2017-02-15

A model for predicting individuals' absolute risk of esophageal adenocarcinoma : moving toward tailored screening and prevention

Xie, Shao-Hua; Lagergren, Jesper

Int J Cancer. 2016;138(12):2813-9.
http://doi.org/10.1002/ijc.29988
http://hdl.handle.net/10616/45524

A Model for Predicting Individuals´ Absolute Risk of Esophageal Adenocarcinoma: Moving towards Tailored Screening and Prevention

Short title: predicting absolute risk of EAC

Authors: Shao-Hua Xie[1], and Jesper Lagergren[1, 2]

Affiliations: [1]Upper Gastrointestinal Surgery, Department of Molecular Medicine and Surgery, Karolinska Institutet, Staockholm, Sweden; and [2] Section of Gastrointestinal Cancer, Division of Cancer Studies, King's College London, London, United Kingdom

Correspondence to: Dr. Shao-Hua Xie, Upper Gastrointestinal Surgery, Department of Molecular Medicine and Surgery, Karolinska Institutet, NS 67, 2nd Floor, Stockholm17176, Sweden. Tel: + 46 8 517 70917; Fax: +46 8 517 7628; Email: shaohua.xie@ki.se

Keywords: esophageal adenocarcinoma; risk prediction; absolute risk; prediction model

Abbreviations: AUC, area under the receiver operating characteristic curve; BMI, body mass index; CI, confidence interval; EAC, esophageal adenocarcinoma; NSAIDs, nonsteroidal anti-inflammatory drugs; PAR, population attributable risk; OR, odds ratio

Novelty and Impact

Esophageal adenocarcinoma (EAC) is characterized by rapidly increasing incidence and poor prognosis, stressing the need for preventive and early detection strategies. However, universal endoscopic screening is unfeasible given the low absolute risk in the population. We developed prediction models for estimating individuals' absolute 5-year risk of EAC. The prediction models had good discriminative accuracy after cross-validation, and limited high-risk groups were identified. Prediction models can guide a move towards tailored prevention and detection of EAC.

Abstract

Esophageal adenocarcinoma (EAC) is characterized by rapidly increasing incidence and poor prognosis, stressing the need for preventive and early detection strategies. We used data from a nationwide population-based case-control study, which included 189 incident cases of EAC and 820 age- and sex-matched control participants, from 1995 through 1997 in Sweden. We developed risk prediction models based on unconditional logistic regression. Candidate predictors included established and readily identifiable risk factors for EAC. The performance of model was assessed by the area under receiver operating characteristic curve (AUC) with cross-validation. The final model could explain 94% of all case patients with EAC (94% population attributable risk) and included terms for gastro-esophageal reflux symptoms or use of antireflux medication, body mass index (BMI), tobacco smoking, duration of living with a partner, previous diagnoses of esophagitis and diaphragmatic hernia, and previous surgery for esophagitis, diaphragmatic hernia, or severe reflux, or gastric or duodenal ulcer. The AUC was 0.84 (95% confidence interval [CI] 0.81-0.87) and slightly lower after cross-validation. A simpler model, based only on reflux symptoms or use of antireflux medication, BMI, and tobacco smoking could explain 91% of the case patients with EAC and had an AUC of 0.82 (95% CI 0.78-0.85). These EAC prediction models showed good discriminative accuracy, but needs to be validated in other populations. These models have the potential for future use in identifying individuals with high absolute risk of EAC in the population, who may be considered for endoscopic screening and targeted prevention.

The incidence of esophageal adenocarcinoma (EAC) has increased rapidly during the past four decades in many Western populations, including North America and Europe, with the highest incidence in the United Kingdom.[1-3] There were 52 000 new patients with EAC (41 000 men and 11 000 women) worldwide in 2012.[4] The incidence of EAC has increased on average by 5% per year since 1970 in Sweden, and even more so during the last 20 years.[1]

EAC is also characterized by poor prognosis with an overall 5-year survival lower than 15%. Tumor stage at diagnosis is by far the strongest prognostic factor,[2, 3, 5] and detection at an early stage would possibly reduce the mortality.[6-8] Endoscopy provides an opportunity of early detection of EAC or its premalignant condition, i.e. Barrett's esophagus with dysplasia. However, universal endoscopic screening is not feasible or justified given the low absolute risk in the population, the risk of complications and the considerable costs. Identifying a limited group of individuals at high absolute risk of EAC for endoscopic screening is a more feasible strategy. Risk prediction modelling combining information on readily identifiable risk factors is a promising approach for selection of individuals with high absolute risk of EAC.[9-12] Unfortunately, such prediction models have rarely been developed for EAC.

Using data from a comprehensive and methodologically rigorous case-control study in Sweden, we aimed to develop a prediction model for valid estimation of the absolute 5-year risk of EAC based on information on a panel of established risk factors, which could identify high-risk individuals who may benefit from tailored endoscopic screening or future prevention strategies.

