2020-03-18

Prediction of individuals at high absolute risk of esophageal squamous cell carcinoma

Wang, Qiao-Li; Lagergren, Jesper; Xie, Shao-Hua

Gastrointest Endosc. 2019 Apr;89(4):726-732.e2.
Elsevier
http://doi.org/10.1016/j.gie.2018.10.025
http://hdl.handle.net/10616/47085

Title: Prediction of individuals at high absolute risk of esophageal squamous cell carcinoma

Short title: Predict risk of esophageal squamous cell carcinoma

Authors: Qiao-Li Wang (MD)[1], Jesper Lagergren(MD, PhD)[1,2], Shao-Hua Xie (PhD)[1]

Affiliations: [1] Upper Gastrointestinal Surgery, Department of Molecular medicine and Surgery, Karolinska Institutet, Karolinska University Hospital, Stockholm, Sweden. [2] School of Cancer and Pharmaceutical Sciences, King's College London, United Kingdom.

Grant support: This work was supported by Swedish Research Council (521-2014-2536; 2015-06275), Swedish Cancer Society (CAN 2015/460), and the United European Research Prize (Jesper Lagergren).

Correspondence: Dr. Qiao-Li Wang, Upper Gastrointestinal Surgery, Department of Molecular medicine and Surgery, Karolinska Institutet, NS 67, 2nd Floor, Stockholm 17176, Sweden, Tel.: +46 8-517 709 40, Fax: +46 8-517 762 80, E-mail: qiaoli.wang@ki.se

Disclosures: The study sponsors had no role in the study design, data collection, analysis, or interpretation, the writing of the report, or the decision to submit the manuscript for publication. All authors declare no conflicts of interest.

Author Contributions: All authors have contributed to and approved the final version of the manuscript. The respective roles of each author are the following: study concept and design: all authors; data analysis and results interpretation: all authors; drafting the manuscript: QW; critical revision of the manuscript for important intellectual content: all authors.

ABSTRACT

Background and Aims: This study aimed to develop a prediction model for identifying individuals at high absolute risk of esophageal squamous cell carcinoma (ESCC) for endoscopic screening at a curable stage based on readily identifiable risk factors.

Methods: This was a nationwide Swedish population-based case-control study, including 167 new cases of ESCC and 820 randomly selected control participants. Odds ratios with 95% confidence intervals (CI) were assessed using multivariable unconditional logistic regression. The discriminative accuracy of the model was assessed by the area under the receiver operating characteristic curve (AUC) with leave-one-out cross-validation. Models for projecting individuals' absolute 5-year risk of ESCC were developed by incorporating the age- and sex-specific incidence rates and competing risk of death from other causes.

Results: A model including the risk factors age, sex, tobacco smoking, alcohol overconsumption, education, duration of living with a partner, and place of residence during childhood generated an AUC of 0.81 (95% CI 0.77-0.84). A model based only on age, sex, tobacco smoking, and alcohol overconsumption obtained a similar AUC (0.79, 95% CI 0.75-0.82). A 5-year follow-up of 355 men aged 70-74 years with over 35 years' smoking and alcohol overconsumption history is needed to detect one ESCC case. The estimated individuals' absolute 5-year risk of ESCC varied according to combinations of risk factors.

Conclusions: This “easy-to-use” risk prediction model showed a good discriminative accuracy and had the potential to identify individuals at high absolute risk of ESCC who might benefit from tailored endoscopic screening and surveillance.

Keywords: Prediction model; esophageal neoplasm; risk assessment; early diagnosis.

