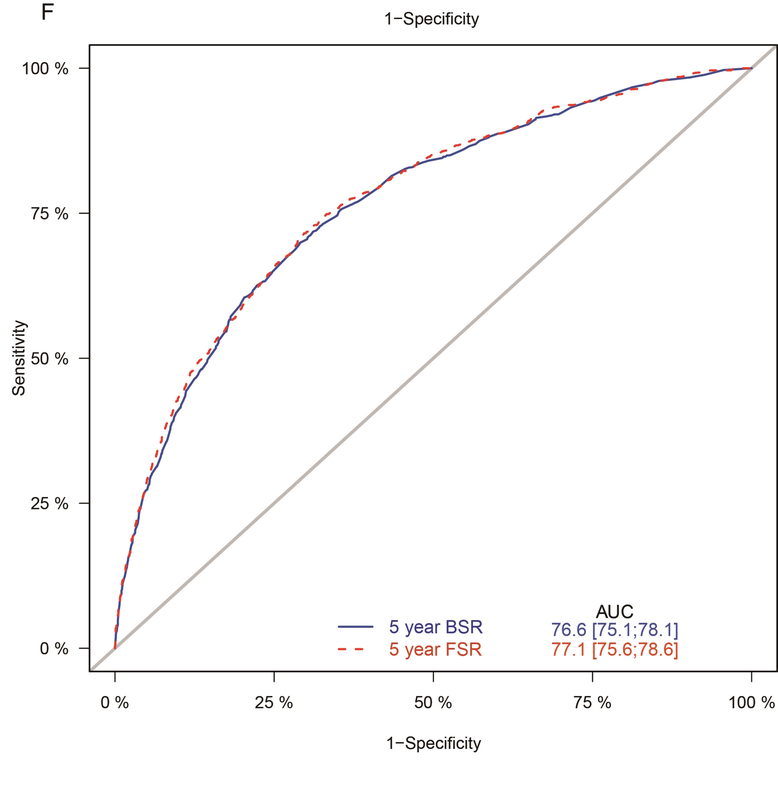




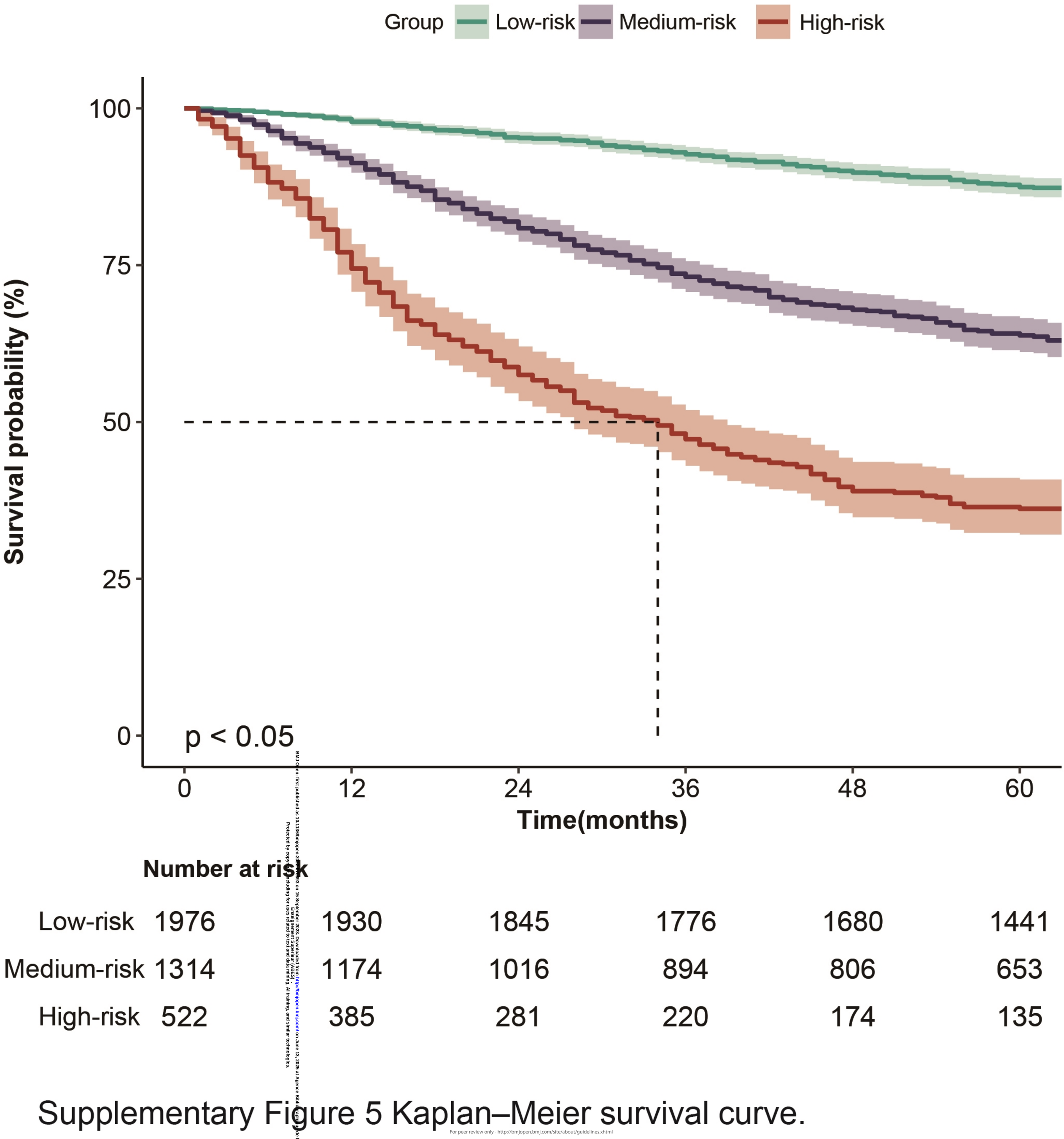








Supplementary Figure 4 Comparison of ROC curves of nomograms created with variables screened by BSR, LASSO, and FSR methods to predict 1-, 3-, and 5-year CSS after surgical treatment of advanced EC. BSR, best subsets regression. LASSO, least absolute shrinkage and selection operator. FSR, forward stepwise regression.



Supplementary Figure 5 Kaplan–Meier survival curve.

Table S1 BSR, LASSO and FSR screening of variables doing predictive models for IDI, NRI comparison.

Index	FSR vs. LASSO			FSR vs. BSR		
	Estimate	95% CI	*P*-value	Estimate	95% CI	*P*-value
IDI						
For 1-year CSS	0.006	0.002-0.012	0.01	0.013	0.006-0.019	<0.05
For 3-year CSS	0.004	0.001-0.008	0.01	0.012	0.006-0.018	<0.05
For 5-year CSS	0.003	0.001-0.007	0.01	0.011	0.006-0.016	<0.05
NRI						
For 1-year CSS	0.119	0.034-0.174	<0.05	0.214	0.113-0.254	<0.05
For 3-year CSS	0.033	(-0.013)-0.112	0.09	0.106	0.044-0.15	<0.05
For 5-year CSS	0.017	(-0.025)-0.099	0.289	0.09	0.053-0.128	<0.05

LASSO, least absolute shrinkage and selection operator. BSR, best subsets regression. FSR, forward stepwise regression. IDI, integrated discrimination improvement. NRI, net reclassification index. CI, confidence interval.

Table S2 IDI, and NRI of the nomogram and the FIGO stage in survival prediction for the advances endometrial carcinoma patients after surgical treatment.

Index	Training cohort			Validation cohort		
	Estimate	95% CI	*P*-value	Estimate	95% CI	*P*-value
IDI (vs. the FIGO stage)						
For 1-year CSS	0.062	0.047-0.084	<0.05	0.071	0.046-0.111	<0.05
For 3-year CSS	0.099	0.084-0.123	<0.05	0.119	0.088-0.155	<0.05
For 5-year CSS	0.112	0.095-0.133	<0.05	0.138	0.103-0.174	<0.05
NRI (vs. the FIGO stage)						
For 1-year CSS	0.364	0.306-0.425	<0.05	0.376	0.293-0.482	<0.05
For 3-year CSS	0.354	0.308-0.395	<0.05	0.352	0.302-0.421	<0.05
For 5-year CSS	0.337	0.292-0.377	<0.05	0.353	0.293-0.419	<0.05

TRIPOD Checklist: Prediction Model Development and Validation

Section/Topic	Item		Checklist Item	Page
Title and abstract				
Title	1	D;V	Identify the study as developing and/or validating a multivariable prediction model, the target population, and the outcome to be predicted.	1
Abstract	2	D;V	Provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusions.	1
Introduction				
Background and objectives	3a	D;V	Explain the medical context (including whether diagnostic or prognostic) and rationale for developing or validating the multivariable prediction model, including references to existing models.	2–3
	3b	D;V	Specify the objectives, including whether the study describes the development or validation of the model or both.	3
Methods				
Source of data	4a	D;V	Describe the study design or source of data (e.g., randomized trial, cohort, or registry data), separately for the development and validation data sets, if applicable.	3
	4b	D;V	Specify the key study dates, including start of accrual; end of accrual; and, if applicable, end of follow-up.	3
Participants	5a	D;V	Specify key elements of the study setting (e.g., primary care, secondary care, general population) including number and location of centres.	3
	5b	D;V	Describe eligibility criteria for participants.	3
	5c	D;V	Give details of treatments received, if relevant.	3–4
Outcome	6a	D;V	Clearly define the outcome that is predicted by the prediction model, including how and when assessed.	
	6b	D;V	Report any actions to blind assessment of the outcome to be predicted.	3–4
Predictors	7a	D;V	Clearly define all predictors used in developing or validating the multivariable prediction model, including how and when they were measured.	3–4
	7b	D;V	Report any actions to blind assessment of predictors for the outcome and other predictors.	3–4
Sample size	8	D;V	Explain how the study size was arrived at.	3–4
Missing data	9	D;V	Describe how missing data were handled (e.g., complete-case analysis, single imputation, multiple imputation) with details of any imputation method.	3–4
Statistical analysis methods	10a	D	Describe how predictors were handled in the analyses.	4
	10b	D	Specify type of model, all model-building procedures (including any predictor selection), and method for internal validation.	4
	10c	V	For validation, describe how the predictions were calculated.	4
	10d	D;V	Specify all measures used to assess model performance and, if relevant, to compare multiple models.	4
	10e	V	Describe any model updating (e.g., recalibration) arising from the validation, if done.	4
Risk groups	11	D;V	Provide details on how risk groups were created, if done.	5
Development vs. validation	12	V	For validation, identify any differences from the development data in setting, eligibility criteria, outcome, and predictors.	4
Results				
Participants	13a	D;V	Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful.	5–6
	13b	D;V	Describe the characteristics of the participants (basic demographics, clinical features, available predictors), including the number of participants with missing data for predictors and outcome.	5
	13c	V	For validation, show a comparison with the development data of the distribution of important variables (demographics, predictors and outcome).	5–6
Model development	14a	D	Specify the number of participants and outcome events in each analysis.	6
	14b	D	If done, report the unadjusted association between each candidate predictor and outcome.	7–9
Model specification	15a	D	Present the full prediction model to allow predictions for individuals (i.e., all regression coefficients, and model intercept or baseline survival at a given time point).	7–9
	15b	D	Explain how to the use the prediction model.	9
Model performance	16	D;V	Report performance measures (with CIs) for the prediction model.	8–9
Model-updating	17	V	If done, report the results from any model updating (i.e., model specification, model performance).	9
Discussion				
Limitations	18	D;V	Discuss any limitations of the study (such as nonrepresentative sample, few events per predictor, missing data).	11–12
Interpretation	19a	V	For validation, discuss the results with reference to performance in the development data, and any other validation data.	10–11
	19b	D;V	Give an overall interpretation of the results, considering objectives, limitations, results from similar studies, and other relevant evidence.	10–11
Implications	20	D;V	Discuss the potential clinical use of the model and implications for future research.	10–

TRIPOD Checklist: Prediction Model Development and Validation

				11
Other information				
Supplementary information	21	D;V	Provide information about the availability of supplementary resources, such as study protocol, Web calculator, and data sets.	12–13
Funding	22	D;V	Give the source of funding and the role of the funders for the present study.	13

*Items relevant only to the development of a prediction model are denoted by D, items relating solely to a validation of a prediction model are denoted by V, and items relating to both are denoted D;V. We recommend using the TRIPOD Checklist in conjunction with the TRIPOD Explanation and Elaboration document.

BMJ Open

Development and validation of a prognostic nomogram for predicting cancer-specific survival in advanced endometrial carcinoma after surgery: a retrospective analysis of the SEER database

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2022-070893.R2
Article Type:	Original research
Date Submitted by the Author:	24-Aug-2023
Complete List of Authors:	Zheng, Chunqin; Shantou Central Hospital, Department of Obstetrics and Gynecology Chen, Weiqiang; Shantou Central Hospital, Department of Anesthesiology Zheng, Zhixiang; Shantou Central Hospital, Department of Obstetrics and Gynecology Liang, Xiaoling; Shantou Central Hospital, Department of Obstetrics and Gynecology Xu, Xiuxia; Shantou Central Hospital, Department of Obstetrics and Gynecology Fang, Danmei; Shantou Central Hospital Department of Obstetrics and Gynecology Ma, Ruijun; Shantou Central Hospital, Department of Obstetrics and Gynecology Fan, Fufang; Shantou Central Hospital Department of Obstetrics and Gynecology Ni, Yanhong; Shantou Central Hospital Department of Obstetrics and Gynecology Zhang, Peili; Shantou Central Hospital Department of Obstetrics and Gynecology Wu, Xuanhua; Shantou Central Hospital, Department of Obstetrics and Gynecology
Primary Subject Heading:	Oncology
Secondary Subject Heading:	Surgery
Keywords:	Gynaecological oncology < GYNAECOLOGY, Surgical pathology < PATHOLOGY, Radiation oncology < RADIOLOGY & IMAGING, Gynaecological oncology < ONCOLOGY, HEALTH ECONOMICS

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Development and validation of a prognostic nomogram for predicting cancer-specific survival in advanced endometrial carcinoma after surgery: a retrospective analysis of the SEER database

Chunqin Zheng[1*], Weiqiang Chen[2], Zhixiang Zheng[1], Xiaoling Liang[1], Xiuxia Xu[1], Danmei Fang[1], Ruijun Ma[1], Fufang Fan[1], Yanhong Ni[1], Peili Zhang[1], Xuanhua Wu[1]

[1]Department of Obstetrics and Gynecology, Shantou Central Hospital, Shantou 515000, China

[2]Department of Anesthesiology, Shantou Central Hospital, Shantou 515000, China.

[*] **Correspondence:** Chunqin Zheng, Department of Obstetrics and Gynecology, Shantou Central Hospital, Shantou 515000, China. 467430374@qq.com; ORCID:0000-0003-1840-9173.

Abstract

Objective We aimed to construct and validate a prognostic nomogram to predict cancer-specific survival (CSS) after surgery in patients with advanced endometrial carcinoma (EC).

Design This study was a retrospective cohort study.

Setting and participants The Surveillance, Epidemiology, and End Results (SEER) database contains cancer incidence and survival data from population-based cancer registries in the USA. A total of 5,445 patients from the SEER database diagnosed with advanced EC between 2004 and 2015 were included and randomized 7:3 into a training cohort (n = 3812) and a validation cohort (n = 1633).

Primary and secondary outcomes: CSS.

Results The nomograms for CSS included 10 variables (positive regional nodes, age, tumor size, FIGO stage, grade, ethnicity, income, radiation, chemotherapy, and historical stage) based on the forward stepwise regression results. They revealed discrimination and calibration using the concordance index (C-index) and area under the time-dependent receiver operating characteristic curve (time-dependent AUC), with a C-index value of 0.7324 (95% confidence interval [CI] = 0.7181–0.7468) and 0.7511 (95% CI = 0.7301–0.7722) for the training and validation cohorts, respectively. Using calibration plots, a high degree of conformance was proven between the predicted and observed results. Additionally, a comparison of the nomogram and FIGO staging based on changes in the C-index, net reclassification index, and integrated discrimination improvement demonstrated that the nomogram was better in accuracy and efficacy.

Conclusions We successfully constructed an accurate and effective nomogram to predict CSS in patients with advanced EC, which may help clinicians determine optimal individualized treatment strategies for patients with advanced EC. The predictive performance of the nomogram was evaluated thoroughly, but only internally. Therefore, further validation using different data sources is warranted in future related studies.

Strengths and limitations of this study

⇒The SEER database is a large database with sufficiently large samples.

⇒The SEER database lacks laboratory test data, which may influence the prognoses of patients with advanced EC.

⇒The chemotherapy and radiotherapy information contained in the SEER database can only be obtained by signing legal agreements that are currently unavailable.

⇒This study may have suffered from selection bias, as all cases were retrieved from the same database.

⇒Our nomogram's predictive performance was evaluated thoroughly, but only internally. Therefore, external validation using different data sources is warranted.

Introduction

Endometrial carcinoma (EC) is the sixth most common cancer in women, with 417,000 new cases diagnosed worldwide in 2020 [1]. There are two histological types of EC [2,3]. Type I tumors include those with grade 1 or 2 endometrioid histological classifications, accounting for approximately 80% of ECs. Type II tumors account for 10–20% of ECs, and include grade 3 endometrioid tumors and tumors with non-endometrioid histology. EC is primarily treated surgically, with radiation and chemotherapy as common adjuvant modalities. For patients with EC who undergo surgery, adjuvant therapy determines disease recurrence for risk stratification based on tumor stage, tumor histology, and other pathologic factors. There is overwhelming evidence that traditional pathological features such as histopathological type, grade, myometrial invasion, and lymphovascular space invasion (LVSI) are imperative for assessing prognosis [4]. Molecular classification in high-grade and/or high-risk ECs shows that POLE-mutated (POLEmut) tumors have an excellent prognosis, p53-abnormal (p53abn) tumors have a poor prognosis, and ECs with mismatch repair deficiency (MMRd) or non-specific molecular profile (NSMP) have an intermediate prognosis [5]. The latest European (ESGO/ESTRO/ESP 2020)/American (NCCN 2020) guidelines combining traditional pathology and The Cancer Genome Atlas (TCGA) molecular groups have proposed a novel risk stratification model: low, intermediate, high–intermediate, high, and advanced metastasis [6]. Generally, the five-year survival rates are 80–90% and 70–80% for stage I and II ECs, respectively, and 20–60% for stage III and IV ECs [7,8]. Stage III and IV ECs are classified as advanced or high-risk ECs. Patients with advanced and recurrent EC have poor prognoses, with an expected 5-year survival rate of <20% [9]. Due to its high mortality rate, a clinical model for predicting the prognosis of patients with advanced EC is necessary. Although the Federation of Gynaecology and Obstetrics (FIGO) staging system has been widely used to predict the survival of EC patients, this approach still suffers from several limitations [10].

A nomogram is a simple visualization tool used by oncologists to predict and quantify patient survival based on multiple variables. Nomograms have been used for patients with EC [11], and Yang et al. published a nomogram for patients with stage IIIC EC following surgery [12]. However, there is no specific prognostic prediction for patients with advanced EC following surgery.

The traditional statistical strategy for EC- adopted variables was significant only on univariate analysis, which led to model overfitting with generally poor results [13]. Certain advanced statistical methodologies may, however, minimize this limitation. These include the best subset regression (BSR), forward stepwise regression (FSR), and least absolute shrinkage and selection operator (LASSO) approaches [14-16]. In this study, we aimed to establish an effective and noninvasive nomogram to predict cancer-specific survival (CSS) in advanced EC following surgery, incorporating advanced statistical methodologies.

Methods

Data sources and patient selection

The Surveillance, Epidemiology, and End Results (SEER) database contains cancer incidence and survival data from population-based cancer registries in the USA. EC case data with complete follow-up records were selected from the 2004–2015 SEER database (SEER Research Plus Data, 17 Registries, November 2021 Sub [2000–2019]) using SEER*Stat V. 8.4.0.1. The inclusion criteria were as follows: primary sites, C54.1-9 and C55.9 [17]; site and morphology, 8380/3 (based on the International Classification of Tumor Diseases for Oncology [ICD-O], Third Edition); histology, 8140-8389 (adenomas and adenocarcinomas); International Federation of Gynecology and Obstetrics (FIGO) stage, III/IV; and therapy, surgical treatment. The exclusion criteria were as follows: (1) undetermined cause of death; (2) undetermined survival time or survival time < 1 month; (3) undetermined tumor size; (4) undetermined lymphadenectomy; (5) unknown regional node status; (6) unknown tumor grade; (7) unknown months from diagnosis to treatment; (8) unknown ethnicity; and (9) unknown median household income. A flowchart of patient screening is shown in Supplemental Figure 1.

Data on variables, including age at diagnosis, year of diagnosis, tumor size, ethnicity, marital status, histologic stage, tumor grade, FIGO stage, lymphadenectomy, regional node positivity, chemotherapy, radiation status, months from diagnosis to treatment, survival time, median household income, and CSS, were collected from the SEER database. The radiation status (with/without radiation) and chemotherapy status (with/without chemotherapy) were of two categories. Marital status was classified as unmarried (single, unmarried, or living with a domestic partner), married, other (divorced, widowed, or separated), or unknown. Grades were associated with each tumor. ICD-O-2 defines grade I as well-differentiated, grade II as moderately differentiated, grade III as poorly differentiated, and grade IV as undifferentiated. According to

the SEER registry, income was examined as aggregate data based on US median income. The median household income is the median household income for the past 12 months, and it was classified into three groups: ≤ 54,999, 55,000–69,999, and ≥70,000. The historical stage was derived from the Collaborative Stage for 2004–2015 and divided into *in situ*, localized, regional, distant, and unknown categories. In the localized stage, an invasive neoplasm is entirely confined to the organ of origin. In the regional stage, a neoplasm has extended 1) beyond the limits of the organ of origin directly into the surrounding organs or tissues, 2) into the regional lymph nodes via the lymphatic system, or 3) into the regional lymph nodes via a combination of extension and regional lymph nodes. In the distant stage, the neoplasm has spread to parts of the body that are remote from the primary tumor. This study categorized lymphadenectomy into two categories: with and without regional lymph node dissection. Failure to perform lymph node dissection included failure to remove or aspirate local lymph nodes, local lymph node biopsy or aspiration, or sentinel lymph node biopsy. Lymph node dissection includes the removal of an unknown number of regional lymph nodes, the removal of 1–3 regional lymph nodes, the removal of ≥ 4 regional lymph nodes, and regional lymph node dissection with anterior lymph node biopsy.

Statistical analysis

X-tile software (Yale University, New Haven, CT, USA) was used to determine the cutoff values for age at diagnosis, tumor size, positive regional nodes, and risk stratification [18]. Statistical analyses were conducted using R 4.0.2 (R Foundation for Statistical Computing, Vienna, Austria; http://www.r-project.org) in the RStudio environment, as well as with Free Statistics 1.8 (Beijing FreeClinical Medical Technology Co., LTD.). CSS was the primary endpoint of this study. The patients were randomly assigned to training and validation cohorts at a 7:3 ratio. Categorical variables are presented as frequencies and proportions. Chi-squared tests were used to compare clinicopathological characteristics between the training and validation cohorts. Statistical significance was set at $P < 0.05$. BSR, FSR, and LASSO were used to select the variables. Significant prognostic factors were identified using the Cox proportional hazards model. A nomogram associated with CSS was constructed and incorporated into the known prognostic factors. The nomogram performance was validated through both training and validation, using the area under the time-dependent receiver operating characteristic curve (time-dependent AUC) to assess its discriminative abilities. Calibration curves were plotted to compare the predicted CSS with the actual CSS after one, three, and five years. The area under the curve (AUC) values ranged from 0.5–1.0, with 0.5 representing random variability and 1.0 representing perfect fit. AUC values g> 0.7 usually indicate rational estimation. The nomogram was compared to the FIGO staging system using the net reclassification index (NRI) and integrated discrimination improvement (IDI). NRI and IDI can be used as alternatives to AUC for assessing the effectiveness of a new risk

prediction model and for determining its effectiveness [19,20]. The Kaplan–Meier method was used to compare the risk stratification of the nomogram.

Patient and public involvement

None.

Results

Patient characteristics

A total of 5,445 patients with advanced EC following surgery were screened from the SEER database according to our inclusion and exclusion criteria. They were randomly allocated into training (n = 3,812) and validation cohorts (n = 1,633) at a 7:3 ratio. Patient characteristics are shown in Table 1. No statistically significant differences were found in the indicators between the two groups (all P > 0.05).