Materials and Methods

Study design, participants, and data collection

This study was based on data from a large nationwide population-based case-control study in Sweden, which has been described in detail elsewhere.[13-17] In brief, the study base included all residents born in Sweden and aged less than 80 years during 1995-1997. All patients with newly diagnosed EAC in the study base were eligible for the study. All 195 (100%) hospital departments involved in the diagnosis or treatment of these patients in Sweden participated in the recruitment of case patients and in the collection of relevant background and clinical data for these patients. The pathology departments in particular were crucial in identifying eligible case patients. Moreover, all 6 regional oncological centers in Sweden contributed to identifying case patients. Strict and nationwide routines for histopathology review of EAC specimens were introduced for the purpose of this study. All tumor specimens were also re-evaluated by one experienced pathologist to further improve the diagnostic accuracy and uniformity. Control subjects were selected randomly from the Registry of the Total Population in Sweden and were frequency matched with the EAC cases by age and sex. All participants underwent computer-aided face-to-face interviews by professional interviewers from the governmental agency Statistics Sweden. The interviewers underwent special training to treat case patients and control subjects in an equal manner, and they were also kept unaware of the study hypotheses. Both written and oral informed consent was obtained from each subject before the interview, and the study was approved by all 6 regional ethical review boards in Sweden.

Model development

The selection of candidate predictors was based on literature review and examination of their population distribution and strength of association with EAC. Candidate predictors included socioeconomic factors (education, duration of living with a partner, and number of children in the household during childhood),[17] gastro-esophageal reflux symptoms and use of anti-reflux medication,[13] anthropometric measurements of body mass index (BMI) and height,[15, 18, 19] tobacco smoking,[16] intake of fruit and vegetables,[20] family history of cancer,[21] use of medications relaxing the lower esophageal sphincter or use of statins,[22] and previous diseases and treatments of the digestive system.[23] Detailed information on definitions and codes of the predictor variables can be found in the corresponding references in the list of references and supplementary materials.

We used a 2-step approach to determine the panel of predictors included in the final models. First, we selected the predictors through an unconditional logistic regression with a forward selection method. Candidate predictors were included 1 by 1 based on their importance scores, which were generated from random forests analysis, from the most to the least important. The random forests analysis measured the importance of candidate predictors by mean decrease in accuracy score, which was estimated by the increase in misclassification for the out-of-bag samples using a data matrix containing the original variables and a vector with randomly permuted outcomes.[24, 25] Predictors stayed in the model if their effects were significant at the level of 0.10. The remaining predictors were examined individually to identify any detectable effects after adjustment for potential confounding factors. If there were multiple measures for one risk factor, selection was based on the Akaike information criterion.[26, 27] We tested pairwise interactions, but since there were no statistically significant interaction terms, we did not include any in the final model. All statistical analyses were performed using SAS 9.4 (SAS Institute, Cary,

NC), except for the random forests analyses, which were conducted in R 3.1.3 (R Foundation for Statistical Computing, Vienna, Austria).

Test of performance

We calculated the area under the receiver operating characteristic curve (AUC) and its 95% confidence interval (CI), which tests the model's ability to discriminate between case patients and control participants, and Somers' D statistic, which measures the strength and direction of associations between predicted probabilities and observed responses.[26, 28] To prevent the problem of over-fitting when performance of the model was assessed with the same dataset as used to build the model, we re-calculated the statistics for model performance using both leave-one-out and 10-fold cross-validation strategies. Such cross-validation processes calculated the unbiased AUC and Somers' D with the predicted probability of each subject or randomly selected group (10% of all individuals) from a model ignoring this subject or group, respectively.[26, 27]

Estimates of absolute 5-year risk

We calculated the absolute 5-year risk of EAC for all possible profiles of risk factors, based on: (1) the estimated relative risk for the individual from the final logistic model; (2) baseline age- and sex-specific incidence rate in the population; (3) the estimated population attributable risk derived from the logistic model; and (4) the age- and sex-specific mortality rate in the population to correct competing risk from death from causes other than esophageal cancer.[9, 27, 29] The detailed algorithm, age- and sex-specific EAC incidence rates, and mortality rates excluding esophageal cancer are provided as supplementary material.

Results

A total of 189 EAC cases and 820 control subjects were successfully interviewed. The participation rates among EAC case patients and control subjects were 87% and 73%, respectively. The majority of the participants were men aged between 60 and 79 years. Reasons for non-participation and basic characteristics of participants are shown in Table 1.

Predictor variables

The full prediction model included the following 8 variables: reflux symptoms or use of antireflux medication until 5 years before interview, BMI 20 years before interview, tobacco smoking status 2 years prior to interview, duration of being married or cohabiting, previous diagnosis of esophagitis or diaphragmatic hernia, previous surgery for gastric or duodenal ulcer, and surgery for esophagitis, diaphragmatic hernia or severe reflux symptoms. The distribution of case patients and control participants and associations between predictors and EAC risk in the final model are presented in Table 2. Reflux and obesity were associated with elevated risk of EAC, with exposure-response patterns. Tobacco smoking, previous diagnoses of esophagitis or diaphragmatic hernia, and surgery for esophagitis, diaphragmatic hernia or severe reflux were associated with increased risks of EAC, while previous ulcer surgery was associated with a decreased EAC risk. The population attributable risk (PAR) combining all these predictors was 0.94. A simpler prediction model included only 3 variables of reflux symptoms or use of antireflux medication, BMI, and tobacco smoking, which resulted in a PAR of 0.91.

Model performance

Table 3 shows the discriminative ability of the two models with and without cross-validation. The AUC statistics for the full model and the simple model were 0.84 (95% CI 0.81-0.87) and 0.82 (95% CI 0.78-0.85), respectively. The cross-validation provided slightly lower AUC statistics, which were 0.82 (95% CI: 0.78, 0.85) and 0.79 (95% CI: 0.75, 0.83) after leave-one-out cross-validation for the full and the simple model, respectively, indicating minimal over-fitting and good discrimination. The receiver operating characteristic curves for the two prediction models are shown in Figure 1, in which the model performance with different pre-specified probability thresholds is assessed. For example, a probability threshold of 10% in the full model would have the sensitivity of 83% and the specificity of 65%, while a threshold of 20% would have the sensitivity of 73% and the specificity of 81%.