INTRODUCTION

Esophageal cancer is the 6th leading cause of cancer-related deaths globally. The two main histological types are esophageal squamous cell carcinoma (ESCC) and adenocarcinoma.[1, 2] ESCC is a dominant cancer in Eastern countries and globally, and despite an increasing incidence of esophageal adenocarcinoma in Western countries, ESCC remains a major type of esophageal cancer also in these countries. In 2012, approximately 38,000 new cases of ESCC occurred in Europe, North America, and Oceania, accounting for around 60% of all cases of esophageal cancer in these regions.[1]

ESCC is characterized by a poor prognosis with a population-based overall 5-year survival rate of less than 15% in Western countries, which has not improved much in recent years.[3, 4] Most patients present with the cardinal symptoms, dysphagia and weight loss, which typically occur first at an advanced tumor stage, when the 5-year survival rate is low.[5-7] Upper endoscopy may enable early detection of ESCC and other upper gastrointestinal cancers, but universal endoscopic screening is not justified, especially not in populations with low incidence rates of upper gastrointestinal cancers.[8]. A more feasible way may be to identify a limited group of individuals with high absolute risk of ESCC, who might benefit from tailored endoscopic screening and surveillance.[2, 9] By combining readily identifiable risk factors, risk prediction modeling is a promising tool for selecting individuals with high absolute cancer risk.[10-12] Yet, such prediction models have rarely been developed for ESCC, and none has been developed in a Western population. Using data from a high-quality nationwide population-based case-control study in Sweden, we aimed to develop a prediction model to estimate individuals' absolute risk of ESCC in five years based on information on readily identifiable risk factors.

METHODS

Study design, participants, and data collection

This study was based on data from a nationwide population-based case-control study in Sweden, which has been described elsewhere in detail.[13] The study encompassed the whole Swedish population aged <80 years, born in Sweden, and living there between December 1, 1994, and December 31, 1997. The main reason for using this older case-control study was its high internal validity, which will be difficult to repeat nowadays. A comprehensive organization was formed to ensure quick ascertainment of newly diagnosed cases. Patients were recruited by a local contact person at any of the 195 (100%) participating hospital departments of general surgery, thoracic surgery, otorhinolaryngology, oncology, or pathology in Sweden and six regional tumor registries, to guarantee every potential case-patient was identified soon after diagnosis. Half of all newly diagnosed ESCC cases (n=228, born on even dates) were eligible, and 167 (73%) of these participated. Only half of all patients with ESCC were included because the incidence of ESCC was higher than that of adenocarcinoma and the case-control study mainly aimed to study risk factors for esophageal adenocarcinoma. The main reasons for non-participation were unwillingness (n=11, 5%) and physical or mental disorders or early death (n=50, 22%). All included patients underwent strict and uniform diagnostic routines, including endoscopy with histological confirmation of all tumors. Almost all (97%) tumor specimens were also re-examined by one experienced pathologist for increased histologic accuracy. Ambiguous cases were reviewed by a panel of researchers who utilized all available information. Eligible control subjects (n=1128) were randomly selected from the Registry of the Total Population in Sweden and were frequency-

matched by age and sex to esophageal adenocarcinoma cases in the original case-control study, and 820 participated (73% participation rate). Unwillingness was the most common reason for non-participation (n=210, 19%). Specially trained professional interviewers from the governmental agency Statistics Sweden conducted computer-aided face-to-face interviews of all case patients and control participants for the collection of a large number of potential risk factors and other relevant variables. The interviewers were kept unaware of the study hypotheses and were trained to treat all participants in an equal manner. Statistical power calculations were conducted, utilizing an α of 0.05 (two-sided), different exposure levers and odds ratios (ORs) of main risk factors, to show sufficient power. The study was approved by all regional ethical committees in Sweden, and informed consent was obtained from each participant before the interview.

Model development

The selection of candidate predictors of ESCC was based on literature review and associations found in the same case-control study. The nine-candidate variables were: age (<50, 50-59, 60-69, or 70-79 years), sex (male or female), tobacco smoking (never-smokers, smokers for 1-20, 21-35, or >35 years), alcohol overconsumption (<70 or ≥70 gram ethanol per week), formal education (≤10 or >10 years), duration of living with a partner (≤10 or >10 years), intake of fruit and vegetables (in tertiles as high, medium, or low), place of residence during childhood (rural or urban), and family history of esophageal or head and neck cancers in first-degree relatives (yes or no). A detailed description of the variable is available as a Supplementary File.