Table 1 The basic characteristics of endometrial carcinoma patients in the study

Variables	Primary cohort (n=5445)	Training cohort (n = 3812)	Validation cohort (n = 1633)	*p*-value
Race, n (%)				0.903
White	4444 (81.6)	3107(81.5)	1337 (81.9)	
Black	351 (6.4)	245(6.4)	106 (6.5)	
Other[a]	650 (11.9)	460 (12.1)	190(11.6)	
Chemotherapy, n (%)				0.899
No	2167 (39.8)	1515(39.7)	652(39.9)	
Yes	3278 (60.2)	2297(60.3)	981(60.1)	
Historic stage[b], n (%)				0.62
Localized	9 (0.2)	5 (0.1)	4 (0.2)	
Regional	3903 (71.7)	2731(71.6)	1172 (71.8)	
Distant	1533 (28.2)	1076(28.2)	457 (28)	
Tumor grade[c], n (%)				0.631
I	1226 (22.5)	853 (22.4)	373(22.8)	
II	2166 (39.8)	1506(39.5)	660(40.4)	
III-IV	2053 (37.7)	1453(38.1)	600(36.7)	
Radiation, n (%)				0.055
No	2659 (48.8)	1894(49.7)	765(46.8)	
Yes	2786 (51.2)	1918(50.3)	868(53.2)	
Marital status, n (%)				0.438
Unmarried	1232 (22.6)	881 (23.1)	351(21.5)	
Married	2675 (49.1)	1855(48.7)	820(50.2)	
Other[d]	1375 (25.3)	967 (25.4)	408 (25)	
Unknown	163 (3.0)	109(2.9)	54(3.3)	
Lymphadenectomy[e], n (%)				0.601
No	70 (1.3)	51 (1.3)	19(1.2)	
Yes	5375 (98.7)	3761(98.7)	1614 (98.8)	
FIGO stage, n (%)				0.11
III	4741 (87.1)	3301(86.6)	1440 (88.2)	
IV	704 (12.9)	511 (13.4)	193(11.8)	

Age of diagnosis, n (%)				0.553
24-64 years	3362 (61.7)	2352(61.7)	1010 (61.8)	
65-75 years	1392 (25.6)	965 (25.3)	427(26.1)	
76-96 years	691 (12.7)	495 (13)	196 (12)	
Regional nodes positive, n (%)				0.447
0	2415 (44.4)	1694(44.4)	721(44.2)	
1-2	1954 (35.9)	1381(36.2)	573(35.1)	
3-82	1076 (19.8)	737 (19.3)	339(20.8)	
Year of diagnosis, n (%)				0.981
2004-2009	2152 (39.5)	1507(39.5)	645(39.5)	
2010-2015	3293 (60.5)	2305(60.5)	988(60.5)	
Tumor size[f], n (%)				0.319
0-35mm	1640 (30.1)	1149(30.1)	491(30.1)	
36-78mm	2847 (52.3)	1974(51.8)	873(53.5)	
79-790mm	958 (17.6)	689 (18.1)	269(16.5)	
Income, n (%)				0.701
≤54,999$	1010 (18.5)	707 (18.5)	303(18.6)	
55,000-69,999$	2168 (39.8)	1505(39.5)	663(40.6)	
≥70,000$	2267 (41.6)	1600 (42)	667(40.8)	
Diagnosis time[g], Mean±SD	1.1±1.2	1.1±1.2	1.1±1.1	0.375

a, American Indian/AK Native, Asian/Pacific Islander.
b, historic stage derived from Collaborative Stage (CS) for 2004-2015. Localized, an invasive neoplasm confined entirely to the organ of origin. Regional, a neoplasm that has extended. Distant, a neoplasm that has spread to parts of the body remote from the primary tumor.
c, ICD-O-2 defines grade I as well differentiated, grade II as moderately differentiated, grade III as poorly differentiated, and grade IV as undifferentiated.
d,divorced, widowed, separated.

e, The article categorizes lymphadenectomy into two categories: those involving regional lymph node dissection and those without it. Without lymphadenectomy includes failure to remove or aspirate local lymph nodes, local lymph node biopsy or aspiration, and sentinel lymph node biopsy only.
Lymphadenectomy includes removal of an unknown number of regional lymph nodes, removal of one
to three regional lymph nodes, removal of four or more regional lymph nodes, and regional lymph
node dissection with anterior lymph node biopsy.
f, Based on X-tile procedure cut-offs.
g, Months from diagnosis to treatment.

Nomogram variable screening

Age, tumor size, regional node positivity, and linear predictors (linear predictor = 0.448 * black ethnicity + 0.166 * other ethnicity - 0.158 * chemotherapy - 0.706 * historical stage regional - 0.702 * historical stage distant + 0.25 * grade II + 0.913 * (grade III–IV) - 0.261 * radiation +

0.977 * FIGO stage IV + 0.471 * (age of diagnosis 65–75 years) + 0.881 * (age of diagnosis 76–96 years) + 0.263 * (tumor size 36–78 mm) + 0.577 * (tumor size 79–790 mm) + 0.317 * (regional nodes positive 1–2) + 0.619 * (regional nodes positive 3–82) - 0.132 * (income $55,000–69,999) - 0.195 * (income ≥ $70,000) - 0.271) were divided into three categories using X-tile software. The best cut-off ages were 64 and 75 years (Online Supplemental Figure 2), the best cut-off tumor sizes were 35 mm and 78 mm (Online Supplemental Figure 2), the best cut-off regional node positivities were 0 and 2 (Online Supplemental Figure 2), and the best cut-off linear predictors were 0.2 and 1.2 (Online Supplemental Figure 2).

BSR, LASSO, and FSR were used to select the variables. The BSR method showed great benefits for variable selection because all possible combinations of variables were calculated and the final selected combination was based on the minimum Bayesian information criterion (BIC). As is shown in Online Supplemental Figure 3A/B, six variables (positive regional nodes, age at diagnosis, tumor size, FIGO stage, grade, and ethnicity) were selected from the variables in the training cohort. Considering that the number of independent variables included in the regression equation should be ~10–15 × the number of ending events, we used LASSO to select the variables. As is shown in Online Supplementary Figure 3C/D, seven variables (positive regional nodes, age at diagnosis, tumor size, FIGO stage, grade, radiation, and income) were selected from the variables in the training cohort. Furthermore, the FSR selected ten variables (positive regional nodes, age at diagnosis, tumor size, FIGO stage, grade, ethnicity, chemotherapy, history, radiation, and income) in the training cohort. As a result (Online Supplementary Figure 4), the discrimination of the FSR was highest in the 1-, 3-, and 5-year training cohorts, with a concordance index (C-index) of 0.808 (95% confidence interval [CI]: 0.786–0.83), 0.787 (95% CI: 0.771–0.802), and 0.771 (95% CI: 0.756–0.786), respectively. Moreover, compared to LASSO and BSR (Online Supplementary Table S1), the 1-, 3-, and 5-year IDIs of FSR were significantly improved (FSR vs. LASSO: 0.006, 0.004, and 0.003, respectively, all $P < 0.05$; FSR vs. BSR: 0.013, 0.012, 0.011, respectively, all $P < 0.05$). Therefore, the nomogram obtained from the FSR was optimal (Figure 1). These 10 variables were obtained from the FSR using multivariate Cox analysis due to their optimal performance for predicting CSS in patients with advanced EC following surgery. The results showed that ethnicity, chemotherapy, historical stage, grade, radiation, FIGO stage, age at diagnosis, tumor size, positive regional nodes, and income were independent prognostic factors in this patient group (Table 2). A nomogram for predicting 1-, 3-, and 5-year CSS was constructed based on these 10 key factors (Figure 1).

Table 2 Univariate and multivariable cox regression analysis of cancer-specific survival.

Variable	Univariate analysis		Multivariate analysis	
	HR	*P*-value	HR	*P*-value

Race				
White	1(Ref)		1(Ref)	
Black	1.88 (1.6~2.21)	<0.001	1.49 (1.26~1.75)	<0.001
Other[a]	1.03 (0.89~1.2)	0.697	1.15 (0.99~1.34)	0.072
Chemotherapy				
No	1(Ref)		1(Ref)	
Yes	1 (0.9~1.1)	0.958	0.84 (0.75~0.93)	0.001
Historic stage[b]				
Localized	1(Ref)		1(Ref)	
Regional	0.41 (0.17~0.98)	0.044	0.32 (0.13~0.78)	0.012
Distant	0.84 (0.35~2.03)	0.705	0.34 (0.14~0.82)	0.016
Tumor grade[c]				
I	1(Ref)		1(Ref)	
II	1.51 (1.28~1.78)	<0.001	1.43 (1.21~1.68)	<0.001
III-IV	3.63 (3.12~4.23)	<0.001	2.79 (2.39~3.26)	<0.001
Radiation				
No	1(Ref)		1(Ref)	
Yes	0.67 (0.61~0.74)	<0.001	0.76 (0.69~0.84)	<0.001
FIGO stage				
III	1(Ref)		1(Ref)	
IV	3.33 (2.98~3.72)	<0.001	2.6 (2.26~3)	<0.001
Age of diagnosis (year)				
24-64	1(Ref)		1(Ref)	
65-75	1.47 (1.32~1.65)	<0.001	1.52 (1.36~1.7)	<0.001
76-96	2.37 (2.08~2.7)	<0.001	2.38 (2.08~2.73)	<0.001
Tumor size (mm)				
0-35	1(Ref)		1(Ref)	
36-78	1.54 (1.36~1.74)	<0.001	1.25 (1.1~1.41)	<0.001
79-790	2.38 (2.07~2.74)	<0.001	1.72 (1.48~2)	<0.001
Regional nodes positive				
Negative	1(Ref)		1(Ref)	
1-2	1.34 (1.2~1.5)	<0.001	1.36 (1.21~1.53)	<0.001
3-82	1.98 (1.76~2.24)	<0.001	1.86 (1.62~2.14)	<0.001
Income				
≤54,999$	1(Ref)		1(Ref)	

55,000-69,999$	0.82 (0.72~0.94)	0.003	0.82 (0.72~0.94)	0.003
≥70,000$	0.72 (0.63~0.82)	<0.001	0.75 (0.65~0.85)	<0.001

a, American Indian/AK Native, Asian/Pacific Islander.

b, historic stage derived from Collaborative Stage (CS) for 2004-2015. Localized, an invasive neoplasm confined entirely to the organ of origin. Regional, a neoplasm that has extended. Distant, a neoplasm that has spread to parts of the body remote from the primary tumor.

c, I, well differentiated; II, moderately differentiated; III, poorly differentiated; IV, undifferentiated.

Nomogram construction and performance

As shown in Figure 1, we developed a nomogram based on FSR to predict one-, three-, and five-year CSS rates. According to the training and validation cohort data, the C-index values were 0.7324 (95% CI = 0.7181–0.7468) and 0.7511 (95% CI = 0.7301–0.7722), respectively. According to Figure 2A/B, the AUC for the prediction of CSS within five years was > 0.7 in both the training and validation cohorts, indicating favorable discrimination. Figure 2C/E/G shows the calibration curves of the 1-, 3- and 5-year CSS for advanced EC following surgery in the training cohort. Figure 2D/F/H shows the calibration curves of the 1-, 3-, and 5-year CSS for advanced EC following surgery in the validation cohort. The dashed black line indicates the ideal reference line, where the predicted probabilities matched the observed survival rates. The red dots represent the performance of the nomogram. The closer the solid red line is to the black dashed line, the more accurately the model predicted survival. As is shown in Figure 2C–H, the calibration curves of the nomogram showed high concordance between the predicted and observed survival probabilities.

Comparative clinical value of the nomogram and FIGO stage

The accuracies of the nomogram and FIGO stage were compared based on changes in the ROC curves and time-dependent AUCs (Figure 3). Compared to the FIGO stage (Online Supplementary Table S2), the 1-, 3-, and 5-year IDI of the nomogram was significantly greater (nomogram vs. FIGO stage: 0.062, 0.099, and 0.112, respectively). Moreover, compared to the FIGO stage (Online Supplementary Table S2), the 1-, 3-, and 5-year NRI of the nomogram was significantly greater (nomogram vs. FIGO stage: 0.364, 0.354, and 0.337, respectively). According to these results, the nomogram predicted the prognosis more accurately than the FIGO stage.

An assessment of the risk of advanced EC following surgery

In addition to the nomogram, we developed a risk stratification system based on the linear predictor cut-off value for each patient in the training cohort. The patients were divided into three groups according to their linear predictors: low risk (≤ 0.2), intermediate risk (0.21–1.2), and high risk (> 1.2). There was a significant difference in CSS between the low-, medium-, and high-risk groups according to our Kaplan–Meier analysis (all P < 0.05, Online Supplementary Figure 5).

Furthermore, according to the nomogram, a total score of ≤ 185 indicated low risk, 185 ≤ 285 indicated medium risk, and > 285 indicated high risk. These results show that the nomogram had excellent risk-stratification capabilities.

Discussion

In this study, we used actual information from patients with advanced EC following surgery. We also developed a prognostic nomogram and risk stratification system using data from the SEER database. The nomogram produced excellent internal and external results, as shown by calibration, C-index, and ROC curves.

Few studies have focused on predicting postoperative CSS in patients with advanced EC. This study focused on postoperative CSS in patients with stage III–IV cancer for two key reasons. First, advanced EC has high prognostic heterogeneity and a poor survival rate, with a five-year survival rate of 20–60% (although different patients have different prognoses). Due to the lack of a reliable model to predict survival in patients with advanced EC following surgery, individualized clinical management and surveillance can be challenging. Second, patients with advanced EC have significantly higher incidence and mortality rates following surgery, leading to confounding bias in prognostic indicators.

EC is usually treated surgically, and postoperative treatment depends on risk factors such as age, tumor stage, myometrial infiltration depth, and histologic grade [21,22]. In this study, a prognostic model after the surgical treatment of advanced EC was constructed based on 10 variables (ethnicity, chemotherapy, historical stage, tumor grade, radiation therapy, FIGO stage, age at diagnosis, tumor size, positive regional nodes, and median household income) screened using FSR. The scores were calculated for each item based on the subtype of each independent prognostic factor. The total score was calculated using scores corresponding to the independent prognostic factors. Each subgroup variable was assigned a score from 0–100 according to its contribution. All enrolled variables were added to generate a total score on the bottom scale, which was then converted to predict CSS. CSS at 1-, 3-, and 5 years was determined by drawing a vertical line on the total score scale, with higher scores indicating a worse prognosis. According to the nomogram, the FIGO stage plays the largest role in prognosis, followed by tumor grade and age at diagnosis.

Cancer grade, histological subtype, tumor size, LVSI, lymph node status, and cervical involvement are vital prognostic factors in patients with EC [23]. In this study, tumor grade, tumor size, and lymph node status were important prognostic factors following surgical treatment for advanced EC. Tumor grade has also been shown to be a prognostic factor in EC [24], and our nomogram indicates that poorly differentiated or undifferentiated tumors have poor prognoses. Conflicting results have been reported concerning the impact of tumor size on survival outcomes. Preoperative ultrasound tumor size was apparently not a prognostic factor for death from any cause in women

with EC [25]. However, tumor size was an independent prognostic factor for recurrence alone [26, 27] and for recurrence and death due to EC [28]. Lymph node metastasis further contributes to poor prognosis in patients with EC; however, there is no consensus on the value and extent of lymph node dissection [29]. In this study, we found that positive lymph nodes could affect the prognosis of surgical treatment for advanced EC, consistent with the findings of previous studies. However, this study did not reflect whether lymph node dissection was beneficial. This may be related to the fact that the population selected in this study underwent lymph node dissection (98.7%), which was not comparable. Compared to women ≥ 65 years, women < 65 years had a significant survival advantage, as indicated by previous studies [30].

Using advanced EC after surgery as a dataset, this study examined factors that could be included in prognostic nomograms. Nomograms combine multiple factors, including demographic and clinicopathological characteristics, into quantitative models that provide better predictions than FIGO staging [31,32]. FIGO staging has traditionally been used to predict the prognosis of women with EC. Staging using this system is closely associated with CSS. However, patients at the same stage have different prognoses. FIGO staging does not consider factors such as age, radiation status, and income, thus resulting in its prognostic heterogeneity. Therefore, we compared nomograms that included more variables. Nomograms generally have better predictive powers than FIGO staging alone due to their positive NRI and IDI scores.

Based on their total nomogram scores, the patients were classified into low-, intermediate-, and high-risk groups. Significant differences were found in CSS among the three risk groups based on Kaplan–Meier analysis (Online Supplementary Figure 5). This nomogram is highly effective in identifying high-risk groups owing to its poor prognosis. Patients with a total score greater than 285 should receive special attention.

To investigate the potential utility of the nomogram in clinical practice, we analyzed data from the SEER database by using a large sample of data representing different population regions. We followed the recommendations of the Transparent Reporting of Individual Prognosis or Diagnosis Multivariate Predictive Model statement [33]. Bootstrapping and cross-validation methods were used to calculate the calibration curves, time-dependent AUCs, and C-index. These positive results show that our nomogram may be useful for assessing the prognosis of patients with advanced EC after surgery.

Although the nomogram performed well, this study had some key limitations. Carbohydrate antigen 125 (CA125) is a tumor marker whose levels are often elevated in patients with malignant tumors such as ovarian epithelial, fallopian tube, and EC, as well as in those with lung and gastrointestinal adenocarcinomas. In the clinical diagnosis and treatment of EC, CA125 levels are often used to monitor disease changes, evaluate treatment effects, and predict prognosis [34].

Studies have shown that CA125 is an important variable in the prognostic prediction model of EC and can significantly improve its accuracy [35]. Human epididymis protein 4 (HE4) is an acidic whey protein first identified in the epithelium of the distal epididymis [36]. It is expressed in the epithelia of several tissues, including the female reproductive tract, and is overexpressed in several cancers [37]. HE4 is strongly associated with survival in patients with EC [38]. ECs have traditionally been classified into two subtypes (1 and 2) based on their histopathological characteristics [2]. However, this classification system lacks reproducibility and yields heterogeneous molecular groups that hamper the advancement and implementation of precision medicine [39, 40]. It is, therefore, being gradually replaced by a clearly defined system based on molecular phenotypes [41]. The TCGA approach results in the molecular stratification of ECs into four distinct molecular groups: DNA polymerase epsilon ultra-mutated classification, which portends a good prognosis; microsatellite instability hypermutated (intermediate prognosis); copy number-low; and copy number-high (which includes p53 mutations and carries the worst prognosis) [41], ESMO 2022 recommends that molecular staging testing should be performed for all ECs, but POLE testing can be omitted for low-risk patients when conditions are limited. However, MMR and p53 testing should still be performed to identify patients with hereditary EC or high-risk factors [42]. LVSI has a prognostic value in patients with EC independent of TCGA signature, age, and adjuvant treatment, increasing the risk of death from any cause [43]. Since data on CA125, HE4, molecular typing, LVSI, hormonal therapy, or immunotherapy was not published in SEER 2004-2015, these variables were not assessed in this study. In addition, the chemotherapy and radiotherapy information in the SEER database can only be obtained by signing certain legal agreements that appeared unavailable at the time. As a result, we were unable to study the relationship between chemotherapy, radiotherapy, targeted therapy, and EC prognosis. Moreover, the study cases derived from the US SEER database were nonrepresentative of regions outside the USA. Finally, although the predictive performance of the nomogram was evaluated thoroughly using internal data, validation using different external data sources is warranted, and further investigation is recommended.

Conclusions

Our nomogram is more accurate, has better clinical utility, and provides better prognostic predictions than FIGO staging for patients with advanced EC after surgery. However, the predictive performance of the nomogram was evaluated using internal data only. Therefore, using different data sources for external validation is warranted, and further investigation is recommended.

Acknowledgments

We thank the SEER project for providing access to their database and the Free Statistics software version 1.7, provided by Beijing FreeClinical Medical Technology Co., Ltd. We gratefully thank Dr. Jie Liu of the Department of Vascular and Endovascular Surgery, Chinese PLA General Hospital, for his contribution to statistical support, study design consultations, and comments regarding the manuscript.

Contributors

CZ: Study conception, data collection, data analysis, interpretation, drafting, critical revision, and final approval of the article. Data collection by RM, XW, and DF. YN, FF, and PZ Data Analysis. ZZ and XL: study conception. WC and XX: critical revision and final approval of the article.

Funding The authors declare no specific grant from any funding agency in the public, commercial, or not-for-profit sectors for this research.

Competing interests None declared.

Patient consent for publication Not applicable.

Ethics approval This study used data from the SEER database approved by the National Institutes of Health Ethics Program. In addition, access request has been approved for the SEER Research Database.

Provenance and peer review Not commissioned, externally peer reviewed.

Data availability Statement Data are available upon reasonable request. Based on the SEER website (https://www. cancer. gov/policies/accessibility), the National Cancer Institute (NCI) provides access to all individuals seeking information on http://www.cancer.gov. Consent for publication The NCI SEER database is publicly available. Data supporting the findings of this study are available from the corresponding author upon request.

Abbreviations

EC: endometrial cancer. SEER: Linked Surveillance, Epidemiology, and End Results. CSS: Cancer-specific survival. CI: confidence interval. FIGO: International Federation of Gynecology and Obstetrics. BSR: Best subset regression. FSR: forward stepwise regression. LASSO: least absolute shrinkage and selection operator. AUC: area under the receiver operating characteristic curve. C-index: concordance index. NRI: net reclassification index. IDI: integrated discrimination improvement. LVSI: lymphovascular space invasion.

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F: **Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries**. *CA Cancer J Clin* 2021, **71**(3):209-249.

2. Bokhman JV: **Two pathogenetic types of endometrial carcinoma**. *Gynecol Oncol* 1983, **15**(1):10-17.

3. Felix AS, Weissfeld JL, Stone RA, Bowser R, Chivukula M, Edwards RP, Linkov F: **Factors associated with Type I and Type II endometrial cancer**. *Cancer Causes Control* 2010, **21**(11):1851-1856.

4. Cho KR, Cooper K, Croce S, Djordevic B, Herrington S, Howitt B, Hui P, Ip P, Koebel M, Lax S *et al*: **International Society of Gynecological Pathologists (ISGyP) Endometrial Cancer Project: Guidelines From the Special Techniques and Ancillary Studies Group**. *Int J Gynecol Pathol* 2019, **38 Suppl 1**(Iss 1 Suppl 1):S114-s122.

5. McAlpine J, Leon-Castillo A, Bosse T: **The rise of a novel classification system for endometrial carcinoma; integration of molecular subclasses**. *J Pathol* 2018, **244**(5):538-549.

6. Concin N, Matias-Guiu X, Vergote I, Cibula D, Mirza MR, Marnitz S, Ledermann J, Bosse T, Chargari C, Fagotti A *et al*: **ESGO/ESTRO/ESP guidelines for the management of patients with endometrial carcinoma**. *Int J Gynecol Cancer* 2021, **31**(1):12-39.

7. Lewin SN, Herzog TJ, Barrena Medel NI, Deutsch I, Burke WM, Sun X, Wright JD: **Comparative performance of the 2009 international Federation of gynecology and**

obstetrics' staging system for uterine corpus cancer. *Obstet Gynecol* 2010, **116**(5):1141-1149.

8. Creasman WT, Odicino F, Maisonneuve P, Quinn MA, Beller U, Benedet JL, Heintz AP, Ngan HY, Pecorelli S: **Carcinoma of the corpus uteri. FIGO 26th Annual Report on the Results of Treatment in Gynecological Cancer**. *Int J Gynaecol Obstet* 2006, **95 Suppl 1**:S105-143.

9. Costa B, Vale N: **Dostarlimab: A Review**. *Biomolecules* 2022, **12**(8).

10. Kasius JC, Pijnenborg JMA, Lindemann K, Forsse D, van Zwol J, Kristensen GB, Krakstad C, Werner HMJ, Amant F: **Risk Stratification of Endometrial Cancer Patients: FIGO Stage, Biomarkers and Molecular Classification**. *Cancers (Basel)* 2021, **13**(22).

11. Zhu L, Sun X, Bai W: **Nomograms for Predicting Cancer-Specific and Overall Survival Among Patients With Endometrial Carcinoma: A SEER Based Study**. *Front Oncol* 2020, **10**:269.

12. Yang XL, Huang H, Kou LN, Lai H, Chen XP, Wu DJ: **Construction and validation of a prognostic model for stage IIIC endometrial cancer patients after surgery**. *Eur J Surg Oncol* 2022, **48**(5):1173-1180.

13. Harrell FE, Jr., Lee KL, Matchar DB, Reichert TA: **Regression models for prognostic prediction: advantages, problems, and suggested solutions**. *Cancer Treat Rep* 1985, **69**(10):1071-1077.

14. Collins GS, Reitsma JB, Altman DG, Moons KG: **Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. The TRIPOD Group**. *Circulation* 2015, **131**(2):211-219.

15. Debray TP, Damen JA, Snell KI, Ensor J, Hooft L, Reitsma JB, Riley RD, Moons KG: **A guide to systematic review and meta-analysis of prediction model performance**. *Bmj* 2017, **356**:i6460.

16. Walter S, Tiemeier H: **Variable selection: current practice in epidemiological studies**. *Eur J Epidemiol* 2009, **24**(12):733-736.

17. Modh A, Ghanem AI, Burmeister C, Rasool N, Elshaikh MA: **Trends in the utilization of adjuvant vaginal brachytherapy in women with early-stage endometrial carcinoma: Results of an updated period analysis of SEER data**. *Brachytherapy* 2016, **15**(5):554-561.

18. Camp RL, Dolled-Filhart M, Rimm DL: **X-tile: a new bio-informatics tool for biomarker assessment and outcome-based cut-point optimization**. *Clin Cancer Res* 2004, **10**(21):7252-7259.

19. Uno H, Tian L, Cai T, Kohane IS, Wei LJ: **A unified inference procedure for a class of measures to assess improvement in risk prediction systems with survival data**. *Stat Med* 2013, **32**(14):2430-2442.

20. Pencina MJ, D'Agostino RB, Sr., Steyerberg EW: **Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers**. *Stat Med* 2011, **30**(1):11-21.

21. Creutzberg CL, van Putten WL, Koper PC, Lybeert ML, Jobsen JJ, Wárlám-Rodenhuis CC, De Winter KA, Lutgens LC, van den Bergh AC, van de Steen-Banasik E *et al*: **Surgery and postoperative radiotherapy versus surgery alone for patients with stage-1 endometrial carcinoma: multicentre randomised trial. PORTEC Study Group. Post**

Operative Radiation Therapy in Endometrial Carcinoma. *Lancet* 2000, **355**(9213):1404-1411.

22. Keys HM, Roberts JA, Brunetto VL, Zaino RJ, Spirtos NM, Bloss JD, Pearlman A, Maiman MA, Bell JG: **A phase III trial of surgery with or without adjunctive external pelvic radiation therapy in intermediate risk endometrial adenocarcinoma: a Gynecologic Oncology Group study**. *Gynecol Oncol* 2004, **92**(3):744-751.

23. Colombo N, Creutzberg C, Amant F, Bosse T, González-Martín A, Ledermann J, Marth C, Nout R, Querleu D, Mirza MR *et al*: **ESMO-ESGO-ESTRO Consensus Conference on Endometrial Cancer: diagnosis, treatment and follow-up**. *Ann Oncol* 2016, **27**(1):16-41.

24. Stålberg K, Kjølhede P, Bjurberg M, Borgfeldt C, Dahm-Kähler P, Falconer H, Holmberg E, Staf C, Tholander B, Åvall-Lundqvist E *et al*: **Risk factors for lymph node metastases in women with endometrial cancer: A population-based, nation-wide register study-On behalf of the Swedish Gynecological Cancer Group**. *Int J Cancer* 2017, **140**(12):2693-2700.

25. Ambrosio M, Raffone A, Alletto A, Cini C, Filipponi F, Neola D, Fabbri M, Arena A, Raimondo D, Salucci P *et al*: **Is preoperative ultrasound tumor size a prognostic factor in endometrial carcinoma patients?** *Front Oncol* 2022, **12**:993629.

