Clinical Gastroenterology and Hepatology 2018;16:1229–1236

ALIMENTARY TRACT

Model for Identifying Individuals at Risk for Esophageal Adenocarcinoma

Andrew T. Kunzmann,[*] Aaron P. Thrift,[‡] Chris R. Cardwell,[*] Jesper Lagergren,[§,‖] Shaohua Xie,[§] Brian T. Johnston,[¶] Lesley A. Anderson,[*] John Busby,[#] Úna C. McMenamin,[*] Andrew D. Spence,[*] and Helen G. Coleman[*,**]

*Cancer Epidemiology and Health Services Research Group, Centre for Public Health, Queen's University Belfast, Belfast, Northern Ireland, United Kingdom; [‡]Section of Epidemiology and Population Sciences, Department of Medicine, Baylor College of Medicine, Houston, Texas; [§]Upper Gastrointestinal Surgery, Department of Molecular Medicine and Surgery, Karolinska Institutet, Karolinska University Hospital, Stockholm, Sweden; [‖]School of Cancer and Pharmaceutical Sciences, King's College London, London, United Kingdom; [¶]Royal Victoria Hospital, Belfast Health and Social Care Trust, Belfast, Northern Ireland, United Kingdom; [#]Centre for Experimental Medicine, Queen's University Belfast, Belfast, Northern Ireland, United Kingdom; and **Centre for Cancer Research and Cell Biology, Queen's University Belfast, Belfast, Northern Ireland, United Kingdom

BACKGROUND & AIMS: **The prognosis for most patients with esophageal adenocarcinoma (EAC) is poor because they present with advanced disease. Models developed to identify patients at risk for EAC and increase early detection have been developed based on data from case–control studies. We analyzed data from a prospective study to identify factors available to clinicians that identify individuals with a high absolute risk of EAC.**

METHODS: **We collected data from 355,034 individuals (all older than 50 years) without a prior history of cancer enrolled in the UK Biobank prospective cohort study from 2006 through 2010; clinical data were collected through September 2014. We identified demographic, lifestyle, and medical factors, measured at baseline, that associated with development of EAC within 5 years using logistic regression analysis. We used these data to create a model to identify individuals at risk for EAC. Model performance was assessed using area under the receiver operating characteristics curve (AUROC), sensitivity, and specificity analyses.**

RESULTS: **Within up to 5 years of follow up, 220 individuals developed EAC. Age, sex, smoking, body mass index, and history of esophageal conditions or treatments identified individuals who developed EAC (AUROC, 0.80; 95% CI, 0.77–0.82). We used these factors to develop a scoring system and identified a point cut off that 104,723 individuals (29.5%), including 170 of the 220 cases with EAC, were above. The scoring system identified individuals who developed EAC with 77.4% sensitivity and 70.5% specificity. The 5-year risk of EAC was 0.16% for individuals with scores above the threshold and 0.02% for individuals with scores below the threshold.**

CONCLUSION: **We combined data on several well-established risk factors that are available to clinicians to develop a system to identify individuals with a higher absolute risk of EAC within 5 years. Studies are needed to evaluate the utility of these factors in a multi-stage, triaged, screening program.**

Keywords: BMI; Upper Gastrointestinal Cancer; Risk-Prediction; Esophagus.

The incidence of esophageal adenocarcinoma (EAC) is increasing in Western populations,[1] and prognosis is poor. Overall survival is <20%.[2] Methods to improve early diagnosis are important[3] because the poor survival is largely attributable to late clinical presentation with advanced disease.[2]

Endoscopy with biopsies for histologic confirmation is the gold standard method of detecting esophageal and gastric cancers.[4,5] However, population-wide endoscopy screening programs are unlikely to be cost-effective or feasible because of the low incidence of upper gastrointestinal cancers and the cost, invasiveness, and psychological burden of endoscopy screening.[6]

Abbreviations used in this paper: **AUROC, area under the receiver operating characteristic curve; BMI, body mass index; CI, confidence interval; EAC, esophageal adenocarcinoma; GERD, gastroesophageal reflux disease; ICD, International Classification of Diseases; NSAIDs, nonsteroidal anti-inflammatory drugs.**

Most current article

1542-3565/$36.00
https://doi.org/10.1016/j.cgh.2018.03.014

Clinical guidelines suggest that individuals with chronic symptomatic gastroesophageal reflux disease (GERD) should be referred for endoscopy screening to identify Barrett's esophagus,[7] a precursor to EAC present in 1%–2% of the adult population.[8] Individuals with Barrett's esophagus are often entered into endoscopy surveillance programs.[9] Use of GERD as the sole initial triage factor for endoscopy is appealing in its simplicity, but delays in obtaining endoscopy are common in part because of a lack of both specificity, because risk of EAC in individuals with GERD is low,[10,11] and sensitivity, because a large proportion of EAC cases (~40%) never report GERD symptoms.[11]

