

Development and validation of prognostic factors for lymph node metastasis in endometrial cancer: A SEER analysis

Xingchen Li (✉ lixingchen90@126.com)

Peking University People's Hospital

Yuan Cheng

Peking University People's Hospital

Yangyang Dong

Peking University People's Hospital

Jingyi Zhou

Peking University People's Hospital

Xiao Yang

Peking University People's Hospital

Boqiang Shen

Peking University People's Hospital

Xiaotong Wu

Peking University People's Hospital

Zhiqi Wang

Peking University People's Hospital

Xiaoping Li

Peking University People's Hospital

Jianliu Wang

Peking University People's Hospital

Research article

Keywords: Endometrial cancer, lymph node metastasis, nomogram, decision curve analysis

Posted Date: April 15th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-22162/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Objective: The purpose of this study was to develop and validate a nomogram that can be used to predict lymph node metastasis (LNM) in patients with endometrial carcinoma (EC).

Methods: Clinical data of EC patients diagnosed between 2004 and 2015 were retrieved from the Surveillance, Epidemiology, and End Results Program (SEER) registry. The nomogram was constructed using independent risk factors chosen using a multivariate logistic regression analysis. Accuracy was validated for both groups using discrimination analysis and calibration curves. The predictive accuracy and clinical value of the nomogram and Mayo criteria were compared using decision curve analysis (DCA).

Results: The final study group consisted of 63,836 women that met specific inclusion criteria. The factors that were identified in the multivariate analysis to be notable predictors of LNM were age, tumor size, histological type, cervical stromal invasion, tumor grade, and myometrial invasion. These risk factors were included in the nomogram. Discriminations of the nomogram and Mayo criteria were 0.848 (95% CI: 0.843-0.853) and 0.806 (95%CI: 0.801-0.812), respectively. In the validation group, the AUC values were 0.847 (95%CI: 0.840-0.857) and 0.804 (95%CI: 0.796-0.813) for the nomogram and the Mayo criteria, respectively ($P < 0.01$). Calibration plots showed that training and validation cohorts were well-calibrated. DCA revealed that by using the nomogram always had a positive net benefit compared to using the Mayo criteria. **Conclusions:** A nomogram was developed to predict LNM in EC patients based on a large population-based analysis. The nomogram showed good performance for predicting LNM in patients with EC.

Introduction

Endometrial cancer (EC) is the most common gynecologic malignancy and the fourth most common of all tumor types in Western countries. According to the American Cancer Society, there are approximately 63,230 new cases of endometrial cancer each year in the United States and 11,350 related deaths ^[1]. Symptoms such as abnormal vaginal bleeding usually present quickly after onset, so most EC patients are diagnosed at an early stage. Early diagnosis can allow for early treatment and good prognosis for EC patients. However, lymph node metastasis (LNM) is a major risk factor for recurrence and metastasis of EC. Para-aortic and pelvic LNM have been observed in 5% and 9% of EC patients, respectively ^[2]. According to the International Federation of Gynecology and Obstetrics (FIGO) staging system, this lymphatic involvement often indicates a poor prognosis ^[3]. Disease-free survival (DFS) for patients at FIGO stages 1 and 2 was 90% for those without LNM, 75% for those with pelvic LNM, and 38% for those with para-aortic LNM ^[4]. Another study reported a 5-year overall survival rate of 53.6% for patients with para-aortal lymph nodal involvement as opposed to 96.5% for those without any LNM ^[5].

A nomogram is a predictive tool that provides the overall probability of a specific outcome ^[6]. Nomograms consist of multiple prognostic factors in a graphical display, for the purpose of simplifying

prediction of a specific event. In general, nomograms used for predicting patient outcomes discriminate patients with a future event from those without. The performance of the nomogram is initially validated by providing discrimination and calibration values [7]. Discrimination evaluates whether the model is able to discriminate patients (distinguish one patient from another) and is generally expressed with area under the curve (AUC) values [8]. Calibration describes how close predicted and actual outcomes are.

Many researchers have attempted to identify risk factors for LNM and predict the probability of LNM by developing nomograms constructed using different risk factors. The Mayo risk stratification model predicts the risk of LNM based on 4 factors: tumor grade, histological type, tumor size, and myometrial invasion. Some studies also have compared the Mayo criteria with other models including nomograms designed as part of their studies [9, 10]. However, due to the small size of patient populations, these studies could not provide crucial information on how small incremental changes in tumor size may affect patient outcomes. To date, there is no published model for predicting lymph node metastasis in EC patients based on a large cohort.

Because of this need, we designed a nomogram to predict LNM based on data from the Surveillance, Epidemiology, and End Results (SEER) database. In this study, we developed and internally validated a parametric model for predicting LNM. The model includes pathological characteristics based on a mathematical algorithm from a large population. Decision curve analysis also was used to estimate the clinical value of the nomogram and compare it with the Mayo criteria.

Materials And Methods

1. Patients and study design

The National Cancer Institute's SEER database on cancer research is freely available to the public upon submission of a signed data-used agreement to the SEER administration. We extracted data for endometrial cancer cases diagnosed between 2004 and 2015 from the SEER registry to use for further analysis. Inclusion criteria for patients were as follows: age at diagnosis > 18 years; and endometrial cancer was pathologically confirmed by histology (histological code: 8140-8389 for EEA, 8440-8499 for SEA). Patients who had a history of prior malignancy, or who had missing information regarding lymph node metastasis, race, marital status, tumor size, histology, myometrial invasion, cervical stromal invasion, or tumor grade were excluded. A total of 63836 patients in the SEER cohort meeting the criteria were selected for further analysis. Patients were then randomly divided into two groups in a 2:1 ratio, to form a training cohort (n=42558) and an internal validation cohort (n=21278). The flow chart used for data selection is shown in Figure. 1. The ethics committee board of Peking University People's Hospital approved the use of patient data for this study.

2. Variables

The patient variables that were evaluated to design the nomogram were as follows: age, race, tumor size, year of diagnosis, marital status, histological type, myometrial invasion, tumor grade, and lymph node

metastasis. Age was categorized subjectively as ≤ 40 years, 41-50 years, 51-60 years, 61-70 years, or >70 years. Tumor size was classified as ≤ 2 cm, 2.1cm-5cm, 5.1cm-10cm, or >10 cm. The histological types were classified as endometrial endometrioid adenocarcinoma (EEA, ICD-O-3 codes: 8140-8389), serous endometrioid adenocarcinoma (SEA, ICD-O-3 codes: 8440-8499), or "other types". Patients were classified as low-risk using the Mayo criteria if they're symptoms were characterized by grade 1 or 2 tumors, less than 50% myometrial invasion, and tumor size ≤ 2 cm. Patients classified as high-risk using the Mayo criteria had tumors with myometrial invasion $>50\%$, grade 3, and tumor size >2 cm.