The final two panels of predictor variables were determined based on 2-step approach logistic regressions.[14] In the first step, the four best-established risk factors for ESCC, i.e., age, sex, tobacco smoking, and alcohol overconsumption were included in the model without any selection procedure. These four variables were included in a simple prediction model. For the more extensive prediction model, selection of additional predictors was based on a stepwise backward selection approach whereby associations were losing statistical significance in the multivariable model at a predefined nominal significance level of 0.1. In the second step, predictors eliminated in the first step were re-entered into the multivariable model one by one to ensure that no omitted predictor statistically significantly improved the goodness of fit in a likelihood ratio test.[14] Potential pairwise interactions between predictors were assessed in a preliminary model, indicating a statistically significant interaction between education and duration of living with a partner. A combination of these two factors was used to categorize participants into the following three groups: both education and partnership for >10 years, only education or partnership for >10 years, or both education and partnership for ≤10 years.

Test of performance

The discriminative ability of the models was assessed by the area under the receiver operating characteristic curve (AUC) and the Somers' D statistic.[15] The AUC measured the model's ability to discriminate ESCC patients from control participants, and the Somers' D statistic assessed the strength and direction of associations between predicted probabilities and observed responses.[15] Because of the possible over-fitting of the model

when the performance was assessed with the same dataset as the one used to develop the model, re-testing of the performance of the models with leave-one-out cross-validation strategy was conducted.[12] In this cross-validation process, the probability for each participant was predicted from a model ignoring this participant and used to calculate the unbiased AUC and Somers' D statistic.

Estimates of absolute 5-year risk

The 5-year absolute risk of ESCC of an individual of a specific profile of risk factors and the corresponding 95% confidence intervals (CIs) were calculated based on: (1) the OR for the individual with the specific risk factors profile from the final logistic model, which was used as an approximation to the relative risk given the rare disease assumption, (2) the baseline age- and sex-specific incidence rate of ESCC in the Swedish population during the study period, (3) the estimated population attributable fraction (PAF) derived from the final logistic model, and (4) the age- and sex-specific mortality rate from reasons other than esophageal cancer in the population, to correct competing risk from mortality.[10-12] The age- and sex-specific incidence rates of ESCC, mortality rates excluding esophageal cancer, and an example of estimating individual's 5-year absolute risk are provided in the Supplementary File.

The detailed algorithm of individual's 5-year risk calculation with age(a), sex(s) and OR(r) was shown as:

$$P(a,s,r) = \frac{b_1 r}{b_1 r + b_2}\left(1 - e^{-5(b_1 r + b_2)}\right) \qquad \text{[Formula 1]}$$

where b_1 are the age- and sex-specific baseline hazard rates of ESCC and b_2 are the age- and sex-specific mortality rates from competing causes.

The age- and sex-specific mortality rates excluding esophageal cancer (b_2) were calculated using the population mortality data from Statistics Sweden minus esophageal cancer specific mortality data from the NORDCAN database.[16] The age- and sex-specific baseline hazard rates of ESCC (b_1) were calculated with Formula 2 below:

$$b_1 = IR_{as} \times (1 - PAF) \qquad \text{[Formula 2]}$$

where IR_{as} are incident rates for specific age and sex from the logistic regression, PAF are population attributable fractions which can be estimated by Formula 3:

$$PAF = 1 - \frac{1}{n}\sum_{i=1}^{n}(\frac{1}{r_i}) \qquad \text{[Formula 3]}$$

where n is the number of total ESCC cases, r_i is the OR for the ith cases estimated from the logistic regression model. All statistical analyses were conducted using the statistical software package SAS 9.4 for Windows (SAS Institute Inc., Cary, NC, USA). This study followed the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) guidelines and checklist.[17, 18]

RESULTS

Participants

Characteristics of the 167 ESCC case patients and the 820 control participants are presented in Table 1. ESCC patients were more often smokers and had higher alcohol intake compared to the control participants.

Predictor variables

The full prediction model included six variables: age, sex, duration of tobacco smoking, alcohol overconsumption, the combined variable education and duration of living with a partner, and place of residence during childhood. The distribution of these variables among case patients and control participants and associations with ESCC risk are shown in Table 1. Tobacco smoking increased the risk of ESCC in a duration-response pattern. Alcohol overconsumption was associated with an increased risk of ESCC (adjusted OR 3.67, 95% CI 2.39-5.62). The PAF combining all six predictors in the full model was 0.95.

