

Risk Models Based On Populations In High-Risk Areas of Oesophageal Cancer In China For Improving The Oesophageal Cancer Diagnosis Rate

Guodong Yang

Affiliated Hospital of North Sichuan Medical College

Xiaogang Yu

Affiliated Hospital of North Sichuan Medical College

Xiaoqi Long

Suining Central Hospital, Sichuan Province, China

Xiaoming Zhang

North Sichuan Medical College

Weiwei Du

Affiliated Hospital of North Sichuan Medical College

Yingbang Wang

Affiliated Hospital of North Sichuan Medical College

Linyi Wu

Affiliated Hospital of North Sichuan Medical College

Zhao Mu

Affiliated Hospital of North Sichuan Medical College

Yuling Chen

Affiliated Hospital of North Sichuan Medical College

Zhen Li

North Sichuan Medical College

Xiaoying Zhang

North Sichuan Medical College

Ke Pu (✉ lp09162718@163.com)

Affiliated Hospital of North Sichuan Medical College

Research

Keywords: oesophageal cancer, nomogram, cancer screening, diagnostic efficiency

Posted Date: September 21st, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-809951/v1>

Abstract

Background: To improve the diagnostic efficiency of early oesophageal cancer, it is of great significance to develop an effective risk prediction model. This study aimed to identify a high-risk population with oesophageal squamous cell carcinoma (ESCC) based on a population screening model.

Methods: From 120 target townships randomly selected from 150 villages selected in Nanchong City, Sichuan Province, China, from Jan 2016 to Sep 2019, a total of 6409 subjects were screened. Each patient underwent standard endoscopy and narrow band imaging (NBI) and iodine staining indicator biopsies to evaluate oesophageal cancer and precancerous lesions. Before endoscopy, the subjects completed a questionnaire about ESCC risk factors. Variables were evaluated by univariate analysis, and variables significantly related to ESCC were extracted by using a logistic regression model. We used the Akaike information criterion to develop the final model structure and the coding form of variables with multiple metrics. We developed two sets of models to define severe dysplasia and above (SDA) and moderate dysplasia and above (MDA) as prognostic events, respectively.

Results: The areas under the receiver operating characteristic curve (AUROC) were 0.896 (95%CI, 0.888-0.903) and 0.825 (95% CI, 0.816-0.835) for our SDA and MDA models, respectively. MDA-related and SDA-related factors included age, sex, cigarette smoking, alcohol drinking, pharyngeal foreign body sensation, swallowing obstruction, pain behind the sternum, and discomfort behind the breastbone.

Conclusions: we developed an easy-to-use model to identify individuals with high risk of dysplasia or oesophageal cancer in high-risk areas of oesophageal cancer in China.

Introduction

Oesophageal cancer (EC) is one of the most common malignant tumours [1–3] with a poor prognosis and low survival rate [4]. Its incidence is ranked eighth in the world for malignant tumours, and it is the fourth most common type of cancer in China [5–7]. When EC was found, most patients were in the stage of progression and lost the best opportunity for treatment. The 5-year survival rate was less than 25% [8]. Early diagnosis of oesophageal squamous cell carcinoma (ESCC) and precancerous lesions can not only reduce mortality and morbidity, but also improve patients' prognosis [8–10]. It is extremely important to study the early diagnosis of early EC and its precancerous lesions, especially for the screening of high-risk populations in areas with a high incidence of EC [11].

Endoscopy can diagnose EC early and increase the proportion of curable cases of oesophageal malignant tumours by up to 30% [12–14]. However, the cost of endoscopy and the effective improvement of cancer patients' survival are uncertain [15, 16]. With increased awareness of people's health and active education and management of diseases in high-incidence areas of EC, gastrointestinal endoscopy has become more common [17]. China has a large population, and endoscopy screening is not cost-effective and feasible even in areas with a high incidence of EC.

Using a risk prediction model to distinguish high-risk individuals in the general population will be an effective strategy for diagnosing ESCC. At present, there are many studies based on upper gastrointestinal tumour prediction [18, 19]. Previous prediction models are often based on hospital case-control studies, which cannot be

based on large-scale population prediction models and have certain limitations for the prediction of individual risks, and the screening model is based on the upper digestive tract (stomach and oesophagus) [20–22]. The screening model cannot satisfy the individual diagnosis of Chinese patients with EC. Therefore, in order to improve the diagnostic efficiency of early EC, it is of great significance to develop an effective risk prediction model. The development of an effective EC risk prediction model must be based on risk factors closely related to the occurrence and development of EC. The current research shows that the risk factors are mainly age, poor nutritional status, low intake of fruits and vegetables [23], family history of EC[24], high-temperature food intake[25], chemical carcinogen exposure[26], cigarette smoking and alcohol consumption [27], and human papilloma virus infection[28]. However, the prediction model of EC constructed with these risk factors is not effective. Studies have shown that the prediction model based on the symptoms of tumour patients has high efficiency. Therefore, we analysed the symptoms of patients with the aim of developing an EC risk prediction model mainly through a large-scale endoscopic screening test. The goal was to develop a method that maximises the use of all information on age, warning symptoms, and life characteristics and is better applied to the diagnosis of ESCC.

Methods

Study design and population

The research design and study criteria are shown in Figure 1. In January 2016, we launched the China Endoscopic Oesophageal Cancer Screening trial in towns and villages in Nanchong, Sichuan Province, China to evaluate and identify high-risk groups of upper gastrointestinal cancer and high-risk groups that should undergo endoscopy. One hundred twenty target towns were randomly selected, and 6409 subjects were screened. Each patient underwent standard endoscopy and narrow band imaging (NBI) and iodine staining indicator biopsy to evaluate for EC and precancerous lesions. Screening and early diagnosis were conducted for high-risk groups of upper digestive tract cancer. Endoscopic minimally invasive treatment of precancerous lesions and early cancers that could be intervened was performed found in time. Inclusion criteria included (1) age ≥ 40 years, male or female, (2) high incidence of upper gastrointestinal cancer but no gastroscopy in the past 5 years, (3) no gastroscopy contraindications such as severe cardiopulmonary dysfunction, (4) No history of diagnosis and surgery for upper gastrointestinal tumors, and (5) willingness to complete this study. Our prediction model was based on 6409 patients in the screening department from January 2016 to September 2019. Examples of endoscopy were constructed and verified.