26. Senol T, Polat M, Ozkaya E, Karateke A: **Tumor Diameter for Prediction of Recurrence, Disease Free and Overall Survival in Endometrial Cancer Cases**. *Asian Pac J Cancer Prev* 2015, **16**(17):7463-7466.

27. Sozzi G, Uccella S, Berretta R, Petrillo M, Fanfani F, Monterossi G, Ghizzoni V, Frusca

T, Ghezzi F, Chiantera V *et al*: **Tumor Size, an Additional Risk Factor of Local Recurrence in Low-Risk Endometrial Cancer: A Large Multicentric Retrospective Study**. *Int J Gynecol Cancer* 2018, **28**(4):684-691.

28. Chattopadhyay S, Cross P, Nayar A, Galaal K, Naik R: **Tumor size: a better independent predictor of distant failure and death than depth of myometrial invasion in International Federation of Gynecology and Obstetrics stage I endometrioid endometrial cancer**. *Int J Gynecol Cancer* 2013, **23**(4):690-697.

29. Sari ME, Yalcin İ, Sahin H, Meydanli MM, Gungor T: **Risk factors for paraaortic lymph node metastasis in endometrial cancer**. *Int J Clin Oncol* 2017, **22**(5):937-944.

30. Chan JK, Cheung MK, Huh WK, Osann K, Husain A, Teng NN, Kapp DS: **Therapeutic role of lymph node resection in endometrioid corpus cancer: a study of 12,333 patients**. *Cancer* 2006, **107**(8):1823-1830.

31. Wu J, Zhang H, Li L, Hu M, Chen L, Xu B, Song Q: **A nomogram for predicting overall survival in patients with low-grade endometrial stromal sarcoma: A population-based analysis**. *Cancer Commun (Lond)* 2020, **40**(7):301-312.

32. Gao J, Wang S, Li F, Xu H, Li X, Yan L, Zhu L: **Nomograms for Predicting Overall Survival and Cancer-Specific Survival of Young Patients with Epithelial Ovarian Cancer: Analysis Based on SEER Program**. *Adv Ther* 2022, **39**(1):257-285.

33. Collins GS, Reitsma JB, Altman DG, Moons KG: **Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement**. *Bmj* 2015, **350**:g7594.

34. Jiang T, Huang L, Zhang S: **Preoperative serum CA125: a useful marker for surgical**

management of endometrial cancer. *BMC Cancer* 2015, **15**:396.

35. Li Q, Kong F, Ma J, Wang Y, Wang C, Yang H, Li Y, Ma X: **Nomograms Based on Fibrinogen, Albumin, Neutrophil-Lymphocyte Ratio, and Carbohydrate Antigen 125 for Predicting Endometrial Cancer Prognosis**. *Cancers (Basel)* 2022, **14**(22).

36. Kirchhoff C, Habben I, Ivell R, Krull N: **A major human epididymis-specific cDNA encodes a protein with sequence homology to extracellular proteinase inhibitors**. *Biol Reprod* 1991, **45**(2):350-357.

37. Galgano MT, Hampton GM, Frierson HF, Jr.: **Comprehensive analysis of HE4 expression in normal and malignant human tissues**. *Mod Pathol* 2006, **19**(6):847-853.

38. Behrouzi R, Barr CE, Crosbie EJ: **HE4 as a Biomarker for Endometrial Cancer**. *Cancers (Basel)* 2021, **13**(19).

39. Gilks CB, Oliva E, Soslow RA: **Poor interobserver reproducibility in the diagnosis of high-grade endometrial carcinoma**. *Am J Surg Pathol* 2013, **37**(6):874-881.

40. de Boer SM, Wortman BG, Bosse T, Powell ME, Singh N, Hollema H, Wilson G, Chowdhury MN, Mileshkin L, Pyman J *et al*: **Clinical consequences of upfront pathology review in the randomised PORTEC-3 trial for high-risk endometrial cancer**. *Ann Oncol* 2018, **29**(2):424-430.

41. Kandoth C, Schultz N, Cherniack AD, Akbani R, Liu Y, Shen H, Robertson AG, Pashtan I, Shen R, Benz CC *et al*: **Integrated genomic characterization of endometrial carcinoma**. *Nature* 2013, **497**(7447):67-73.

42. Oaknin A, Bosse TJ, Creutzberg CL, Giornelli G, Harter P, Joly F, Lorusso D, Marth C, Makker V, Mirza MR *et al*: **Endometrial cancer: ESMO Clinical Practice Guideline for**

diagnosis, treatment and follow-up. *Ann Oncol* 2022, **33**(9):860-877.

43. Raffone A, Travaglino A, Raimondo D, Neola D, Maletta M, Santoro A, Insabato L, Casadio P, Fanfani F, Zannoni GF *et al*: **Lymphovascular space invasion in endometrial carcinoma: A prognostic factor independent from molecular signature**. *Gynecol Oncol* 2022, **165**(1):192-197.

Figure Legend:

Figure 1 - Nomograms for predicting 1-,3-, and 5-year cancer-specific survival.
* American Indian/AK Native, Asian/Pacific Islander.

Figure 2 - Time-dependent AUC and calibration curves of the nomogram. (A-B) Time-dependent AUC of using the nomogram to predict cancer-specific survival probability within 5 years in the training and validation cohorts. The red line represents AUC = 0.7, which is considered ideal. (C, E, G) Calibration curves of 1-year, 3-year, and 5-year CSS for advanced EC post-surgery in the training cohort. (D, F, H) Calibration curves of 1-year, 3-year, and 5-year CSS for advanced EC post-surgery in the validation cohort. The black dashed line indicates the ideal reference line where predicted probabilities match the observed survival rates. The red dots represent the performance of the nomogram. The closer the solid red line is to the black dashed line, the more accurately the model predicts survival. AUC: area under the time-dependent receiver operating characteristic curves; CSS: cancer-specific survival. EC: endometrial carcinoma.

Figure 3 - Compares the accuracy of the nomograms and FIGO stage based on the changes in the ROC curves and the time-dependent AUC.

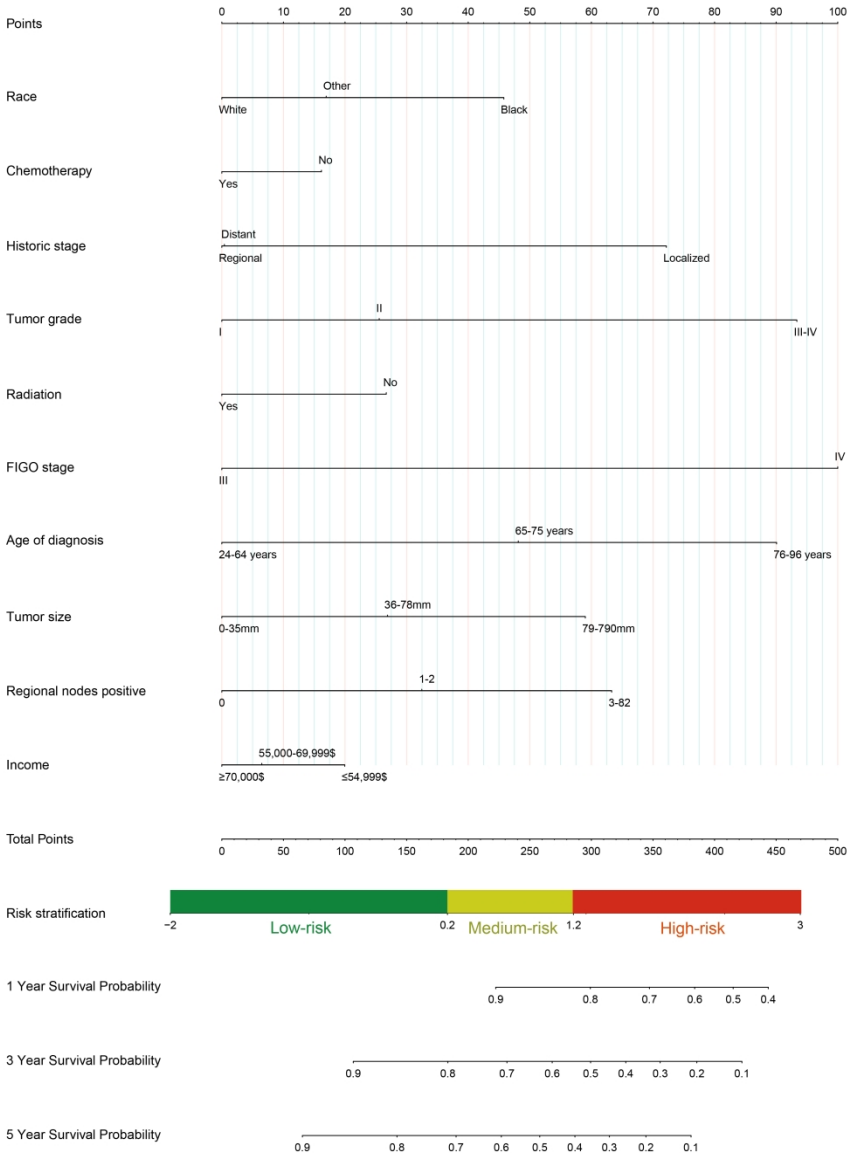


Figure 1 - Nomograms for predicting 1-,3-, and 5-year cancer-specific survival.
* American Indian/AK Native, Asian/Pacific Islander.

288x385mm (300 x 300 DPI)

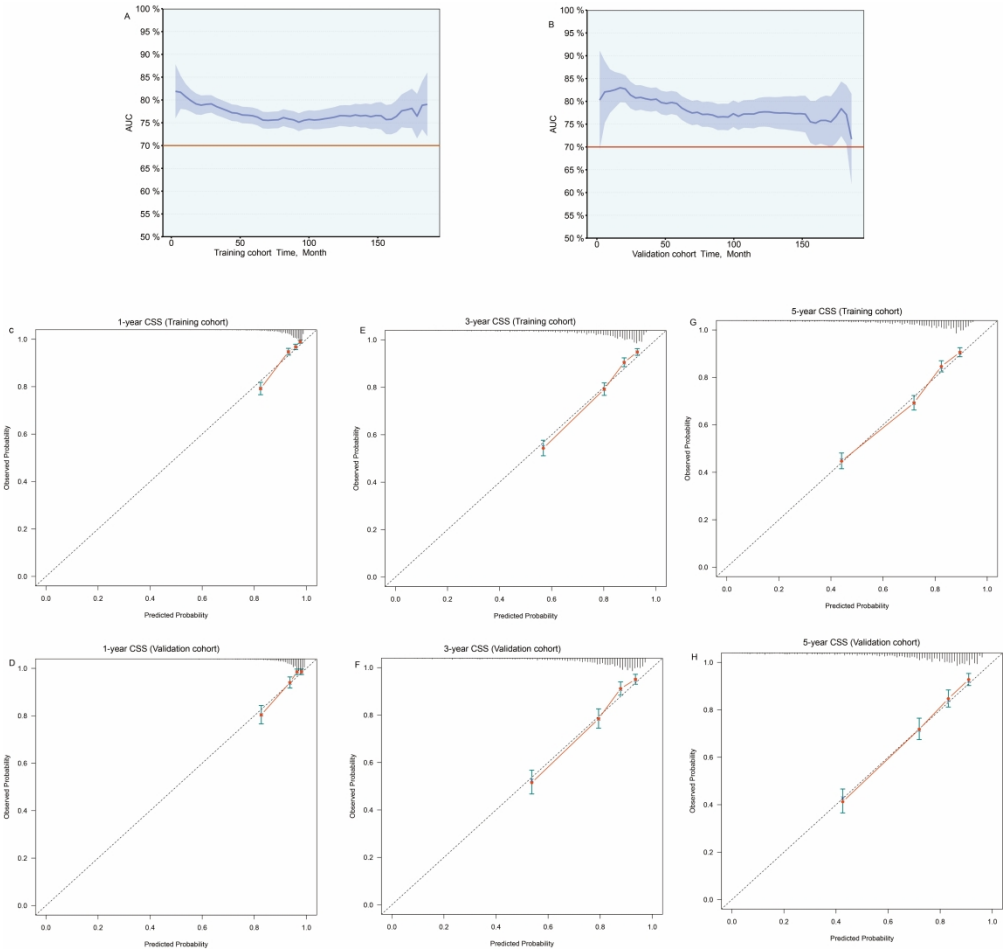


Figure 2 - Time-dependent AUC and calibration curves of the nomogram. (A-B) Time-dependent AUC of using the nomogram to predict cancer-specific survival probability within 5 years in the training and validation cohorts. The red line represents AUC = 0.7, which is considered ideal. (C, E, G) Calibration curves of 1-year, 3-year, and 5-year CSS for advanced EC post-surgery in the training cohort. (D, F, H) Calibration curves of 1-year, 3-year, and 5-year CSS for advanced EC post-surgery in the validation cohort. The black dashed line indicates the ideal reference line where predicted probabilities match the observed survival rates. The red dots represent the performance of the nomogram. The closer the solid red line is to the black dashed line, the more accurately the model predicts survival. AUC: area under the time-dependent receiver operating characteristic curves; CSS: cancer-specific survival. EC: endometrial carcinoma.

508x481mm (300 x 300 DPI)

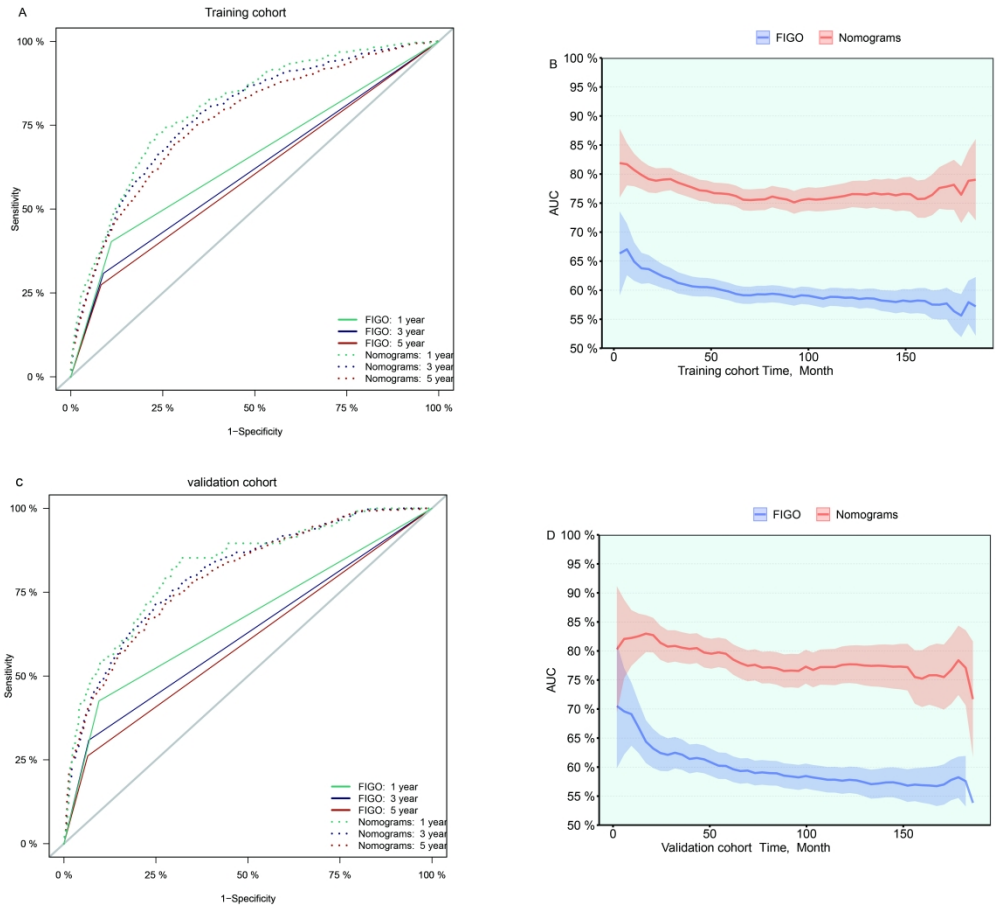
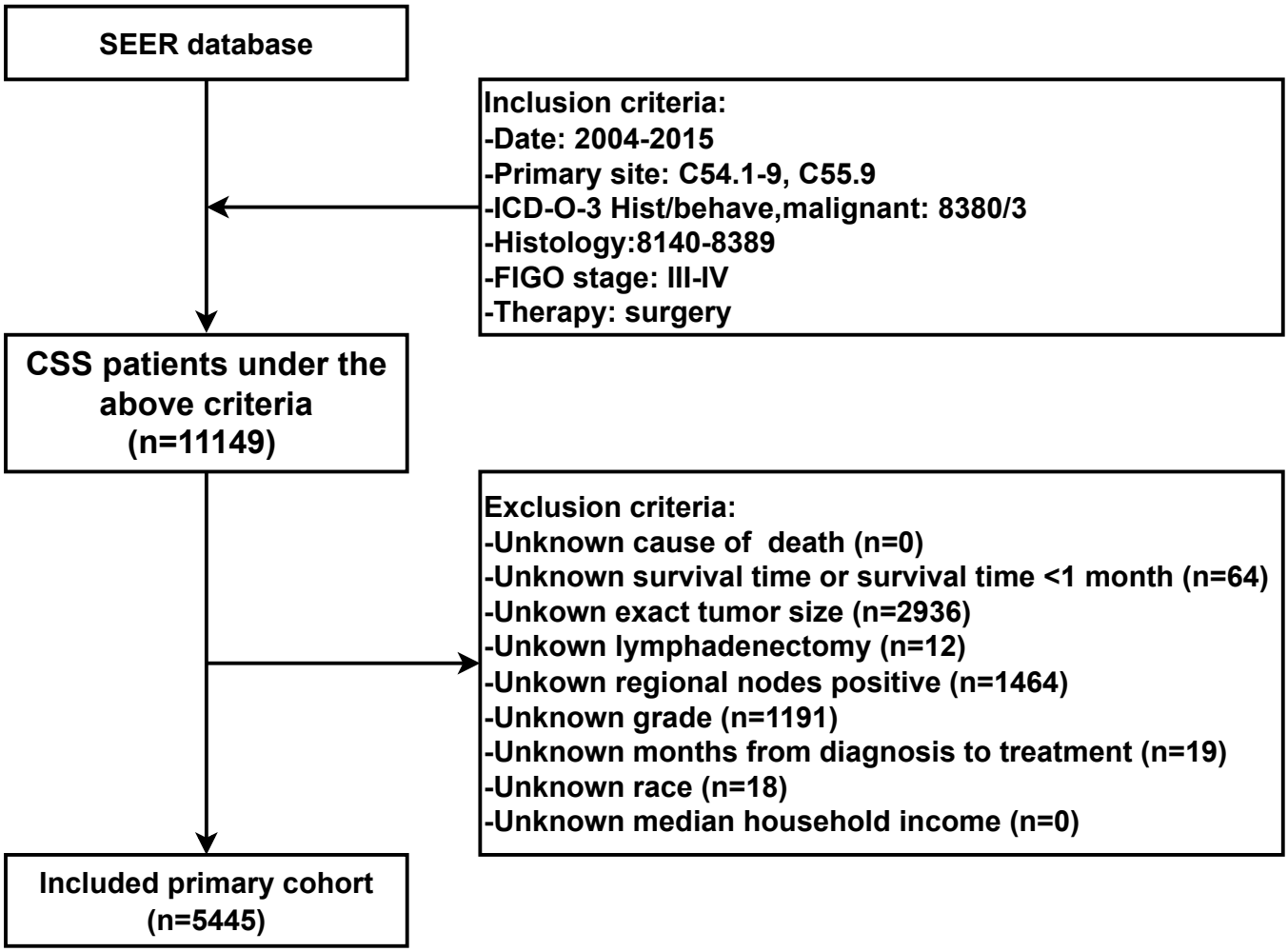
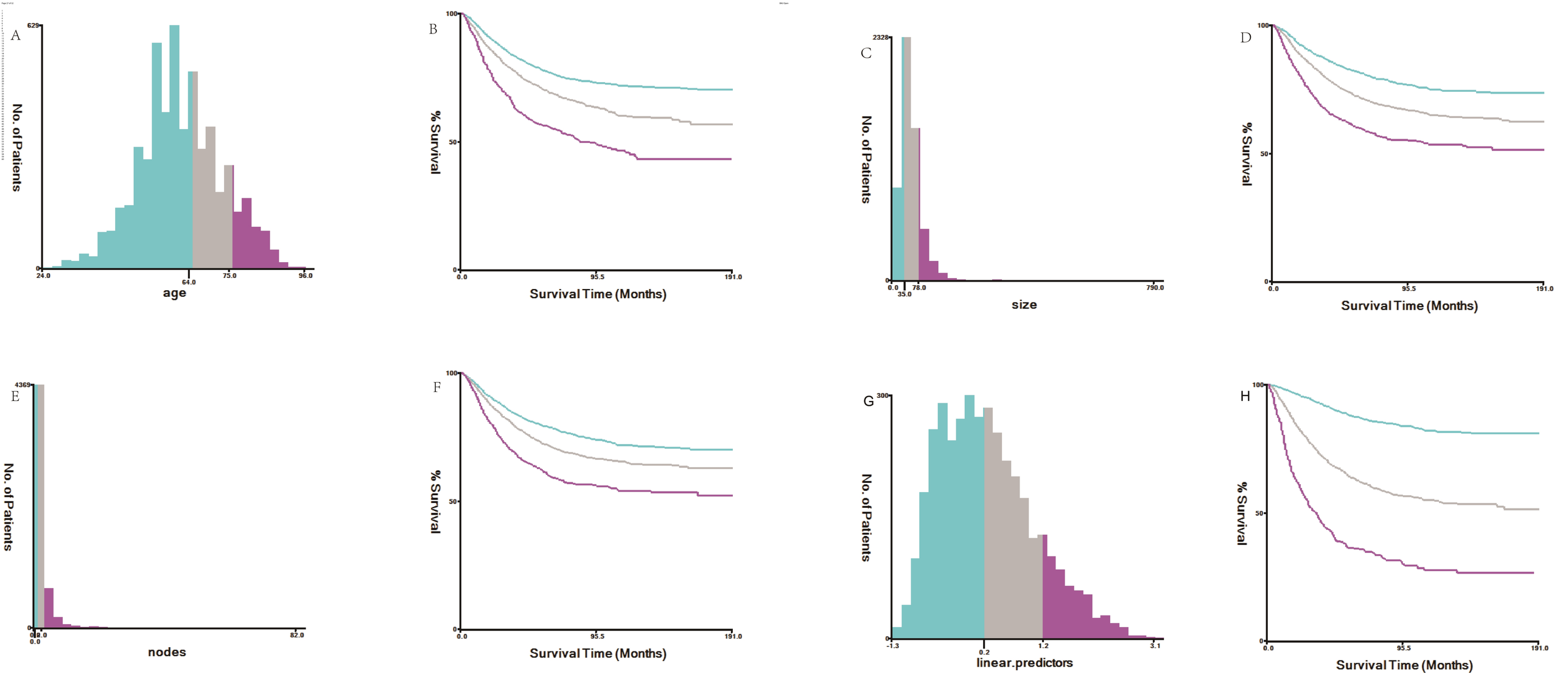


Figure 3 - Compares the accuracy of the nomograms and FIGO stage based on the changes in the ROC curves and the time-dependent AUC.