Automated clinical risk prediction tools using data easily accessible to primary care physicians to flag high-risk individuals may offer a lower cost method to improve early detection and are being investigated for other conditions.[12] Established risk factors including body mass index (BMI), smoking status, and GERD may be useful for risk prediction of EAC[13–15] and Barrett's esophagus.[16–18] However, existing studies have relied on age and sex matched case-control data, which may be subject to recall bias and ignore age and sex in improving the risk prediction, which are strong risk factors.[19] An assessment of traditional risk factors may be useful as triage directly to endoscopy, in which case risk prediction for any upper gastrointestinal cancers diagnosed via endoscopy would be most informative. However, it seems likely that EAC risk prediction will require a multistage screening strategy,[20] of which some steps may be specific to EAC.

We aimed to use prospective data from the UK Biobank to develop a risk-prediction model that is based on a combination of factors widely available to clinicians that may predict risk of EAC development within 5 years. Secondary analyses aimed to develop and assess a risk-prediction model that predicts risk of all upper gastrointestinal cancers detected via endoscopy.

Methods

Study Design

This cohort study used prospective data from the UK Biobank, which recruited 502,640 men and women aged 40–69 years from 1 of 22 centers located across England, Scotland, and Wales between 2006 and 2010.[21] Approximately 9.2 million individuals registered with the National Health Service living within a 25-mile radius of 1 of the 22 centers were invited to participate. The response rate was 5.5% (n = 503,325).[22] Included in the present study were individuals aged ≥50 years (because upper gastrointestinal cancers are rare in those aged <50), without a history of cancer (excluding non-melanoma skin cancer) at or before baseline or within 6 months after baseline (to exclude diagnostic delays) and with complete information on relevant risk factors.

The UK Biobank was approved by the North West Multi-Centre Research Ethics Committee, and all participants provided written informed consent.

Assessment and Classification of Candidate Predictor Variables

Participants were asked to complete electronic touchscreen questionnaires at baseline, which inquired about a wide range of potential risk factors for chronic diseases, and have anthropometric measurements taken (Supplementary Methods). Self-reported current medication use, medical history, and surgical history were assessed via the electronic touchscreen questionnaire and were verified during a face-to-face interview with a trained nurse with responses matched to a coding tree, where possible, by a doctor.

Candidate predictor variables widely available to clinicians, such as age and BMI, were identified and categorized a priori from the literature (Table 1).[23–27] Additional candidate predictor variables not thought to be widely available to clinicians, such as waist: hip ratio and others that do not have as strong evidence for an association with EAC, were treated separately and only included if they added independent predictive value (Supplementary Table 1).

Outcome Assessment

The UK Biobank is regularly linked to UK cancer registry data from the Health and Social Care Information Centre (in England and Wales), the Scottish Cancer Registry (in Scotland), and death records from the UK Office of National Statistics. Cancer data were provided up until September 30, 2014. Newly diagnosed cancers were classified by site according to International Classification of Diseases, 10th version (ICD/10) and histology (ICD-O morphology codes). Primary EACs (ICD/10 C15, with ICD-O 8140–8573) diagnosed between 6 months (because of potential diagnostic delays) and 5 years from baseline were the main outcome of interest. Secondary outcomes included all primary upper gastrointestinal cancers (ICD/10 C15 and C16), esophageal cancer (ICD-10 C15, regardless of histology), gastric cancer (ICD-10 C16, regardless of histology), and esophageal squamous cell carcinoma (ICD-10 C15, ICD-O 8050–8082) diagnosed between 6 months and 5 years from baseline. Information on tumor stage was not available.

Statistical Analysis

Stepwise logistic regression was used to estimate factors widely available to clinicians associated with risk of EAC within 5 years ($P < .1$).

Points-based models were created from the coefficient-based model by dividing the coefficient of

Table 1. Characteristics of Study Population in Terms of Candidate Predictor Variables at the Baseline Assessment Visit Who Did or Did Not Develop Esophageal Adenocarcinoma Within 5 Years