Samples and data collection

All patients underwent endoscopy and were requested to stop use of proton pump inhibitors or H₂ receptor antagonists for at least 2 weeks before the examination to avoid disturbing the gastrointestinal malignancies[29]. Endoscopy was performed with ordinary white light, NBI, and Lugol iodine staining. The oesophagus and stomach were examined, and biopsies from all suspected lesions were obtained. If no abnormality was found, a pathological biopsy of the middle of the oesophagus was performed (28-33cm from the incisors at the 6 o'clock position), and then the biopsy tissue was fixed in 10% formaldehyde, embed in paraffin, cut into 5-mm thick sections, and stained with hematoxylin-eosin[19].

Two pathologists from North Sichuan Medical College independently diagnosed the pathological section, and to avoid bias, the pathologists were not aware of the patients' medical history, endoscopy results, pathological outcomes, and predictive factor outcomes. Diagnoses included normal, oesophagitis, basal cell hyperplasia, moderate dysplasia and above (MDA), severe dysplasia and above (SDA), carcinoma in situ (CIS), squamous cell CIS, invasive squamous cell carcinoma, and adenocarcinoma

Before endoscopy, the patients completed a questionnaire about ESCC risk factors. The questionnaire included age, body mass index (BMI), cigarette smoking status, alcohol drinking status, high-temperature food intake, preserved food intake, pharyngeal foreign body sensation, swallowing obstruction, pain behind the sternum, discomfort behind the breastbone, and family history of ESCC (Supplementary materials).

Statistical analysis

Structure determination

First, unconditional univariate logistic regression analysis was performed using the above observation factors for preliminary screening, followed by multivariate logistic regression modelling and unconditional logistic regression analysis to calculate the odds ratio (OR) and 95% confidence interval (CI) of each factor to establish a regression model. The final model structure of the variables was determined using the Akaike information criterion (AIC). Interactions were tested exhaustively in a multivariate model, and no statistically significant differences were found. Because of the strict logical inspection of our survey system, there were no missing data for the variables in the questionnaire.

Model development and performance testing

We developed two sets of models in this study, one for SDA (including lesions of severe dysplasia and higher grade lesions) and the other for MDA (including moderate dysplasia and higher grade lesions), defined as the ending event [30]. In the SDA model, if diagnosed with severe dysplasia, CIS, or ESCC, the patients were classified as case groups. In the MDA model, moderate dysplasia was also regarded as a result event. The discriminant ability between the multivariate prediction model and the simple age model was compared, and the area under the receiver operating characteristic curve (AUC) and the DeLong test were used to evaluate the performance of these models [31]. Leave-one-out cross-validation was adopted to evaluate the generalisation error based on the predicted probabilities of each patient from models built on all the remaining patients.

The data were statistically analysed using SAS9.4 statistical software (SAS Institute), and MDA and SDA were analysed separately. Normally distributed measurement data are expressed as mean \pm standard deviation ($\bar{X} \pm S$) according to the Student t-test. The measurement data of skewed distribution are expressed as median and quartile M (25th quartile, 75th quartile) according to the Mann-Whitney U test. The classification data are expressed as the number of cases and percentile (n (%)) according to the chi-square test or Fisher exact probability method. Meaningful variables were incorporated from single-factor analysis into multivariate logistic regression to explore the risk factors of EC. Medcalc software (Medcalc) was used to draw the receiver operating characteristic (ROC) curve, and R software (The R Foundation) was used to develop the nomogram. $P < 0.05$ was considered statistically significant.

Ethics Statement

This study was approved by the ethics committee of the Affiliated Hospital of North Sichuan Medical College. Written informed consent was obtained from the patients involved in this study or their close relatives. The research data were analysed anonymously, and personal identifiers were removed.

Results

Patients' characteristics

Table 1 shows patients' demographic characteristics. From 2016 to 2019, 6409 patients were included, and the average age was 59.70 ± 7.94 years (maximum age, 78 years; minimum age, 40 years). Because of medical restrictions in China, patients' family history of ESCC was unknown and was not included in the data.

Model development and validation

Age, cigarette smoking, alcohol drinking, pharyngeal foreign body sensation, swallowing obstruction, pain behind the sternum, and discomfort behind the breastbone were included into multivariate logistic regression (stepwise regression method). The results showed that age, cigarette smoking, alcohol drinking, pharyngeal foreign body sensation, swallowing obstruction, pain behind the sternum, and discomfort behind the breastbone were risk factors of SDA. According to AIC, we first constructed the SDA model for the joint prediction model and compared it with the simple age model and simple BMI model. The ROC curve showed that age plays a major role in predicting high-grade oesophageal lesions, and by including all other risk factors and alarm symptoms, the prediction model was significantly better than other simple predictors ($AUC_{\text{FULL model for SDA}}=0.896$ versus [vs.] $AUC_{\text{Simple age model for SDA}}=0.634$, $P<0.001$; $AUC_{\text{Full model for SDA}}=0.896$ vs $AUC_{\text{Simple BMI model for SDA}}=0.527$, $P<0.001$) (Figure 2). The sensitivity of the joint prediction model (83.5%) was higher than the sensitivities of the simple age model (76%) and simple BMI model (30.4%) (Figure 3).

In addition to the SDA model, we constructed a model to predict MDA lesions in this population. Age, sex, cigarette smoking, alcohol drinking, pharyngeal foreign body sensation, swallowing obstruction, pain behind the sternum, and discomfort behind the breastbone were also risk factors of MDA (Tables S1 and S2). In the SDA and MDA models, pharyngeal foreign body sensation, swallowing obstruction, pain behind the sternum, and discomfort behind the breastbone had a high correlation with SDA and MDA.

We tried to determine whether the SDA model could also distinguish patients with moderate dysplasia, and the results showed that its ability to predict moderate dysplasia was moderate (Figure S1A, S1B). We developed a nomogram based on the structure of our model (Figure 4). By using this nomogram, patients can quickly determine the risk rate of oesophageal squamous cell carcinoma according to their own level of risk factors, and it can improve the detection rate of screening and remind patients with risk factors of EC to correct bad habits (e.g., smoking and drinking).

Discussion

We developed a risk prediction model of ESCC and its precancerous lesions that has the potential to be applied in secondary prevention projects. It can help patients make preliminary judgements, and it can improve the

detection rate of ESCC and precancerous lesions. Our results provide important information for further clarifying the impact of clinical symptoms on early warning and prognosis of EC.