3. Statistical analyses

All the categorical variables were described as frequencies and percentages. Univariate and multivariate logistic regression analyses were used to identify independent risk factors predictive of LNM and to develop the nomogram in the training cohort. All variables in the univariate analysis with $P < 0.05$ were considered statistically significant and selected for multivariate analysis. Candidate variables with $P < 0.05$ were selected using a backward stepwise selection from the full multivariate model. The nomogram was then constructed using these candidate variables.

The performance of the nomogram was assessed in both the training and validation groups by calculating discrimination and calibration criteria [11]. Discrimination was quantified using an area under the receiver operating characteristic (ROC) curve. The area under the curve (AUC) is a summary measure of the ROC that reflects the ability of a test to discriminate the outcomes across all possible levels of positivity. AUC values range from 0 to 1, and a model is considered to have a poor, fair or good performance if the AUC value is between 0.5 to 0.6, 0.6 to 0.7 or greater than 0.7, respectively. A calibration plot was generated to visualize how far the predictions were from the actual outcomes, displaying mean nomogram-based predictions in training and validation cohorts on the horizontal axis versus actual observed LNM probabilities.

Finally, decision curve analysis was performed to quantify the clinical usefulness of the model. Such analyses can determine the ability of a model to predict fine-scale outcomes based on a set of risk parameters. A model that performs well in the decision curve analysis has a higher net benefit than a model that simply classifies all patients as having the predicted outcome or no (zero) patients as having the outcome. Decision curve analysis can also be used to compare the net benefits of multiple models.

All analyses were performed using SPSS 21.0 and R software version 3.4.4 (<https://www.r-project.org/>), using the *rms*, *presence/absence*, and *decision curve* packages. $P < 0.05$ was considered statistically significant.

Results:

1. Clinical characteristics of patients

The data from a total of 63836 patients were included in the study. Out of the patients, 42558 patients were placed within the training cohort, while 21278 were placed within a validation cohort. **Figure. 1** shows a schematic of the screening process. The mean ages of patients within the training and validation sets were 62.41 ± 11.62 years and 62.43 ± 11.58 years, respectively. Tumor size was 6.58 ± 15.63 cm in the training cohort and 6.56 ± 15.62 cm in the validation group. In the training cohort, most patients (88.45%) were negative for LNM. The age of 3.49% patients was less than 40 years old, 10.11% patients were between 41 and 50 years old, 30.79% were between 51 and 60 years old, 31.86% were between 61 and 70 years old, and 23.75% were older than 71 years old. Most of the patients in both cohorts were white (82.40%), married (53.79%), and their tumor size was between 2cm to 5cm (49.36%). The pathological characteristics of the majority included EEA (87.27%), no myometrial invasion (65.10%), no cervical stromal invasion (79.80%) and tumors classified as grade 1 (41.50%). The two sets showed similar results for nearly all variables. Table 1 shows the details of demographic and pathological characteristics of the patients in the two cohorts.

2. Risk factors for lymph node metastasis

Our univariate analysis considered age at diagnosis, marital status, race, tumor size, histological type, myometrial invasion, cervical stromal invasion, and tumor grade as potential risk factors for LNM from the training cohort data. After multiple logistic regression analysis, it was found that independent risk factors associated with LNM including age at diagnosis, tumor size, histological type, myometrial invasion, cervical stromal invasion, and tumor grade (**Table 2**). Among these independent risk factors, cervical stromal invasion was considered as a major predictor (OR=6.09, 95% CI 5.50-6.76, $P<0.001$). Other factors considered to be predictors of LNM included age (OR=1.01, 95% CI 1.01-1.02, $P<0.001$), tumor size 2-5cm (OR=1.51, 95% CI 1.34-1.70, $P<0.001$), tumor size 5-10cm (OR=2.71, 95% CI 2.39-3.06, $P<0.001$), tumor size ≥ 10 cm (OR=3.38, 95% CI 2.90-3.95, $P<0.001$), histological type SEA (OR=1.78, 95% CI 1.61-1.97, $P<0.001$) and histological types other than EEA or SEA (OR=1.33, 95% CI 1.18-1.50, $P<0.001$), positive myometrial invasion (OR=2.77, 95% CI 2.18-3.52, $P<0.001$), tumor grade 2 (OR=2.27, 95% CI 2.04-2.53, $P<0.001$) and tumor grade 3 (OR=4.68, 95% CI 4.21-5.20, $P<0.001$). Analysis in the validation set revealed the same independent risk factors for LNM.

3. Design and validation of the nomogram

Based on the independent risk factors identified in the multivariate regression analysis, we designed a nomogram to predict LNM in EC patients (**Figure. 2**). Among the variables considered in the predictive model, cervical stromal invasion was identified to be the most important predictive factor for the LNM nomogram. Point assignments and predictive scores for each variable in the nomogram models were calculated, with the total score corresponding to a predicted probability of LNM. The performance of the final model was assessed through discrimination and calibration analyses. Based on these analyses, the nomogram had an AUC value of 0.848 (95%CI: 0.843-0.853) for the training group, as compared with 0.806 (95%CI: 0.801-0.812) for the Mayo criteria ($P<0.01$; **Figure. 3A**). In the validation group, the AUC value was 0.847 (95%CI: 0.840-0.857) for the nomogram and 0.804 (95%CI: 0.796-0.813) for the Mayo

criteria, respectively ($P < 0.01$; **Figure. 3B**). The nomogram showed discrimination majority to the Mayo criteria for both the training and validation cohorts. The calibration curves for predicting LNM demonstrated that the nomogram was well-calibrated for both the training (**Figure. 4A**) and validation groups (**Figure. 4B**).

4. Optimal threshold of the nomogram

Each patient was assigned a score using the calibrated nomogram. Then, an optimal cut-off value of 200 points was selected to maximize sensitivity and specificity of average scores in the ROC curve. Patients from training and validation cohorts were divided into low-risk (score < 200 points) and high-risk (score ≥ 200 points) groups. The performance of the nomogram stratification was compared to the Mayo criteria for predicting LNM. The nomogram showed better discrimination than the Mayo criteria in both the training (nomogram: AUC=0.754, 95%CI; Mayo: AUC=0.716, 95%CI; $P < 0.01$; **Figure. 5A**) and the validation groups (nomogram: AUC=0.751, 95%CI; Mayo: AUC=0.714, 95%CI; $P < 0.01$; **Figure. 5B**). In the training group, the LNM rates were 4.80% and 34.0% in low-risk and high risk groups, respectively, according to the nomogram, and 5.7% and 26.4%, respectively, according to the Mayo criteria (**Table 4**). In the validation cohort, the predicted rates of LNM were 4.8% and 33.7% in the low-risk and high-risk groups, respectively, according to the nomogram, and 5.6% and 25.9%, respectively, according to the Mayo criteria (**Table 5**).