Absolute 5-year risk of esophageal adenocarcinoma

The absolute 5-year risks of EAC for individuals with various combinations of risk factors can be easily calculated in a Microsoft Excel worksheet which is provided as a supplementary material. Table 4 presents the estimated absolute 5-year risks of EAC associated with selected combinations of risk factors calculated with the simple model, while the estimated absolute 5-year risks of EAC for all possible profiles of risk factors in men aged 50 years or above are shown in a supplementary figure. The magnitude of risk varied greatly across combinations of risk factors. The absolute 5-year risk of EAC ranged from 5.2/100 000 to 533/100 000 in men aged 50 years or above with weekly reflux symptoms for 5 years or longer, depending on the combinations of BMI, tobacco smoking, and use of antireflux medication (supplementary figure 1). The highest absolute 5-year risk of EAC (533/100 000) was observed in male smokers who were aged 70-74 years, had suffered from weekly reflux for at least 5 years with antireflux

medication, and had a BMI over 25.5 kg/m^2, indicating that 188 individuals needed to be surveyed to detect one EAC over 5 years in this group (Table 4).

Discussion

This study indicates that prediction models can be constructed for assessing individuals' absolute 5-year risk of EAC. Both a more extensive and a simpler model had good performance in discriminative ability as assessed by AUC with cross-validation. A combination of reflux symptoms or use of antireflux medications and high BMI were important predictors of EAC risk, but the risk also heavily depended on sex, age, tobacco smoking, living with a partner, and past medical history associated with EAC risk. The estimated absolute 5-year risk of EAC varied greatly across different profiles of risk factors, which highlights the relevance of the models in predicting individuals' absolute risk of EAC.

A few predictors warrant some explanation. Never being married or having cohabited for at least one year increased the risk of EAC, which might be explained by the hypothesis that marriage increases social support and income and reduces risky behavior and stress, and thus, contributes to a better health.[17, 30] A lower risk of EAC by having a gastric or duodenal ulcer operation could be explained by the inverse associations between EAC risk and two common causes of peptic ulcer diseases, i.e. *Helicobacter pylori* infection and use of non-steroidal anti-inflammatory drugs (NSAIDs).[2] Thus, inclusion of these predictors in our model is biologically plausible and consistent with the existing evidence, which is crucial for the robustness of the prediction model.

Strengths of this study include the population-based design with high participation rates and a well-defined study base, strict random sampling of population control subjects, rapid and complete case ascertainment throughout the whole nation, uniform histological confirmation of all case patients, and personal interviews with all study participants. Moreover, the performance of developed risk prediction models was assessed by cross-validation, which may have reduced the risk of over-fitting.

Similar to other risk models, the developed models in this study were based on a case-control design, which might be subject to information misclassification on risk factors. However, the misclassification was less likely to be differential among cases and controls, since participants were unaware of etiological hypotheses and treated equally by professional interviewers who did not work in healthcare. Furthermore, we also interviewed 167 patients of squamous-cell carcinoma of the esophagus and observed strongly divergent associations between major predictors and risk of esophageal malignancies by histological subtypes, which argue against differential misclassification of exposures, i.e. recall bias.[13, 15-17] We were not able to examine the contribution of use of NSAIDs to predicting the risk of EAC because of lack of such data. However, use of NSAIDs is associated with only a moderately altered risk of EAC,[2] and it might have been partly assessed by the variable surgery for gastric or duodenal ulcer. Therefore, it was unlikely that our models lost predicting precision to any great extent. A further limitation is that we did not include age and sex in the logistic models as EAC cases and control subjects were frequency-matched by these two variables.

With data from a case-control study, we assessed the performance of the models only in terms of discriminative ability, instead of calibration of predicted risk as performed in cohort studies.[28] In addition, since there remains a risk of over-estimation of the model performance, these models need to be validated by independent external populations. Furthermore, although the sample size of this study was calculated for examining the effects of major risk factors of EAC, there is some uncertainty regarding the statistical power of the absolute risk of EAC.

To our knowledge, there are only two previous prediction model developed to estimate the absolute risk of EAC.[27, 31] Only one of them estimated the risk in the general population based on a case-control study in Australia [27], while the other one predicted the age- and sex-specific EAC incidence in American white non-Hispanics merely depending on reflux symptoms.[31]

Compared to our study, the Australian study had a larger sample size (364 EAC cases and 1580 controls), but was more vulnerable for selection bias as suggested by a lower coverage of EAC patients from the source population (around 35% of all incident EAC cases), and lower participation rates (70% in case patients and 51% in control subjects). The final model from the Australian study included terms for BMI in the previous year, frequency of reflux symptoms or use of antireflux medication, tobacco smoking status, education, and frequency of NSAIDs use. The performance of the model was slightly lower than our model, as indicated by the AUC of 0.75 after cross-validation, and although it increased to 0.85 after addition of alarm symptoms (dysphagia and unexplained weight loss), this might not be entirely accurate for a prediction model to include symptoms of the EAC itself. When we assessed the alarm symptoms of dysphagia and chest pain in the past 5 years, the alarm symptoms were inversely associated with EAC risk after adjustment for major risk factors, suggesting the existence of over-adjustment. Therefore, we did not include alarm symptoms in our models. As acknowledged by the authors, a potential limitation of the Australian study was that BMI during the year prior to interview was included in the model, which did not allow for a latency period between exposures and the onset of EAC. In our study, we assessed BMI estimated at different time points, including 20 years before the interview, at the age of 20 years, the highest and the lowest during adulthood, and found that BMI 20 years before the interview fitted the model best. Compared with existing risk prediction models for other types of cancer,[9, 29, 32, 33] the two models for EAC presented in this study seem to have more favorable discriminative accuracy as indicated by AUC statistics. In addition, the Australian model identified variables similar to the simple model in the present study, i.e., BMI, reflux symptoms and medication, and tobacco smoking, suggesting that these factors are likely to be important in any model for predicting EAC risk as they have been identified in two independent studies in geographically different populations. Yet, there is a need

for more research assessing risk prediction models for EAC in different populations, for validation and refinement of existing models or derivation of alternatives.