Kunzmann et al. [3] established a predictive model of oesophageal adenocarcinoma based on life risk factors through a prospective study. Age, sex, smoking, BMI, and history of oesophageal treatment were included in the final model (AUC = 0.80, 95%CI 0.77–0.82) used to predict patients with a higher absolute risk of oesophageal adenocarcinoma within 5 years. Ireland et al. [32] compared 120 Barrett oesophagus patients with 235 healthy individuals by age, sex, history of reflux, family history of reflux, history of hypertension, weekly alcoholic beverages, and BMI, and found significant differences; they established a Barrett oesophagus risk prediction model for the Australian population. The model's AUC was 0.82(95% CI 0.78–0.87), showing good discrimination.

In this study, the SDA and MDA models included the same risk factors age, cigarette smoking, alcohol drinking, pharyngeal foreign body sensation, swallowing obstruction, pain behind the sternum, and discomfort behind the breastbone. The MDA model's AUC indicated that it had high discrimination ability. Other prediction models have been based on patients' living habits and living environment as predictive factors, such as pesticide exposure and use coal or wood as the main cooking fuel[30]. During the questionnaire, subjects could not accurately select answers, and because of economic and medical restrictions in China, the prevalence of a family history of EC in patients older than 60 years of age cannot be accurately informed, making it difficult to apply prediction models. Our research model predictors can accurately reflect the patients and be used easily in screening practise. Compared with the hospital-based case-control design model, the main advantage of this study is that our prediction model was based on a large population-level screening procedure. Patients who participated in the screening were randomly selected from Nanchong City, Sichuan Province, a high-incidence area of EC. They were able to describe the distribution of EC and precancerous lesions in the general population. We used screening cases including severe dysplasia, CIS, and ESCC as the result of model construction. Early malignant lesions are the target of ESCC screening work [33]. Compared with the model established by the case-control study with advanced ESCC as the outcome event, our model is more suitable for early detection of ESCC and its malignant lesions.

The present study showed that SDA and MDA are significantly related to smoking and drinking. Compared with non-smokers, smokers have a significantly increased risk of SDA and MDA, which is consistent with findings of previous case-control and cohort studies [34, 35]. Drinking alcohol produces carcinogenic metabolites, and heavy drinking especially increases the accumulation of such metabolites in the body, which in turn increases the risk of ESCC [36, 37].

The characteristics and criteria of early symptoms of EC and further clarifying the relationship between these symptoms and the course and pathological characteristics of EC are key to determining whether these symptoms can be used for early warning of EC. This strategy is also one of the most effective, low-cost, non-invasive, and easily accepted and popularised EC prevention and control methods in the absence of effective molecular markers for early warning and diagnosis of EC. Western literature reported that the main early warning symptoms of oesophageal adenocarcinoma are swallowing/choking [38]. Current research on symptoms and tumours, lack of symptoms, and lack of awareness of the risk of alarm symptoms are risk factors for poor tumour prognosis [2]. The most common presenting symptoms of EC are dysphagia and weight loss. Other symptoms include odynophagia, upper gastrointestinal bleeding, hoarseness, and respiratory

symptoms[39].Based on large-scale screening tests, we summarised the warning symptoms of early oesophageal tumours. Although the value of each alarm symptom was relatively low, it may be possible to avoid unnecessary endoscopy by using various symptoms combined with patient risk factors to make a comprehensive judgement.

Some previous studies have also evaluated the value of age and warning symptoms in predicting cancer risk in patients with dyspepsia [20–22].These studies used age and warning symptoms to predict upper gastrointestinal cancer. Bai and colleagues studied the predictive value of alarm symptoms and age for upper gastrointestinal malignancies in China, and found that age or any alarm symptoms have limited value. In their study, alarm symptoms were highly specific, but the sensitivity was low. However, most of their discussions were based on the positive diagnostic likelihood ratio for each symptom, and they did not use all predictors to build models to predict the risk limits of upper gastrointestinal malignancies [21], Fransen[40] and colleagues conducted a meta-analysis and found a limited diagnostic value for each alarm symptom. They recommended that the alarm symptoms may be related to other factors. Combined use may be a better tool for selecting high-risk patients. However, they could not test their hypothesis. Numans and colleagues[20] used the calculated total score to establish a risk prediction model and showed that the classic warning symptoms through the risk prediction model are useful predictors of upper gastrointestinal malignancies. However, their model is somewhat complicated and contains multiple variables, so it is somewhat unstable. The above studies all predicted upper gastrointestinal tumours, which is not practical for high incidence areas of EC in China. In our risk prediction model, various symptoms and patients' life habits can be used to diagnose EC.

Disease risk prediction models can provide personalised estimates of ESCC risk based on personal baseline data [41]. The use of risk prediction models to distinguish high-risk groups from the general population will help formulate highly effective ESCC prevention and treatment strategies [42]. The first method of using this model is that when the resources for EC are high and endoscopic screening is limited, it is urgent to identify more patients with early malignant lesions, such as the national population with ESCC in some high-risk areas of EC in China's screening plan [43]. A nomogram can be used as a predictor to assess the risk > 0.8 , which can maximise the detection rate of malignant lesions. Compared with general screening, the detection rate of SDA and MDA lesions will increase. The second method is under the conditions of unlimited resources, such as the environment that usually exists in clinical and scientific research. Therefore, it is possible to select subjects with a risk > 0.5 for screening to ensure high sensitivity.

This study has a couple limitations. Although our study screened more than 6409 participants, the study population was still limited, and the screening population was divided according to whether patients participated in screening, causing biased selection. Multi-centre external verification and calibration are required in future studies. Because of regional differences in EC, validation studies conducted in other populations are also crucial for the generalisation of the model. Prospective collection of more malignant events and dynamic observation of exposure to predictive factors are also essential to improving these models. In conclusions, our EC high-risk population screening model constructed based on screening population information has good discriminatory and calibration capabilities, and it can be used to assess the risk of local EC. Screening high-risk populations may make EC screening more cost-effective. Additionally, the model can be applied to other Asian populations with similar socioeconomic or lifestyle characteristics, who may also benefit from our ESCC risk prediction model.

Abbreviations

AIC, Akaike information criterion; AUC, area under the receiver operating characteristic curve; BMI, body mass index; OR, odds ratio; CI, confidence interval; CIS, carcinoma in situ; ESCC, oesophageal squamous cell carcinoma; MDA, moderate dysplasia and above; ROC, receiver operating characteristic; SDA, severe dysplasia and above; NBI, narrow band imaging.