5. Decision curve analysis

The decision curve analysis results for the nomogram and Mayo models are shown in **Supplementary Figure 1A** (training cohort) and **Supplementary Figure 1B** (validation cohort). For predicted probability thresholds between 0% and nearly 60%, the nomogram showed a positive net benefit for both cohorts.

Discussion

Endometrial cancer (EC) is one of the most common types of gynecologic malignancy. It is estimated that 63.4 out of every 100,000 people in China in 2015 were diagnosed as EC ^[12]. One of the most important prognostic factors for EC is the presence of LMN. Nevertheless, there is an ongoing debate concerning practicality of lymphadenectomy. Some studies have shown that traditional lymphadenectomy may not improve DFS or overall survival, while other studies have suggested the opposite ^[13-15]. Therefore, it is crucial to distinguish patients with low risk of LNM from those with high risk. Currently, most of the studies have followed the Mayo criteria for predicting LNM risk and compared it with different models. According to one study based on the Mayo criteria ^[16], 78.9% of the studied patients were at high-risk for nodal metastasis, but the actual LMN rate was only 6.4%. Thus, almost 70% patients without LMN were over-treated. Multiple retrospective studies have shown that a low-risk subset of EC patients have a low overall risk of lymph node involvement ^[17-19]. Mariani et al. performed a large retrospective study of EC patients, which determined that patients with myometrial invasion $< 50\%$, tumor grade 1 or 2, and tumor size < 2 cm were at a low risk for lymph node involvement ^[20]. Vargas et al. analyzed 19329 patients with EC and the results suggested that EC patients for which the Mayo criteria

predicted a low risk of LMN, did have a low incidence of metastasis [16]. However, the scale of the patients in these studies was relatively small. Therefore, our goal was to conduct a large-cohort study to identify risk factors for LNM and develop accurate risk stratification models to separate patients with EC into low-risk and high-risk categories. The benefit of this model will be preventing patients from unnecessary lymph node dissection and the associated surgical morbidity.

Nomograms and clinicopathologic variables have been used by some researchers to estimate risk of metastasis. In the current study, we adopted a large-scale population-based SEER database to develop and validate a convenient nomogram for doctors to make individualized predictions of LNM in patients with EC. Data were collected for a total of 63836 EC patients and randomly divided into two groups. Several prognostic factors were identified for the SEER patients in this study, including age at diagnosis, marital status, race, tumor size, histological type, myometrial invasion, cervical stromal invasion, and tumor grade. Further multivariate logistic regression analysis revealed six factors that were predictive of LNM, including age at diagnosis, tumor size, histological type, myometrial invasion, cervical stromal invasion, and tumor grade. These six factors were incorporated in the nomogram design. The results showed that the nomogram provided good discrimination for both the training (AUC = 0.848; 95%CI: 0.843–0.853) and validation groups (AUC = 0.847; 95%CI: 0.840–0.857). The nomogram also performed significantly better than the Mayo criteria for predicting LNM in both the training (AUC = 0.806; 95%CI: 0.801–0.812, $P < 0.001$) and validation groups (AUC = 0.804; 95%CI: 0.796–0.813, $P < 0.001$). Furthermore, calibration results indicated that the predictions made using the nomogram fit well with observations for both groups. We used 200 points as the cut-off value based on Youden's index [21] and divided the patients from both cohorts into low-risk (score < 200 points) and high-risk groups (score \geq 200 points). The performance comparison of our nomogram and Mayo criteria for predicting LNM was verified. The nomogram showed a better discrimination than the Mayo criteria in both training (with AUC of 0.754, 95% CI, 0.72–0.83 vs. 0.716, 95% CI, 0.60–0.67; $P < 0.001$) and validation cohorts (with AUC of 0.751, 95% CI, 0.66–0.75, vs. 0.714, 95% CI, 0.54–0.59; $P < 0.001$) for the AUC values were greater for the nomogram than for the Mayo criteria. Decision curve analysis showed that the use of our nomogram had a positive benefit compared to the Mayo criteria. The lymph node metastasis rates were 4.80% and 34.0% in low-risk and high-risk groups, respectively, according to the nomogram, while the rates were 5.7% and 26.4%, respectively, according to the Mayo criteria.

Several previous studies have reported different risk-scoring nomograms for predicting LNM. Sofiane constructed a predictive model to identify high-risk LNM patients in early-stage type 1 EC [22]. Korean Gynecologic Oncology Group (KGOG) developed a preoperative assessment of LNM in EC that included features of MRI results and serum CA125 levels [23]. Lymph-vascular space invasion (LVSI) also has been shown to have a great impact on LNM in some studies [24, 25]. However, these studies had smaller populations, and thus, their predictive values were lower than the nomogram presented here.

Some research also has concentrated on comparing different models. Gokhan et al. compare the Mayo criteria and Milwaukee risk stratification models and found that the Mayo model was more accurate for

predicting LNM than the Milwaukee model [26]. Tuomi et al. compared the performance characteristics of three risk-stratification models, (Mayo, Helsinki and Milwaukee models) and found that these models had similar accuracies for predicting lymphatic dissemination in EC patients [27].

The nomogram may have performed better than the Mayo model for predicting LNM due to the differences in predictive value assigned to two specific factors in the models: age at diagnosis and cervical stromal invasion (the other four factors performed the same in both models). Age at diagnosis had only a small impact on LNM according to the nomogram algorithm, however, cervical stromal invasion was shown to be a strong predictor of LNM. A previous study from our institute also revealed that cervical stromal invasion is an independent risk factor for LNM (in addition to LVSI, tumor grade, and myometrial invasion) in EC patients and may have an important role in predicting LNM. [28]. Another study also suggested that cervical stromal invasion was useful in estimating LNM risk for EC patients and directing therapeutic strategies [29]. This may be because the degree of metastasis and invasion in a particular EC patient is a good predictive indicator of LNM. The stronger the invasive ability of cancer cells, the greater the likelihood of LNM. What's more, according to the 2014 FIGO staging system, cervical stromal invasion causes the cancer to be classified as stage II and the next step in the progression is local and/or regional spread of the cancer [30].