Clinicians have increasingly been performing upper endoscopy on patients with reflux symptoms for the purpose of early detection of EAC or a premalignant Barrett's mucosa. Although endoscopy in patients with reflux diseases would capture over 90% of EAC cases arising from these patients who undergo the endoscopy, unselective endoscopic surveillance of reflux patients remains unfeasible, given the high prevalence of reflux symptoms (20%) and the low absolute risk (~20/100 000 person-years of EAC). Furthermore, around 40% of EAC cases have no reflux symptoms and would have been "lost" using such a strategy.[34] The Clinical Guidelines Committee of the American College of Physicians proposed upper endoscopy in men aged >50 years with long-lasting reflux symptoms and other risk factors.[35] However, as acknowledged in the guideline, the recommendations were based mainly on expert opinions, which largely warranted proper scientific evaluation to support evidence-based practice.

Risk prediction models can provide individualized estimates of absolute risk of EAC based on personal information. This would help clinicians and patients to determine their practice regarding endoscopy. Public health researchers and decision-makers may want to stratify the population based on the absolute risks estimated from valid prediction models, and design endoscopic screening programs in stratums with high absolute risk of EAC, or target future interventions for prevention. Such efforts will contribute to a move towards more tailored detection and prevention of EAC. However, necessary thresholds of predicted risk for clinical and public health practice still need to be carefully determined based on further investigations balancing the predicted absolute risk of esophageal adenocarcinoma, costs of related clinical practice, potential benefits for patients, as well as risk of complications patients may experience. Our model may also be applied to other populations with a similar ethnic and social background,

e.g. other Nordic and European populations, if further validated. Particularly, the simple model included information only on reflux symptoms or use of antireflux medication, BMI, and tobacco smoking, which may be readily available in similar studies and easily captured in routine medical records. Thus, this simple model may be more widely used in public health and clinical practice, as well as for external validation.

In summary, the prediction models for EAC based on information on readily identifiable risk factors can be used to estimate individuals' absolute 5-year risk of EAC. The developed models had good discriminative accuracy, but need to be validated in other populations. By identifying individuals with high absolute risk of EAC, prediction models would be useful for a more tailored detection of EAC.

Acknowledgements

This work is supported by the Swedish Research Council (SIMSAM) [D0547801] and the Swedish Cancer Society [14 0322] to JL.

References

1. Edgren G, Adami HO, Weiderpass E, et al. A global assessment of the oesophageal adenocarcinoma epidemic. Gut 2013;62(10):1406-14.
2. Lagergren J, Lagergren P. Recent developments in esophageal adenocarcinoma. CA Cancer J Clin 2013;63(4):232-48.
3. Rustgi AK, El-Serag HB. Esophageal carcinoma. N Engl J Med 2014;371(26):2499-509.
4. Arnold M, Soerjomataram I, Ferlay J, et al. Global incidence of oesophageal cancer by histological subtype in 2012. Gut 2015;64(3):381-7.
5. Lagarde SM, ten Kate FJ, Reitsma JB, et al. Prognostic factors in adenocarcinoma of the esophagus or gastroesophageal junction. J Clin Oncol 2006;24(26):4347-55.
6. Lagergren J, Lagergren P. Oesophageal cancer. BMJ 2010;341:c6280.
7. Bird-Lieberman EL, Fitzgerald RC. Early diagnosis of oesophageal cancer. Br J Cancer 2009;101(1):1-6.
8. Lao-Sirieix P, Fitzgerald RC. Screening for oesophageal cancer. Nat Rev Clin Oncol 2012;9(5):278-87.
9. 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(12):951-61.
10. Win AK, Macinnis RJ, Hopper JL, et al. Risk prediction models for colorectal cancer: a review. Cancer Epidemiol Biomarkers Prev 2012;21(3):398-410.
11. van Dieren S, Beulens JW, Kengne AP, et al. Prediction models for the risk of cardiovascular disease in patients with type 2 diabetes: a systematic review. Heart 2012;98(5):360-9.
12. Thrift AP, Whiteman DC. Can we really predict risk of cancer? Cancer Epidemiol 2013;37(4):349-52.
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(11):825-31.
14. Lagergren J, Wang Z, Bergstrom R, et al. Human papillomavirus infection and esophageal cancer: a nationwide seroepidemiologic case-control study in Sweden. J Natl Cancer Inst 1999;91(2):156-62.
15. Lagergren J, Bergstrom R, Nyren O. Association between body mass and adenocarcinoma of the esophagus and gastric cardia. Ann Intern Med 1999;130(11):883-90.
16. 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(3):340-6.
17. Jansson C, Johansson AL, Nyren O, et al. Socioeconomic factors and risk of esophageal adenocarcinoma: a nationwide Swedish case-control study. Cancer Epidemiol Biomarkers Prev 2005;14(7):1754-61.
18. Lagergren K, Mattsson F, Lagergren J. Abdominal fat and male excess of esophageal adenocarcinoma. Epidemiology 2013;24(3):465-6.
19. Thrift AP, Risch HA, Onstad L, et al. Risk of esophageal adenocarcinoma decreases with height, based on consortium analysis and confirmed by mendelian randomization. Clin Gastroenterol Hepatol 2014;12(10):1667-76 e1.
20. Terry P, Lagergren J, Hansen H, et al. Fruit and vegetable consumption in the prevention of oesophageal and cardia cancers. Eur J Cancer Prev 2001;10(4):365-9.
21. Lagergren J, Ye W, Lindgren A, et al. Heredity and risk of cancer of the esophagus and gastric cardia. Cancer Epidemiol Biomarkers Prev 2000;9(7):757-60.
22. Lagergren J, Bergstrom R, Adami HO, et al. Association between medications that relax the lower esophageal sphincter and risk for esophageal adenocarcinoma. Ann Intern Med 2000;133(3):165-75.
23. Freedman J, Lagergren J, Bergstrom R, et al. Cholecystectomy, peptic ulcer disease and the risk of adenocarcinoma of the oesophagus and gastric cardia. Br J Surg 2000;87(8):1087-93.