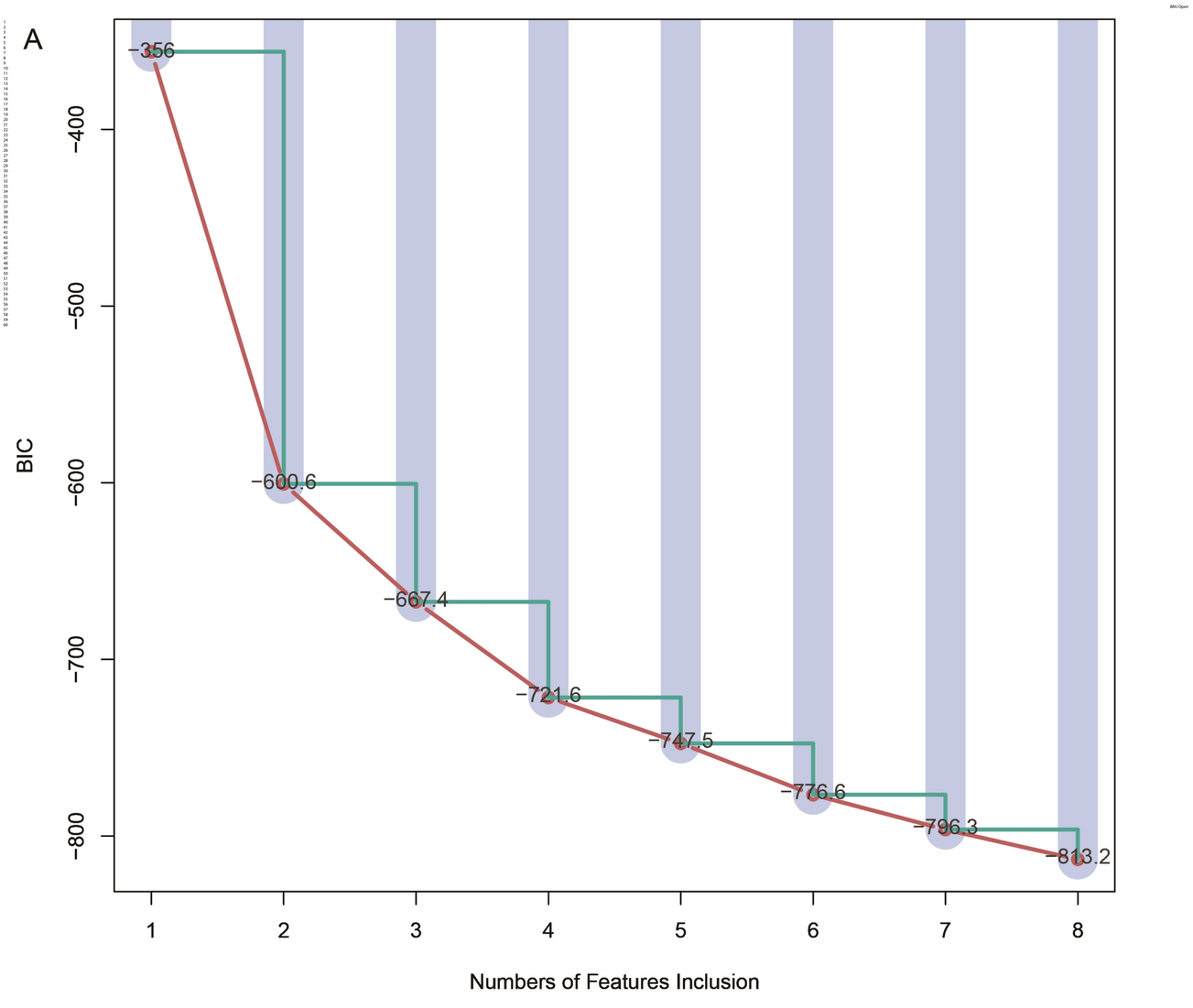
431x388mm (300 x 300 DPI)

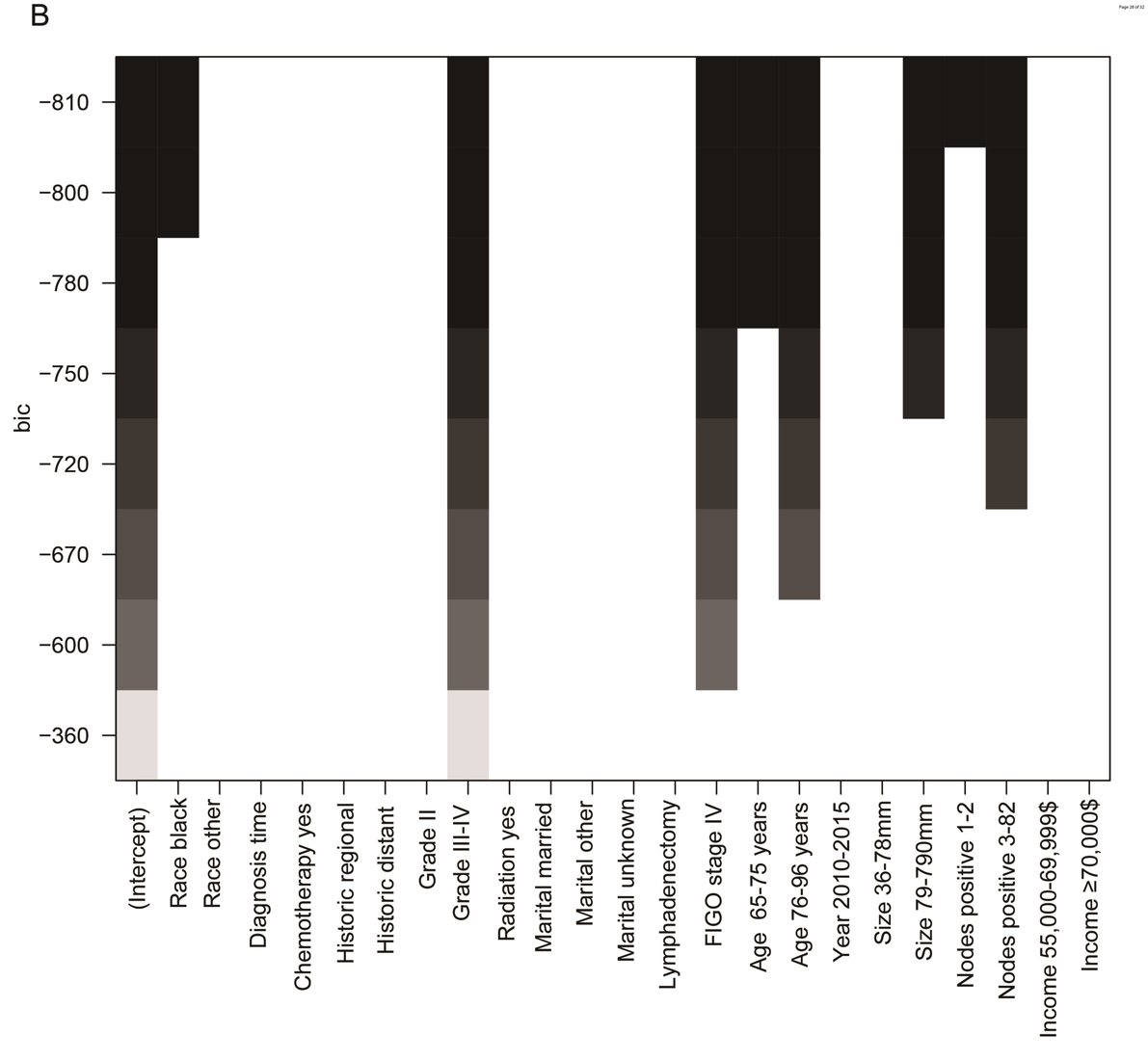


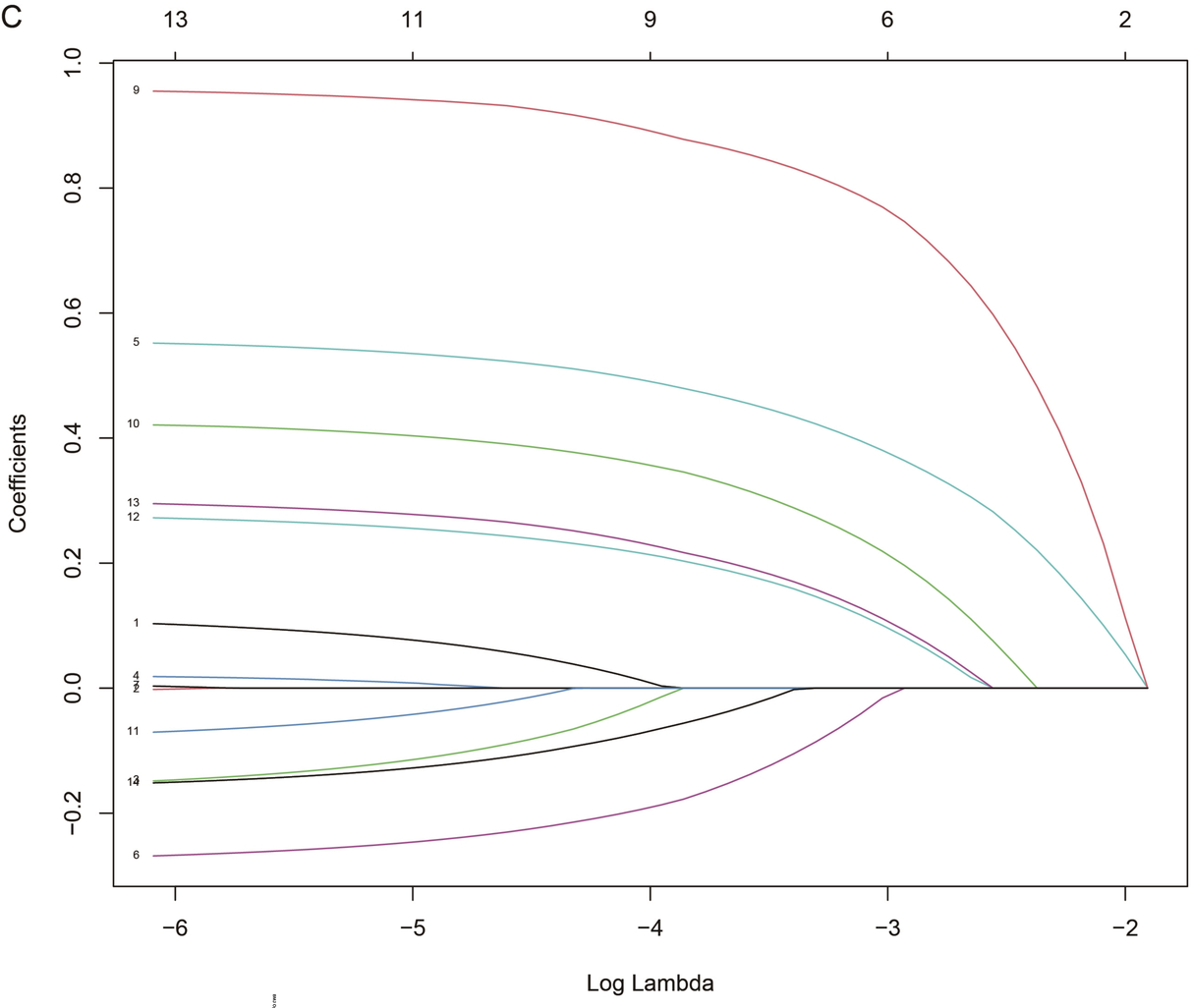
Supplementary Figure 1 Flowchart of the study



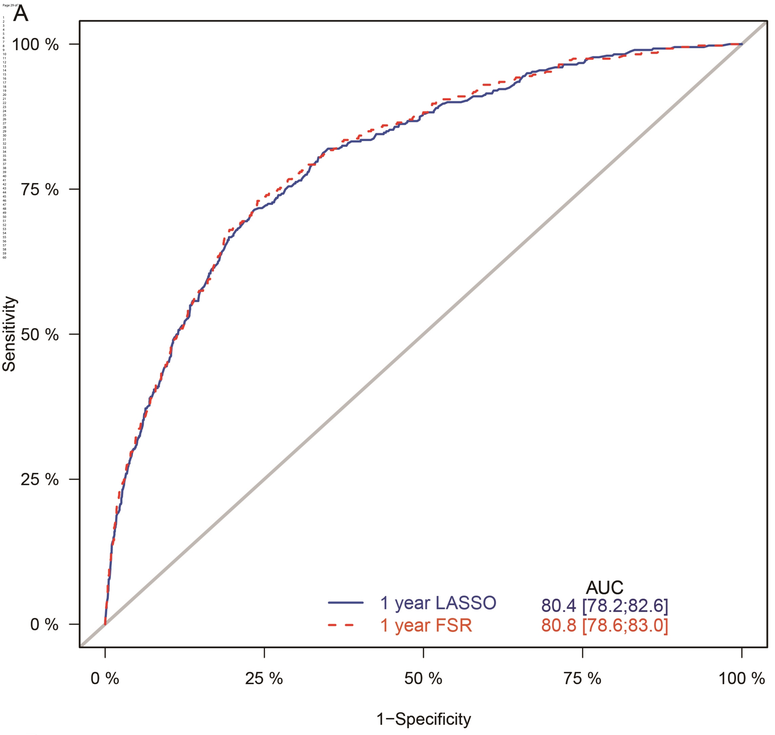
Supplementary Figure 2 Cutoff values for age (A,B), tumor size (C,D), regional nodes positive (E,F), and linear predictors (G,H) that were assessed by X-tile analysis.

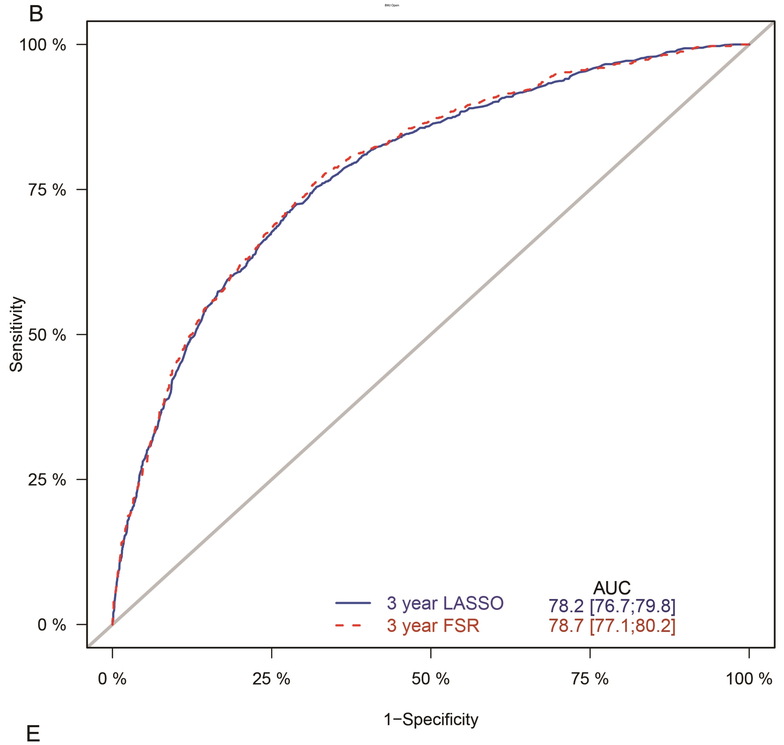


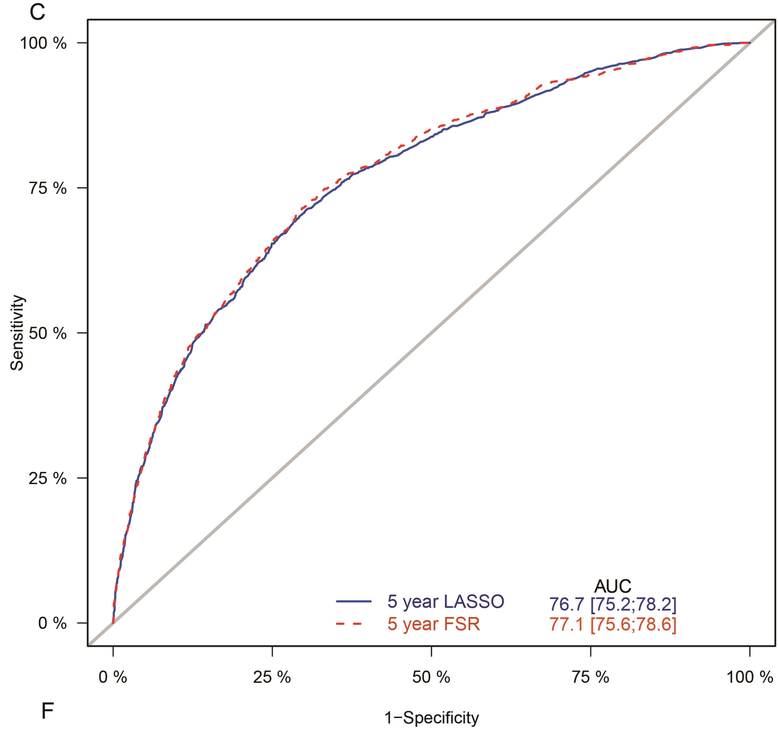


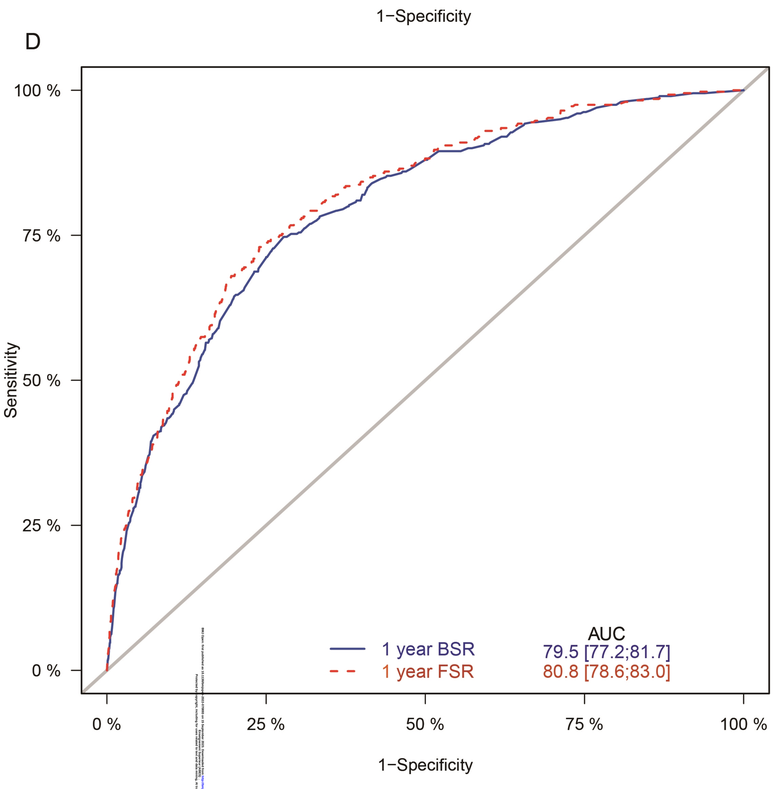


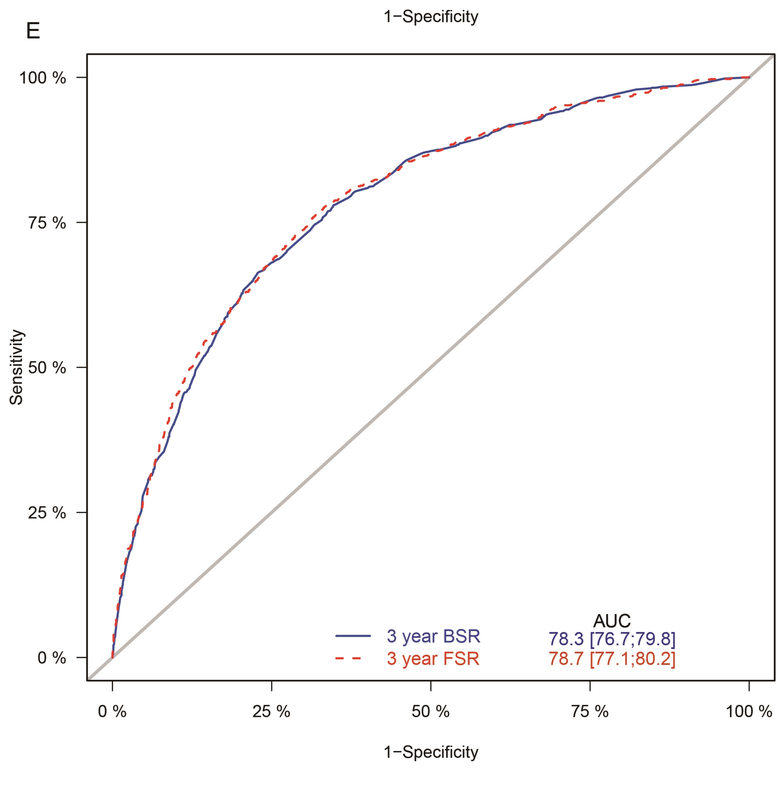
Supplementary Figure 3 Variables selection methods. (A, B) The selection of variables using the BSR method. (C) The LASSO coefficient profile of 14-related variables in primary cohort. (D) 10-fold cross-validation (CV) for tuning parameter (λ) selection.

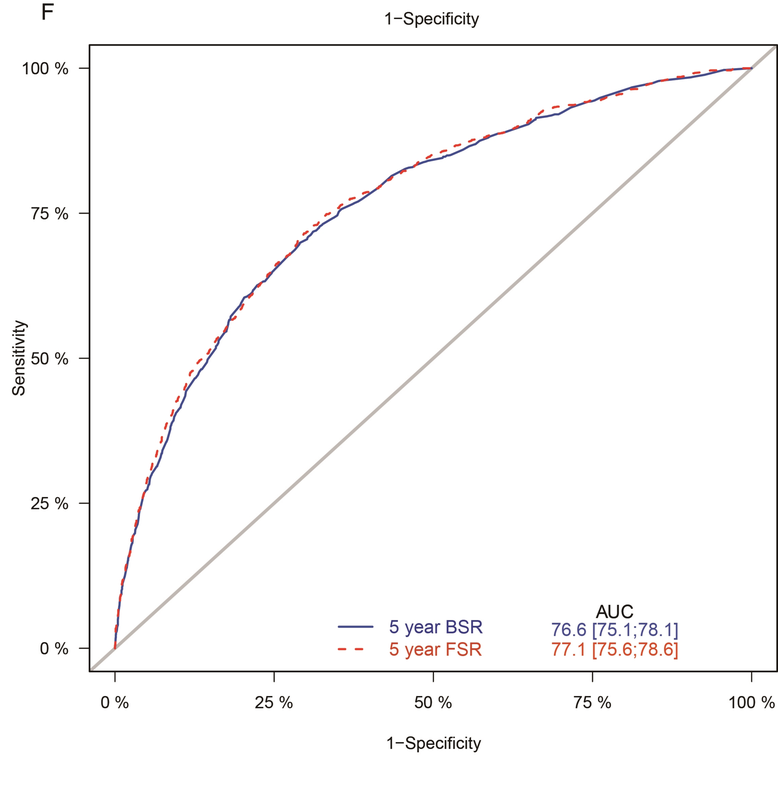




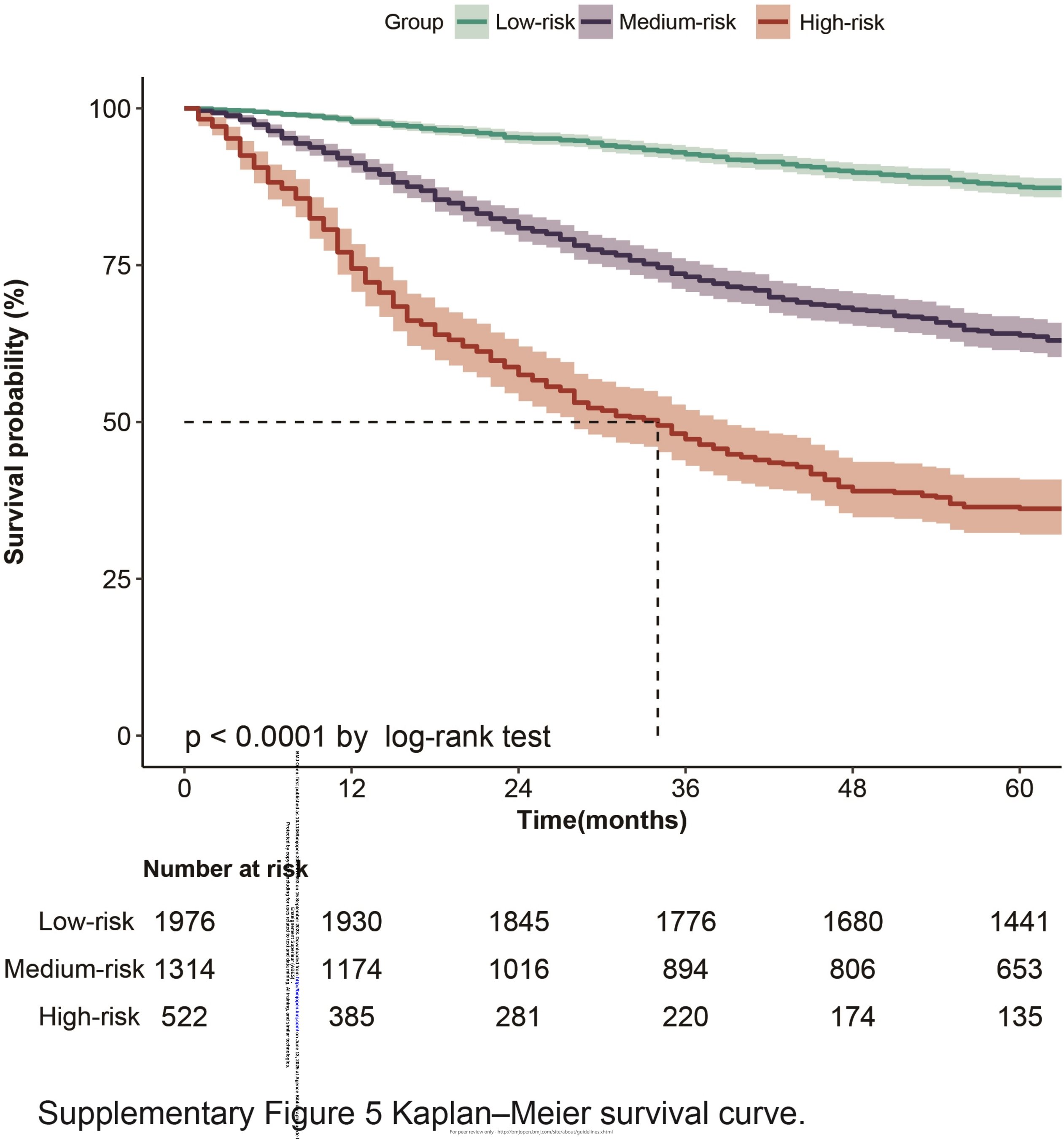








Supplementary Figure 4 Comparison of ROC curves of nomograms created with variables screened by BSR, LASSO, and FSR methods to predict 1-, 3-, and 5-year CSS after surgical treatment of advanced EC. BSR, best subsets regression. LASSO, least absolute shrinkage and selection operator. FSR, forward stepwise regression.



Supplementary Figure 5 Kaplan–Meier survival curve.

Table S1 BSR, LASSO and FSR screening of variables doing predictive models for IDI, NRI comparison.

Index	FSR vs. LASSO			FSR vs. BSR		
	Estimate	95% CI	*P*-value	Estimate	95% CI	*P*-value
IDI						
For 1-year CSS	0.006	0.002-0.012	0.01	0.013	0.006-0.019	<0.001
For 3-year CSS	0.004	0.001-0.008	0.01	0.012	0.006-0.018	<0.001
For 5-year CSS	0.003	0.001-0.007	0.01	0.011	0.006-0.016	<0.001
NRI						
For 1-year CSS	0.119	0.034-0.174	<0.001	0.214	0.113-0.254	<0.001
For 3-year CSS	0.033	(-0.013)-0.112	0.09	0.106	0.044-0.15	<0.001
For 5-year CSS	0.017	(-0.025)-0.099	0.289	0.09	0.053-0.128	<0.001

LASSO, least absolute shrinkage and selection operator. BSR, best subsets regression. FSR, forward stepwise regression. IDI, integrated discrimination improvement. NRI, net reclassification index. CI, confidence interval.

Table S2 IDI, and NRI of the nomogram and the FIGO stage in survival prediction for the advances endometrial carcinoma patients after surgical treatment.