There were several limitations in the present study. First, the study was conducted retrospectively and some selection bias may have occurred. The second limitation was that several critical prognostic factors, such as LVSI and menopausal status were unavailable in the SEER database. Third, prospective datasets need to be used to externally validate the nomogram developed in this study. Addressing these areas should be the focus of future research.

Declarations

Funding

This study was funded by the grants from the National Natural Science Foundation of China (Grant No 81874108 and 81802607)

Conflicts of Interest

We declare that we have no conflict of interest.

Acknowledgements

We especially appreciate Jiayang Jin, PhD of Peking University People's Hospital Rheumatology department for study design consultations and editing the manuscript.

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. *CA Cancer J Clin.* 2018;68(1):7–30.
2. Chan JK, Kapp DS. Role of complete lymphadenectomy in endometrioid uterine cancer. *Lancet Oncol.* 2007;8(9):831–41.
3. Pecorelli S. Revised FIGO staging for carcinoma of the vulva, cervix, and endometrium. *Int J Gynaecol Obstet.* 2009;105(2):103–4.
4. Morrow CP, Bundy BN, Homesley HD, Creasman WT, Hornback NB, Kurman R, et al. Doxorubicin as an adjuvant following surgery and radiation therapy in patients with high-risk endometrial carcinoma, stage I and occult stage II: a Gynecologic Oncology Group Study. *Gynecol Oncol.* 1990;36(2):166–71.
5. Todo Y, Takeshita S, Okamoto K, Yamashiro K, Kato H. Implications of para-aortic lymph node metastasis in patients with endometrial cancer without pelvic lymph node metastasis. *J Gynecol Oncol.* 2017;28(5):e59.
6. Iasonos A, Schrag D, Raj GV, Panageas KS. How to build and interpret a nomogram for cancer prognosis. *J Clin Oncol.* 2008;26(8):1364–70.
7. Balachandran VP, Gonen M, Smith JJ, DeMatteo RP. Nomograms in oncology: more than meets the eye. *Lancet Oncol.* 2015;16(4):e173–80.
8. Florkowski CM. Sensitivity, specificity, receiver-operating characteristic (ROC) curves and likelihood ratios: communicating the performance of diagnostic tests. *Clin Biochem Rev.* 2008;29(Suppl 1):83–7.
9. Kilts TP, Glaser GE, Langstraat CL, Kumar A, Weaver AL, McGree ME, et al. Comparing risk stratification criteria for predicting lymphatic dissemination in endometrial cancer. *Gynecol Oncol.* 2019;155(1):21–6.
10. Yang B, Shan B, Xue X, Wang H, Shan W, Ning C, et al. Predicting Lymph Node Metastasis in Endometrial Cancer Using Serum CA125 Combined with Immunohistochemical Markers PR and Ki67, and a Comparison with Other Prediction Models. *PLoS One.* 2016;11(5):e0155145.
11. Steyerberg EW, Eijkemans MJ, Harrell FE, Jr, Habbema JD. Prognostic modelling with logistic regression analysis: a comparison of selection and estimation methods in small data sets. *Stat Med.* 2000;19(8):1059–79.
12. Chen W, Zheng R, Baade PD, Zhang S, Zeng H, Bray F, et al. Cancer statistics in China, 2015. *CA Cancer J Clin.* 2016;66(2):115–32.
13. Group AS, Kitchener H, Swart AM, Qian Q, Amos CP, Parmar MK. Efficacy of systematic pelvic lymphadenectomy in endometrial cancer (MRC ASTEC trial): a randomised study. *Lancet.* 2009;373(9658):125–36.
14. Coronado PJ, Rychlik A, Martinez-Maestre MA, Baquedano L, Fasero M, Garcia-Arreza A, et al. Role of lymphadenectomy in intermediate-risk endometrial cancer: a matched-pair study. *J Gynecol Oncol.* 2018;29(1):e1.
15. Venigalla S, Chowdhry AK, Shalowitz DI. Survival implications of staging lymphadenectomy for non-endometrioid endometrial cancers. *Gynecol Oncol.* 2018;149(3):531–8.

16. Vargas R, Rauh-Hain JA, Clemmer J, Clark RM, Goodman A, Growdon WB, et al. Tumor size, depth of invasion, and histologic grade as prognostic factors of lymph node involvement in endometrial cancer: a SEER analysis. *Gynecol Oncol.* 2014;133(2):216–20.
17. Convery PA, Cantrell LA, Di Santo N, Broadwater G, Modesitt SC, Secord AA, et al. Retrospective review of an intraoperative algorithm to predict lymph node metastasis in low-grade endometrial adenocarcinoma. *Gynecol Oncol.* 2011;123(1):65–70.
18. Milam MR, Java J, Walker JL, Metzinger DS, Parker LP, Coleman RL, et al. Nodal metastasis risk in endometrioid endometrial cancer. *Obstet Gynecol.* 2012;119(2 Pt 1):286–92.
19. Shah C, Johnson EB, Everett E, Tamimi H, Greer B, Swisher E, et al. Does size matter? Tumor size and morphology as predictors of nodal status and recurrence in endometrial cancer. *Gynecol Oncol.* 2005;99(3):564–70.
20. Mariani A, Webb MJ, Keeney GL, Haddock MG, Calori GP, Podratz KC. Low-risk corpus cancer: is lymphadenectomy or radiotherapy necessary? *Am J Obstet Gynecol.* 2000;182(6):1506–19.
21. Youden WJ. Index for rating diagnostic tests. *Cancer.* 1950;3(1):32–5.
22. Bendifallah S, Canlorbe G, Laas E, Huguet F, Coutant C, Hudry D, et al. A Predictive Model Using Histopathologic Characteristics of Early-Stage Type 1 Endometrial Cancer to Identify Patients at High Risk for Lymph Node Metastasis. *Ann Surg Oncol.* 2015;22(13):4224–32.
23. Kang S, Nam JH, Bae DS, Kim JW, Kim MH, Chen X, et al. Preoperative assessment of lymph node metastasis in endometrial cancer: A Korean Gynecologic Oncology Group study. *Cancer.* 2017;123(2):263–72.
24. Koskas M, Bassot K, Graesslin O, Aristizabal P, Barranger E, Clavel-Chapelon F, et al. Impact of lymphovascular space invasion on a nomogram for predicting lymph node metastasis in endometrial cancer. *Gynecol Oncol.* 2013;129(2):292–7.
25. AlHilli MM, Podratz KC, Dowdy SC, Bakkum-Gamez JN, Weaver AL, McGree ME, et al. Risk-scoring system for the individualized prediction of lymphatic dissemination in patients with endometrioid endometrial cancer. *Gynecol Oncol.* 2013;131(1):103–8.
26. Boyraz G, Atalay FO, Salman MC, Usubutun A, Erturk A, Gultekin M, et al. Comparison of Mayo and Milwaukee Risk Stratification Models for Predicting Lymph Node Metastasis in Endometrial Cancer. *Int J Gynecol Cancer.* 2018;28(5):869–74.
27. Tuomi T, Pasanen A, Leminen A, Butzow RL, Loukovaara M. Prediction of lymphatic dissemination in endometrioid endometrial cancer: Comparison of three risk-stratification models in a single-institution cohort. *Gynecol Oncol.* 2017;144(3):510–4.
28. Dong Y, Cheng Y, Tian W, Zhang H, Wang Z, Li X, et al. An Externally Validated Nomogram for Predicting Lymph Node Metastasis of Presumed Stage I and II Endometrial Cancer. *Front Oncol* 2019, 9(1218).
29. Pollom EL, Conklin CM, von Eyben R, Folkins AK, Kidd EA. Nomogram to Predict Risk of Lymph Node Metastases in Patients With Endometrioid Endometrial Cancer. *Int J Gynecol Pathol.* 2016;35(5):395–401.