24. Lunetta KL, Hayward LB, Segal J, et al. Screening large-scale association study data: exploiting interactions using random forests. BMC Genet 2004;5:32.
25. Wu IC, Zhao Y, Zhai R, et al. Association between polymorphisms in cancer-related genes and early onset of esophageal adenocarcinoma. Neoplasia 2011;13(4):386-92.
26. SAS Institute Inc. SAS/STAT 9.2 User's Guide. Cary, NC, USA: SAS Institute Inc.; 2008.
27. Thrift AP, Kendall BJ, Pandeya N, et al. A model to determine absolute risk for esophageal adenocarcinoma. Clin Gastroenterol Hepatol 2013;11(2):138-44 e2.
28. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. Epidemiology 2010;21(1):128-38.
29. 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(22):3590-6.
30. Johnson NJ, Backlund E, Sorlie PD, et al. Marital status and mortality: the national longitudinal mortality study. Ann Epidemiol 2000;10(4):224-38.
31. Rubenstein JH, Scheiman JM, Sadeghi S, et al. Esophageal adenocarcinoma incidence in individuals with gastroesophageal reflux: synthesis and estimates from population studies. Am J Gastroenterol 2011;106(2):254-60.
32. Gail MH, Brinton LA, Byar DP, et al. Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. J Natl Cancer Inst 1989;81(24):1879-86.
33. D'Amelio AM, Jr., Cassidy A, Asomaning K, et al. Comparison of discriminatory power and accuracy of three lung cancer risk models. Br J Cancer 2010;103(3):423-9.
34. Vaughan TL, Fitzgerald RC. Precision prevention of oesophageal adenocarcinoma. Nat Rev Gastroenterol Hepatol 2015;12(4):243-8.
35. Shaheen NJ, Weinberg DS, Denberg TD, et al. Upper endoscopy for gastroesophageal reflux disease: best practice advice from the clinical guidelines committee of the American College of Physicians. Ann Intern Med 2012;157(11):808-16.