Declarations

Acknowledgements

None

Contributions

Guodong Yang and Xiaogang Yu contributed equally to this work. Xiaoqi Long carried out the statistical analysis and drafted the manuscript. Xiaogang Yu participated in the study design and coordination and helped to draft the manuscript. Weiwei Du, Yingbang Wang, Linyi Wu, Zhao Mu, Yuling Chen, Xiaoming Zhang and Zhen Li contributed the acquisition of clinical data. Xiaoying Zhang and Ke Pu helped the statistical analysis. All authors read and approved the final manuscript

Funding

This study was supported by the Sichuan provincial Health and Family Planning Commission Foundation (18PJ386) and Construction of Key Experimental Platform for Pancreatic Diseases in Sichuan Province(18SXHZ0371)

Availability of data and materials

The datasets used and/or analyzed during the present study are available from the corresponding author on reasonable request.

Ethics approval and consent to participate

This study was approved by the Ethical Committee of University Hospital Rostock, and the written informed consent was collected from each participant.

Consent for publication

Not applicable

Conflict of interest

The authors have no conflict of interest to declare

References

1. Hua Z, Zheng X, Xue H, Wang J, Yao J. Long-term trends and survival analysis of esophageal and gastric cancer in Yangzhong, 1991–2013. *PLoS One*. 2017;12:e0173896.
2. Dregan A, Møller H, Charlton J, Gulliford MC. Are alarm symptoms predictive of cancer survival?: population-based cohort study. *Br J Gen Pract*. 2013;63:e807–12.
3. Kunzmann AT, Thrift AP, Cardwell CR, Lagergren J, Xie S, Johnston BT, Anderson LA, Busby J, McMenamin ÚC, Spence AD, Coleman HG. Model for Identifying Individuals at Risk for Esophageal Adenocarcinoma. *Clin Gastroenterol Hepatol*. 2018;16:1229–36.e4.
4. Tang WR, Fang JY, Wu KS, Shi XJ, Luo JY, Lin K. Epidemiological characteristics and prediction of esophageal cancer mortality in China from 1991 to 2012. *Asian Pac J Cancer Prev*. 2014;15:6929–34.
5. Hur C, Choi SE, Kong CY, Wang GQ, Xu H, Polydorides AD, Xue LY, Perzan KE, Tramontano AC, Richards-Kortum RR, Anandasabapathy S. High-resolution microendoscopy for esophageal cancer screening in China: A cost-effectiveness analysis. *World J Gastroenterol*. 2015;21:5513–23.
6. Chen XX, Zhong Q, Liu Y, Yan SM, Chen ZH, Jin SZ, Xia TL, Li RY, Zhou AJ, Su Z, Huang YH, Huang QT, Huang LY, Zhang X, Zhao YN, Yun JP, Wu QL, Lin DX, Bai F, Zeng MS. Genomic comparison of esophageal squamous cell carcinoma and its precursor lesions by multi-region whole-exome sequencing. *Nat Commun*. 2017;8:524.
7. Sun X, Zhao D, Liu Y, Liu Y, Yuan Z, Wang J, Xue F. The long-term spatial-temporal trends and burden of esophageal cancer in one high-risk area: A population-registered study in Feicheng, China. *PLoS One*. 2017;12:e0173211.
8. Otterstatter MC, Brierley JD, De P, Ellison LF, Macintyre M, Marrett LD, Semenciw R, Weir HK. Esophageal cancer in Canada: trends according to morphology and anatomical location. *Can J Gastroenterol*. 2012;26:723–7.
9. Shimamura Y, Ikeya T, Marcon N, Mosko JD. Endoscopic diagnosis and treatment of early esophageal squamous neoplasia. *World J Gastrointest Endosc*. 2017;9:438–47.
10. Liang H, Fan JH, Qiao YL. Epidemiology, etiology, and prevention of esophageal squamous cell carcinoma in China. *Cancer Biol Med*. 2017;14:33–41.
11. Malhotra GK, Yanala U, Ravipati A, Follet M, Vijayakumar M, Are C. Global trends in esophageal cancer. *J Surg Oncol*. 2017;115:564–79.
12. Dickerson LM, King DE. Evaluation and management of nonulcer dyspepsia. *Am Fam Physician*. 2004;70:107–14.
13. Axon AT, Bell GD, Jones RH, Quine MA, McCloy RF. Guidelines on appropriate indications for upper gastrointestinal endoscopy. Working Party of the Joint Committee of the Royal College of Physicians of London, Royal College of Surgeons of England, Royal College of Anaesthetists, Association of Surgeons, the British Society of Gastroenterology, and the Thoracic Society of Great Britain, *BMJ*. 310(1995) 853–856.
14. Jansen M, Banks M. Early detection and risk stratification of gastric cancer are likely to be refined with biopsies targeted through high-resolution-enhanced imaging. *Gut*. 2020;69:1.
15. Kumagai K, Nakatsu M, Liu SH, Haba T, Takizawa N, Gonda H, Urabe M, Hayashida Y, Takezoe K, Maekawa K. [Detection of resection line in early gastric cancer]. *Gan No Rinsho*. 1984;30:1379–83.
16. Delaney BC, Wilson S, Roalfe A, Roberts L, Redman V, Wearn A, Briggs A, Hobbs FD. Cost effectiveness of initial endoscopy for dyspepsia in patients over age 50 years: a randomised controlled trial in primary care.

Lancet. 2000;356:1965–9.