The simple prediction model included the four variables age, sex, duration of tobacco smoking, and alcohol overconsumption. The ORs for these factors in the simple model were similar to those in the full model (Table 1). The PAF of these four predictors was 0.85.

Model performance

The receiver operating characteristic curves for the two models are shown in Figure 1. The AUC statistics without cross-validation were 0.81 (95% CI 0.77-0.84) for the full model and 0.79 (95% CI 0.75-0.82) for the simple model (Table 2). The AUCs slightly decreased after leave-one-out cross-validation to 0.78 (95% CI 0.75-0.82) in the full model and 0.75 (95% CI 0.71-0.79) in the simple model (Table 2).

Absolute 5-year risk of esophageal squamous cell carcinoma

The projected absolute 5-year risks in individuals with different combinations of risk factors of ESCC in the simple model are illustrated in heat charts (Figure 2). In addition, Table 3 shows the estimated absolute 5-year risk of ESCC associated with selected combinations of risk factors. The estimated absolute 5-year risk of ESCC varied greatly across combinations of risk factors, ranging from 2.2/100,000 to 281.4/100,000 in men aged ≥50 years and from 0.4/100,000 to 126.6/100,000 in women aged ≥50 years (Figure 2). The highest absolute risk (281.4/100,000) was found in men aged 70-74 years who had smoked for over 35 years and overconsumed alcohol. In this group, 355 individuals are needed to survey, e.g., by endoscopy or other screening or surveillance tools, to detect one case of ESCC within five years (Table 3, Supplementary Figure 1). Individuals' 5-year absolute risk of ESCC can be calculated by both the simple and the full model calculators in the Supplementary Files.

DISCUSSION

This study developed a six-variable model and a four-variable model to predict individuals' absolute risk of developing ESCC within the next five years based on information of readily identifiable risk factors. Both models had good performances regarding discriminative accuracy. The estimated absolute 5-year risk of ESCC varied greatly across combinations of risk factors and was heavily dependent on the risk factors age, sex, tobacco smoking, and alcohol overconsumption.

To the best of our knowledge, this is the first study that developed a prediction model for the absolute risk of ESCC in a Western population. A few risk prediction models of ESCC or serious dysplasia of the esophagus have been developed in some Asian countries where ESCC seems to be associated with partly different risk factors than in Western populations. By pooling known risk factors (age, sex, tobacco smoking, and alcohol drinking) with genetic information (25 single nucleotide polymorphisms), a prediction model of ESCC from a Chinese population had an AUC of 0.71, which was lower than the AUCs in the presently developed models,[19] and collecting and analysing genetic material might not be practical in the clinical setting. Using several potential risk factors (age, ethnicity, tobacco smoking, opium use, education, marital status, oral health, family history, tea temperature, and water source), a prediction model of ESCC based on a hospital-based case-control study in Iran generated an AUC of 0.77, which is similar to the four-variable model in the current study.[20] A model for predicting the risk of severe esophageal dysplasia and higher-grade leisions has recently been developed from a cross-sectional study of a Chinese population.[21] Using nine variables (age, family history of ESCC, cigarette smoking, body mass index ≤22, pesticide exposure, irregular eating

habits, intake of high-temperature foods, rapid ingestion of meals, and ingestion of leftover food during summer months) in participants aged >60 years, an AUC of 0.68 was indicated. With five variables (age, use of coal or wood as the main source of cooking fuel, body mass index ≤22, unexplained epigastric pain, and rapid ingestion of meals), the model for participants aged ≤60 had an AUC of 0.80. However, those models included ESCC symptoms (epigastric pain), which might limit their application in a screening program. Compared with the models developed in Asian populations, the model developed in the present study is the simplest, but still with the highest discrimination accuracy as measured by the AUC, and none of the previous studies estimated individuals' absolute risk of ESCC.