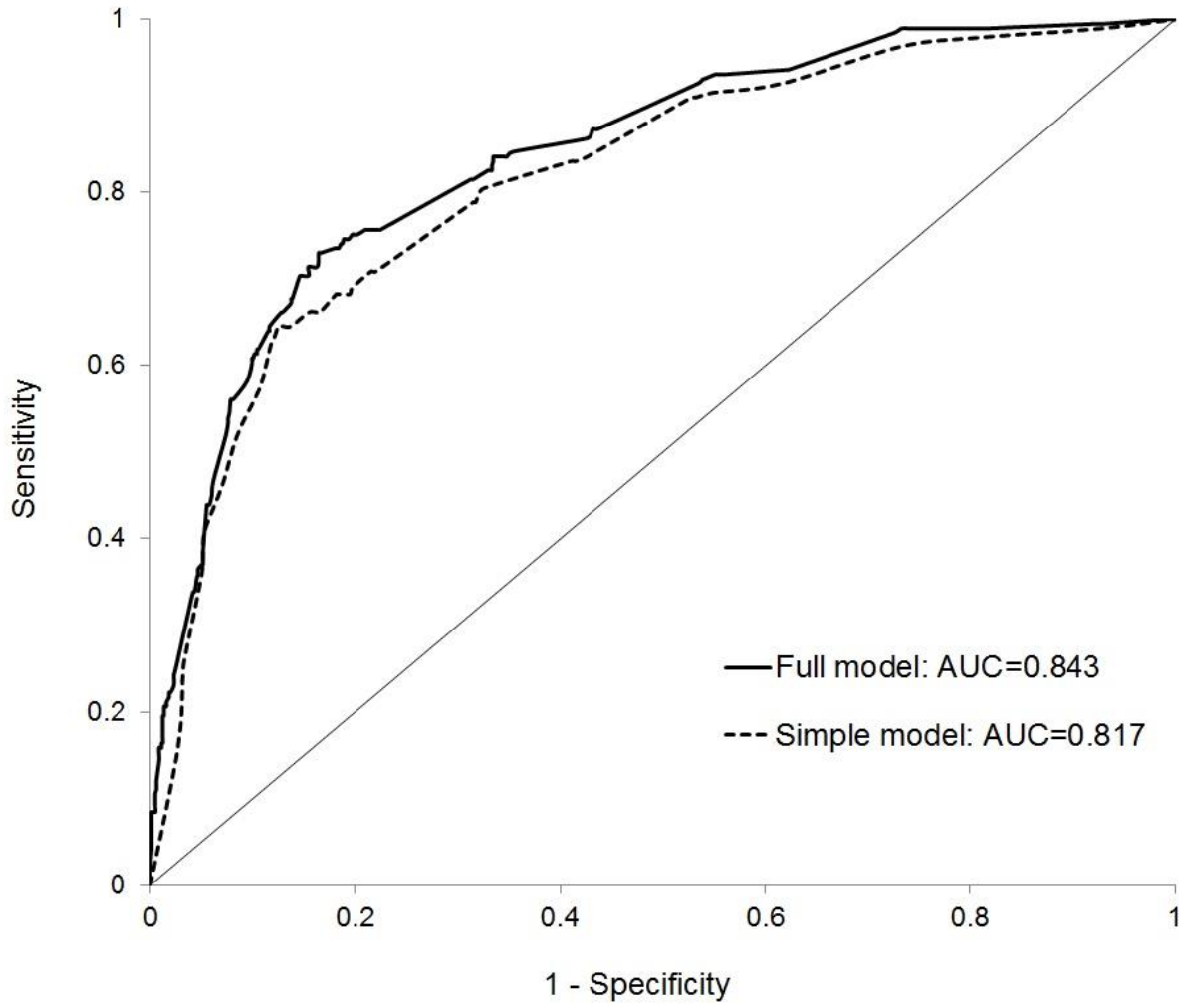


Figure 1. The receiver operating characteristic (ROC) curve based on a more extensive model as well a simpler model.

Table 1. Participation and characteristics of study subjects

Variables	Controls	Cases
No. of participants (% of all eligible)	820 (73)	189 (88)
Reasons for non-participation, N (% of all eligible)		
Unwillingness	210 (19)	2 (1)
Physical/mental disorders or early death	70 (6)	25 (12)
Age, years		
< 50	48 (6)	7 (4)
50-59	161 (20)	31 (16)
60-69	245 (30)	61 (32)
70-79	366 (45)	90 (48)
Sex, N (%)		
Male	679 (83)	165 (87)
Female	141 (17)	24 (13)
Education, years		
< 10	499 (61)	142 (75)
10-12	161 (20)	24 (13)
> 12	160 (19)	23 (12)

Table 2. Estimated ORs from unconditional logistic regressions

Variables	Controls n (%)	Cases n (%)	Crude OR (95% CI)	Adjusted OR[a] (95% CI)	Adjusted OR [b] (95% CI)
Reflux symptoms or use of antireflux medications					
No weekly reflux, no medication	602 (73.4)	50 (26.5)	1.00 (reference)	1.00 (reference)	1.00 (reference)
With weekly reflux < 5 years, no medication	6 (0.7)	2 (1.1)	3.99 (0.78, 20.27)	1.63 (0.27, 9.84)	2.42 (0.46, 12.72)
With weekly reflux ≥ 5 years, no medication	30 (3.7)	14 (7.4)	5.58 (2.78, 11.21)	6.19 (2.95, 12.98)	5.94 (2.87, 12.27)
No weekly reflux, with medication	79 (9.6)	26 (13.8)	3.94 (2.32, 6.68)	3.21 (1.81, 5.70)	3.68 (2.12, 6.37)
With weekly reflux < 5 years, with medication	16 (2.0)	6 (3.2)	4.49 (1.68, 11.97)	4.10 (1.26, 13.27)	4.19 (1.48, 11.86)
With weekly reflux ≥ 5 years, with medication	87 (10.6)	91 (48.2)	12.51 (8.29, 18.89)	8.01 (5.06, 12.67)	10.91 (7.11, 16.75)
BMI 20 years before interview, quartiles (kg/m^2) [c]					
First (men < 22.3; women < 21.1)	205 (25.1)	12 (6.4)	1.00 (reference)	1.00 (reference)	1.00 (reference)
Second (men 22.3-23.9; women 21.1-22.4)	207 (25.4)	26 (13.8)	2.15 (1.05, 4.37)	2.21 (1.01, 4.86)	2.06 (0.98, 4.35)
Third (men 24.0-25.5; women 22.5-24.2)	203 (24.9)	53 (28.0)	4.46 (2.32, 8.60)	3.79 (1.82, 7.92)	3.51 (1.76, 7.01)
Fourth (men > 25.5; women >24.2)	201 (24.6)	98 (51.9)	8.33 (4.44, 15.64)	7.47 (3.68, 15.17)	7.22 (3.71, 14.06)
Tobacco smoking	495 (60.4)	132 (70.0)	1.52 (1.08, 2.13)	1.62 (1.08, 2.44)	1.44 (0.98, 2.13)
Living with a partner for less than one year	44 (5.4)	26 (13.8)	2.80 (1.68, 4.68)	3.50 (1.86, 6.59)	-
Previously diagnosed esophagitis	15 (1.8)	21 (11.1)	6.68 (3.37, 13.22)	3.08 (1.22, 7.79)	-
Previously diagnosed diaphragmatic hernia	37(4.5)	47 (24.9)	6.97 (4.37, 11.11)	2.56 (1.41, 4.66)	-
Having a specific gastrointestinal operation previously					
Gastric/duodenal ulcer operation	31 (3.78)	4 (2.12)	0.55 (0.19, 1.57)	0.32 (0.10, 1.05)	-
Operation for esophagitis, diaphragmatic hernia, or severe reflux	6 (0.7)	14 (7.4)	10.80 (4.09, 28.49)	3.00 (0.96, 9.41)	-
Population attributable risk				0.94	0.91

BMI: body mass index; CI: confidence interval; OR: odds ratio

[a] Adjusted estimates in the final model including all listed variables; [b] Adjusted estimates in the simple model including reflux symptoms and/or use of antireflux medications, BMI, and tobacco smoking; [c] Four controls with missing data on BMI were excluded from analyses.