17. Tan L, Feng J, Zhao Q, Chen P, Yang G. Preoperative endoscopic titanium clip placement facilitates intraoperative localization of early-stage esophageal cancer or severe dysplasia. *World J Surg Oncol*. 2017;15:145.
18. Chang J, Huang Y, Wei L, Ma B, Miao X, Li Y, Hu Z, Yu D, Jia W, Liu Y, Tan W, He Z, Ke Y, Wu T, Shen H, Zeng Y, Wu C, Lin D. Risk prediction of esophageal squamous-cell carcinoma with common genetic variants and lifestyle factors in Chinese population. *Carcinogenesis*. 2013;34:1782–6.
19. He Z, Zhao Y, Guo C, Liu Y, Sun M, Liu F, Wang X, Guo F, Chen K, Gao L, Ning T, Pan Y, Li Y, Zhang S, Lu C, Wang Z, Cai H, Ke Y. Prevalence and risk factors for esophageal squamous cell cancer and precursor lesions in Anyang, China: a population-based endoscopic survey. *Br J Cancer*. 2010;103:1085–8.
20. Numans ME, van der Graaf Y, de Wit NJ, de Melker RA. How useful is selection based on alarm symptoms in requesting gastroscopy? An evaluation of diagnostic determinants for gastro-oesophageal malignancy. *Scand J Gastroenterol*. 2001;36:437–43.
21. Bai Y, Li ZS, Zou DW, Wu RP, Yao YZ, Jin ZD, Ye P, Li SD, Zhang WJ, Du YQ, Zhan XB, Liu F, Gao J, Xu GM. Alarm features and age for predicting upper gastrointestinal malignancy in Chinese patients with dyspepsia with high background prevalence of *Helicobacter pylori* infection and upper gastrointestinal malignancy: an endoscopic database review of 102,665 patients from 1996 to 2006. *Gut*. 59(2010) 722–728.
22. Kapoor N, Bassi A, Sturgess R, Bodger K. Predictive value of alarm features in a rapid access upper gastrointestinal cancer service. *Gut*. 2005;54:40–5.
23. Tran GD, Sun XD, Abnet CC, Fan JH, Dawsey SM, Dong ZW, Mark SD, Qiao YL, Taylor PR. Prospective study of risk factors for esophageal and gastric cancers in the Linxian general population trial cohort in China. *Int J Cancer*. 2005;113:456–63.
24. Wei WQ, Abnet CC, Lu N, Roth MJ, Wang GQ, Dye BA, Dong ZW, Taylor PR, Albert P, Qiao YL, Dawsey SM. Risk factors for oesophageal squamous dysplasia in adult inhabitants of a high risk region of China. *Gut*. 54(2005) 759–763.
25. Islami F, Boffetta P, Ren JS, Pedoeim L, Khatib D, Kamangar F. High-temperature beverages and foods and esophageal cancer risk—a systematic review. *Int J Cancer*. 2009;125:491–524.
26. Roth MJ, Strickland KL, Wang GQ, Rothman N, Greenberg A, Dawsey SM. High levels of carcinogenic polycyclic aromatic hydrocarbons present within food from Linxian, China may contribute to that region's high incidence of oesophageal cancer. *Eur J Cancer*. 1998;34:757–8.
27. Fan Y, Yuan JM, Wang R, Gao YT, Yu MC. Alcohol, tobacco, and diet in relation to esophageal cancer: the Shanghai Cohort Study. *Nutr. Cancer*. 2008;60:354–63.
28. Guo F, Liu Y, Wang X, He Z, Weiss NS, Madeleine MM, Liu F, Tian X, Song Y, Pan Y, Ning T, Yang H, Shi X, Lu C, Cai H, Ke Y. Human papillomavirus infection and esophageal squamous cell carcinoma: a case-control study. *Cancer Epidemiol. Biomarkers Prev*. 2012;21:780–5.
29. Bramble MG, Suvakovic Z, Hungin AP. Detection of upper gastrointestinal cancer in patients taking antisecretory therapy prior to gastroscopy. *Gut*. 2000;46:464–7.
30. Liu M, Liu Z, Cai H, Guo C, Li X, Zhang C, Wang H, Hang D, Liu F, Deng Q, Yang X, Yuan W, Pan Y, Li J, Zhang C, Shen N, He Z, Ke Y. A Model To Identify Individuals at High Risk for Esophageal Squamous Cell

- Carcinoma and Precancerous Lesions in Regions of High Prevalence in China. *Clin Gastroenterol Hepatol*. 2017;15:1538–46.e7.
31. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach, *Biometrics*44(1988) 837–845.
 32. [French comment on. article A deep neural network improves endoscopic detection of early gastric cancer without blind spots]. *Endoscopy*. 2019;51:608–9.
 33. Roshandel G, Nourouzi A, Pourshams A, Semnani S, Merat S, Khoshnia M. Endoscopic screening for esophageal squamous cell carcinoma. *Arch Iran Med*. 2013;16:351–7.
 34. Ishikawa A, Kuriyama S, Tsubono Y, Fukao A, Takahashi H, Tachiya H, Tsuji I. Smoking, alcohol drinking, green tea consumption and the risk of esophageal cancer in Japanese men. *J Epidemiol*. 2006;16:185–92.
 35. Huxley R, Jamrozik K, Lam TH, Barzi F, Ansary-Moghaddam A, Jiang CQ, Suh I, Woodward M. Asia Pacific Cohort Studies Collaboration, Impact of smoking and smoking cessation on lung cancer mortality in the Asia-Pacific region. *Am J Epidemiol*. 2007;165:1280–6.
 36. Dong J, Thrift AP. Alcohol, smoking and risk of oesophago-gastric cancer, *Best Pract. Res Clin Gastroenterol*. 2017;31:509–17.
 37. Tramacere I, Pelucchi C, Bagnardi V, Rota M, Scotti L, Islami F, Corrao G, Boffetta P, La Vecchia C, Negri E. A meta-analysis on alcohol drinking and esophageal and gastric cardia adenocarcinoma risk. *Ann Oncol*. 2012;23:287–97.
 38. Grotenhuis BA, van Hagen P, Wijnhoven BP, Spaander MC, Tilanus HW, van Lanschot JJ. Delay in diagnostic workup and treatment of esophageal cancer. *J Gastrointest Surg*. 2010;14:476–83.
 39. Cortortntestagnosti,Villaseñor CalocaR. [Esophageal cancer]. *Rev Gastroenterol Mex*. 1997;62:149–59.
 40. Fransen GA, Janssen MJ, Muris JW, Laheij RJ, Jansen JB. Meta-analysis: the diagnostic value of alarm symptoms for upper gastrointestinal malignancy. *Aliment Pharmacol Ther*. 2004;20:1045–52.
 41. Moons KG, Altman DG, Reitsma JB, Ioannidis JP, Macaskill P, Steyerberg EW, Vickers AJ, Ransohoff DF, Collins GS. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med*. 2015;162:W1–73.
 42. Zhao P, Dai M, Chen W, Li N. Cancer trends in China. *Jpn J Clin Oncol*. 2010;40:281–5.
 43. Wynants L, Collins GS, Van Calster B. Key steps and common pitfalls in developing and validating risk models,*BJOG*.124(2017) 423–432.