Strengths of this study include the use of data from a large and population-based study with high participation rates and excellent internal validity, based on random sampling of control participants, thorough histological ascertainment of all case-patients, and personal interviews with all participants. The cross-validation strategy used to evaluate the model performance partially counteracted over-fitting. The study also has limitations. Due to the case-control study design, as opposed to cohort studies, calibration of predicted absolute risk against observed risk was not possible.[14] Moreover, given the low absolute incidence rate of ESCC (<2/100,000 person-year in 1994-1997 in Sweden),[22] the study contained too few control participants (n=820) for such analysis. Uncertainty in the recalling of past exposures is a threat with the case-control design. However, the risk factor profile was very different in patients with ESCC compared with the 189 patients with esophageal adenocarcinoma who also were interviewed as part of the same case-control study, which argues against recall bias.[13, 23-25] Although this study had sufficient

statistical power to detect associations between major risk factors and ESCC risk, there remains uncertainty in the precision of the predicted absolute risk of ESCC. Even though good internal validation was achieved, an over-estimation of the model was possible, and external validation in independent populations is needed before any clinical application can be recommended.

Risk prediction models provide objective evidence-based estimates of individuals' absolute risk of ESCC, and the information regarding the included risk factors is readily available and easily collected in clinical practice, public health settings, and external validation studies. The overall 5-year survival in ESCC is below 15%, and the survival rate is extremely low in patients diagnosed at late tumor stages.[4] Detection at an early stage, through close follow-up in high-risk individuals, would probably reduce the mortality. Therefore, if confirmed in future research, the model presented in this study could help clinicians in their decision-making regarding possible tailored endoscopy surveillance.[26] Additionally, with the help of this model, public health care professionals and policymakers may estimate the future disease burden of ESCC in the population and plan appropriate prevention programs, including possible targeted screening programs for ESCC, in groups at high absolute risk. However, the thresholds for any screening or surveillance program need to be carefully determined, for which several aspects should be considered, including the absolute risk (incidence) in the target population, costs, and benefits and harm for the individuals. Notably, applying our four-variable simple model in the Swedish population in 1995-1997 suggested that 355 individuals are needed to be surveyed to detect one ESCC case in 5 years even among the highest risk category, which might not be cost-effective. However, the model may be more feasible in

populations with higher incidence rates, if externally validated, or with simpler and more affordable screening and surveillance tools in the future. Given the similarity in the etiology of ESCC, the present model might also be applicable for other Western countries, especially the Nordic countries. Choosing a risk threshold or evaluating the cost-effectiveness is beyond the scope of this study.

In summary, this study identified a model of good discriminative accuracy based on readily identifiable risk factors for predicting the absolute 5-year risk of ESCC in a Western population. The model needs to be validated in independent populations but has the potential for future use in identifying individuals at high absolute risk of ESCC who may benefit from tailored endoscopic surveillance for early tumor detection at a curable stage.