Tables

Table 1 The demographics and pathological characteristics of patients in training and validation cohorts

	Training cohort	Validation cohort
Total number	42558	21278
	Mean±SD	Mean±SD
Age (years)	62.41 ± 11.62	62.41 ± 11.58
Tumor size (cm)	6.58 ± 15.63	6.56 ± 15.62
	N (%)	N (%)
Lymph node metastasis		
Negative	37641 (88.45%)	18874 (88.70%)
Positive	4917 (11.55%)	2404 (11.30%)
Age (years)		
≤40	1484 (3.49%)	740 (3.48%)
41-50	4301 (10.11%)	2152 (10.11%)
51-60	13104 (30.79%)	6550 (30.78%)
61-70	13560 (31.86%)	6782 (31.87%)
≥71	10109 (23.75%)	5054 (23.75%)
Race		
White	35067 (82.40%)	17660 (82.99%)
Black	3757 (8.83%)	1830 (8.60%)
Others	1470 (3.45%)	616 (2.90%)
Unknown	2264 (5.32%)	1172 (5.51%)
Marital status		
Married	22894 (53.79%)	11448 (53.80%)
Single	8194 (19.25%)	4087 (19.21%)
Others	11470 (26.95%)	5743 (26.99%)
Tumor size (cm)		
≤2	10435 (24.52%)	5219 (24.53%)
2-5	21006 (49.36%)	10510 (49.39%)
5-10	8655 (20.34%)	4351 (20.45%)
≥10	2462 (5.79%)	1198 (5.63%)
Histological type		
EEA	37142 (87.27%)	18634 (87.57%)
SEA	3248 (7.63%)	1588 (7.46%)
Others	2168 (5.09%)	1056 (4.96%)
Myometrial invasion		
Negative	27721 (65.10%)	13792 (64.82%)
Positive	14837 (34.90%)	7486 (35.18%)
Cervical stromal invasion		
Negative	33960 (79.80%)	17039 (80.08%)
Positive	8598 (20.20%)	4239 (19.92%)
Tumor grade		
G1	17663 (41.50%)	8975 (42.18%)
G2	12771 (30.01%)	6280 (29.51%)
G3	12124 (28.49%)	6023 (28.31%)

Table 2: Univariate and multivariate analysis of lymph node metastasis in the training cohort

	Univariate analysis		Multivariate analysis	
	OR(95%CI)	P value	OR(95%CI)	P value
Age	1.01 (1.01, 1.01)	<0.001	1.01 (1.01, 1.02)	<0.001
Marital status				
Married	1.0			
Single	1.34 (1.25, 1.45)	<0.001		
Others	1.33 (1.24, 1.42)	<0.001		
Race				
White	1.0			
Black	1.70 (1.51, 1.92)	<0.001		
Others	1.25 (1.11, 1.42)	0.003		
Unknown	0.27 (0.08, 0.84)	0.024		
Tumor size				
≤2cm	1.0		1.0	
2-5cm	2.68 (2.40, 2.99)	<0.001	1.51 (1.34, 1.70)	<0.001
5-10cm	7.65 (6.83, 8.56)	<0.001	2.71 (2.39, 3.06)	<0.001
≥10cm	7.72 (6.73, 8.87)	<0.001	3.38 (2.90, 3.95)	<0.001
Histological type				
EEA	1.0		1.0	
SEA	3.88 (3.57, 4.22)	<0.001	1.78 (1.61, 1.97)	<0.001
Others	3.30 (2.98, 3.66)	<0.001	1.33 (1.18, 1.50)	<0.001
Myometrial invasion				
Negative	1.0		1.0	
Positive	17.24 (13.84, 21.47)	<0.001	2.77 (2.18, 3.52)	<0.001
Cervical stromal invasion				
Negative	1.0		1.0	
Positive	13.49 (12.30, 14.79)	<0.001	6.09 (5.50, 6.76)	<0.001
Tumor grade				
G1	1.0		1.0	
G2	3.41 (3.08, 3.79)	<0.001	2.27 (2.04, 2.53)	<0.001
G3	11.13 (10.13, 12.24)	<0.001	4.68 (4.21, 5.20)	<0.001

OR, odds ratio; 95%CI, 95% confidence interval

Table 3: Univariate and multivariate analysis of lymph node metastasis in the validation cohort

	Univariate analysis		Multivariate analysis	
	OR(95%CI)	P value	OR(95%CI)	P value
Age	1.01 (1.01, 1.02)	<0.001	1.01 (1.01, 1.02)	<0.001
Marital status				
Married	1.0			
Single	1.20 (1.07, 1.34)	0.019		
Others	1.37 (1.24, 1.51)	<0.001		
Race				
White	1.0			
Black	1.82 (1.41, 2.18)	<0.001		
Others	1.17 (1.04, 1.86)	0.005		
Unknown	0.15 (0.06, 0.36)	0.038		
Tumor size				
≤2cm	1.0		1.0	
2-5cm	3.48 (2.93, 4.13)	<0.001	3.48 (2.93, 4.13)	<0.001
5-10cm	9.15 (7.68, 10.90)	<0.001	9.15 (7.68, 10.90)	<0.001
≥10cm	9.58 (7.77, 11.80)	<0.001	9.58 (7.77, 11.80)	<0.001
Histological type				
EEA	1.0		1.0	
SEA	4.09 (3.63, 4.61)	<0.001	4.09 (3.63, 4.61)	<0.001
Others	2.88 (2.47, 3.35)	<0.001	2.88 (2.47, 3.35)	<0.001
Myometrial invasion				
Negative	1.0		1.0	
Positive	19.98 (14.22, 28.08)	<0.001	19.98 (14.22, 28.08)	<0.001
Cervical stromal invasion				
Negative	1.0		1.0	
Positive	12.96 (11.39, 14.76)	<0.001	12.96 (11.39, 14.76)	<0.001
Tumor grade				
G1	1.0		1.0	
G2	2.99 (2.59, 3.45)	<0.001	2.99 (2.59, 3.45)	<0.001
G3	9.96 (8.75, 11.34)	<0.001	9.96 (8.75, 11.34)	<0.001