Table 3. Statistics for the performance of developed logistic risk-prediction models

Model	Original without cross-validation		Leave-one-out cross-validation		10-fold cross-validation	
	AUC (95% CI)	Somers' D	AUC (95% CI)	Somers' D	AUC (95% CI)	Somers' D
Full model	0.843 (0.811, 0.874)	0.685	0.818 (0.784, 0.852)	0.636	0.828 (0.795, 0.860)	0.655
Simple model	0.817 (0.783, 0.852)	0.635	0.791 (0.754, 0.828)	0.582	0.804 (0.769, 0.839)	0.608

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

Table 4. Estimated absolute 5-year risks for esophageal adenocarcinoma with selected profiles of risk factors

Profile	Sex	Age, years	Weekly reflux symptoms	Antireflux medication	BMI 20 years ago, kg/m^2	Tobacco smoking	Absolute 5-year risk, 1/100 000	Number of individuals needed to survey to detect one case
1	Male	50-54	No	No	22.3-23.9	Never	1.8	55316
2	Male	50-54	5 years or more	No	22.3-23.9	Never	5.2	19231
3	Male	50-54	5 years or more	No	> 25.5	Ever	54.2	1845
4	Male	50-54	less than 5 years	Yes	24.0-25.5	Ever	16.3	6137
5	Male	60-64	5 years or more	Yes	> 25.5	Ever	226.2	442
6	Male	60-64	5 years or more	Yes	> 25.5	Never	157.0	637
7	Male	60-64	less than 5 years	No	> 25.5	Ever	50.3	1989
8	Male	70-74	5 years or more	Yes	> 25.5	Ever	533.0	188
9	Female	60-64	5 years or more	Yes	> 25.5	Ever	24.3	4114
10	Female	70-74	5 years or more	Yes	> 25.5	Ever	68.9	1450

BMI: body mass index

Supplementary methods 1. Description of predictor variables

1. Reflux symptoms or use of antireflux medications

Questions were asked about recurrent heartburn and regurgitation, which are the two major symptoms of gastro-esophageal reflux, in terms of having symptoms or not, frequency, and duration. We also asked subjects whether they had taken any medicine for the symptoms of heartburn or regurgitation. We disregarded symptoms that had occurred less than five years before the interview to avoid collecting data on symptoms and related medications caused by EAC.

To avoid the problem of multi-collinearity when closely correlated variables are entered into the logistic model, we constructed one variable combining information on both frequency and duration of reflux symptoms, together with use of antireflux medications. This variable categorized subjects into six groups and was entered into the models as a dummy variable.

2. BMI 20 years before the interview

Subjects were asked to report their height and weight 20 years before the interview. BMI was calculated as body weight in kilograms divided by the square of body height in meters (kg/m^2). To obtain an independent measure of body fat, we asked each interviewee to choose the picture that best resembled his or her body build 20 years ago from a pictogram that showed nine somatotypes ranging from very lean to grossly obese. The Spearman correlation coefficient with BMI is 0.5 to 0.8. Subjects were categorized into four groups using cut-off points for sex-specific quartiles in controls.

3. Tobacco smoking

Subjects were asked about their lifetime smoking history of cigarettes, cigars, and pipes. Tobacco users were defined as individuals smoking regularly (at least one cigarette per day or at least one cigar or pipe per week) or taking a quid of snuff at least once a week for at least six months. Smoking status was determined by tobacco usage two years before the interview.

4. Living with a partner

Subjects were asked whether they had ever been married or “sambo” (Swedish term for cohabitation) for at least one year.

5. Medical history

Subjects were asked about their past medical history, including separate terms of various gastrointestinal diseases and operations, until 5 years before the interview.

Supplementary methods 2. Estimation of absolute 5-year risks of EAC

We calculated the absolute 5-year risks for all possible profiles of risk factors, based on the following information:

1. Relative risk for the individual

The relative risk associated with a specific profile of risk factors was calculated as the product of the odds ratios for individual risk factors.

2. Baseline age- and sex- specific incidence rates

We obtained the age- and sex-specific incidence rates of EAC from the Swedish Cancer Register, which are presented in the supplementary table.

3. Population attributable risk of the model

The population attributable risk of the model was calculated by the following formula:

$$Population\ attributable\ risk = 1 - \frac{1}{x}\sum_{i=1}^{x}\left(\frac{1}{r_i}\right)$$

where x was the number of EAC cases, r_i was the relative risk for the ith case estimated from the logistic regression model (*Bruzzi et al. Am J Epidemiol 1985;122:904-913*).

4. The age- and sex-specific mortality rates excluding esophageal cancer

We calculated the age- and sex-specific mortality rates excluding esophageal cancer using the population mortality data from Statistic Sweden, and the mortality data from Nordcan database.

For an individual with the age of t (in five-year groups), sex of s (1=male, 2=female), and relative risk of r, we first calculated the baseline hazard as:

$$b_{1(t,s)} = IR_{(t,s)} \times (1 - population\ attributable\ risk)$$

where $IR_{(t,s)}$ was the age- and sex-specific incidence rate of EAC in the population.