REFERENCES

1. Arnold M, Soerjomataram I, Ferlay J, et al. Global incidence of oesophageal cancer by histological subtype in 2012. Gut 2015;64:381-387.
2. Lagergren J, Smyth E, Cunningham D, et al. Oesophageal cancer. Lancet 2017;390:2383-2396.
3. Lagergren J, Mattsson F. Diverging trends in recent population-based survival rates in oesophageal and gastric cancer. PLoS One 2012;7:e41352.
4. Kauppila JH, Mattsson F, Brusselaers N, et al. Prognosis of oesophageal adenocarcinoma and squamous cell carcinoma following surgery and no surgery in a nationwide Swedish cohort study. BMJ Open 2018;8:e021495. Epub 2018/05/12.
5. Pennathur A, Gibson MK, Jobe BA, et al. Oesophageal carcinoma. Lancet 2013;381:400-412.
6. Enzinger PC, Mayer RJ. Esophageal cancer. N Engl J Med 2003;349:2241-2252.
7. O'Connell J B, Maggard MA, Liu JH, et al. A report card on outcomes for surgically treated gastrointestinal cancers: are we improving? J Surg Res 2004;121:214-221.
8. Lao-Sirieix P, Fitzgerald RC. Screening for oesophageal cancer. Nat Rev Clin Oncol 2012;9:278-287. Epub 2012/03/21.
9. Xie SH, Lagergren J. A model for predicting individuals' absolute risk of esophageal adenocarcinoma: Moving toward tailored screening and prevention. Int J Cancer 2016;138:2813-2819.
10. Fears TR, Guerry Dt, Pfeiffer RM, et al. Identifying individuals at high risk of melanoma: a practical predictor of absolute risk. J Clin Oncol 2006;24:3590-3596.
11. Matsuno RK, Costantino JP, Ziegler RG, et al. Projecting individualized absolute invasive breast cancer risk in Asian and Pacific Islander American women. J Natl Cancer Inst 2011;103:951-961.
12. Thrift AP, Kendall BJ, Pandeya N, et al. A model to determine absolute risk for esophageal adenocarcinoma. Clin Gastroenterol Hepatol 2013;11:138-144 e132.
13. Lagergren J, Bergstrom R, Lindgren A, et al. Symptomatic gastroesophageal reflux as a risk factor for esophageal adenocarcinoma. N Engl J Med 1999;340:825-831.
14. Moons KG, Kengne AP, Woodward M, et al. Risk prediction models: I. Development, internal validation, and assessing the incremental value of a new (bio)marker. Heart 2012;98:683-690.
15. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. Epidemiology 2010;21:128-138.
16. Engholm G, Ferlay J, Christensen N, et al. NORDCAN--a Nordic tool for cancer information, planning, quality control and research. Acta Oncol 2010;49:725-736. Epub 2010/05/25.

17. 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.

18. Moons KG, Altman DG, Reitsma JB, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. Ann Intern Med 2015;162:W1-73.

19. Chang J, Huang Y, Wei L, et al. Risk prediction of esophageal squamous-cell carcinoma with common genetic variants and lifestyle factors in Chinese population. Carcinogenesis 2013;34:1782-1786.

20. Etemadi A, Abnet CC, Golozar A, et al. Modeling the risk of esophageal squamous cell carcinoma and squamous dysplasia in a high risk area in Iran. Arch Iran Med 2012;15:18-21.

21. Liu MF, Liu Z, Cai H, et al. 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;15:1538-1546.

22. Wang QL, Xie SH, Wahlin K, et al. Global time trends in the incidence of esophageal squamous cell carcinoma. Clin Epidemiol 2018;10:717-728. Epub 2018/06/29.

23. Lagergren J, Bergstrom R, Lindgren A, et al. The role of tobacco, snuff and alcohol use in the aetiology of cancer of the oesophagus and gastric cardia. Int J Cancer 2000;85:340-346.

24. Lagergren J, Bergstrom R, Nyren O. Association between body mass and adenocarcinoma of the esophagus and gastric cardia. Ann Intern Med 1999;130:883-890.

25. Lagergren J, Wang ZH, Bergstrom R, et al. Human papillomavirus infection and esophageal cancer: a nationwide seroepidemiologic case-control study in Sweden. J Natl Cancer Inst 1999;91:156-162.

26. NICE. Suspected cancer recognition and referral: symptoms and findings. 2018.

Table 1 Characteristics of controls and cases with esophageal squamous cell carcinoma from a Swedish case-control study with estimated odds ratios (ORs) with 95% confidence intervals (CIs) of esophageal squamous cell carcinoma