	No esophageal adenocarcinoma, n (%)	Esophageal adenocarcinoma, n (%)
Total	354,814 (100.0)	220 (100.0)
Age (*y*)[23]		
50–<55	72,499 (20.4)	16 (7.3)
55–<60	85,090 (24.0)	39 (17.7)
60–<65	111,494 (31.4)	75 (34.1)
65+	85,731 (24.2)	90 (40.9)
Sex[23]		
Female	190,846 (53.8)	35 (15.9)
Male	163,968 (46.2)	185 (84.1)
Smoking status[23]		
Never	188,024 (53)	56 (25.5)
Former	132,871 (37.4)	117 (53.2)
Current	33,919 (9.6)	47 (21.4)
Body mass index (kg/m^2)[23]		
<25	111,426 (31.4)	35 (15.9)
25–<30	154,262 (43.5)	102 (46.4)
30–<35	64,366 (18.1)	56 (25.5)
35+	24,760 (7.0)	27 (12.3)
Esophageal condition[a,23]		
No	306,216 (86.3)	161 (73.2)
Yes	48,598 (13.7)	59 (26.8)
Gastric condition[b,24]		
No	348,463 (98.2)	210 (95.5)
Yes	6351 (1.8)	10 (4.5)
Diabetes[25]		
No	334,093 (94.2)	194 (88.2)
Yes	20,721 (5.8)	26 (11.8)
Hypertension[c]		
No	140,275 (39.5)	55 (25.0)
Yes	214,539 (60.5)	165 (75.0)
Height[23]		
Above UK average[d]	178,480 (50.3)	92 (41.8)
Below UK average[d]	176,334 (49.7)	128 (58.2)
Asthma inhaler use [26,27]		
No	329,864 (93.0)	197 (89.5)
Yes (current)	24,950 (7.0)	23 (10.5)
Statin use[23]		
No	283,430 (79.9)	142 (64.5)
Yes (current)	71,384 (20.1)	78 (35.5)
NSAID use[23]		
No	249,903 (70.4)	143 (65.0)
Yes (current)	104,911 (29.6)	77 (35.0)

NSAID, nonsteroidal anti-inflammatory drug.
[a]Esophageal conditions included self-reported history of gastroesophageal reflux disease, Barrett's esophagus, hiatus hernia, or esophageal stricture and/or esophageal fundoplication or hiatus hernia surgery and/or anti-reflux medication use (none or any).
[b]Gastric conditions included self-reported history of gastric ulcers, gastritis/gastric erosions, pyloric stenosis, and/or gastric surgery for gastric/peptic/duodenal ulcer surgery, gastrectomy, and/or *Helicobacter pylori* eradication therapy (proton pump inhibitors and clarithromycin and/or amoxicillin).
[c]Medical history or systolic blood pressure >140 mm Hg at baseline.
[d]Sex-specific UK average: 1.753 m for men and 1.619 m for women.

each variable by the smallest coefficient in the model and rounding to the nearest 0.5 to allow ease of calculation without a computer and easier to interpret cutoffs.[28] For example, the coefficient for men was 1.64, and the smallest coefficient in the model was 0.40 (BMI, 25 to <30 kg/m^2), so men were assigned 4 points (1.64/0.40, then rounded to nearest 0.5).

Diagnostic accuracy was quantified for the simple factor coefficient- and points- based model, additional factor coefficient- and points- based model by using the area under the receiver operating characteristic (AUROC) curve with 95% confidence interval (CI). An internally validated AUROC was calculated by using bootstrap methods described by Steyerberg et al,[29] accounting for optimism in model selection (including the stepwise selection procedure) and performance. Goodness-of-fit was assessed by using Hosmer-Lemeshow tests and calibration curves.

AUROC, sensitivity, specificity, Youden's index (sensitivity + specificity −1), risk of EAC within 5 years (equivalent to positive predictive value), and number of referrals for additional screens per cancer correctly predicted were assessed for individuals above each points-based cutoff threshold.

To assess whether a more complex model increased model performance (AUROC), additional factors not widely available to clinicians or that may require additional examination were added to the initial model in turn (excluding individuals with missing data) and then with all factors significant at $P < .1$ added to a stepwise selection model.

A priori sensitivity/secondary analyses assessed the AUROC and stepwise model selection when the following:

- including individuals with a history of non-upper gastrointestinal cancers;
- conducting multiple imputation for missing variables (using 10 imputations and combined using Rubin's rules);
- using different cancer follow-up periods;
- using different age periods;
- using different categories for smoking and BMI and separate variables for each esophageal condition;
- stratifying by age, sex, BMI, smoking history, and esophageal condition status at baseline to check model performance for importance patient subgroups; and
- excluding individuals who reported a history of Barrett's esophagus or esophagitis because these may already be undergoing endoscopic surveillance.

Further secondary analyses assessed the discriminative ability of the risk-prediction model identified in the primary analyses for total upper gastrointestinal cancer, because the tool could be used as triage to

Table 2. Risk Factors Associated With Esophageal Adenocarcinoma in a Stepwise Logistic Regression and the Points Assigned for the Points-based Model

	No esophageal adenocarcinoma	Esophageal adenocarcinoma	Adjusted odds ratio (95% confidence interval)	Points[a]
Age (*y*)				
50–55	72,499	16	1.00 (reference)	0
55–60	85,090	39	1.99 (1.11–3.56)	1.5
60–65	111,494	75	2.76 (1.60–4.74)	2.5
65+	85,731	90	4.03 (2.36–6.89)	3.5
Sex				
Female	190,846	35	1.00 (reference)	0
Male	163,968	185	5.16 (3.58–7.44)	4.0
Body mass index (kg/m^2)				
<25	188,024	56	1.00 (reference)	0
25–<30	132,871	117	1.50 (1.02–2.21)	1.0
30–<35	33,919	47	1.91 (1.25–2.94)	1.5
35+			2.97 (1.79–4.94)	2.5
Smoking status	111,426	35		
Never	154,262	102	1.00 (reference)	0
Former	64,366	56	2.03 (1.47–2.80)	2.0
Current	24,760	27	3.83 (2.59–5.66)	3.5
Esophageal condition				
No	306,216	161	1.00 (reference)	0
Yes	48,598	59	1.88 (1.39–2.54)	1.5