OR, odds ratio; 95%CI, 95% confidence interval

Table 4: Performance of the nomogram and Mayo criteria scoring system in predicting lymph node metastasis in the training cohort

Performance parameter	Mayo	Nomogram	P (compare)
ROC area(AUC)	0.7163	0.7540	<0.001
95%CI low	0.7092	0.7472	
95%CI high	0.7233	0.7608	
Accuracy	0.7824	0.8104	
Specificity	0.8023	0.8273	
Sensitivity	0.6303	0.6807	
Positive-LR	3.1882	3.9419	
Negative-LR	0.4608	0.3859	
Positive-PV	0.2640	0.3399	
Negative-PV	0.9432	0.9520	

AUC, area under the receiver operating characteristic curve; LR, likelihood ratio; PV, predictive value

Table 5: Performance of the nomogram and Mayo criteria scoring system in predicting lymph node metastasis in the validation cohort

Performance parameter	Mayo	nomogram	<i>P</i> (compare)
ROC area(AUC)	0.7143	0.7511	<0.001
95%CI low	0.7055	0.7413	
95%CI high	0.7256	0.7609	
Accuracy	0.7835	0.8138	
Specificity	0.8034	0.8321	
Sensitivity	0.6277	0.6701	
Positive-LR	3.1925	3.9912	
Negative-LR	0.4634	0.3964	
Positive-PV	0.2591	0.3370	
Negative-PV	0.9443	0.9519	

AUC, area under the receiver operating characteristic curve; LR, likelihood ratio; PV, predictive value

Figures

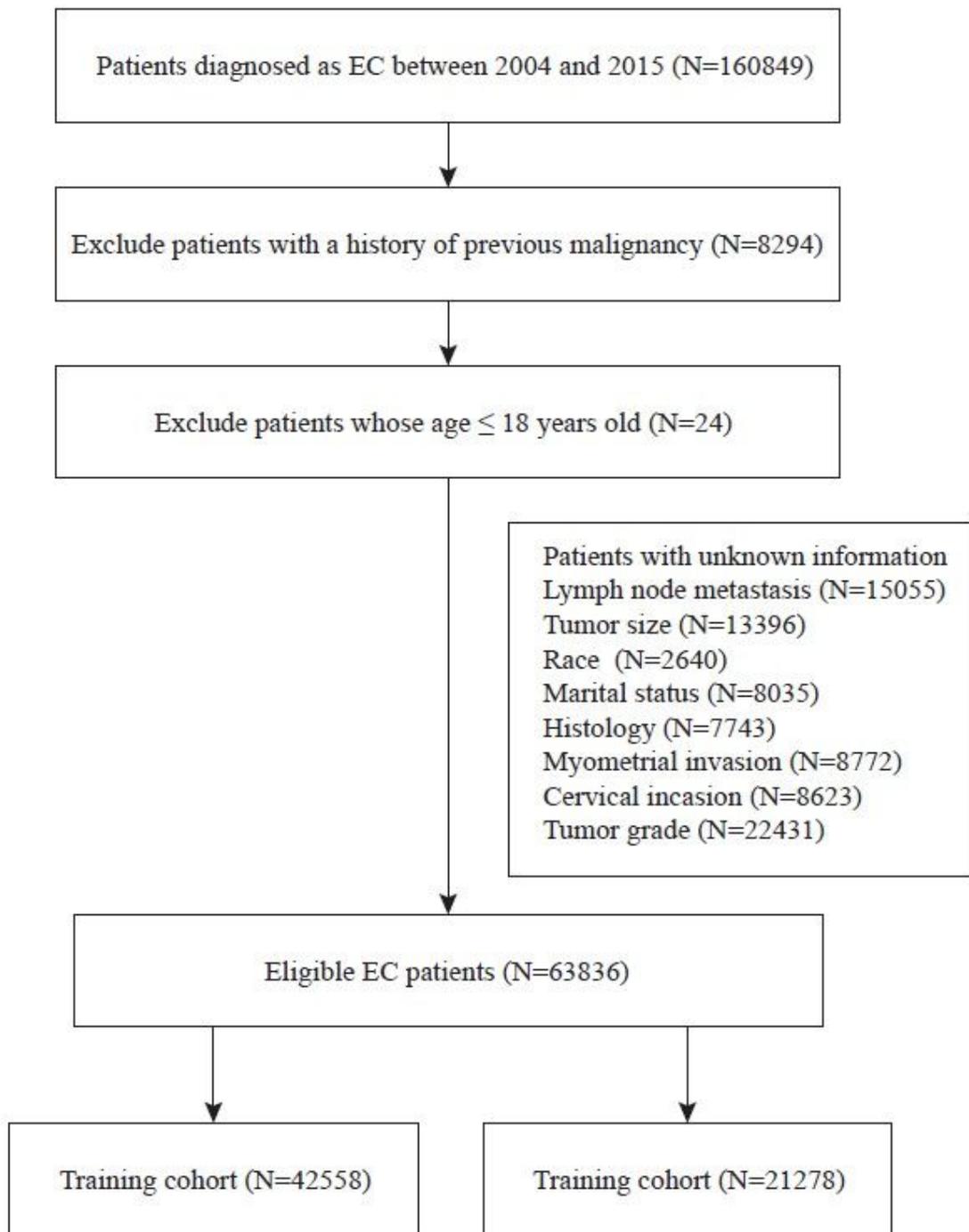


Figure 1

Flow chart for screening eligible patients in training and validation cohort.