Index	Training cohort			Validation cohort		
	Estimate	95% CI	*P*-value	Estimate	95% CI	*P*-value
IDI (vs. the FIGO stage)						
For 1-year CSS	0.062	0.047-0.084	<0.001	0.071	0.046-0.111	<0.001
For 3-year CSS	0.099	0.084-0.123	<0.001	0.119	0.088-0.155	<0.001
For 5-year CSS	0.112	0.095-0.133	<0.001	0.138	0.103-0.174	<0.001
NRI (vs. the FIGO stage)						
For 1-year CSS	0.364	0.306-0.425	<0.001	0.376	0.293-0.482	<0.001
For 3-year CSS	0.354	0.308-0.395	<0.001	0.352	0.302-0.421	<0.001
For 5-year CSS	0.337	0.292-0.377	<0.001	0.353	0.293-0.419	<0.001

TRIPOD Checklist: Prediction Model Development and Validation

Section/Topic	Item		Checklist Item	Page
Title and abstract				
Title	1	D;V	Identify the study as developing and/or validating a multivariable prediction model, the target population, and the outcome to be predicted.	1
Abstract	2	D;V	Provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusions.	1
Introduction				
Background and objectives	3a	D;V	Explain the medical context (including whether diagnostic or prognostic) and rationale for developing or validating the multivariable prediction model, including references to existing models.	2–3
	3b	D;V	Specify the objectives, including whether the study describes the development or validation of the model or both.	3
Methods				
Source of data	4a	D;V	Describe the study design or source of data (e.g., randomized trial, cohort, or registry data), separately for the development and validation data sets, if applicable.	3
	4b	D;V	Specify the key study dates, including start of accrual; end of accrual; and, if applicable, end of follow-up.	3
Participants	5a	D;V	Specify key elements of the study setting (e.g., primary care, secondary care, general population) including number and location of centres.	3
	5b	D;V	Describe eligibility criteria for participants.	3
	5c	D;V	Give details of treatments received, if relevant.	3–4
Outcome	6a	D;V	Clearly define the outcome that is predicted by the prediction model, including how and when assessed.	
	6b	D;V	Report any actions to blind assessment of the outcome to be predicted.	3–4
Predictors	7a	D;V	Clearly define all predictors used in developing or validating the multivariable prediction model, including how and when they were measured.	3–4
	7b	D;V	Report any actions to blind assessment of predictors for the outcome and other predictors.	3–4
Sample size	8	D;V	Explain how the study size was arrived at.	3–4
Missing data	9	D;V	Describe how missing data were handled (e.g., complete-case analysis, single imputation, multiple imputation) with details of any imputation method.	3–4
Statistical analysis methods	10a	D	Describe how predictors were handled in the analyses.	4
	10b	D	Specify type of model, all model-building procedures (including any predictor selection), and method for internal validation.	4
	10c	V	For validation, describe how the predictions were calculated.	4
	10d	D;V	Specify all measures used to assess model performance and, if relevant, to compare multiple models.	4
	10e	V	Describe any model updating (e.g., recalibration) arising from the validation, if done.	4
Risk groups	11	D;V	Provide details on how risk groups were created, if done.	5
Development vs. validation	12	V	For validation, identify any differences from the development data in setting, eligibility criteria, outcome, and predictors.	4
Results				
Participants	13a	D;V	Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful.	5–6
	13b	D;V	Describe the characteristics of the participants (basic demographics, clinical features, available predictors), including the number of participants with missing data for predictors and outcome.	5
	13c	V	For validation, show a comparison with the development data of the distribution of important variables (demographics, predictors and outcome).	5–6
Model development	14a	D	Specify the number of participants and outcome events in each analysis.	6
	14b	D	If done, report the unadjusted association between each candidate predictor and outcome.	7–9
Model specification	15a	D	Present the full prediction model to allow predictions for individuals (i.e., all regression coefficients, and model intercept or baseline survival at a given time point).	7–9
	15b	D	Explain how to the use the prediction model.	9
Model performance	16	D;V	Report performance measures (with CIs) for the prediction model.	8–9
Model-updating	17	V	If done, report the results from any model updating (i.e., model specification, model performance).	9
Discussion				
Limitations	18	D;V	Discuss any limitations of the study (such as nonrepresentative sample, few events per predictor, missing data).	11–12
Interpretation	19a	V	For validation, discuss the results with reference to performance in the development data, and any other validation data.	10–11
	19b	D;V	Give an overall interpretation of the results, considering objectives, limitations, results from similar studies, and other relevant evidence.	10–11
Implications	20	D;V	Discuss the potential clinical use of the model and implications for future research.	10–

TRIPOD Checklist: Prediction Model Development and Validation

				11
Other information				
Supplementary information	21	D;V	Provide information about the availability of supplementary resources, such as study protocol, Web calculator, and data sets.	12–13
Funding	22	D;V	Give the source of funding and the role of the funders for the present study.	13

*Items relevant only to the development of a prediction model are denoted by D, items relating solely to a validation of a prediction model are denoted by V, and items relating to both are denoted D;V. We recommend using the TRIPOD Checklist in conjunction with the TRIPOD Explanation and Elaboration document.

BMJ Open

Development and validation of a prognostic nomogram for predicting cancer-specific survival in advanced endometrial carcinoma after surgery: a retrospective analysis of the SEER database

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2022-070893.R3
Article Type:	Original research
Date Submitted by the Author:	31-Aug-2023
Complete List of Authors:	Zheng, Chunqin; Shantou Central Hospital, Department of Obstetrics and Gynecology Chen, Weiqiang; Shantou Central Hospital, Department of Anesthesiology Zheng, Zhixiang; Shantou Central Hospital, Department of Obstetrics and Gynecology Liang, Xiaoling; Shantou Central Hospital, Department of Obstetrics and Gynecology Xu, Xiuxia; Shantou Central Hospital, Department of Obstetrics and Gynecology Fang, Danmei; Shantou Central Hospital Department of Obstetrics and Gynecology Ma, Ruijun; Shantou Central Hospital, Department of Obstetrics and Gynecology Fan, Fufang; Shantou Central Hospital Department of Obstetrics and Gynecology Ni, Yanhong; Shantou Central Hospital Department of Obstetrics and Gynecology Zhang, Peili; Shantou Central Hospital Department of Obstetrics and Gynecology Wu, Xuanhua; Shantou Central Hospital, Department of Obstetrics and Gynecology
Primary Subject Heading:	Oncology
Secondary Subject Heading:	Surgery
Keywords:	Gynaecological oncology < GYNAECOLOGY, Surgical pathology < PATHOLOGY, Radiation oncology < RADIOLOGY & IMAGING, Gynaecological oncology < ONCOLOGY, HEALTH ECONOMICS

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Development and validation of a prognostic nomogram for predicting cancer-specific survival in advanced endometrial carcinoma after surgery: a retrospective analysis of the SEER database

Chunqin Zheng[1*], Weiqiang Chen[2], Zhixiang Zheng[1], Xiaoling Liang[1], Xiuxia Xu[1], Danmei Fang[1], Ruijun Ma[1], Fufang Fan[1], Yanhong Ni[1], Peili Zhang[1], Xuanhua Wu[1]

[1]Department of Obstetrics and Gynecology, Shantou Central Hospital, Shantou 515000, China
[2]Department of Anesthesiology, Shantou Central Hospital, Shantou 515000, China.

[*]Correspondence to:
Chunqin Zheng
Department of Obstetrics and Gynecology, Shantou Central Hospital, Shantou 515000, China
Email: 467430374@qq.com
ORCID: 0000-0003-1840-9173

Abstract

Objective We aimed to construct and validate a prognostic nomogram to predict cancer-specific survival (CSS) after surgery in patients with advanced endometrial carcinoma (EC).

Design Retrospective cohort study.

Setting and participants The Surveillance, Epidemiology, and End Results (SEER) database contains cancer incidence and survival data from population-based cancer registries in the USA. A total of 5,445 patients from the SEER database diagnosed with advanced EC between 2004 and 2015 were included and randomized 7:3 into a training cohort (n=3812) and a validation cohort (n=1633).

Outcome measure: CSS.

Results The nomograms for CSS included 10 variables (positive regional nodes, age, tumor size, FIGO stage, grade, ethnicity, income, radiation, chemotherapy, and historical stage) based on the forward stepwise regression results. They revealed discrimination and calibration using the concordance index (C-index) and area under the time-dependent receiver operating characteristic curve (time-dependent AUC), with a C-index value of 0.7324 (95% confidence interval [CI] = 0.7181–0.7468) and 0.7511 (95% CI = 0.7301–0.7722) for the training and validation cohorts, respectively. Using calibration plots, a high degree of conformance was shown between the predicted and observed results. Additionally, a comparison of the nomogram and FIGO staging based on changes in the C-index, net reclassification index, and integrated discrimination

improvement demonstrated that the nomogram had better accuracy and efficacy.

Conclusions We successfully constructed an accurate and effective nomogram to predict CSS in patients with advanced EC, which may help clinicians determine optimal individualized treatment strategies for patients with advanced EC. The predictive performance of the nomogram was evaluated thoroughly, but only internally. Therefore, further validation using different data sources is warranted in future related studies.

Strengths and limitations of this study

- The SEER database is a large database with sufficiently large numbers of samples.
- The SEER database lacks laboratory test data, which may influence the prognoses of patients with advanced EC.
- The chemotherapy and radiotherapy information contained in the SEER database can only be obtained by signing legal agreements that are currently unavailable.
- This study may have suffered from selection bias, as all cases were retrieved from the same database.
- Our nomogram's predictive performance was evaluated thoroughly, but only internally; external validation using different data sources is warranted.

Introduction

Endometrial carcinoma (EC) is the sixth most common cancer in women, with 417,000 new cases diagnosed worldwide in 2020 [1]. There are two histological types of EC [2,3]. Type I tumors include those with grade 1 or 2 endometrioid histological classifications, accounting for approximately 80% of ECs. Type II tumors account for 10–20% of ECs, and include grade 3 endometrioid tumors and tumors with non-endometrioid histology. EC is primarily treated surgically, with radiation and chemotherapy as common adjuvant modalities. For patients with EC who undergo surgery, adjuvant therapy determines disease recurrence for risk stratification based on tumor stage, tumor histology, and other pathologic factors. There is overwhelming evidence that traditional pathological features such as histopathological type, grade, myometrial invasion, and lymphovascular space invasion (LVSI) are imperative for assessing prognosis [4]. Molecular classification in high-grade and/or high-risk ECs shows that POLE-mutated (POLEmut) tumors have an excellent prognosis, p53-abnormal (p53abn) tumors have a poor prognosis, and ECs with mismatch repair deficiency (MMRd) or non-specific molecular profile (NSMP) have an intermediate prognosis [5]. The latest European (ESGO/ESTRO/ESP 2020)/American (NCCN 2020) guidelines combining traditional pathology and The Cancer Genome Atlas (TCGA) molecular groups have proposed a novel risk stratification model: low, intermediate, high–

intermediate, high, and advanced metastasis [6]. Generally, the five-year survival rates are 80–90% and 70–80% for stage I and II ECs, respectively, and 20–60% for stage III and IV ECs [7,8]. Stage III and IV ECs are classified as advanced or high-risk ECs. Patients with advanced and recurrent EC have poor prognoses, with an expected 5-year survival rate of <20% [9]. Due to its high mortality rate, a clinical model for predicting the prognosis of patients with advanced EC is necessary. Although the Federation of Gynaecology and Obstetrics (FIGO) staging system has been widely used to predict the survival of EC patients, this approach still suffers from several limitations [10].

A nomogram is a simple visualization tool used by oncologists to predict and quantify patient survival based on multiple variables. Nomograms have been used for patients with EC [11], and Yang et al. published a nomogram for patients with stage IIIC EC following surgery [12]. However, there is no specific prognostic prediction for patients with advanced EC following surgery.

The traditional statistical strategy for EC- adopted variables was significant only on univariate analysis, which led to model overfitting with generally poor results [13]. Certain advanced statistical methodologies may, however, minimize this limitation. These include the best subset regression (BSR), forward stepwise regression (FSR), and least absolute shrinkage and selection operator (LASSO) approaches [14-16]. In this study, we aimed to establish an effective and noninvasive nomogram to predict cancer-specific survival (CSS) in advanced EC following surgery, incorporating advanced statistical methodologies.

Methods

Data sources and patient selection

The Surveillance, Epidemiology, and End Results (SEER) database contains cancer incidence and survival data from population-based cancer registries in the USA. EC case data with complete follow-up records were selected from the 2004–2015 SEER database (SEER Research Plus Data, 17 Registries, November 2021 Sub [2000–2019]) using SEER*Stat V. 8.4.0.1. The inclusion criteria were as follows: primary sites, C54.1-9 and C55.9 [17]; site and morphology, 8380/3 (based on the International Classification of Tumor Diseases for Oncology [ICD-O], Third Edition); histology, 8140-8389 (adenomas and adenocarcinomas); International Federation of Gynecology and Obstetrics (FIGO) stage, III/IV; and therapy, surgical treatment. The exclusion criteria were as follows: (1) undetermined survival time or survival time < 1 month; (2) undetermined tumor size; (3) undetermined lymphadenectomy; (4) unknown regional node status; (5) unknown tumor grade; (6) unknown months from diagnosis to treatment; (7) unknown ethnicity; and (8) unknown median household income. A flowchart of patient screening is shown in Supplemental Figure 1.

Data on variables, including age at diagnosis, year of diagnosis, tumor size, ethnicity, marital status, histologic stage, tumor grade, FIGO stage, lymphadenectomy, regional node positivity, chemotherapy, radiation status, months from diagnosis to treatment, survival time, median household income, and CSS, were collected from the SEER database. The radiation status (with/without radiation) and chemotherapy status (with/without chemotherapy) were of two categories. Marital status was classified as unmarried (single, unmarried, or living with a domestic partner), married, other (divorced, widowed, or separated), or unknown. Grades were associated with each tumor. ICD-O-2 defines grade I as well-differentiated, grade II as moderately differentiated, grade III as poorly differentiated, and grade IV as undifferentiated. According to the SEER registry, income was examined as aggregate data based on US median income. The median household income is the median household income for the past 12 months, and it was classified into three groups: ≤ 54,999, 55,000–69,999, and ≥70,000. The historical stage was derived from the Collaborative Stage for 2004–2015 and divided into in situ, localized, regional, distant, and unknown categories. In the localized stage, an invasive neoplasm is entirely confined to the organ of origin. In the regional stage, a neoplasm has extended 1) beyond the limits of the organ of origin directly into the surrounding organs or tissues, 2) into the regional lymph nodes via the lymphatic system, or 3) into the regional lymph nodes via a combination of extension and regional lymph nodes. In the distant stage, the neoplasm has spread to parts of the body that are remote from the primary tumor. This study categorized lymphadenectomy into two categories: with and without regional lymph node dissection. Failure to perform lymph node dissection included failure to remove or aspirate local lymph nodes, local lymph node biopsy or aspiration, or sentinel lymph node biopsy. Lymph node dissection includes the removal of an unknown number of regional lymph nodes, the removal of 1–3 regional lymph nodes, the removal of ≥ 4 regional lymph nodes, and regional lymph node dissection with anterior lymph node biopsy.

Statistical analysis

X-tile software (Yale University, New Haven, CT, USA) was used to determine the cutoff values for age at diagnosis, tumor size, positive regional nodes, and risk stratification [18]. Statistical analyses were conducted using R 4.0.2 (R Foundation for Statistical Computing, Vienna, Austria; http://www.r-project.org) in the RStudio environment, as well as with Free Statistics 1.8 (Beijing FreeClinical Medical Technology Co., Ltd.). CSS was the primary endpoint of this study. The patients were randomly assigned to training and validation cohorts at a 7:3 ratio. Categorical variables are presented as frequencies and proportions. Chi-squared tests were used to compare clinicopathological characteristics between the training and validation cohorts. Statistical significance was set at $P < 0.05$. BSR, FSR, and LASSO were used to select the variables.

Significant prognostic factors were identified using the Cox proportional hazards model. A nomogram associated with CSS was constructed and incorporated into the known prognostic factors. The nomogram performance was validated through both training and validation, using the area under the time-dependent receiver operating characteristic curve (time-dependent AUC) to assess its discriminative abilities. Calibration curves were plotted to compare the predicted CSS with the actual CSS after one, three, and five years. The area under the curve (AUC) values ranged from 0.5–1.0, with 0.5 representing random variability and 1.0 representing perfect fit. AUC values g> 0.7 usually indicate rational estimation. The nomogram was compared to the FIGO staging system using the net reclassification index (NRI) and integrated discrimination improvement (IDI). NRI and IDI can be used as alternatives to AUC for assessing the effectiveness of a new risk prediction model and for determining its effectiveness [19,20]. The Kaplan–Meier method was used to compare the risk stratification of the nomogram.

Patient and public involvement

None.

Results

Patient characteristics

A total of 5,445 patients with advanced EC following surgery were screened from the SEER database according to our inclusion and exclusion criteria. They were randomly allocated into training (n = 3,812) and validation cohorts (n = 1,633) at a 7:3 ratio. Patient characteristics are shown in Table 1. No statistically significant differences were found in the indicators between the two groups (all P > 0.05).

Table 1. The basic characteristics of endometrial carcinoma patients in the study

Variables	Primary cohort (n=5445)	Training cohort (n = 3812)	Validation cohort (n = 1633)	*p*-value
Race, n (%)				0.903
White	4444 (81.6)	3107(81.5)	1337 (81.9)	
Black	351 (6.4)	245(6.4)	106 (6.5)	
Other[a]	650 (11.9)	460 (12.1)	190(11.6)	
Chemotherapy, n (%)				0.899
No	2167 (39.8)	1515(39.7)	652(39.9)	
Yes	3278 (60.2)	2297(60.3)	981(60.1)	
Historic stage[b], n (%)				0.62
Localized	9 (0.2)	5 (0.1)	4 (0.2)	
Regional	3903 (71.7)	2731(71.6)	1172 (71.8)	
Distant	1533 (28.2)	1076(28.2)	457 (28)	

Tumor grade[c], n (%)				0.631
I	1226 (22.5)	853 (22.4)	373(22.8)	
II	2166 (39.8)	1506(39.5)	660(40.4)	
III-IV	2053 (37.7)	1453(38.1)	600(36.7)	
Radiation, n (%)				0.055
No	2659 (48.8)	1894(49.7)	765(46.8)	
Yes	2786 (51.2)	1918(50.3)	868(53.2)	
Marital status, n (%)				0.438
Unmarried	1232 (22.6)	881 (23.1)	351(21.5)	
Married	2675 (49.1)	1855(48.7)	820(50.2)	
Other[d]	1375 (25.3)	967 (25.4)	408 (25)	
Unknown	163 (3.0)	109(2.9)	54(3.3)	
Lymphadenectomy[e], n (%)				0.601
No	70 (1.3)	51 (1.3)	19(1.2)	
Yes	5375 (98.7)	3761(98.7)	1614 (98.8)	
FIGO stage, n (%)				0.11
III	4741 (87.1)	3301(86.6)	1440 (88.2)	
IV	704 (12.9)	511 (13.4)	193(11.8)	
Age of diagnosis, n (%)				0.553
24-64 years	3362 (61.7)	2352(61.7)	1010 (61.8)	
65-75 years	1392 (25.6)	965 (25.3)	427(26.1)	
76-96 years	691 (12.7)	495 (13)	196 (12)	
Regional nodes positive, n (%)				0.447
0	2415 (44.4)	1694(44.4)	721(44.2)	
1-2	1954 (35.9)	1381(36.2)	573(35.1)	
3-82	1076 (19.8)	737 (19.3)	339(20.8)	
Year of diagnosis, n (%)				0.981
2004-2009	2152 (39.5)	1507(39.5)	645(39.5)	
2010-2015	3293 (60.5)	2305(60.5)	988(60.5)	
Tumor size[f], n (%)				0.319
0-35mm	1640 (30.1)	1149(30.1)	491(30.1)	
36-78mm	2847 (52.3)	1974(51.8)	873(53.5)	
79-790mm	958 (17.6)	689 (18.1)	269(16.5)	
Income, n (%)				0.701
≤54,999$	1010 (18.5)	707 (18.5)	303(18.6)	
55,000-69,999$	2168 (39.8)	1505(39.5)	663(40.6)	
≥70,000$	2267 (41.6)	1600 (42)	667(40.8)	
Diagnosis time[g], Mean±SD	1.1±1.2	1.1±1.2	1.1±1.1	0.375

a, American Indian/AK Native, Asian/Pacific Islander.
b, historic stage derived from Collaborative Stage (CS) for 2004-2015. Localized, an invasive neoplasm confined entirely to the organ of origin. Regional, a neoplasm that has extended. Distant, a neoplasm that has spread to parts of the body remote from the primary tumor.
c, ICD-O-2 defines grade I as well differentiated, grade II as moderately differentiated, grade III as poorly differentiated, and grade IV as undifferentiated.
d, divorced, widowed, separated.

e, the article categorizes lymphadenectomy into two categories: those involving regional lymph node dissection and those without it. Without lymphadenectomy includes failure to remove or aspirate local lymph nodes, local lymph node biopsy or aspiration, and sentinel lymph node biopsy only.Lymphadenectomy includes removal of an unknown number of regional lymph nodes, removal of one to three regional lymph nodes, removal of four or more regional lymph nodes, and regional lymph node dissection with anterior lymph node biopsy.

f, Based on X-tile procedure cut-offs.

g, Months from diagnosis to treatment.

Nomogram variable screening

Age, tumor size, regional node positivity, and linear predictors (linear predictor = 0.448 * black ethnicity + 0.166 * other ethnicity - 0.158 * chemotherapy - 0.706 * historical stage regional - 0.702 * historical stage distant + 0.25 * grade II + 0.913 * (grade III–IV) - 0.261 * radiation + 0.977 * FIGO stage IV + 0.471 * (age of diagnosis 65–75 years) + 0.881 * (age of diagnosis 76–96 years) + 0.263 * (tumor size 36–78 mm) + 0.577 * (tumor size 79–790 mm) + 0.317 * (regional nodes positive 1–2) + 0.619 * (regional nodes positive 3–82) - 0.132 * (income $55,000–69,999) - 0.195 * (income ≥ $70,000) - 0.271) were divided into three categories using X-tile software. The best cut-off ages were 64 and 75 years (Online Supplemental Figure 2), the best cut-off tumor sizes were 35 mm and 78 mm (Online Supplemental Figure 2), the best cut-off regional node positivities were 0 and 2 (Online Supplemental Figure 2), and the best cut-off linear predictors were 0.2 and 1.2 (Online Supplemental Figure 2).

BSR, LASSO, and FSR were used to select the variables. The BSR method showed great benefits for variable selection because all possible combinations of variables were calculated and the final selected combination was based on the minimum Bayesian information criterion (BIC). As is shown in Online Supplemental Figure 3A/B, six variables (positive regional nodes, age at diagnosis, tumor size, FIGO stage, grade, and ethnicity) were selected from the variables in the training cohort. Considering that the number of independent variables included in the regression equation should be ~10–15 × the number of ending events, we used LASSO to select the variables. As is shown in Online Supplementary Figure 3C/D, seven variables (positive regional nodes, age at diagnosis, tumor size, FIGO stage, grade, radiation, and income) were selected from the variables in the training cohort. Furthermore, the FSR selected ten variables (positive regional nodes, age at diagnosis, tumor size, FIGO stage, grade, ethnicity, chemotherapy, history, radiation, and income) in the training cohort. As a result (Online Supplementary Figure 4), the discrimination of the FSR was highest in the 1-, 3-, and 5-year training cohorts, with a concordance index (C-index) of 0.808 (95% confidence interval [CI]: 0.786–0.83), 0.787 (95% CI: 0.771–0.802), and 0.771 (95% CI: 0.756–0.786), respectively. Moreover, compared to LASSO and BSR (Online

Supplementary Table S1), the 1-, 3-, and 5-year IDIs of FSR were significantly improved (FSR vs. LASSO: 0.006, 0.004, and 0.003, respectively, all P < 0.05; FSR vs. BSR: 0.013, 0.012, 0.011, respectively, all P < 0.05). Therefore, the nomogram obtained from the FSR was optimal (Figure 1). These 10 variables were obtained from the FSR using multivariate Cox analysis due to their optimal performance for predicting CSS in patients with advanced EC following surgery. The results showed that ethnicity, chemotherapy, historical stage, grade, radiation, FIGO stage, age at diagnosis, tumor size, positive regional nodes, and income were independent prognostic factors in this patient group (Table 2). A nomogram for predicting 1-, 3-, and 5-year CSS was constructed based on these 10 key factors (Figure 1).

Table 2. Univariate and multivariable cox regression analysis of cancer-specific survival

Variable	Univariate analysis		Multivariate analysis	
	HR	*P*-value	HR	*P*-value
Race				
White	1(Ref)		1(Ref)	
Black	1.88 (1.6~2.21)	<0.001	1.49 (1.26~1.75)	<0.001
Other[a]	1.03 (0.89~1.2)	0.697	1.15 (0.99~1.34)	0.072
Chemotherapy				
No	1(Ref)		1(Ref)	
Yes	1 (0.9~1.1)	0.958	0.84 (0.75~0.93)	0.001
Historic stage[b]				
Localized	1(Ref)		1(Ref)	
Regional	0.41 (0.17~0.98)	0.044	0.32 (0.13~0.78)	0.012
Distant	0.84 (0.35~2.03)	0.705	0.34 (0.14~0.82)	0.016
Tumor grade[c]				
I	1(Ref)		1(Ref)	
II	1.51 (1.28~1.78)	<0.001	1.43 (1.21~1.68)	<0.001
III-IV	3.63 (3.12~4.23)	<0.001	2.79 (2.39~3.26)	<0.001
Radiation				
No	1(Ref)		1(Ref)	
Yes	0.67 (0.61~0.74)	<0.001	0.76 (0.69~0.84)	<0.001
FIGO stage				
III	1(Ref)		1(Ref)	
IV	3.33 (2.98~3.72)	<0.001	2.6 (2.26~3)	<0.001
Age of diagnosis (year)				
24-64	1(Ref)		1(Ref)	
65-75	1.47 (1.32~1.65)	<0.001	1.52 (1.36~1.7)	<0.001
76-96	2.37 (2.08~2.7)	<0.001	2.38 (2.08~2.73)	<0.001
Tumor size (mm)				
0-35	1(Ref)		1(Ref)	
36-78	1.54 (1.36~1.74)	<0.001	1.25 (1.1~1.41)	<0.001
79-790	2.38 (2.07~2.74)	<0.001	1.72 (1.48~2)	<0.001
Regional nodes positive				
Negative	1(Ref)		1(Ref)	
1-2	1.34 (1.2~1.5)	<0.001	1.36 (1.21~1.53)	<0.001
3-82	1.98 (1.76~2.24)	<0.001	1.86 (1.62~2.14)	<0.001
Income				

≤54,999$	1(Ref)		1(Ref)	
55,000-69,999$	0.82 (0.72~0.94)	0.003	0.82 (0.72~0.94)	0.003
≥70,000$	0.72 (0.63~0.82)	<0.001	0.75 (0.65~0.85)	<0.001

a, American Indian/AK Native, Asian/Pacific Islander.

b, historic stage derived from Collaborative Stage (CS) for 2004-2015. Localized, an invasive neoplasm confined entirely to the organ of origin. Regional, a neoplasm that has extended. Distant, a neoplasm that has spread to parts of the body remote from the primary tumor.

c, I, well differentiated; II, moderately differentiated; III, poorly differentiated; IV, undifferentiated.