NOTE. Factors excluded during stepwise selection: height, respiratory sympathomimetic use, statin medication use, diabetes, hypertension, or high baseline blood pressure.
[a]Points calculated by dividing the coefficient of each variable in a coefficient-based model by the smallest coefficient in the model (0.41) and rounding to the nearest 0.5 to allow ease of calculation without a computer and easier to interpret cutoffs.

endoscopy, which could detect all types of gastric and esophageal cancers. For comparison we also used stepwise logistic regression to estimate factors widely available to clinicians associated with risk of total upper gastrointestinal cancer within 5 years ($P < .1$) and assessed the discriminative ability of a new points-based model.

Analyses were conducted by using Stata/SE statistical software (version 14.1; College Station, TX).

Results

Participants

There were 502,640 participants in the UK Biobank, of whom 117,891 (23.5%) were excluded because they were aged younger than 50 years, 30,665 (6.1%) were excluded because of a history of cancer (or cancer within 6 months of baseline), and 4060 were excluded because of missing data (0.8%). This left 355,034 (70.7%) for inclusion in the final study cohort, among whom 220 individuals were diagnosed with EAC within 5 years. Individuals diagnosed with EAC were more likely to be older, male, smoke (current or former), have a higher BMI, have an existing esophageal or gastric condition, diabetes or hypertension, be below UK average height for their sex, and use nonsteroidal anti-inflammatory drugs (NSAIDs), statins, or asthma inhalers (Table 1). Mean follow-up time was 4.8 years (standard deviation, 0.6).

Non-participants

Individuals excluded because of incomplete data for some of the candidate predictor variables tended to be older, male, smoke (current), obese, tall, have an existing esophageal or gastric condition, diabetes or hypertension, or use NSAIDs, statins, or asthma inhalers.

Esophageal Adenocarcinoma Risk-Prediction Model: Coefficient-based Model

After applying the multi-phase stepwise procedure, the final coefficient-based model for predicting EAC development within 5 years included age at baseline, sex, BMI, smoking status, and history of diagnosis or treatment for esophageal conditions (Table 2). This model had good discrimination, with AUROC of 0.80 (95% CI, 0.77–0.82; Figure 1*A*). We found little evidence of overfitting in internal validation where the model showed equally good discriminatory ability (internally validated AUROC of 0.79). The performance of the model was statistically good by the goodness-of-fit test (Hosmer-Lemeshow test, χ^2 statistic = 6.58, $P = .58$) and the calibration curve (Figure 1*B*).

Esophageal Adenocarcinoma Risk-Prediction Model: Points-based Model

A points-based model assigned additional points on the basis of age (55–60 years, 1.5; 60–65 years, 2.5;