Tables

Table 1
Patients'baseline characteristics

Baseline	Description n(%)
Gender	
Male	3127(48.79)
Female	3282(51.21)
Age	
≤ 60	3179(49.60)
> 60	3230(50.40)
BMI	
< 18.5	154(2.40)
18.5–23.9	3251(50.73)
≥ 23.9	3004(46.87)
MDA	
No	6102(95.21)
Yes	307(4.79)
SDA	
No	6330(98.77)
Yes	79(1.23)
Abbreviation: BMI, body mass index; MDA, moderate dysplasia and above; SDA, severe dysplasia and above	

Table 2
SDA univariate coefficients analysis

variable	Total	SDA		Statistics	P
		- (n = 6330)	+ (n = 79)		
Sex n(%)				$\chi^2 = 2.137$	0.144
Male	3127(48.79)	3082(48.69)	45(56.96)		
Female	3282(51.21)	3248(51.31)	34(43.04)		
Age n(%)				$\chi^2 = 11.823$	< 0.001
≤ 60	3179(49.60)	3155(49.84)	24(30.38)		
> 60	3230(50.40)	3175(50.16)	55(69.62)		
BMI n(%)				Z=-0.455	0.649
< 18.5	154(2.40)	152(2.40)	2(2.53)		
18.5–23.9	3251(50.73)	3209(50.70)	42(53.16)		
≥ 23.9	3004(46.87)	2969(46.90)	35(44.30)		
Cigarette smoker n(%)				-	< 0.001
Never	5738(89.53)	5698(90.02)	40(50.63)		
< 20 pys	479(7.47)	463(7.31)	16(20.25)		
≥ 20 pys	192(3.00)	169(2.67)	23(29.11)		
Alcohol drinker n(%)				-	< 0.001
Never	5580(87.07)	5547(87.63)	33(41.77)		
100g/d	524(8.18)	497(7.85)	27(34.18)		
> 100g/d	305(4.76)	286(4.52)	19(24.05)		
high-temperature food intake n(%)				$\chi^2 = 1.642$	0.440
Never	1753(27.35)	1734(27.39)	19(24.05)		
< 2 Times/week	1866(29.12)	1846(29.16)	20(25.32)		
≥ 2 Times/week	2790(43.53)	2750(43.44)	40(50.63)		
Preserved food n(%)				$\chi^2 = 3.272$	0.195

BMI, body mass index; SDA, severe dysplasia and above

variable	Total	SDA		Statistics	P
		- (n = 6330)	+ (n = 79)		
Never	1560(24.34)	1543(24.38)	17(21.52)		
< 2 Times/week	2233(34.84)	2211(34.93)	22(27.85)		
≥ 2 Times/week	2616(40.82)	2576(40.70)	40(50.63)		
Pharyngeal foreign body sensation n(%)				-	< 0.001
No	6284(98.05)	6217(98.21)	67(84.81)		
Yes	125(1.95)	113(1.79)	12(15.19)		
Swallowing obstruction n(%)				-	< 0.001
No	6302(98.33)	6244(98.64)	58(73.42)		
Yes	107(1.67)	86(1.36)	21(26.58)		
Discomfort behind the breastbone n(%)				-	< 0.001
No	6290(98.14)	6221(98.28)	69(87.34)		
Yes	119(1.86)	109(1.72)	10(12.66)		
Pain behind the sternum n(%)				-	< 0.001
No	6299(98.28)	6231(98.44)	68(86.08)		
Yes	110(1.72)	99(1.56)	11(13.92)		
BMI, body mass index; SDA, severe dysplasia and above					

Table 3
Multivariate coefficient analysis of SDA

Feature	β	S.E	Wald	P	OR	95%CI	
						Lower Limit	Upper limit
Constant	-6.021	0.262	529.656	< 0.001			
Age							
≤ 60	Ref						
> 60	0.615	0.267	5.307	0.021	1.849	1.096	3.119
Cigarette smoker							
Never	Ref						
< 20 pys	0.963	0.359	7.193	0.007	2.620	1.296	5.297
≥ 20 pys	1.887	0.356	28.033	< 0.001	6.601	3.282	13.274
Alcohol drinker							
Never	Ref						
100g/d	1.233	0.339	13.237	< 0.001	3.431	1.766	6.666
> 100g/d	1.084	0.372	8.481	0.004	2.957	1.426	6.135
pharyngeal foreign body sensation							
No	Ref						
Yes	1.712	0.375	20.819	< 0.001	5.539	2.655	11.557
swallowing obstruction							
No	Ref						
Yes	2.297	0.312	54.319	< 0.001	9.944	5.399	18.318
discomfort behind the breastbone							
No	Ref						
Yes	1.765	0.389	20.561	< 0.001	5.844	2.725	12.534
pain behind the sternum							

BMI, body mass index; SDA, severe dysplasia and above; OR, odds ratio; CI, confidence interval; SE, standard error; Ref, reference

Feature	β	S.E	Wald	P	OR	95%CI	
						Lower Limit	Upper limit
No	Ref						
Yes	2.174	0.389	31.244	< 0.001	8.796	4.104	18.855

BMI, body mass index; SDA, severe dysplasia and above; OR, odds ratio; CI, confidence interval; SE, standard error; Ref, reference

Table 4

Comparison of ROC curves between the SDA joint prediction model and simple age and simple BMI prediction models

Variable	AUC(95%CI)	SE	Sensitivity (95%CI)	Specificity (95%CI)	Youden Index	Cut off value	DeLong-test	
							Z	P-value
Joint prediction model	0.896(0.888–0.903)	0.021	0.835(0.735–0.909)	0.867(0.859–0.876)	0.703	0.013		
Simple age prediction model	0.634(0.622–0.646)	0.030	0.760(0.650–0.849)	0.466(0.454–0.478)	0.225	58	7.648	< 0.001
Simple BMI prediction model	0.527(0.515–0.540)	0.034	0.304(0.205–0.418)	0.764(0.753–0.774)	0.067	21	2.481	< 0.001

BMI, body mass index; SDA, severe dysplasia and above; CI, confidence interval; SE, standard error; AUC, area under the receiver operating characteristic curve; ROC, receiver operating characteristic