Nomogram construction and performance

As shown in Figure 1, we developed a nomogram based on FSR to predict one-, three-, and five-year CSS rates. According to the training and validation cohort data, the C-index values were 0.7324 (95% CI = 0.7181–0.7468) and 0.7511 (95% CI = 0.7301–0.7722), respectively. According to Figure 2A/B, the AUC for the prediction of CSS within five years was > 0.7 in both the training and validation cohorts, indicating favorable discrimination. Figure 2C/E/G shows the calibration curves of the 1-, 3- and 5-year CSS for advanced EC following surgery in the training cohort. Figure 2D/F/H shows the calibration curves of the 1-, 3-, and 5-year CSS for advanced EC following surgery in the validation cohort. The dashed black line indicates the ideal reference line, where the predicted probabilities matched the observed survival rates. The red dots represent the performance of the nomogram. The closer the solid red line is to the black dashed line, the more accurately the model predicted survival. As is shown in Figure 2C–H, the calibration curves of the nomogram showed high concordance between the predicted and observed survival probabilities.

Comparative clinical value of the nomogram and FIGO stage

The accuracies of the nomogram and FIGO stage were compared based on changes in the ROC curves and time-dependent AUCs (Figure 3). Compared to the FIGO stage (Online Supplementary Table S2), the 1-, 3-, and 5-year IDI of the nomogram was significantly greater (nomogram vs. FIGO stage: 0.062, 0.099, and 0.112, respectively). Moreover, compared to the FIGO stage (Online Supplementary Table S2), the 1-, 3-, and 5-year NRI of the nomogram was significantly greater (nomogram vs. FIGO stage: 0.364, 0.354, and 0.337, respectively). According to these results, the nomogram predicted the prognosis more accurately than the FIGO stage.

Assessment of the risk of advanced EC following surgery

In addition to the nomogram, we developed a risk stratification system based on the linear predictor cut-off value for each patient in the training cohort. The patients were divided into three groups according to their linear predictors: low risk (≤ 0.2), intermediate risk (0.21–1.2), and high risk (>

1.2). There was a significant difference in CSS between the low-, medium-, and high-risk groups according to our Kaplan–Meier analysis (all $P < 0.05$, Online Supplementary Figure 5). Furthermore, according to the nomogram, a total score of ≤ 185 indicated low risk, 185 ≤ 285 indicated medium risk, and > 285 indicated high risk. These results show that the nomogram had excellent risk-stratification capabilities.

Discussion

In this study, we used actual information from patients with advanced EC following surgery. We also developed a prognostic nomogram and risk stratification system using data from the SEER database. The nomogram produced excellent internal and external results, as shown by calibration, C-index, and ROC curves.

Few studies have focused on predicting postoperative CSS in patients with advanced EC. This study focused on postoperative CSS in patients with stage III–IV cancer for two key reasons. First, advanced EC has high prognostic heterogeneity and a poor survival rate, with a five-year survival rate of 20–60% (although different patients have different prognoses). Due to the lack of a reliable model to predict survival in patients with advanced EC following surgery, individualized clinical management and surveillance can be challenging. Second, patients with advanced EC have significantly higher incidence and mortality rates following surgery, leading to confounding bias in prognostic indicators.

EC is usually treated surgically, and postoperative treatment depends on risk factors such as age, tumor stage, myometrial infiltration depth, and histologic grade [21,22]. In this study, a prognostic model after the surgical treatment of advanced EC was constructed based on 10 variables (ethnicity, chemotherapy, historical stage, tumor grade, radiation therapy, FIGO stage, age at diagnosis, tumor size, positive regional nodes, and median household income) screened using FSR. The scores were calculated for each item based on the subtype of each independent prognostic factor. The total score was calculated using scores corresponding to the independent prognostic factors. Each subgroup variable was assigned a score from 0–100 according to its contribution. All enrolled variables were added to generate a total score on the bottom scale, which was then converted to predict CSS. CSS at 1-, 3-, and 5 years was determined by drawing a vertical line on the total score scale, with higher scores indicating a worse prognosis. According to the nomogram, the FIGO stage plays the largest role in prognosis, followed by tumor grade and age at diagnosis.

Cancer grade, histological subtype, tumor size, LVSI, lymph node status, and cervical involvement are vital prognostic factors in patients with EC [23]. In this study, tumor grade, tumor size, and lymph node status were important prognostic factors following surgical treatment for

advanced EC. Tumor grade has also been shown to be a prognostic factor in EC [24], and our nomogram indicates that poorly differentiated or undifferentiated tumors have poor prognoses. Conflicting results have been reported concerning the impact of tumor size on survival outcomes. Preoperative ultrasound tumor size was apparently not a prognostic factor for death from any cause in women with EC [25]. However, tumor size was an independent prognostic factor for recurrence alone [26, 27] and for recurrence and death due to EC [28]. Lymph node metastasis further contributes to poor prognosis in patients with EC; however, there is no consensus on the value and extent of lymph node dissection [29]. In this study, we found that positive lymph nodes could affect the prognosis of surgical treatment for advanced EC, consistent with the findings of previous studies. However, this study did not reflect whether lymph node dissection was beneficial. This may be related to the fact that the population selected in this study underwent lymph node dissection (98.7%), which was not comparable. Compared to women ≥ 65 years, women < 65 years had a significant survival advantage, as indicated by previous studies [30].

Using advanced EC after surgery as a dataset, this study examined factors that could be included in prognostic nomograms. Nomograms combine multiple factors, including demographic and clinicopathological characteristics, into quantitative models that provide better predictions than FIGO staging [31,32]. FIGO staging has traditionally been used to predict the prognosis of women with EC. Staging using this system is closely associated with CSS. However, patients at the same stage have different prognoses. FIGO staging does not consider factors such as age, radiation status, and income, thus resulting in its prognostic heterogeneity. Therefore, we compared nomograms that included more variables. Nomograms generally have better predictive powers than FIGO staging alone due to their positive NRI and IDI scores.

Based on their total nomogram scores, the patients were classified into low-, intermediate-, and high-risk groups. Significant differences were found in CSS among the three risk groups based on Kaplan–Meier analysis (Online Supplementary Figure 5). This nomogram is highly effective in identifying high-risk groups owing to its poor prognosis. Patients with a total score greater than 285 should receive special attention.

To investigate the potential utility of the nomogram in clinical practice, we analyzed data from the SEER database by using a large sample of data representing different population regions. We followed the recommendations of the Transparent Reporting of Individual Prognosis or Diagnosis Multivariate Predictive Model statement [33]. Bootstrapping and cross-validation methods were used to calculate the calibration curves, time-dependent AUCs, and C-index. These positive results show that our nomogram may be useful for assessing the prognosis of patients with advanced EC after surgery.

Although the nomogram performed well, this study had some key limitations. Carbohydrate

antigen 125 (CA125) is a tumor marker whose levels are often elevated in patients with malignant tumors such as ovarian epithelial, fallopian tube, and EC, as well as in those with lung and gastrointestinal adenocarcinomas. In the clinical diagnosis and treatment of EC, CA125 levels are often used to monitor disease changes, evaluate treatment effects, and predict prognosis [34]. Studies have shown that CA125 is an important variable in the prognostic prediction model of EC and can significantly improve its accuracy [35]. Human epididymis protein 4 (HE4) is an acidic whey protein first identified in the epithelium of the distal epididymis [36]. It is expressed in the epithelia of several tissues, including the female reproductive tract, and is overexpressed in several cancers [37]. HE4 is strongly associated with survival in patients with EC [38]. ECs have traditionally been classified into two subtypes (1 and 2) based on their histopathological characteristics [2]. However, this classification system lacks reproducibility and yields heterogeneous molecular groups that hamper the advancement and implementation of precision medicine [39, 40]. It is, therefore, being gradually replaced by a clearly defined system based on molecular phenotypes [41]. The TCGA approach results in the molecular stratification of ECs into four distinct molecular groups: DNA polymerase epsilon ultra-mutated classification, which portends a good prognosis; microsatellite instability hypermutated (intermediate prognosis); copy number-low; and copy number-high (which includes p53 mutations and carries the worst prognosis) [41], ESMO 2022 recommends that molecular staging testing should be performed for all ECs, but POLE testing can be omitted for low-risk patients when conditions are limited. However, MMR and p53 testing should still be performed to identify patients with hereditary EC or high-risk factors [42]. LVSI has a prognostic value in patients with EC independent of TCGA signature, age, and adjuvant treatment, increasing the risk of death from any cause [43]. Since data on CA125, HE4, molecular typing, LVSI, hormonal therapy, or immunotherapy was not published in SEER 2004-2015, these variables were not assessed in this study. In addition, the chemotherapy and radiotherapy information in the SEER database can only be obtained by signing certain legal agreements that appeared unavailable at the time. As a result, we were unable to study the relationship between chemotherapy, radiotherapy, targeted therapy, and EC prognosis. Moreover, the study cases derived from the US SEER database were nonrepresentative of regions outside the USA. Finally, although the predictive performance of the nomogram was evaluated thoroughly using internal data, validation using different external data sources is warranted, and further investigation is recommended.

Conclusions

Our nomogram is more accurate, has better clinical utility, and provides better prognostic predictions than FIGO staging for patients with advanced EC after surgery. However, the

predictive performance of the nomogram was evaluated using internal data only. Therefore, using different data sources for external validation is warranted, and further investigation is recommended.

Acknowledgements

We thank the SEER project for providing access to their database and the Free Statistics software version 1.8, provided by Beijing FreeClinical Medical Technology Co., Ltd. We gratefully thank Dr. Jie Liu of the Department of Vascular and Endovascular Surgery, Chinese PLA General Hospital, for his contribution to statistical support, study design consultations, and comments regarding the manuscript.

Contributors

CZ: Study conception, data collection, data analysis, interpretation, drafting, critical revision, and final approval of the article. Data collection by RM, XW, and DF. YN, FF, and PZ Data Analysis. ZZ and XL: study conception. WC and XX: critical revision and final approval of the article.

Funding

The authors declare no specific grant from any funding agency in the public, commercial, or not-for-profit sectors for this research.

Competing interests

None declared.

Patient consent for publication

Not applicable.

Ethics approval

This study used data from the SEER database, approved by the National Institutes of Health Ethics Program. Access request has been approved for the SEER Research Database. Ethics approval was not required for the present analysis.

Provenance and peer review

Not commissioned, externally peer reviewed.

Data availability statement

The data source, SEER, is accessible via the SEER website (https://www. cancer. gov/policies/accessibility). The SEER database is publicly available. Data supporting the findings of this study are available from the corresponding author upon request.

Abbreviations

EC: endometrial cancer. SEER: Linked Surveillance, Epidemiology, and End Results. CSS: Cancer-specific survival. CI: confidence interval. FIGO: International Federation of Gynecology and Obstetrics. BSR: Best subset regression. FSR: forward stepwise regression. LASSO: least absolute shrinkage and selection operator. AUC: area under the receiver operating characteristic curve. C-index: concordance index. NRI: net reclassification index. IDI: integrated discrimination improvement. LVSI: lymphovascular space invasion.

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F: Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* 2021, 71(3):209-249.
2. Bokhman JV: Two pathogenetic types of endometrial carcinoma. *Gynecol Oncol* 1983, 15(1):10-17.
3. Felix AS, Weissfeld JL, Stone RA, Bowser R, Chivukula M, Edwards RP, Linkov F: Factors associated with Type I and Type II endometrial cancer. *Cancer Causes Control* 2010, 21(11):1851-1856.
4. Cho KR, Cooper K, Croce S, Djordevic B, Herrington S, Howitt B, Hui P, Ip P, Koebel M, Lax S *et al*: International Society of Gynecological Pathologists (ISGyP) Endometrial Cancer Project: Guidelines From the Special Techniques and Ancillary Studies Group. *Int J Gynecol Pathol* 2019, 38 Suppl 1(Iss 1 Suppl 1):S114-s122.
5. McAlpine J, Leon-Castillo A, Bosse T: The rise of a novel classification system for endometrial carcinoma; integration of molecular subclasses. *J Pathol* 2018, 244(5):538-549.
6. Concin N, Matias-Guiu X, Vergote I, Cibula D, Mirza MR, Marnitz S, Ledermann J, Bosse T, Chargari C, Fagotti A *et al*: ESGO/ESTRO/ESP guidelines for the management of patients with endometrial carcinoma. *Int J Gynecol Cancer* 2021, 31(1):12-39.
7. Lewin SN, Herzog TJ, Barrena Medel NI, Deutsch I, Burke WM, Sun X, Wright JD: Comparative performance of the 2009 international Federation of gynecology and obstetrics' staging system for uterine corpus cancer. *Obstet Gynecol* 2010,

116(5):1141-1149.

8. Creasman WT, Odicino F, Maisonneuve P, Quinn MA, Beller U, Benedet JL, Heintz AP, Ngan HY, Pecorelli S: Carcinoma of the corpus uteri. FIGO 26th Annual Report on the Results of Treatment in Gynecological Cancer. *Int J Gynaecol Obstet* 2006, 95 Suppl 1:S105-143.

9. Costa B, Vale N: Dostarlimab: A Review. *Biomolecules* 2022, 12(8).

10. Kasius JC, Pijnenborg JMA, Lindemann K, Forsse D, van Zwol J, Kristensen GB, Krakstad C, Werner HMJ, Amant F: Risk Stratification of Endometrial Cancer Patients: FIGO Stage, Biomarkers and Molecular Classification. *Cancers (Basel)* 2021, 13(22).

11. Zhu L, Sun X, Bai W: Nomograms for Predicting Cancer-Specific and Overall Survival Among Patients With Endometrial Carcinoma: A SEER Based Study. *Front Oncol* 2020, 10:269.

12. Yang XL, Huang H, Kou LN, Lai H, Chen XP, Wu DJ: Construction and validation of a prognostic model for stage IIIC endometrial cancer patients after surgery. *Eur J Surg Oncol* 2022, 48(5):1173-1180.

13. Harrell FE, Jr., Lee KL, Matchar DB, Reichert TA: Regression models for prognostic prediction: advantages, problems, and suggested solutions. *Cancer Treat Rep* 1985, 69(10):1071-1077.

14. Collins GS, Reitsma JB, Altman DG, Moons KG: Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. The TRIPOD Group. *Circulation* 2015, 131(2):211-219.

15. Debray TP, Damen JA, Snell KI, Ensor J, Hooft L, Reitsma JB, Riley RD, Moons KG:

A guide to systematic review and meta-analysis of prediction model performance. *Bmj* 2017, 356:i6460.

16. Walter S, Tiemeier H: Variable selection: current practice in epidemiological studies. *Eur J Epidemiol* 2009, 24(12):733-736.

17. Modh A, Ghanem AI, Burmeister C, Rasool N, Elshaikh MA: Trends in the utilization of adjuvant vaginal brachytherapy in women with early-stage endometrial carcinoma: Results of an updated period analysis of SEER data. *Brachytherapy* 2016, 15(5):554-561.

18. Camp RL, Dolled-Filhart M, Rimm DL: X-tile: a new bio-informatics tool for biomarker assessment and outcome-based cut-point optimization. *Clin Cancer Res* 2004, 10(21):7252-7259.

19. Uno H, Tian L, Cai T, Kohane IS, Wei LJ: A unified inference procedure for a class of measures to assess improvement in risk prediction systems with survival data. *Stat Med* 2013, 32(14):2430-2442.

20. Pencina MJ, D'Agostino RB, Sr., Steyerberg EW: Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. *Stat Med* 2011, 30(1):11-21.

21. Creutzberg CL, van Putten WL, Koper PC, Lybeert ML, Jobsen JJ, Wárlám-Rodenhuis CC, De Winter KA, Lutgens LC, van den Bergh AC, van de Steen-Banasik E *et al*: Surgery and postoperative radiotherapy versus surgery alone for patients with stage-1 endometrial carcinoma: multicentre randomised trial. PORTEC Study Group. Post Operative Radiation Therapy in Endometrial Carcinoma. *Lancet* 2000,

355(9213):1404-1411.

22. Keys HM, Roberts JA, Brunetto VL, Zaino RJ, Spirtos NM, Bloss JD, Pearlman A, Maiman MA, Bell JG: A phase III trial of surgery with or without adjunctive external pelvic radiation therapy in intermediate risk endometrial adenocarcinoma: a Gynecologic Oncology Group study. *Gynecol Oncol* 2004, 92(3):744-751.

23. Colombo N, Creutzberg C, Amant F, Bosse T, González-Martín A, Ledermann J, Marth C, Nout R, Querleu D, Mirza MR *et al*: ESMO-ESGO-ESTRO Consensus Conference on Endometrial Cancer: diagnosis, treatment and follow-up. *Ann Oncol* 2016, 27(1):16-41.

24. Stålberg K, Kjølhede P, Bjurberg M, Borgfeldt C, Dahm-Kähler P, Falconer H, Holmberg E, Staf C, Tholander B, Åvall-Lundqvist E *et al*: Risk factors for lymph node metastases in women with endometrial cancer: A population-based, nation-wide register study-On behalf of the Swedish Gynecological Cancer Group. *Int J Cancer* 2017, 140(12):2693-2700.

25. Ambrosio M, Raffone A, Alletto A, Cini C, Filipponi F, Neola D, Fabbri M, Arena A, Raimondo D, Salucci P *et al*: Is preoperative ultrasound tumor size a prognostic factor in endometrial carcinoma patients? *Front Oncol* 2022, 12:993629.

26. Senol T, Polat M, Ozkaya E, Karateke A: Tumor Diameter for Prediction of Recurrence, Disease Free and Overall Survival in Endometrial Cancer Cases. *Asian Pac J Cancer Prev* 2015, 16(17):7463-7466.

27. Sozzi G, Uccella S, Berretta R, Petrillo M, Fanfani F, Monterossi G, Ghizzoni V, Frusca T, Ghezzi F, Chiantera V *et al*: Tumor Size, an Additional Risk Factor of Local

Recurrence in Low-Risk Endometrial Cancer: A Large Multicentric Retrospective Study. *Int J Gynecol Cancer* 2018, 28(4):684-691.

28. Chattopadhyay S, Cross P, Nayar A, Galaal K, Naik R: Tumor size: a better independent predictor of distant failure and death than depth of myometrial invasion in International Federation of Gynecology and Obstetrics stage I endometrioid endometrial cancer. *Int J Gynecol Cancer* 2013, 23(4):690-697.

29. Sari ME, Yalcin İ, Sahin H, Meydanli MM, Gungor T: Risk factors for paraaortic lymph node metastasis in endometrial cancer. *Int J Clin Oncol* 2017, 22(5):937-944.

30. Chan JK, Cheung MK, Huh WK, Osann K, Husain A, Teng NN, Kapp DS: Therapeutic role of lymph node resection in endometrioid corpus cancer: a study of 12,333 patients. *Cancer* 2006, 107(8):1823-1830.

31. Wu J, Zhang H, Li L, Hu M, Chen L, Xu B, Song Q: A nomogram for predicting overall survival in patients with low-grade endometrial stromal sarcoma: A population-based analysis. *Cancer Commun (Lond)* 2020, 40(7):301-312.

32. Gao J, Wang S, Li F, Xu H, Li X, Yan L, Zhu L: Nomograms for Predicting Overall Survival and Cancer-Specific Survival of Young Patients with Epithelial Ovarian Cancer: Analysis Based on SEER Program. *Adv Ther* 2022, 39(1):257-285.

33. Collins GS, Reitsma JB, Altman DG, Moons KG: Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *Bmj* 2015, 350:g7594.

34. Jiang T, Huang L, Zhang S: Preoperative serum CA125: a useful marker for surgical management of endometrial cancer. *BMC Cancer* 2015, 15:396.

35. Li Q, Kong F, Ma J, Wang Y, Wang C, Yang H, Li Y, Ma X: Nomograms Based on Fibrinogen, Albumin, Neutrophil-Lymphocyte Ratio, and Carbohydrate Antigen 125 for Predicting Endometrial Cancer Prognosis. *Cancers (Basel)* 2022, 14(22).

36. Kirchhoff C, Habben I, Ivell R, Krull N: A major human epididymis-specific cDNA encodes a protein with sequence homology to extracellular proteinase inhibitors. *Biol Reprod* 1991, 45(2):350-357.

37. Galgano MT, Hampton GM, Frierson HF, Jr.: Comprehensive analysis of HE4 expression in normal and malignant human tissues. *Mod Pathol* 2006, 19(6):847-853.

38. Behrouzi R, Barr CE, Crosbie EJ: HE4 as a Biomarker for Endometrial Cancer. *Cancers (Basel)* 2021, 13(19).

39. Gilks CB, Oliva E, Soslow RA: Poor interobserver reproducibility in the diagnosis of high-grade endometrial carcinoma. *Am J Surg Pathol* 2013, 37(6):874-881.

40. de Boer SM, Wortman BG, Bosse T, Powell ME, Singh N, Hollema H, Wilson G, Chowdhury MN, Mileshkin L, Pyman J *et al*: Clinical consequences of upfront pathology review in the randomised PORTEC-3 trial for high-risk endometrial cancer. *Ann Oncol* 2018, 29(2):424-430.

41. Kandoth C, Schultz N, Cherniack AD, Akbani R, Liu Y, Shen H, Robertson AG, Pashtan I, Shen R, Benz CC *et al*: Integrated genomic characterization of endometrial carcinoma. *Nature* 2013, 497(7447):67-73.

42. Oaknin A, Bosse TJ, Creutzberg CL, Giornelli G, Harter P, Joly F, Lorusso D, Marth C, Makker V, Mirza MR *et al*: Endometrial cancer: ESMO Clinical Practice Guideline for diagnosis, treatment and follow-up. *Ann Oncol* 2022, 33(9):860-877.

43. Raffone A, Travaglino A, Raimondo D, Neola D, Maletta M, Santoro A, Insabato L, Casadio P, Fanfani F, Zannoni GF *et al*: Lymphovascular space invasion in endometrial carcinoma: A prognostic factor independent from molecular signature. *Gynecol Oncol* 2022, 165(1):192-197.

FIGURE LEGENDS

Figure 1. Nomograms for predicting 1-,3-, and 5-year cancer-specific survival

*American Indian/AK Native, Asian/Pacific Islander.

Figure 2. Time-dependent AUC and calibration curves of the nomogram

(A-B) Time-dependent AUC of using the nomogram to predict cancer-specific survival probability within 5 years in the training and validation cohorts. The red line represents AUC = 0.7, which is considered ideal. (C, E, G) Calibration curves of 1-year, 3-year, and 5-year CSS for advanced EC post-surgery in the training cohort. (D, F, H) Calibration curves of 1-year, 3-year, and 5-year CSS for advanced EC post-surgery in the validation cohort. The black dashed line indicates the ideal reference line where predicted probabilities match the observed survival rates. The red dots represent the performance of the nomogram. The closer the solid red line is to the black dashed line, the more accurately the model predicts survival. AUC: area under the time-dependent receiver operating characteristic curves; CSS: cancer-specific survival. EC: endometrial carcinoma.

Figure 3. Comparison of the accuracy of the nomograms and FIGO stage based on changes in the ROC curves and the time-dependent AUC

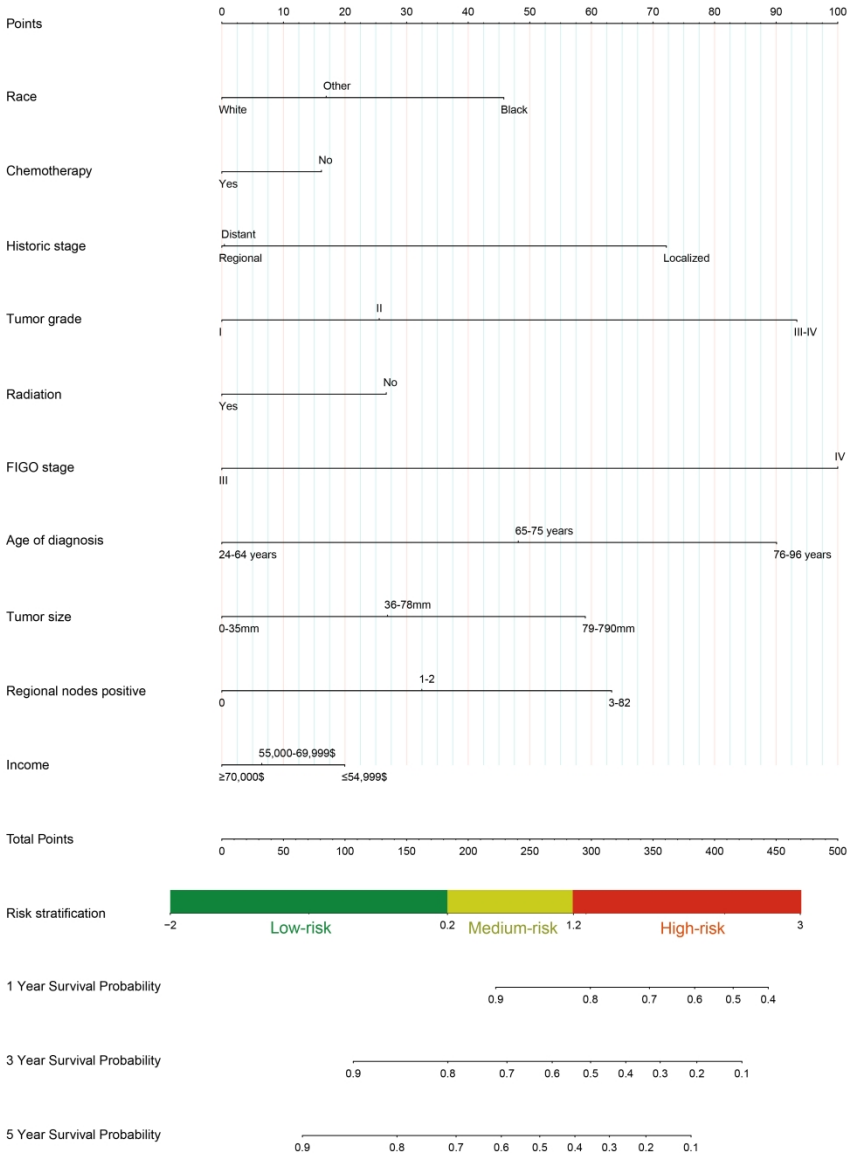


Figure 1 - Nomograms for predicting 1-,3-, and 5-year cancer-specific survival.
* American Indian/AK Native, Asian/Pacific Islander.