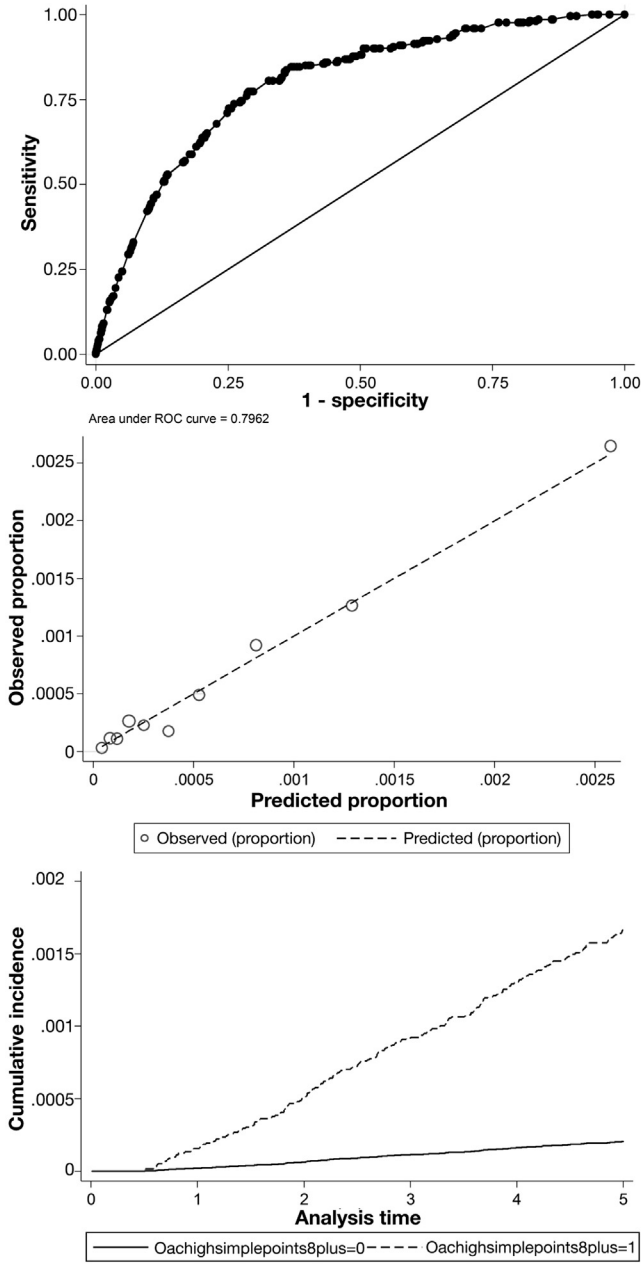

Figure 1. (*A*) ROC curve for the based model for predicting risk of EAC within 5 years. (*B*) Hosmer-Lemeshow calibration plot: observed versus predicted proportion, based on risk score, of individuals with EAC within 5 years. (*C*) Cumulative proportion developing EAC from 6 months to 5 years after baseline based on risk score (<8 points versus 8+ points).

65+ years, 3.5), sex (males, 4), smoking status (former, 2; current, 3.5), BMI (>25–30, 1; 30–<35, 1.5; 35+, 2.5), and history of esophageal conditions or treatment (1.5) (Table 2, Figure 2). The AUROC for the points-based model was similar (0.80; 95% CI, 0.77–0.82) to that of the coefficient-based model. The discriminative performances at each points-based cutoff threshold are provided in Table 3. A cutoff threshold of 8+ points with the highest Youden's index (0.48) had a sensitivity of 77.5%, a specificity of 70.5%, and a positive predictive value of 0.16% (Figure 1*C*) and would mean 612 referrals for

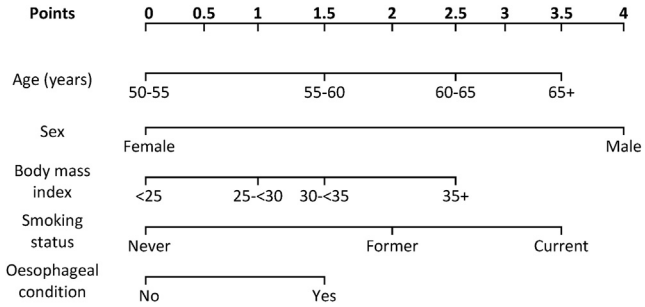

Figure 2. Nomogram for assigning points (out of total of 15) to help identify individuals at higher risk of EAC within 5 years.

further screening for every EAC predicted, with 29.5% (104,723 individuals) of the cohort (59.7% of men and 3.5% of women) deemed high risk. Of the individuals above the 8-point cutoff threshold, 76.4% (79,970 individuals) had not reported an esophageal medical history.

Sensitivity Analyses

No improvement in AUROC was apparent when additional variables not widely available to clinicians such as smoking status by pack-years and abdominal obesity were added to the model (Supplementary Table 2). The AUROC was similar in various sensitivity analyses, including when separating esophageal conditions by type and after multiple imputation (Supplementary Tables 3 and 4).

Upper Gastrointestinal Cancers (Gastric and Esophageal Cancers)

The factors selected in the model for upper gastrointestinal cancer (gastric or esophageal cancers) within 5 years were the same as the model selected for EAC (age, sex, BMI, smoking status, history of esophageal conditions or treatments). The AUROC for predicting any upper gastrointestinal cancer of 0.74 (95% CI, 0.72–0.76) or for predicting risk of any gastric cancer within 5 years of 0.72 (95% CI, 0.69–0.76) were lower than that observed for EAC. The AUROC for predicting risk of esophageal cancer of any histologic type within 5 years was 0.76 (95% CI, 0.73–0.79) (Supplementary Table 3).

Because of similarity in structure, for simplicity, we report results by using the EAC points system. The 8-point cutoff again had the highest AUROC for any upper gastrointestinal cancer (0.68; 95% CI, 0.66–0.71), sensitivity was 66.3%, specificity was 70.5%, and positive predictive value was 0.31% cancer. This would represent 312 referrals for screening for every upper gastrointestinal cancer predicted (compared with 612 for EAC alone) (Supplementary Table 5).