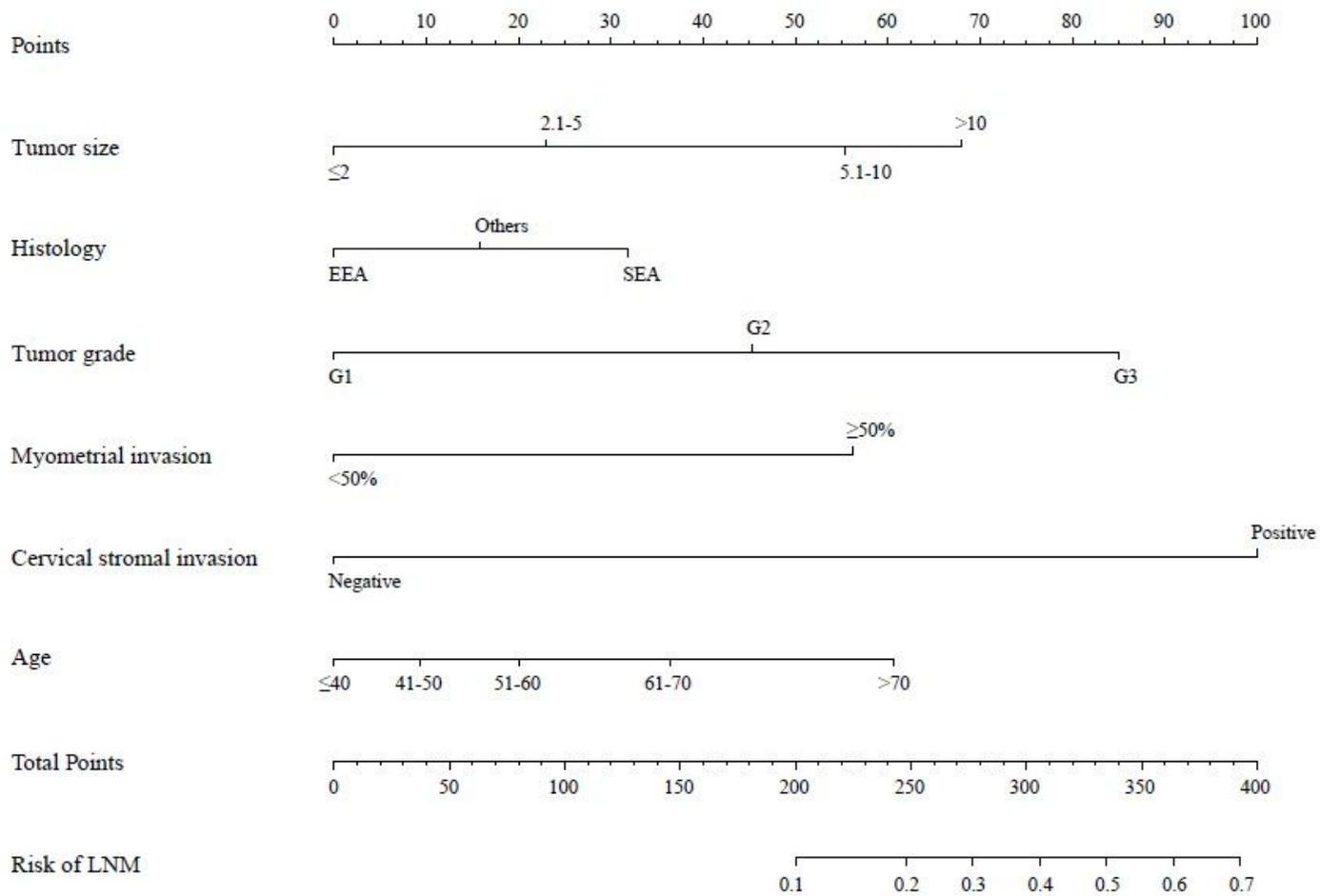


Figure 2

Nomogram to predict lymph node metastasis for EC patients.

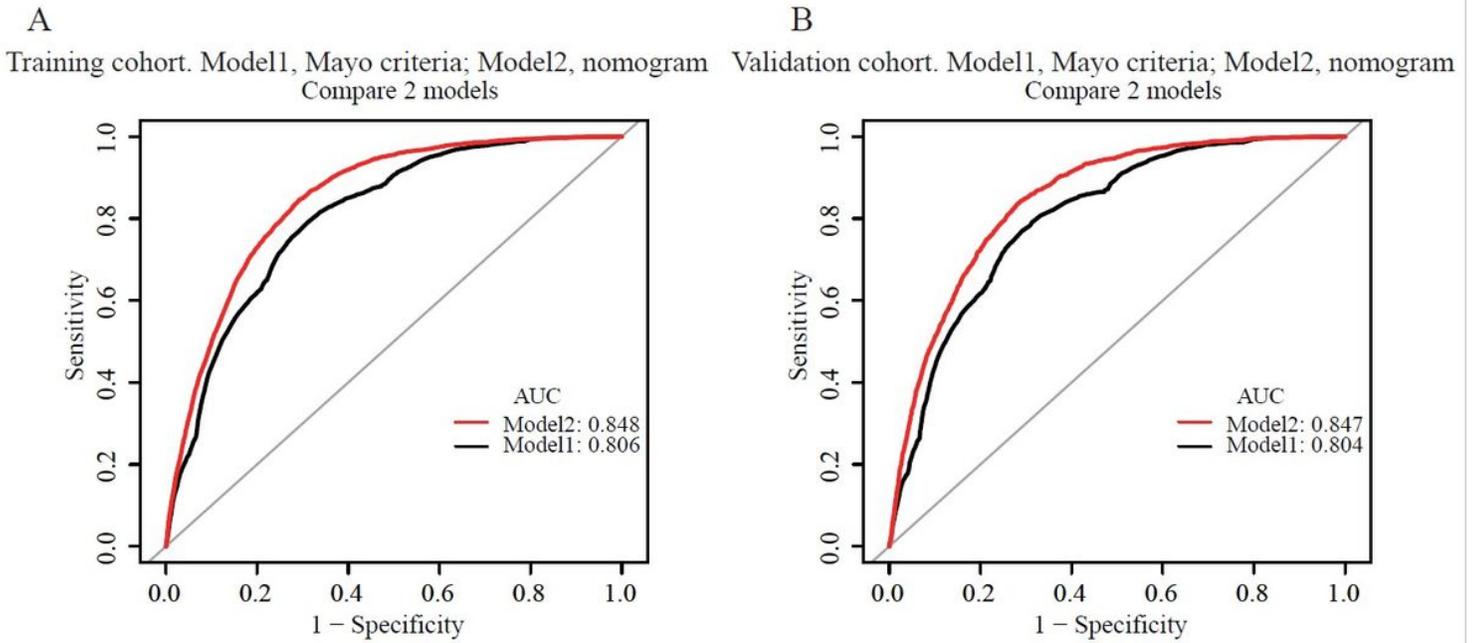


Figure 3

Area under the receiver operating characteristic curves (AUCs) of nomogram and Mayo criteria for the prediction of lymph node metastasis in EC in (A) training cohort and (B) validation cohort.