We estimated the absolute risk of EAC over 5 years as:

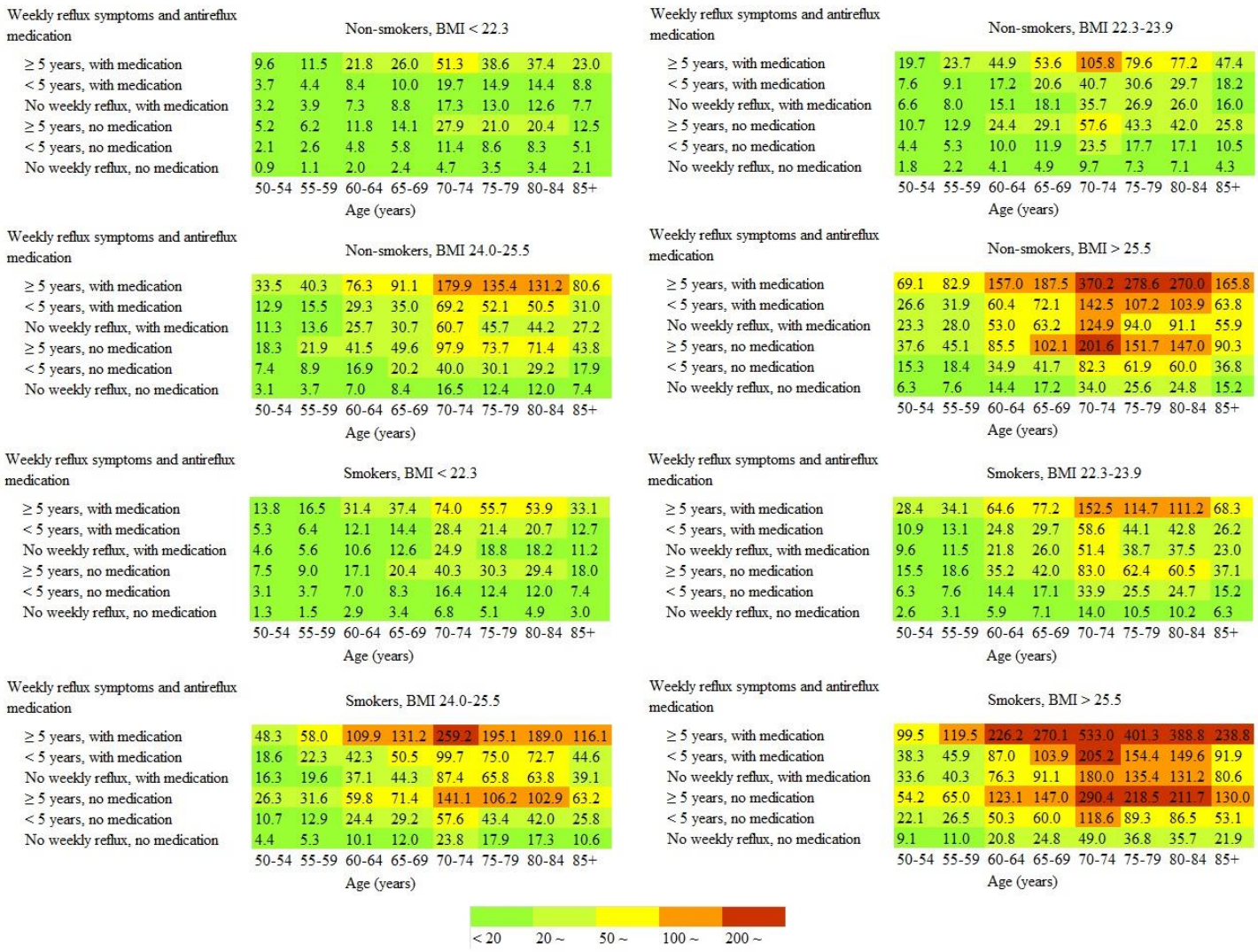
$$P(t,s,r) = \frac{b_{1(t,s)}r}{b_{1(t,s)}r + b_{2(t,s)}} \times \left(1 - e^{-5\left(b_{1(t,s)}r + b_{2(t,s)}\right)}\right)$$

where $b_{2(t,s)}$ was the age- and sex-specific mortality rate from competing causes.

Supplementary table S1. The age- and sex-specific incidence rates of esophageal adenocarcinoma and population mortality rates excluding esophageal cancer in Sweden, 1995-1997 (1/100 000)

Age, years	Men		Women	
	EAC incidence	Mortality	EAC incidence	Mortality
30-34	0.1	83.9	0.0	43.4
35-39	0.3	119.0	0.0	60.4
40-44	0.3	175.8	0.0	98.6
45-49	1.1	267.0	0.0	164.4
50-54	2.0	429.5	0.1	280.9
55-59	2.4	665.6	0.3	408.4
60-64	4.6	1188.7	0.5	661.3
65-69	5.6	1989.4	0.5	1050.7
70-74	11.4	3317.4	1.4	1798.8
75-79	9.0	5701.3	2.4	3174.1
80-84	9.7	10038.5	4.1	6158.1
85+	7.6	21051.2	3.1	15948.1

Data source: Swedish Cancer Register, Statistics Sweden, and NordCan database.



Supplementary figure S1. Absolute 5-year risks of esophageal adenocarcinoma (1/100 000) estimated from a model based on various profiles of the major risk factors (reflux symptoms or use of antireflux medications, body mass index, and tobacco smoking) in men aged 50 years or above.