Discussion

This cohort study using prospective data from the UK Biobank identified how combining established risk

Table 3. Statistics for Performance of a Points-based Esophageal Adenocarcinoma Risk-Prediction Model at Different Points-based Cutoffs

Points cutoff	Patients deemed high risk (%)	True cancers predicted (out of 220)	Sensitivity	Specificity	Youden's index	Absolute 5-year risk of esophageal adenocarcinoma per 100,000[a]	Referrals for additional screens per esophageal adenocarcinoma[b] predicted[c]	Odds ratio (above vs below cutoff) (95% confidence interval)
5+	222,686 (62.7)	203	92.31	37.28	0.30	91.16	1096.98	7.13 (4.37–11.64)
6+	181,069 (51.0)	198	90.05	48.99	0.39	109.35	914.49	8.69 (5.61–13.45)
7+	143,516 (40.5)	187	85.07	59.53	0.45	130.30	767.47	8.38 (5.80–12.11)
8+	104,723 (29.5)	170	77.38	70.49	0.48	162.33	616.02	8.17 (5.97–11.18)
9+	72,840 (20.5)	140	63.80	79.48	0.43	192.20	520.29	6.83 (5.19–8.97)
10+	47,230 (13.3)	114	52.04	86.69	0.39	241.37	414.30	7.07 (5.43–9.20)
11+	22,982 (6.5)	66	30.32	93.52	0.24	287.18	348.21	6.28 (4.72–8.36)
12+	8,821 (2.5)	33	15.38	97.51	0.13	374.11	267.30	7.12 (4.95–10.25)

[a]In individuals above points cutoff.
[b]Within 5 years of baseline assessment.
[c]Not all cancers predicted will be detected during follow-up screens.

factors can aid risk prediction of EAC. The risk predictors identified included age at baseline, sex, tobacco smoking status, BMI, and a history of esophageal conditions or treatments including GERD (Table 2). The model was well-calibrated, and the discriminative performance was unchanged after internal validation using bootstrapping.

The factors included in the risk-prediction model are broadly consistent with factors identified in risk-prediction models of Barrett's esophagus and EAC by using case-control methodology,[13,14,16,18] which included BMI, smoking status, and esophageal conditions. However, the association between previous esophageal conditions and EAC risk was weaker than in previous studies,[13,14] perhaps because of the lack of a specific question on gastroesophageal reflux symptoms or the minimization of recall bias in the current study. This indicates that the findings are robust to methodological differences. Other potential risk factors for EAC, ie, NSAIDs, were not confirmed to have utility in risk prediction, although this does not exclude a role for such factors in disease etiology. The discriminative ability of the model was similar after accounting for optimism using bootstrap samples and in analyses stratified by age, BMI, smoking history, and history of esophageal conditions or treatments, suggesting the finding may be robust to changing characteristics of the model or population. The discriminative ability was lower when stratified by sex, which highlights the importance of sex as a predictive factor.

Additional factors tested that may not be so widely available to clinicians and may require face-to-face assessment, ie, waist: hip ratio, did not improve the model performance in terms of AUROC. Therefore, it may not be necessary to make the additional effort it may take clinicians to collect information on these factors. Thus, the risk-prediction model could be used to develop an automated algorithm linked to clinical records to flag high-risk individuals to clinicians for screening or lifestyle advice, which could lower the cost and time of administering the test.

Despite low cost of administration, the absolute 5-year risk of EAC in individuals above the 8-point cutoff threshold (162.3 per 100,000) is still fairly low, which may limit cost-effectiveness of referring these individuals for further screening. This is especially true if the test is used to triage directly to endoscopy, even when considering the higher absolute 5-year risk of any upper gastrointestinal cancer that can be detected by endoscopy in individuals above this threshold (313.20 per 100,000). Nevertheless, it could be argued that these estimates are higher than the estimated absolute 5-year incidence in GERD patients of ~100 per 100,000 in a recent review,[30] in whom guidelines recommend referral for endoscopy.[7]

This study focused on a risk-prediction tool for triage to an intermediate step such as blood test, breath tests, Cytosponge,[31] or capsule endoscopies.[32] These additional triage steps could improve specificity and better identify high-risk individuals in whom endoscopic screening may be cost-effective. Future studies should

assess whether blood-based biomarkers could further enhance the specificity of this clinical risk prediction model, as has been demonstrated for Barrett's esophagus risk prediction.[15,17] Full health economic modeling would be required to assess the cost-effectiveness of changes to screening practices.

Strengths and Limitations

A main strength is the cohort design using prospective data for this relatively rare cancer, which minimized recall bias and allowed the predictive value of age and sex to be estimated (which are typically matched on within case-control studies).

The self-report medical history without a specific question on gastroesophageal reflux symptoms lowers the accuracy of reporting of GERD history. The medication use was self-reported, which could limit the accuracy.[33] However, the follow-up interview with a trained health professional should reduce misreporting of medical history or medication use. We also did not have information on degree or duration of medication use for reflux symptoms, which added predictive value to a similar model in a previous study.[14] The generalizability of the UK Biobank to the general population has been criticized because of the healthy participant effect.[22] Further studies could validate the findings of the current study by using electronic clinical record databases, where symptom history may be better captured, because this would better reflect the level of information available to clinicians and be more generalizable.