288x385mm (300 x 300 DPI)

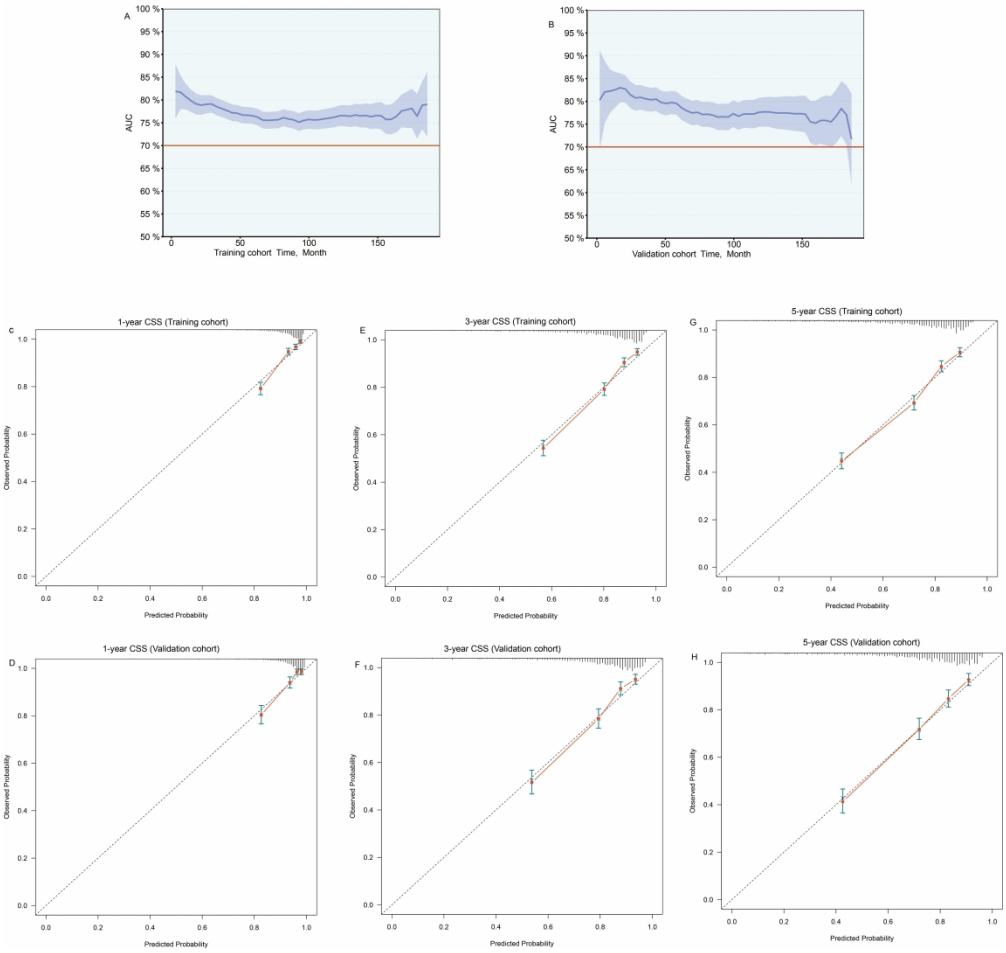


Figure 2 - Time-dependent AUC and calibration curves of the nomogram. (A-B) Time-dependent AUC of using the nomogram to predict cancer-specific survival probability within 5 years in the training and validation cohorts. The red line represents AUC = 0.7, which is considered ideal. (C, E, G) Calibration curves of 1-year, 3-year, and 5-year CSS for advanced EC post-surgery in the training cohort. (D, F, H) Calibration curves of 1-year, 3-year, and 5-year CSS for advanced EC post-surgery in the validation cohort. The black dashed line indicates the ideal reference line where predicted probabilities match the observed survival rates. The red dots represent the performance of the nomogram. The closer the solid red line is to the black dashed line, the more accurately the model predicts survival. AUC: area under the time-dependent receiver operating characteristic curves; CSS: cancer-specific survival. EC: endometrial carcinoma.

508x481mm (300 x 300 DPI)

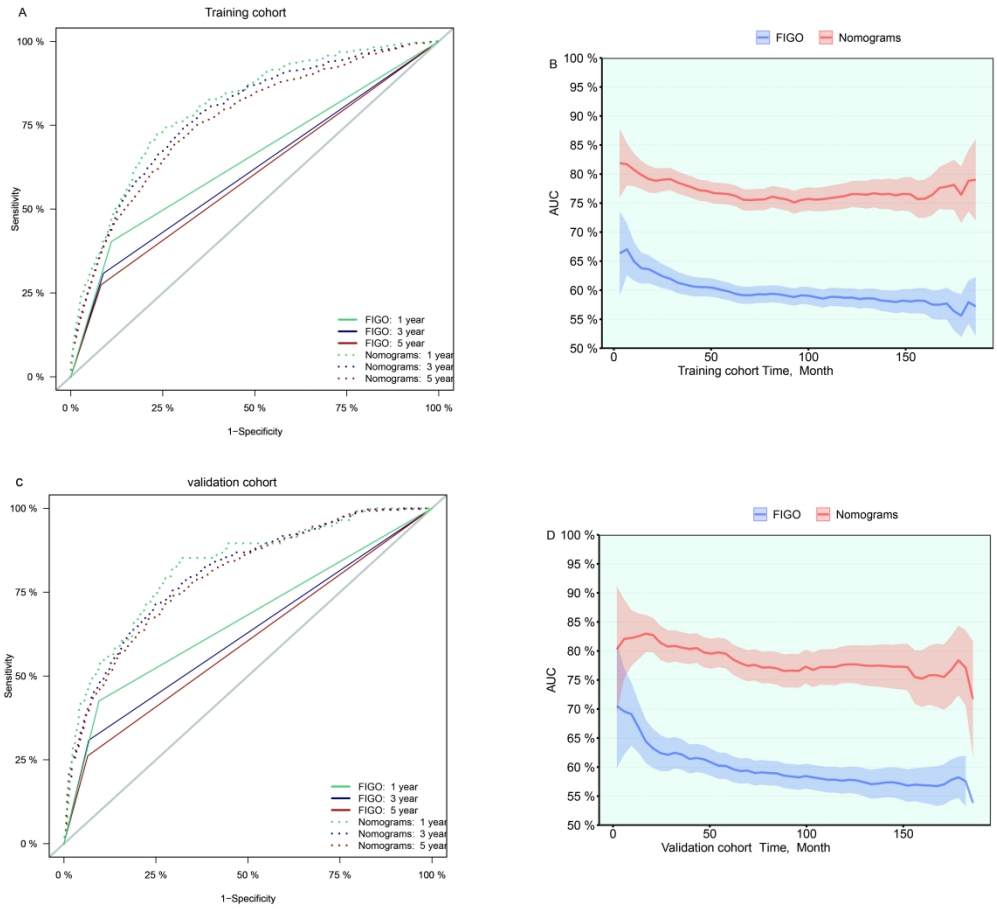
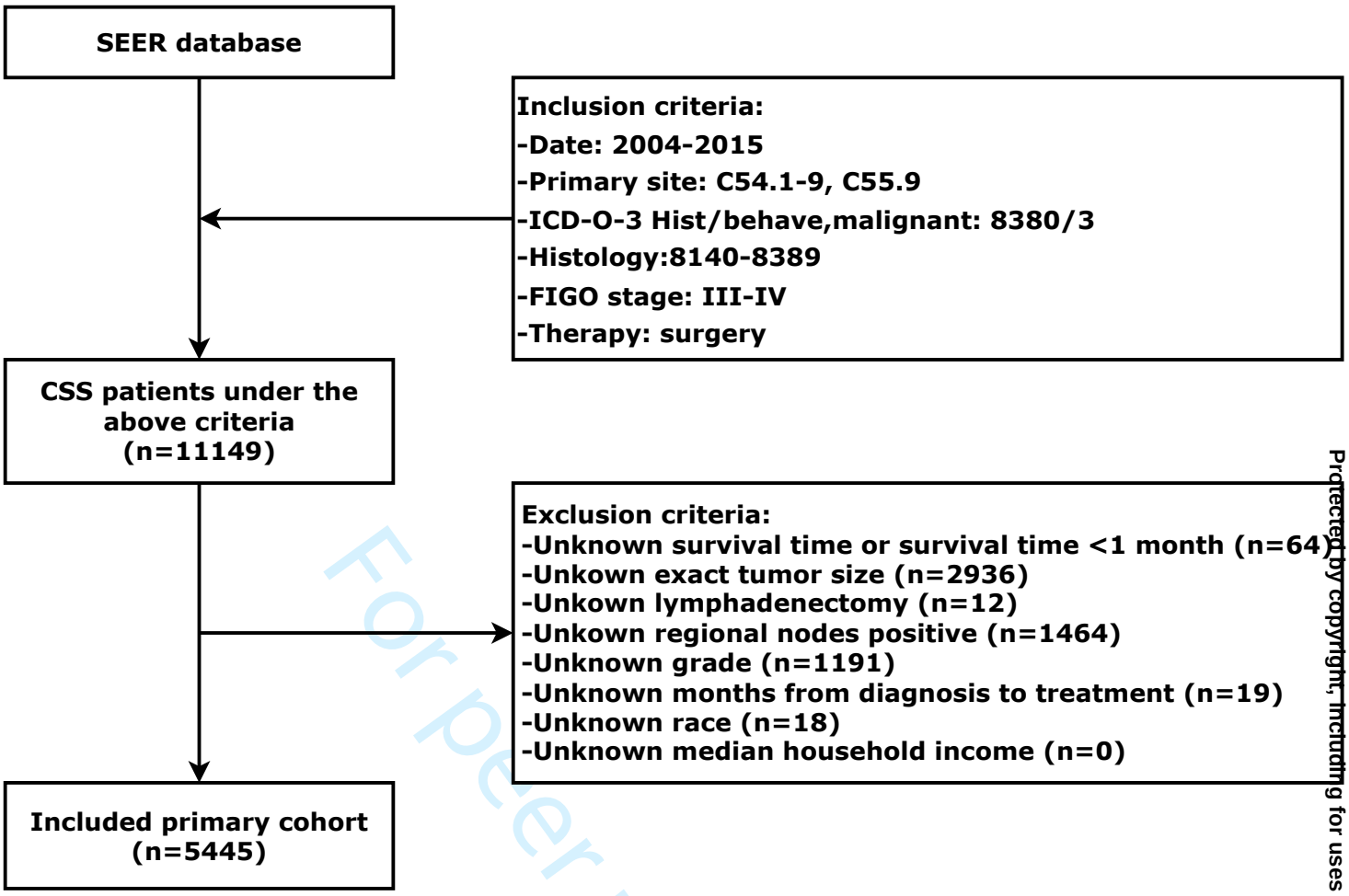
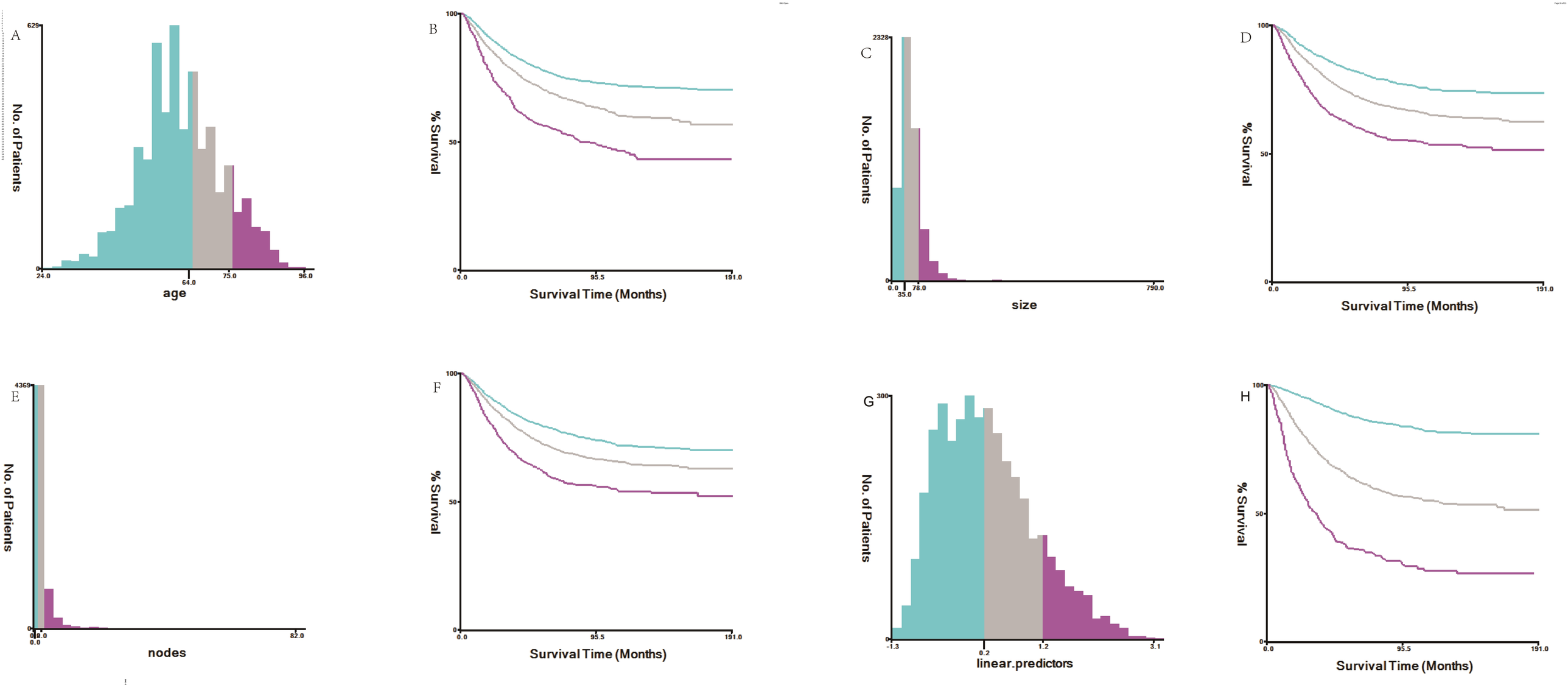


Figure 3 - Compares the accuracy of the nomograms and FIGO stage based on the changes in the ROC curves and the time-dependent AUC.

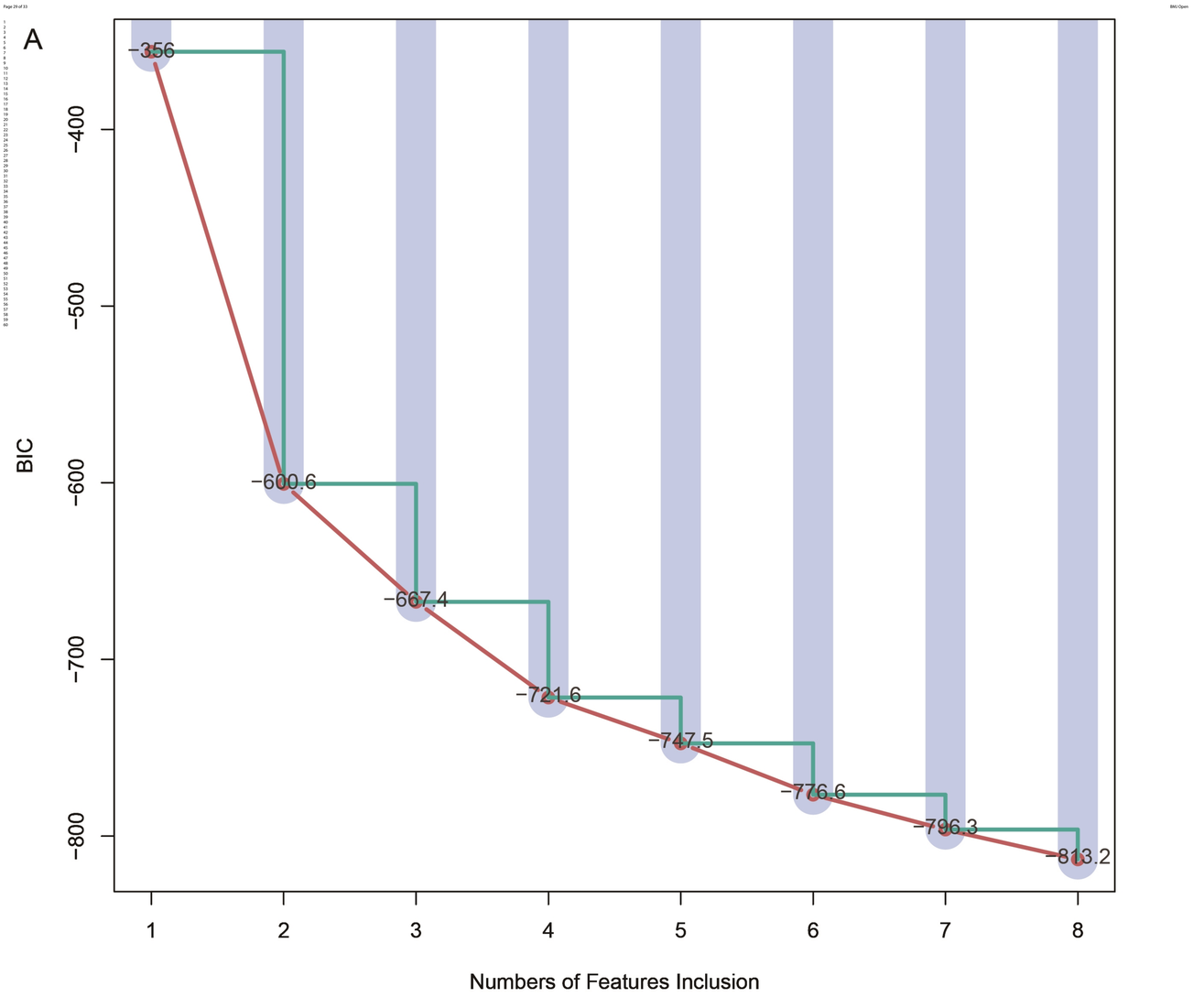
431x388mm (300 x 300 DPI)

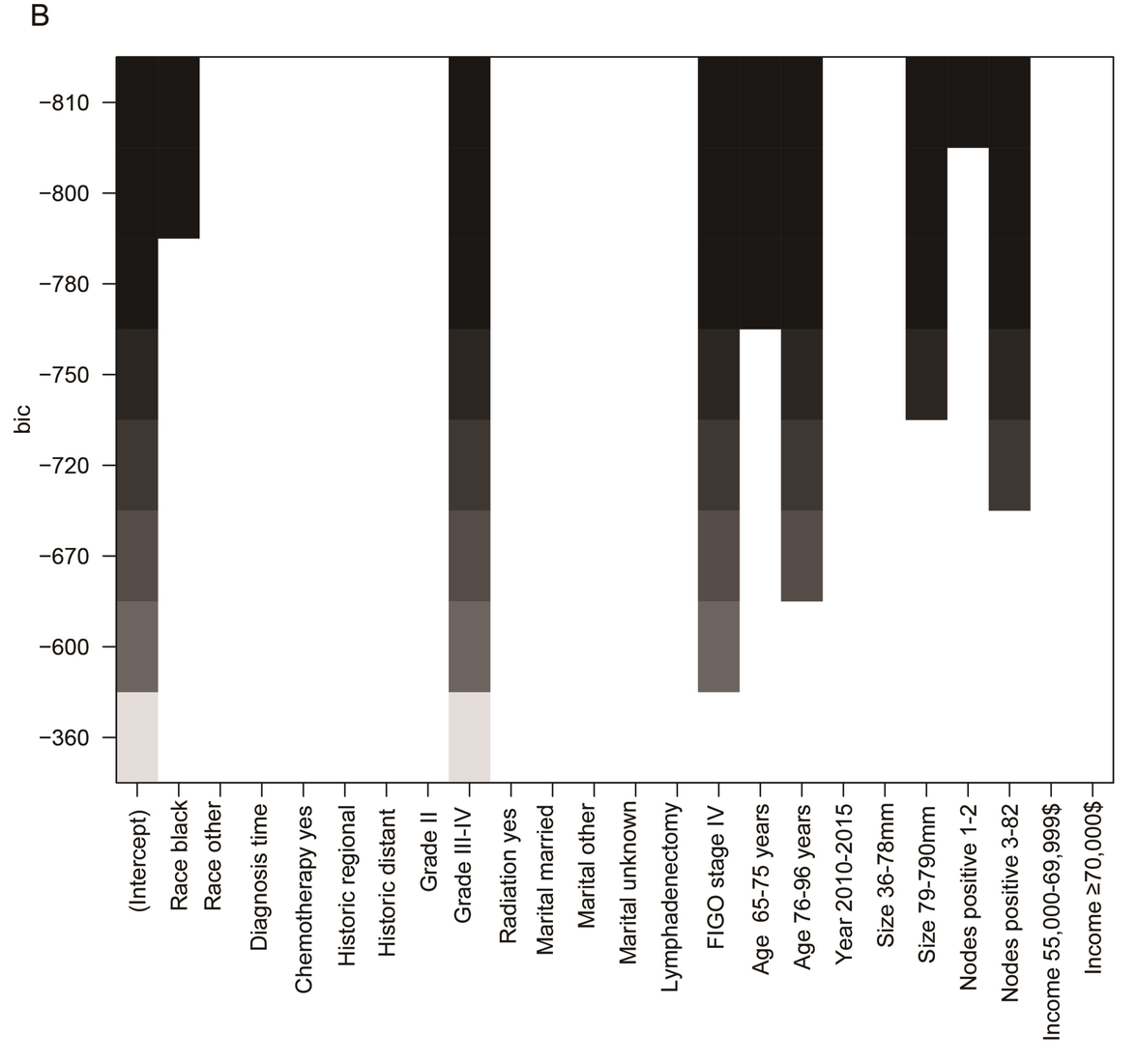


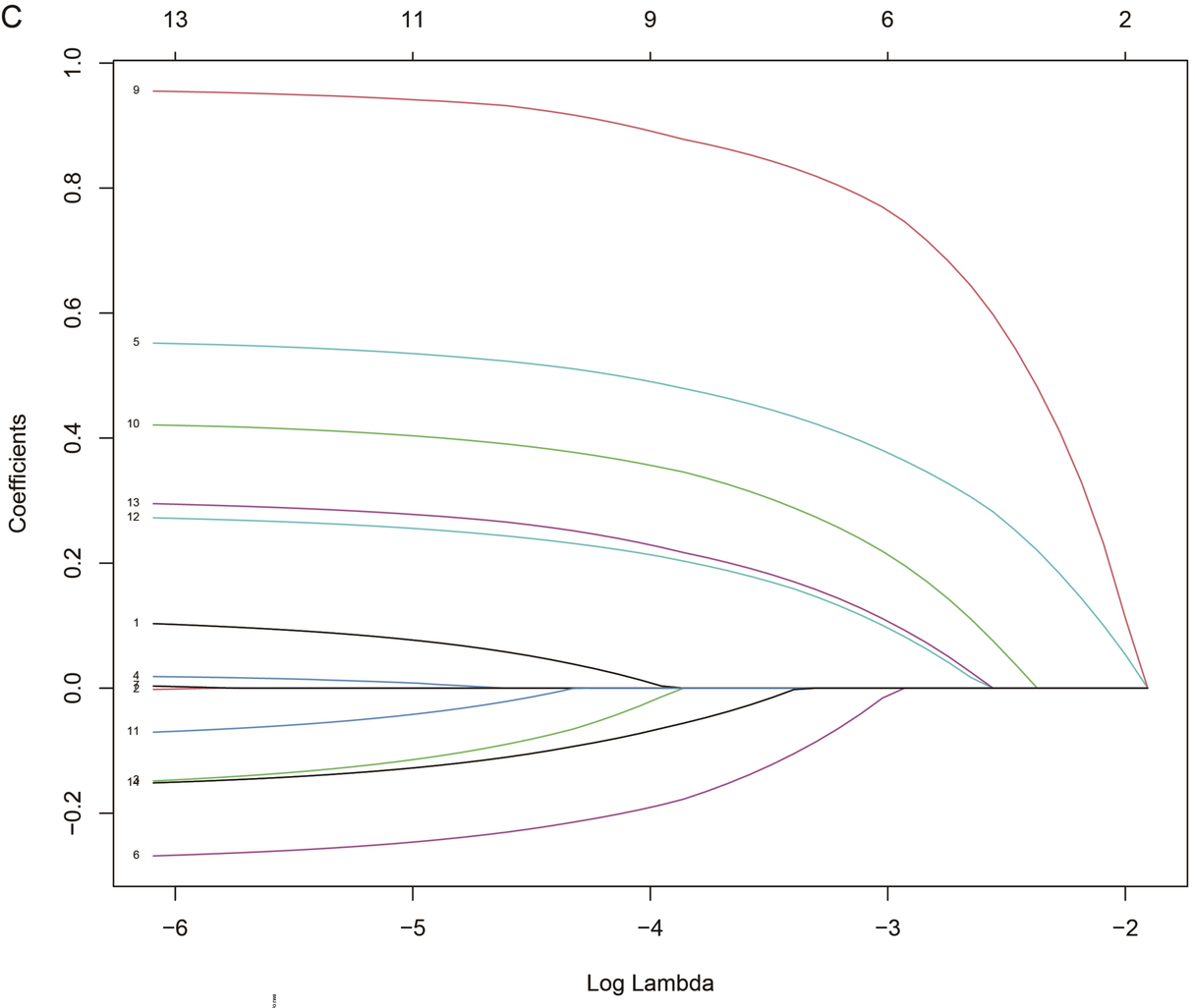
Supplementary Figure 1 Flowchart of the study.