Figures

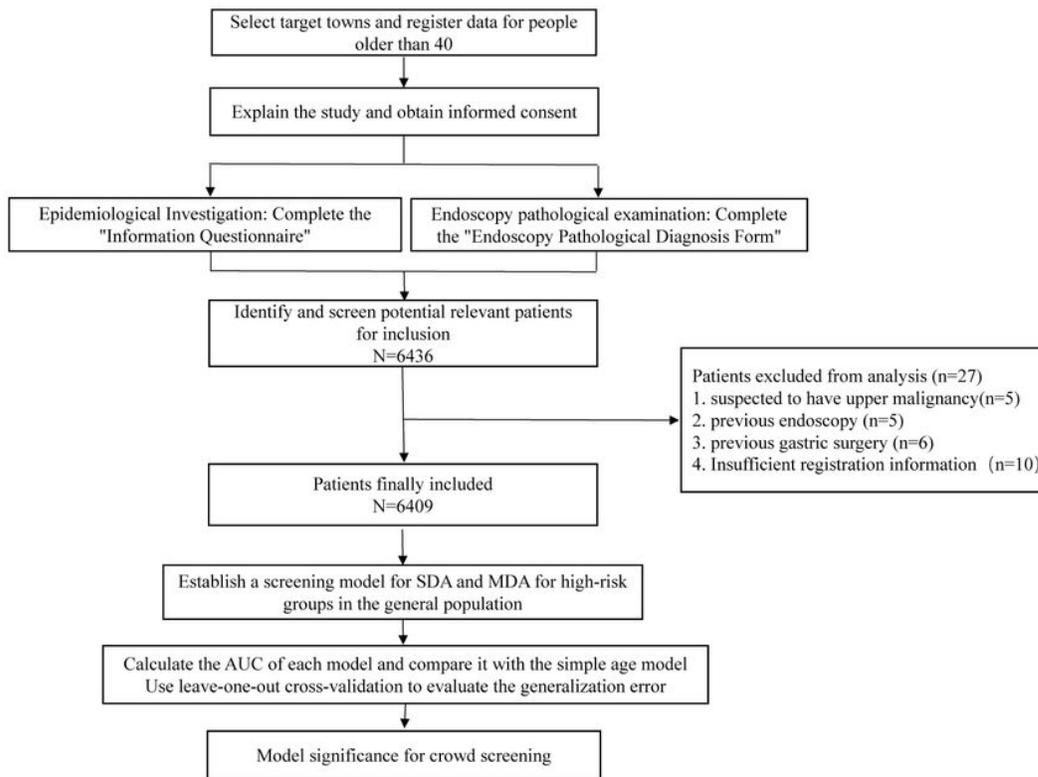


Figure 1

Research flow chart MDA, moderate dysplasia and above; SDA, severe dysplasia and above; AUC, area under the receiver operating characteristic curve

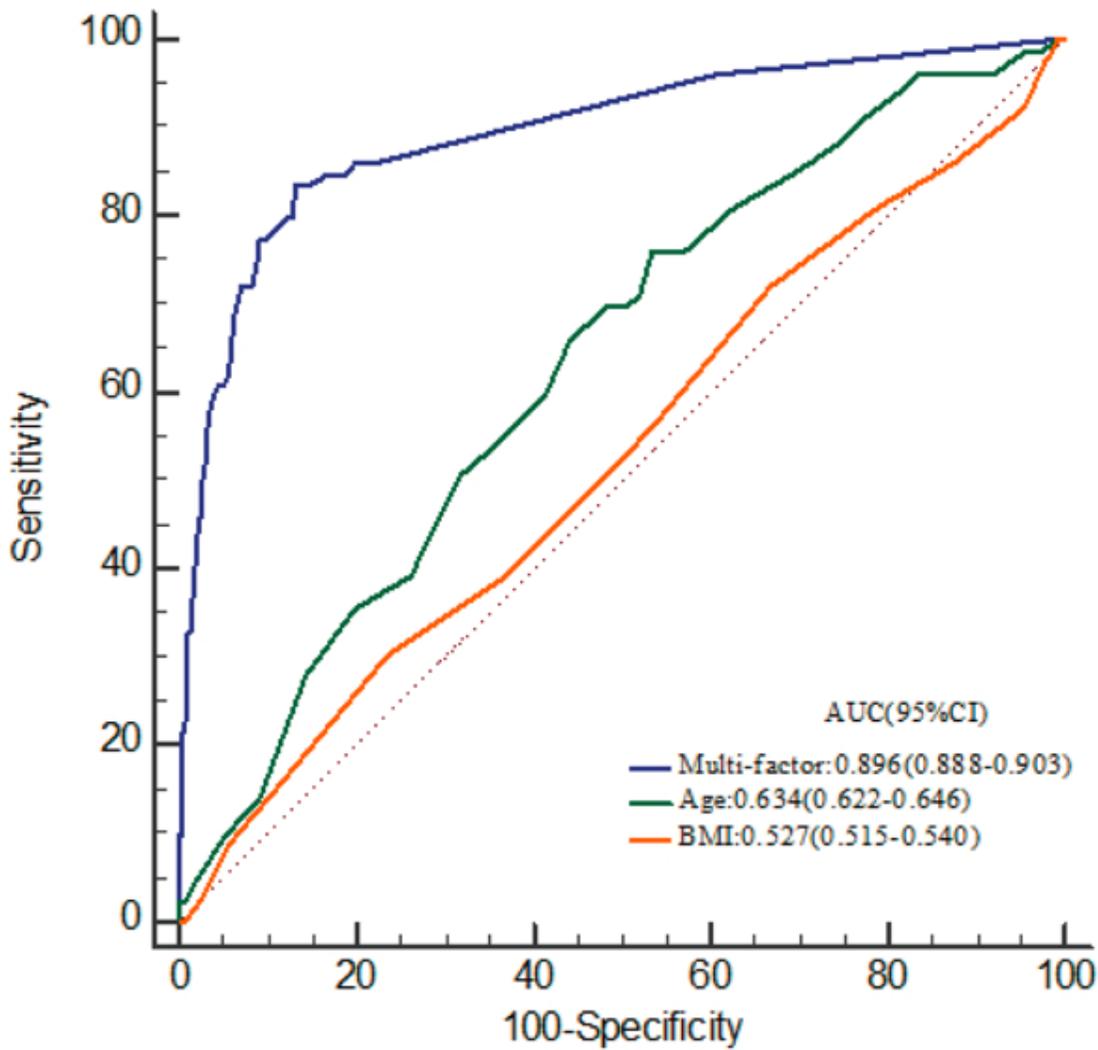


Figure 2

Comparison of the receiver operating characteristic curves between the SDA joint prediction model and simple age and simple BMI prediction models. The performance of the SDA prediction model for 6409 early cancer screenings in rural areas of Nanchong, Sichuan Province, China from 2016 to 2019 is shown. SDA, severe dysplasia; BMI, body mass index; AUC, area under the receiver operating characteristic curve; CI, confidence interval