The medical history data provided information on Barrett's esophagus or esophagitis rather than on either condition alone. Individuals with Barrett's esophagus or esophagitis remained in the primary analyses because esophagitis offers a potentially useful source of EAC risk prediction. A sensitivity analysis in which individuals with Barrett's esophagus or esophagitis were excluded did not alter the results, suggesting any potential detection bias because of endoscopic surveillance in some Barrett's esophagus patients was minimal.

Conclusion

In summary, a list of established risk factors including age, sex, BMI, smoking status, and esophageal conditions could aid risk prediction of EAC. These factors are consistent with previous risk-prediction studies, although the points attributed and positive predictive values for specific cutoffs require external validation.

Supplementary Material

Note: To access the supplementary material accompanying this article, visit the online version of *Clinical Gastroenterology and Hepatology* at www.cghjournal.org, and at https://doi.org/10.1016/j.cgh.2018.03.014.

References

1. Thrift AP, Whiteman DC. The incidence of esophageal adenocarcinoma continues to rise: analysis of period and birth cohort effects on recent trends. Ann Oncol 2012;23:3155–3162.
2. Cancer research UK. Oesophageal cancer incidence statistics. Available at: http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/oesophageal-cancer/incidence#heading-Three. Accessed October 27, 2016.
3. Britton J, Gadeke L, Lovat L, et al. Research priority setting in Barrett's oesophagus and gastro-oesophageal reflux disease. Lancet Gastroenterol Hepatol 2017;2:824–831.
4. Lagergren J, Smyth E, Cunningham D, et al. Oesophageal cancer. Lancet 2017;390:2383–2396.
5. Yao K. The endoscopic diagnosis of early gastric cancer. Ann Gastroenterol 2013;26:11–22.
6. Gerson LB, Groeneveld PW, Triadafilopoulos G. Cost-effectiveness model of endoscopic screening and surveillance in patients with gastroesophageal reflux disease. Clin Gastroenterol Hepatol 2004;2:868–879.
7. NICE. Gastro-oesophageal reflux disease and dyspepsia in adults: investigation and management—guidance and guidelines. NICE; 2014.
8. Ronkainen J, Aro P, Storskrubb T, et al. Prevalence of Barrett's esophagus in the general population: an endoscopic study. Gastroenterology 2005;129:1825–1831.
9. Sharma P, Sidorenko EI. Are screening and surveillance for Barrett's oesophagus really worthwhile? Gut 2005;54:I27–I32.
10. Coleman HG, Bhat SK, Murray LJ, et al. Symptoms and endoscopic features at Barrett's esophagus diagnosis: implications for neoplastic progression risk. Am J Gastroenterol 2014;109:527–534.
11. Lagergren J, Bergström R, Lindgren A, et al. Symptomatic gastroesophageal reflux as a risk factor for esophageal adenocarcinoma. N Engl J Med 1999;340:825–831.
12. Hippisley-Cox J, Coupland C. Development and validation of QMortality risk prediction algorithm to estimate short term risk of death and assess frailty: cohort study. BMJ 2017;358:j4208.
13. Thrift AP, Kendall BJ, Pandeya N, et al. A model to determine absolute risk for esophageal adenocarcinoma. Clin Gastroenterol Hepatol 2013;11:138–144.e2.
14. 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.
15. Dong J, Buas MF, Gharahkhani P, et al. Determining risk of Barrett's esophagus and esophageal adenocarcinoma based on epidemiologic factors and genetic variants. Gastroenterology 2018;154:1273–1281. e3.
16. Thrift AP, Kendall BJ, Pandeya N, et al. A clinical risk prediction model for Barrett esophagus. Cancer Prev Res 2012;5:1115–1123.
17. Thrift AP, Garcia JM, El-Serag HB. A multibiomarker risk score helps predict risk for Barrett's esophagus. Clin Gastroenterol Hepatol 2014;12:1267–1271.
18. Thrift AP, Vaughan TL, Anderson LA, et al. External validation of the Michigan Barrett's esophagus prediction tool. Clin Gastroenterol Hepatol 2017;15:1124–1126.
19. Mathieu LN, Kanarek NF, Tsai H-L, et al. Age and sex differences in the incidence of esophageal adenocarcinoma: results from the Surveillance, Epidemiology, and End Results (SEER) Registry (1973-2008). Dis Esophagus 2014;27:757–763.