Variables	**Controls n (%)**	**Cases n (%)**	**Crude OR (95%CI)**	**Adjusted OR (95%CI)** [a]	**Adjusted OR (95%CI)** [b]
Total	820 (100)	167 (100)			
Age (years)					
<50	48 (6)	3 (2)	0.37 (0.11,1.22)	0.64 (0.18,2.31)	0.46 (0.13,1.60)
50-59	161 (20)	35 (21)	1.28 (0.82,2.02)	1.66 (0.96,2.89)	1.28 (0.76,2.15)
60-69	245 (30)	67 (40)	1.61 (1.10,2.36)	1.33 (0.86,2.06)	1.23 (0.80,1.88)
70-79	366 (44)	62 (37)	1.00 (reference)	1.00 (reference)	1.00 (reference)
Sex					
Male	679 (83)	120 (72)	1.00 (reference)	1.00 (reference)	1.00 (reference)
Female	141 (17)	47 (28)	1.89 (1.29,2.77)	5.01 (3.07,8.18)	4.35 (2.70,7.00)
Duration of smoking (years)					
Non-smokers	325 (40)	22 (13)	1.00 (reference)	1.00 (reference)	1.00 (reference)
1-20	155 (19)	21 (13)	2.00 (1.07,3.75)	2.09 (1.06,4.12)	2.17 (1.12,4.22)
21-35	156 (19)	27 (16)	2.56 (1.41,4.63)	2.62 (1.36,5.05)	2.64 (1.40,4.98)
>35	184 (22)	97 (58)	7.79 (4.74,12.80)	8.62 (4.91,15.14)	8.41 (4.87,14.52)
Alcohol consumption (gram/week)					
≤70	642 (78)	89 (53)	1.00 (reference)	1.00 (reference)	1.00 (reference)
>70	178 (22)	78 (47)	3.16 (2.24,4.47)	3.67 (2.39,5.62)	3.24 (2.15,4.89)
Education and duration of living with a partner (years)					
Education>10 and partnership>10	302 (37)	37 (22)	1.00 (reference)	1.00 (reference)	-
Education or partnership>10	462 (56)	110 (66)	1.94 (1.30,2.90)	2.10 (1.32,3.36)	-
Education≤10 and partnership≤10	56 (7)	20 (12)	2.92 (1.58,5.39)	3.10 (1.52,6.32)	-
Place of residence during childhood					
Urban	313 (38)	58 (35)	1.00 (reference)	1.00 (reference)	-
Rural	505 (62)	107 (65)	1.14 (0.81,1.62)	1.53 (1.01,2.33)	-
Population attributable fraction	-	-	-	0.95	0.85

[a] Adjusted for all listed variables in the six-variable model; [b] Adjusted for the other three listed variables in the four-variable model.

Table 2 Performance of prediction models for individuals' risk of esophageal squamous cell carcinoma

	Original without cross-validation		Leave-one-out cross-validation	
Model	**AUC (95% CI)**	**Somers' D**	**AUC (95% CI)**	**Somers' D**
Six-variable model	0.81 (0.77,0.84)	0.61	0.78 (0.75,0.82)	0.57
Four-variable model	0.79 (0.75,0.82)	0.57	0.75 (0.71,0.79)	0.50

AUC: area under the receiver operating characteristic curve; CI: confidence interval.

Table 3 Estimated absolute 5-year risks of esophageal squamous cell carcinoma in a Swedish population associated with selected profiles of risk factor combinations from a model including age, sex, tobacco smoking, and alcohol overconsumption

Profile no.	Sex	Age (years)	Smoking (years)	Alcohol overconsumption [a]	Absolute 5-year risk, 1/100,000	Number of individuals needed to survey to detect one case
1	Male	50-54	Non-smokers	No	2.2	46,440
2	Male	50-54	1-20	No	4.7	21,391
3	Male	50-54	21-35	No	5.7	17,618
4	Male	50-54	>35	No	18.1	5,521
5	Male	50-54	>35	Yes	58.7	1,702
6	Male	60-64	21-35	No	19.5	5,137
7	**Male**	**60-64**	**>35**	**No**	**62.1**	**1,610**
8	Male	60-64	>35	Yes	201.3	497
9	Male	70-74	Non-smokers	No	10.3	9,683
10	Male	70-74	>35	No	86.8	1,152
11	Male	70-74	>35	Yes	281.4	355
12	Female	50-54	1-20	Yes	2.8	35,521
13	Female	60-64	>35	Yes	45.4	2,201
14	Female	70-74	Non-smokers	No	3.8	26,150
15	Female	70-74	>35	Yes	104.3	959

[a] Alcohol overconsumption was defined as alcohol consumption >70 gram/week.

FIGURE LEGENDS:

Figure 1: The receiver operating characteristic curves based on a six-variable full model and a four-variable simple model. AUC= the area under the receiver operating characteristic curve.

Figure 2: Absolute 5-year risks of esophageal squamous cell carcinoma per 100,000 person-years, estimated from the four-variable simple model individuals aged 50 years or above.