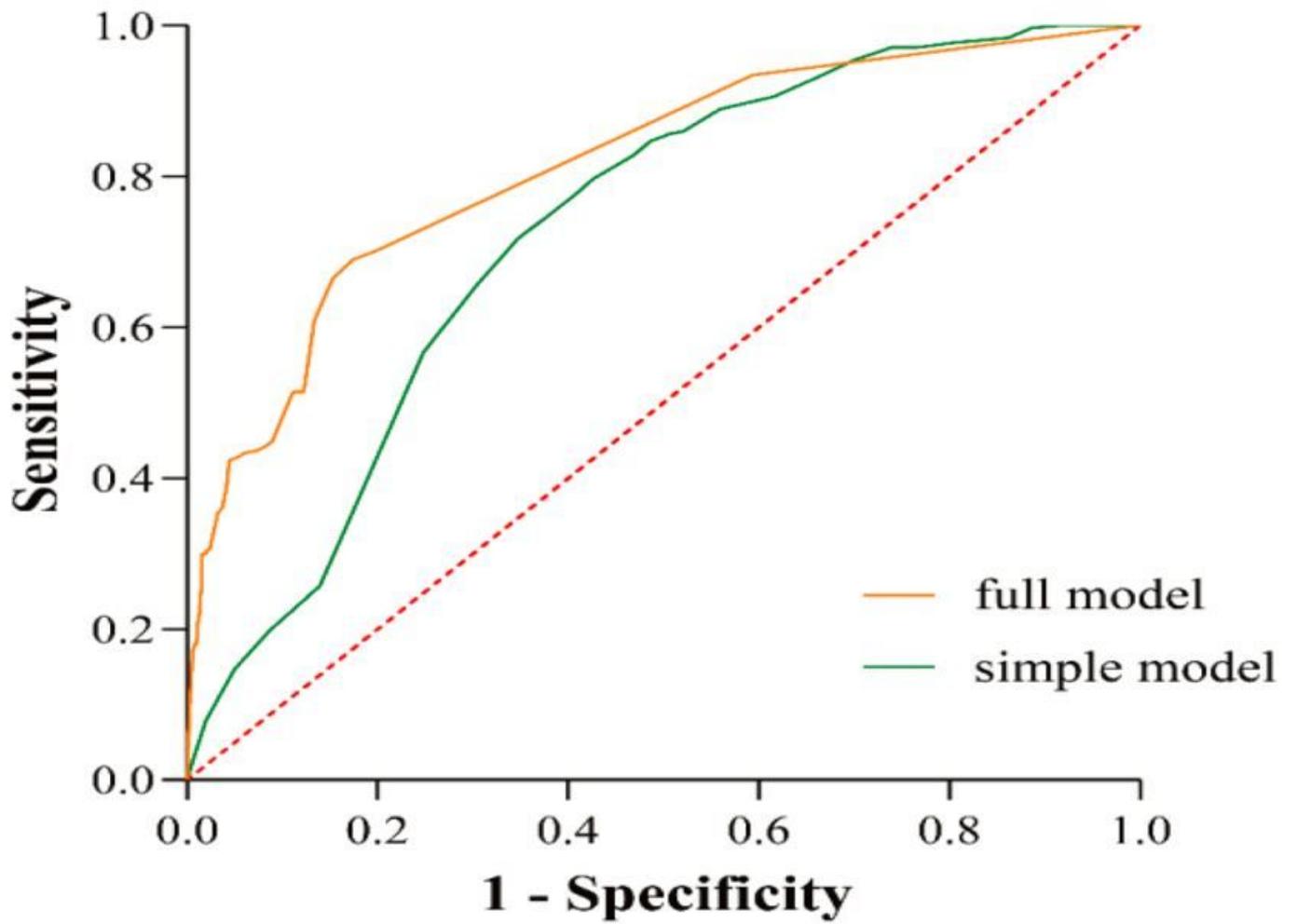


Figure 3

Comparison of the joint prediction model and simple age predication model (leave-one-out cross-validation).

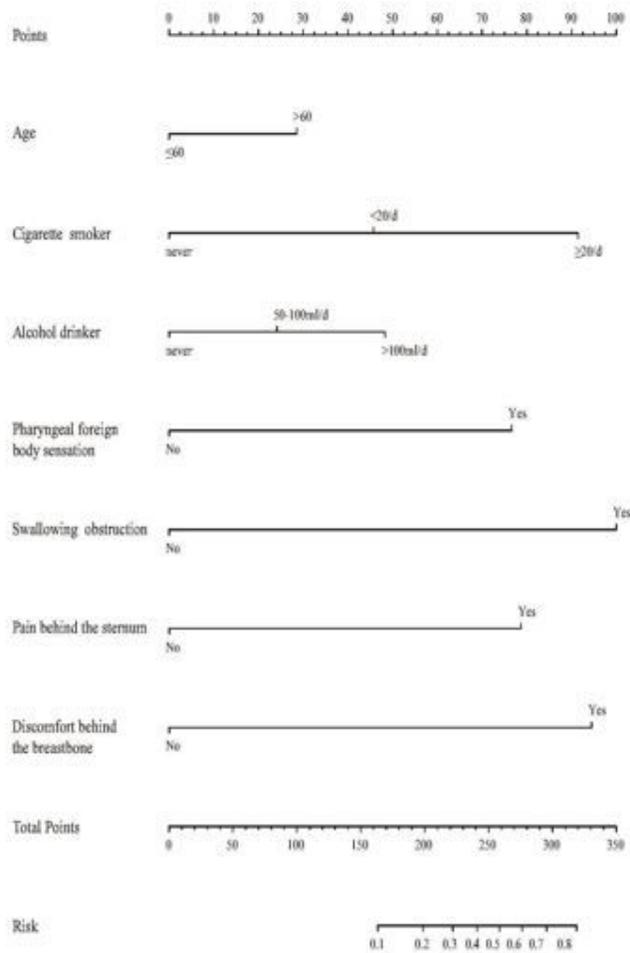


Figure 4

Nomogram for severe dysplasia

Supplementary Files

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

- [Supplementarymaterias1.doc](#)