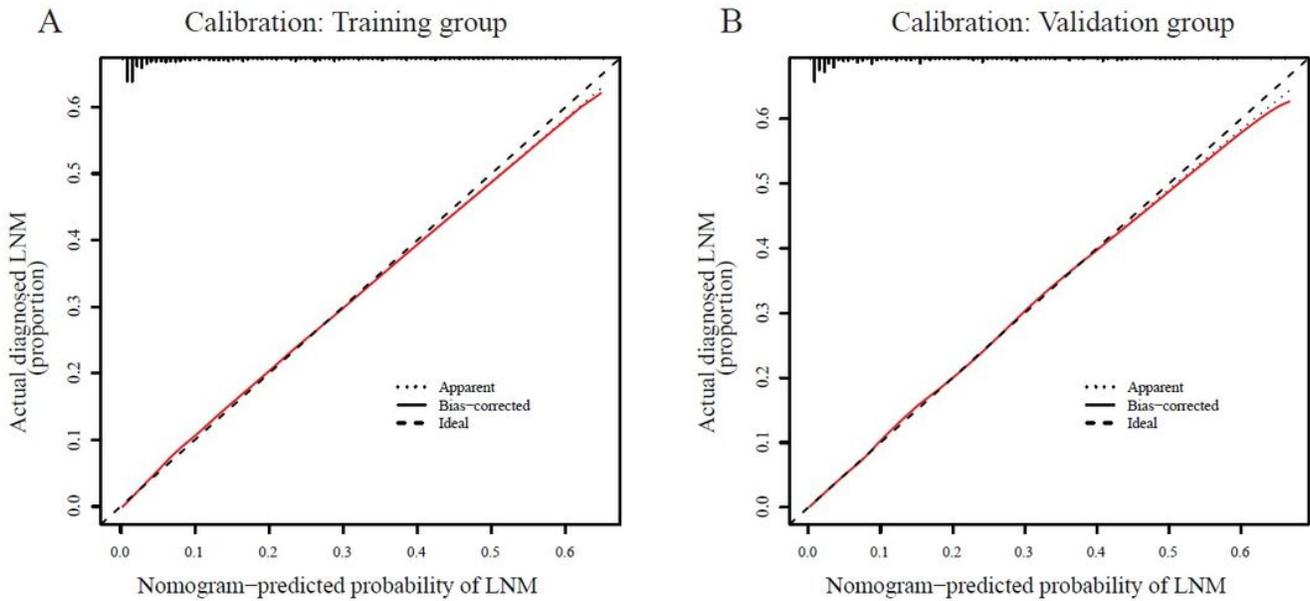


Figure 4

Calibration plots of the nomogram to predict lymph node metastasis in (A) training cohort and (B) validation cohort. In case of perfect calibration, all groups of predicted probabilities fit close to the blue diagnosed line, corresponding to an intercept of 0 and slope of 1 for the calibration plot.

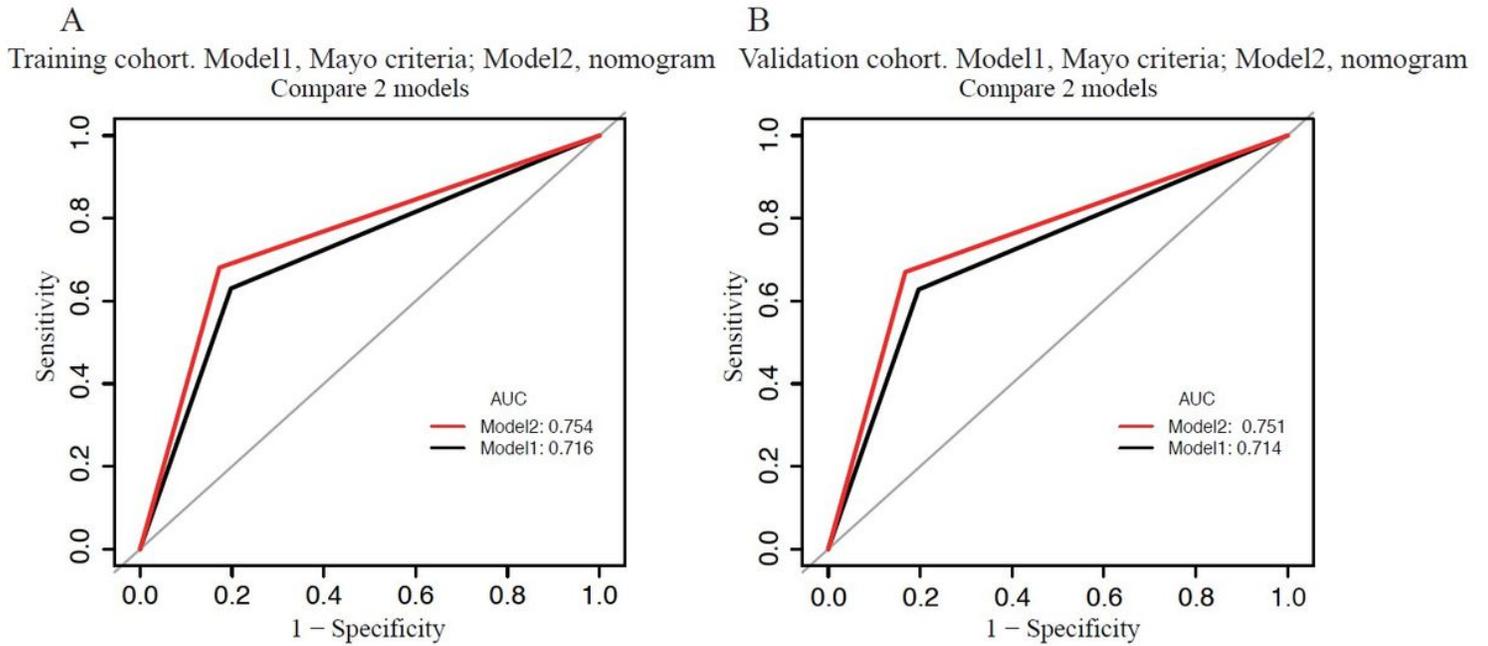


Figure 5

Receiver operating characteristic curve of classification based on nomogram score and Mayo criteria in (A) training cohort and (B) validation cohort.

Supplementary Files

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

- [Figure6.pdf](#)