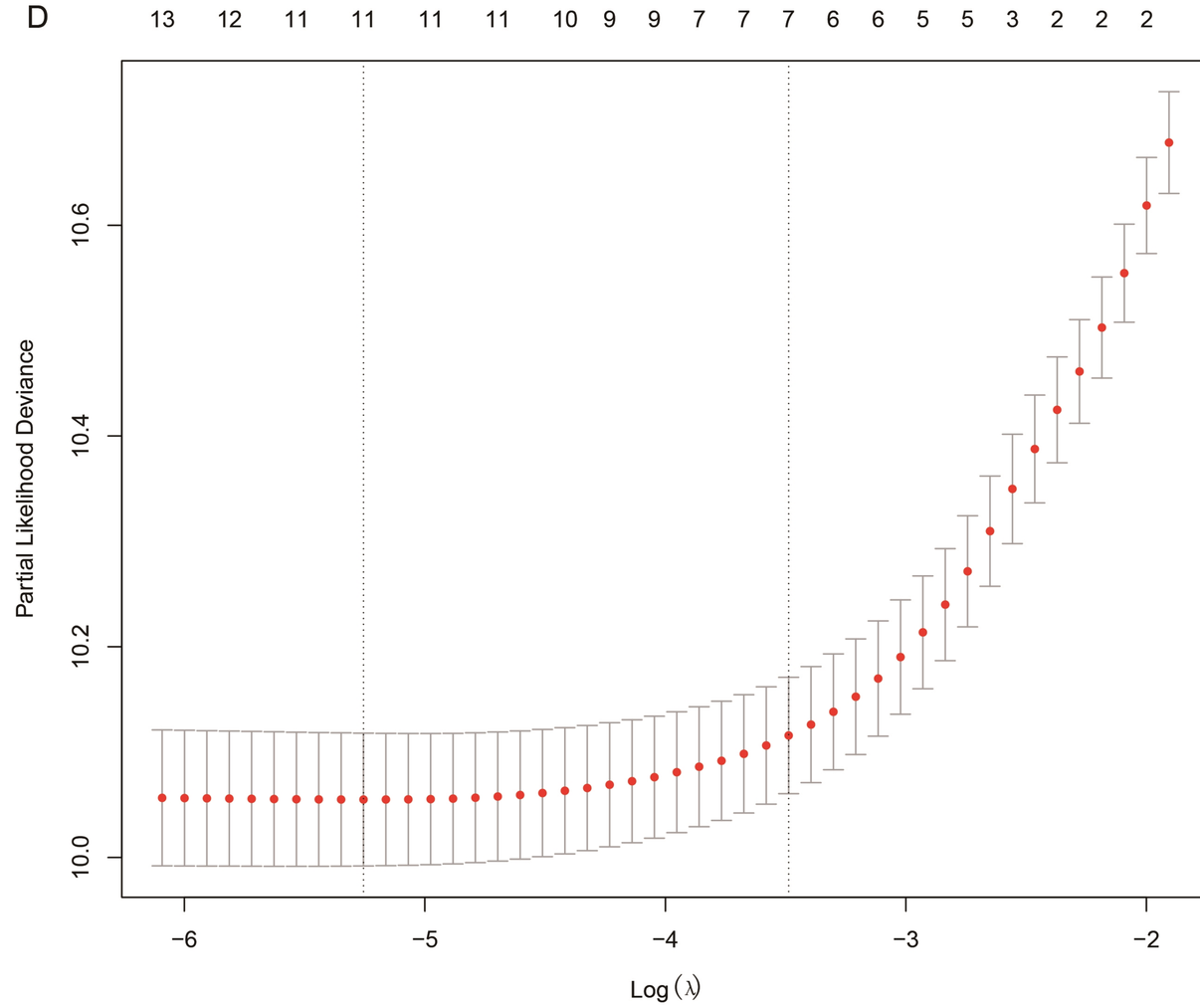


Supplementary Figure 2 Cutoff values for age (A,B), tumor size (C,D), regional nodes positive (E,F), and linear predictors (G,H) that were assessed by X-tile analysis.

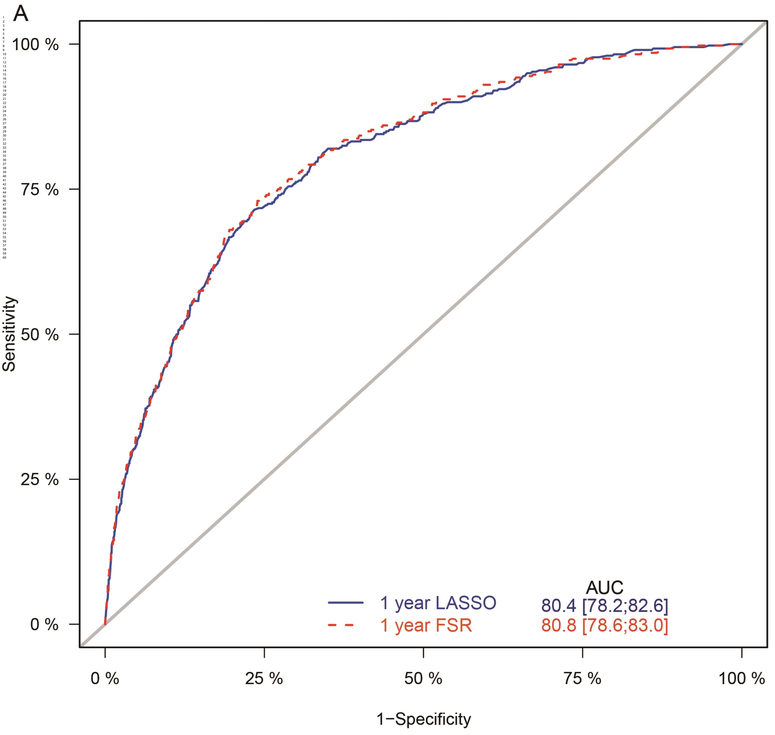


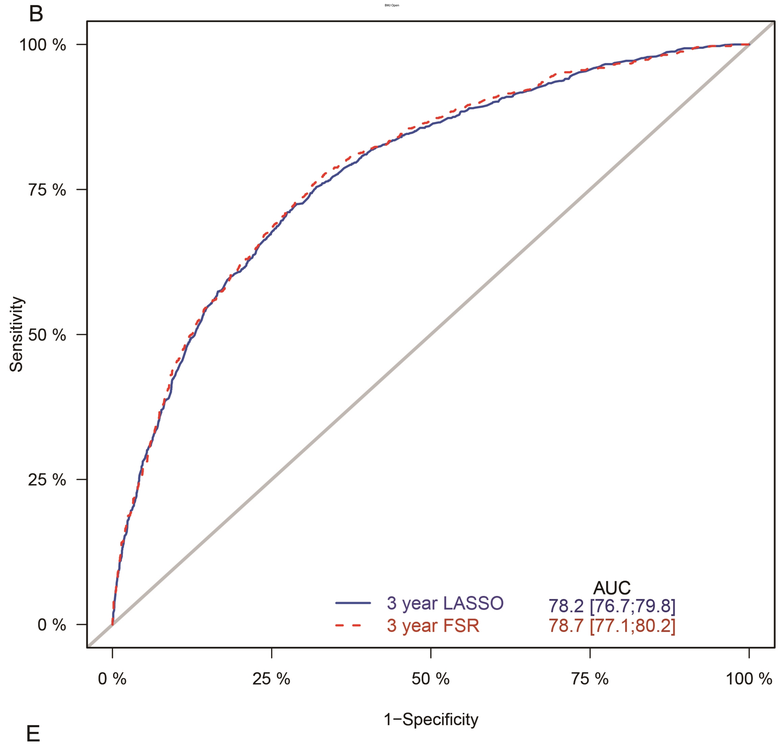


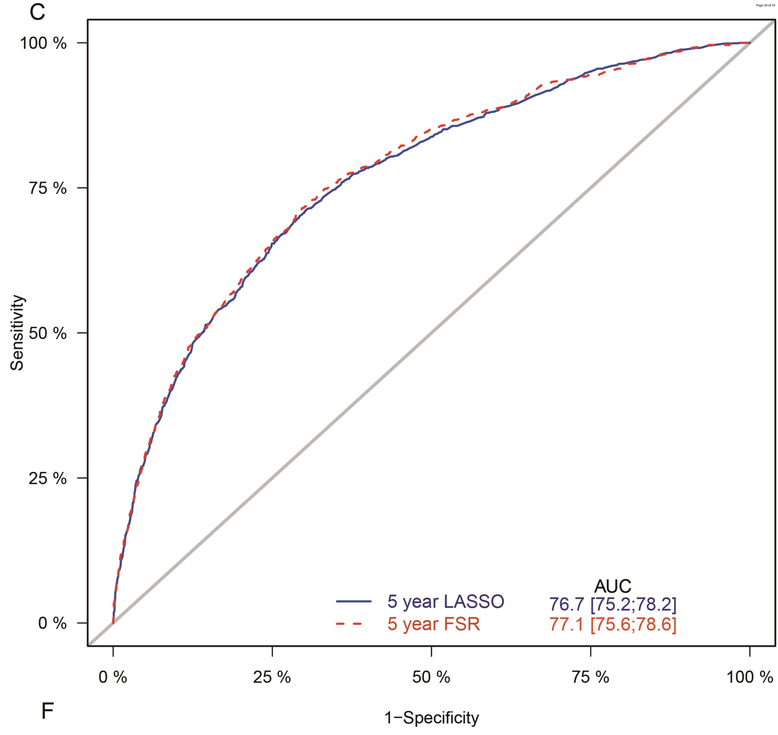


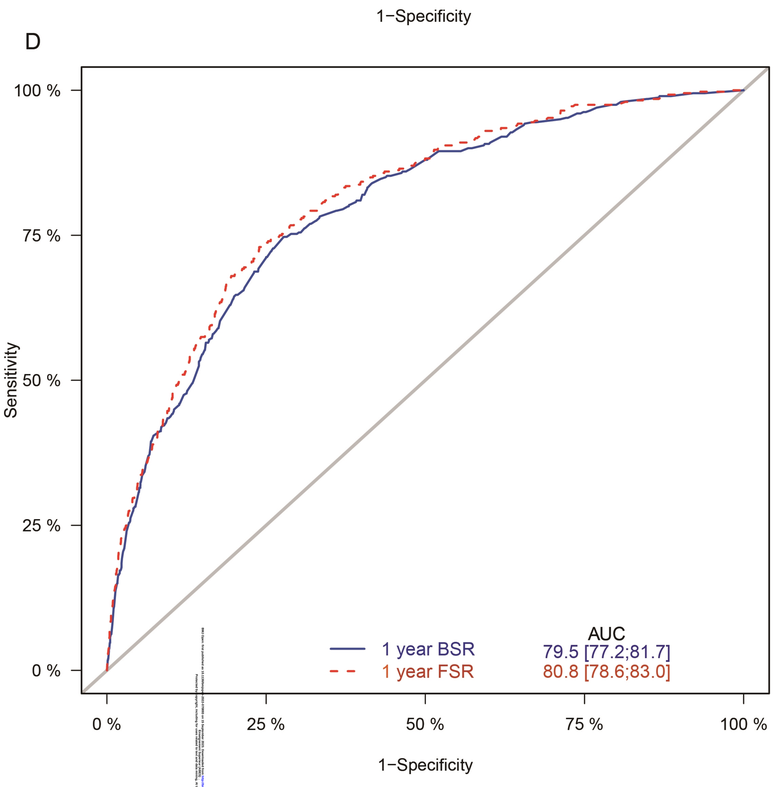


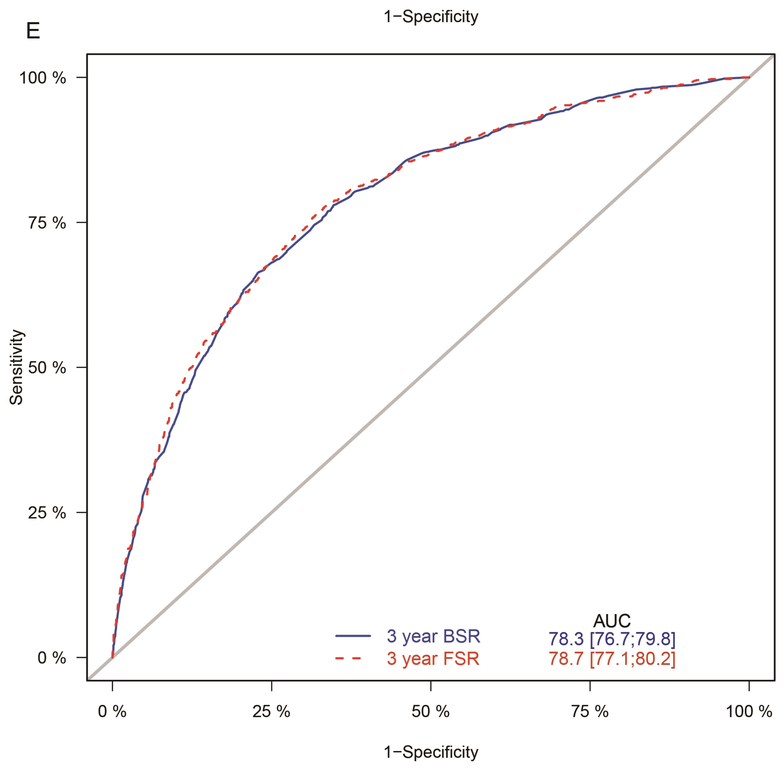
Supplementary Figure 3 Variables selection methods. (A, B) The selection of variables using the BSR method. (C) The LASSO coefficient profile of 14-related variables in primary cohort. (D) 10-fold cross-validation (CV) for tuning parameter (λ) selection.

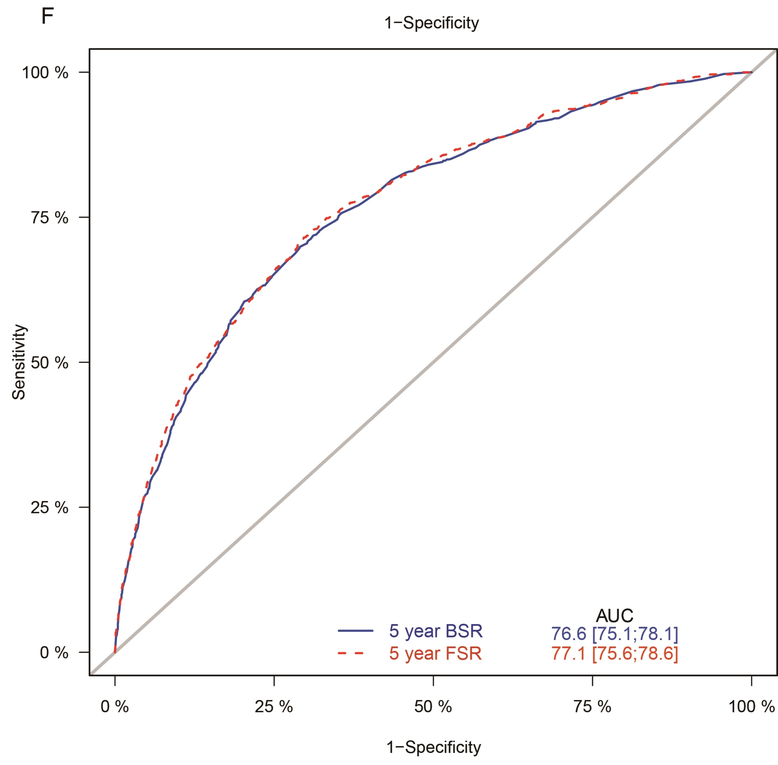




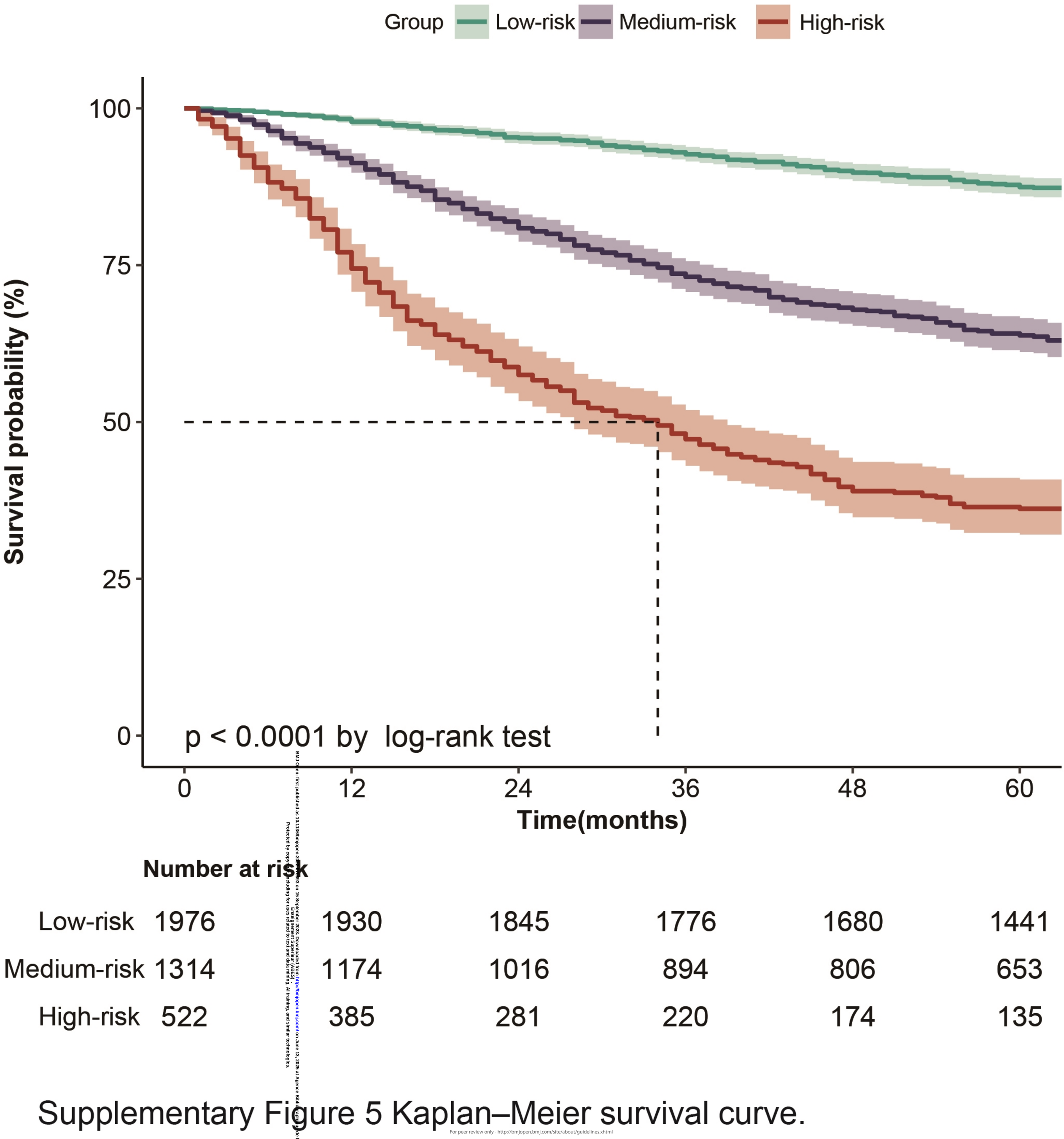








Supplementary Figure 4 Comparison of ROC curves of nomograms created with variables screened by BSR, LASSO, and FSR methods to predict 1-, 3-, and 5-year CSS after surgical treatment of advanced EC. BSR, best subsets regression. LASSO, least absolute shrinkage and selection operator. FSR, forward stepwise regression.



Supplementary Figure 5 Kaplan–Meier survival curve.

Table S1 BSR, LASSO and FSR screening of variables doing predictive models for IDI, NRI comparison.

Index	FSR vs. LASSO			FSR vs. BSR		
	Estimate	95% CI	*P*-value	Estimate	95% CI	*P*-value
IDI						
For 1-year CSS	0.006	0.002-0.012	0.01	0.013	0.006-0.019	<0.001
For 3-year CSS	0.004	0.001-0.008	0.01	0.012	0.006-0.018	<0.001
For 5-year CSS	0.003	0.001-0.007	0.01	0.011	0.006-0.016	<0.001
NRI						
For 1-year CSS	0.119	0.034-0.174	<0.001	0.214	0.113-0.254	<0.001
For 3-year CSS	0.033	(-0.013)-0.112	0.09	0.106	0.044-0.15	<0.001
For 5-year CSS	0.017	(-0.025)-0.099	0.289	0.09	0.053-0.128	<0.001

LASSO, least absolute shrinkage and selection operator. BSR, best subsets regression. FSR, forward stepwise regression. IDI, integrated discrimination improvement. NRI, net reclassification index. CI, confidence interval.

Table S2 IDI, and NRI of the nomogram and the FIGO stage in survival prediction for the advances endometrial carcinoma patients after surgical treatment.

Index	Training cohort			Validation cohort		
	Estimate	95% CI	*P*-value	Estimate	95% CI	*P*-value
IDI (vs. the FIGO stage)						
For 1-year CSS	0.062	0.047-0.084	<0.001	0.071	0.046-0.111	<0.001
For 3-year CSS	0.099	0.084-0.123	<0.001	0.119	0.088-0.155	<0.001
For 5-year CSS	0.112	0.095-0.133	<0.001	0.138	0.103-0.174	<0.001
NRI (vs. the FIGO stage)						
For 1-year CSS	0.364	0.306-0.425	<0.001	0.376	0.293-0.482	<0.001
For 3-year CSS	0.354	0.308-0.395	<0.001	0.352	0.302-0.421	<0.001
For 5-year CSS	0.337	0.292-0.377	<0.001	0.353	0.293-0.419	<0.001

TRIPOD Checklist: Prediction Model Development and Validation

Section/Topic	Item		Checklist Item	Page
Title and abstract				
Title	1	D;V	Identify the study as developing and/or validating a multivariable prediction model, the target population, and the outcome to be predicted.	1
Abstract	2	D;V	Provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusions.	1
Introduction				
Background and objectives	3a	D;V	Explain the medical context (including whether diagnostic or prognostic) and rationale for developing or validating the multivariable prediction model, including references to existing models.	2–3
	3b	D;V	Specify the objectives, including whether the study describes the development or validation of the model or both.	3
Methods				
Source of data	4a	D;V	Describe the study design or source of data (e.g., randomized trial, cohort, or registry data), separately for the development and validation data sets, if applicable.	3
	4b	D;V	Specify the key study dates, including start of accrual; end of accrual; and, if applicable, end of follow-up.	3
Participants	5a	D;V	Specify key elements of the study setting (e.g., primary care, secondary care, general population) including number and location of centres.	3
	5b	D;V	Describe eligibility criteria for participants.	3
	5c	D;V	Give details of treatments received, if relevant.	3–4
Outcome	6a	D;V	Clearly define the outcome that is predicted by the prediction model, including how and when assessed.	
	6b	D;V	Report any actions to blind assessment of the outcome to be predicted.	3–4
Predictors	7a	D;V	Clearly define all predictors used in developing or validating the multivariable prediction model, including how and when they were measured.	3–4
	7b	D;V	Report any actions to blind assessment of predictors for the outcome and other predictors.	3–4
Sample size	8	D;V	Explain how the study size was arrived at.	3–4
Missing data	9	D;V	Describe how missing data were handled (e.g., complete-case analysis, single imputation, multiple imputation) with details of any imputation method.	3–4
Statistical analysis methods	10a	D	Describe how predictors were handled in the analyses.	4
	10b	D	Specify type of model, all model-building procedures (including any predictor selection), and method for internal validation.	4
	10c	V	For validation, describe how the predictions were calculated.	4
	10d	D;V	Specify all measures used to assess model performance and, if relevant, to compare multiple models.	4
	10e	V	Describe any model updating (e.g., recalibration) arising from the validation, if done.	4
Risk groups	11	D;V	Provide details on how risk groups were created, if done.	5
Development vs. validation	12	V	For validation, identify any differences from the development data in setting, eligibility criteria, outcome, and predictors.	4
Results				
Participants	13a	D;V	Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful.	5–6
	13b	D;V	Describe the characteristics of the participants (basic demographics, clinical features, available predictors), including the number of participants with missing data for predictors and outcome.	5
	13c	V	For validation, show a comparison with the development data of the distribution of important variables (demographics, predictors and outcome).	5–6
Model development	14a	D	Specify the number of participants and outcome events in each analysis.	6
	14b	D	If done, report the unadjusted association between each candidate predictor and outcome.	7–9
Model specification	15a	D	Present the full prediction model to allow predictions for individuals (i.e., all regression coefficients, and model intercept or baseline survival at a given time point).	7–9
	15b	D	Explain how to the use the prediction model.	9
Model performance	16	D;V	Report performance measures (with CIs) for the prediction model.	8–9
Model-updating	17	V	If done, report the results from any model updating (i.e., model specification, model performance).	9
Discussion				
Limitations	18	D;V	Discuss any limitations of the study (such as nonrepresentative sample, few events per predictor, missing data).	11–12
Interpretation	19a	V	For validation, discuss the results with reference to performance in the development data, and any other validation data.	10–11
	19b	D;V	Give an overall interpretation of the results, considering objectives, limitations, results from similar studies, and other relevant evidence.	10–11
Implications	20	D;V	Discuss the potential clinical use of the model and implications for future research.	10–

TRIPOD Checklist: Prediction Model Development and Validation

				11
Other information				
Supplementary information	21	D;V	Provide information about the availability of supplementary resources, such as study protocol, Web calculator, and data sets.	12–13
Funding	22	D;V	Give the source of funding and the role of the funders for the present study.	13

*Items relevant only to the development of a prediction model are denoted by D, items relating solely to a validation of a prediction model are denoted by V, and items relating to both are denoted D;V. We recommend using the TRIPOD Checklist in conjunction with the TRIPOD Explanation and Elaboration document.