20. Rubenstein JH, Thrift AP. Risk factors and populations at risk: selection of patients for screening for Barrett's oesophagus. Best Pract Res Clin Gastroenterol 2015;29:41–50.
21. Allen N, Sudlow C, Downey P, et al. UK Biobank: current status and what it means for epidemiology. Heal Policy Technol 2012; 1:123–126.
22. Fry A, Littlejohns TJ, Sudlow C, et al. Comparison of sociodemographic and health-related characteristics of UK Biobank participants with those of the general population. Am J Epidemiol 2017;186:1026–1034.
23. Coleman HG, Xie S-H, Lagergren J. The epidemiology of esophageal adenocarcinoma. Gastroenterology 2018; 154:390–405.
24. Islami F, Kamangar F. Helicobacter pylori and esophageal cancer risk: a meta-analysis. Cancer Prev Res (Phila) 2008;1:329–338.
25. Xu B, Zhou X, Li X, et al. Diabetes mellitus carries a risk of esophageal cancer: a meta-analysis. Medicine (Baltimore) 2017; 96:e7944.
26. Lagergren J, Bergström R, Adami H-O, et al. Association between medications that relax the lower esophageal sphincter and risk for esophageal adenocarcinoma. Ann Intern Med 2000;133:165.
27. Vaughan TL, Farrow DC, Hansten PD, et al. Risk of esophageal and gastric adenocarcinomas in relation to use of calcium channel blockers, asthma drugs, and other medications that promote gastroesophageal reflux. Cancer Epidemiol Biomarkers Prev 1998;7:749–756.
28. Hay AD, Sterne JAC, Hood K, et al. Improving the diagnosis and treatment of urinary tract infection in young children in primary care: results from the DUTY Prospective Diagnostic Cohort Study. Ann Fam Med 2016;14:325–336.
29. 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.
30. Vaughan TL, Fitzgerald RC. Precision prevention of oesophageal adenocarcinoma. Nat Rev Gastroenterol Hepatol 2015; 12:243–248.
31. Freeman M, Offman J, Walter FM, et al. Acceptability of the Cytosponge procedure for detecting Barrett's oesophagus: a qualitative study. BMJ Open 2017;7:e013901.
32. Fernandez-Urien I, Carretero C, Armendariz R, et al. Esophageal capsule endoscopy. World J Gastroenterol 2008; 14:5254–5260.
33. West SL, Savitz DA, Koch G, et al. Recall accuracy for prescription medications: self-report compared with database information. Am J Epidemiol 1995;142:1103–1112.

Reprint requests
Address requests for reprints to: Andrew T. Kunzmann, PhD, Institute of Clinical Sciences Block B, Queen's University Belfast, Royal Victoria Hospital, Belfast, Northern Ireland BT12 6BA. e-mail: a.kunzmann@qub.ac.uk; fax: +44-28-90235900.

Conflicts of interest
The authors disclose no conflicts.

Funding
Supported by Ochre charity (Registered charity number: SC032343).

Supplementary Table 1. Characteristics of Study Population[a] in Terms of Additional Candidate Predictor Variables Not Widely Available to Clinicians at the Baseline Assessment Visit Who Did or Did Not Develop Esophageal Adenocarcinoma Within 5 Years

	No esophageal adenocarcinoma n (%)	Esophageal adenocarcinoma n (%)
Total	219,651 (100)	132 (100)
Abdominal obesity[b]		
No	108,766 (49.5)	50 (37.9)
Yes	110,885 (50.5)	82 (62.1)
Weight loss in last year		
No	174,760 (79.6)	102 (77.3)
Yes	44,891 (20.4)	30 (22.7)
Alcohol intake		
Never	8546 (3.9)	3 (2.3)
Current >0–14 units/wk	119,428 (54.4)	58 (43.9)
Current >14 units/wk	84,334 (38.4)	64 (48.5)
Former	7343 (3.3)	7 (5.3)
Physical activity[c]		
Low	24,296 (11.1)	19 (14.4)
Medium	111,932 (51)	61 (46.2)
High	83,423 (38)	52 (39.4)
Fruit and vegetable intake		
<5 pieces per day	125,643 (57.2)	90 (68.2)
5+ pieces per day	94,008 (42.8)	42 (31.8)
Smoking status by pack-years		
Never	118,122 (53.8)	36 (27.3)
Former, <20 pack-years	59,935 (27.3)	36 (27.3)
Former, 20+ pack-years	21,521 (9.8)	33 (25)
Current, <20 pack-years	9430 (4.3)	7 (5.3)
Current, >20 pack-years	10,643 (4.8)	20 (15.2)

[a]When the sample is limited to 219,989 individuals with complete data.
[b]International Diabetes Federation guidelines for abdominal obesity (>94 cm in men and >80 cm in women).
[c]Low, medium, or high physical activity measured by using validated short-form International Physical Activity Questionnaire.

Supplementary Table 2. Performance of Model for Risk Prediction of Esophageal Adenocarcinoma When Adding or Replacing Factors to the Original Model in 219,989 Individuals With Complete Data

	AUROC (95% CI)	*P* value (likelihood ratio test)	Included in stepwise selection
Original model[a]	0.78 (0.74–0.81)	NA	NA
Additional factors			
Abdominal obesity[b]	0.77 (0.74–0.81)	.13	x
Weight loss in last year[c]	0.78 (0.74–0.81)	.57	x
Alcohol intake[d]	0.78 (0.74–0.81)	.68	x
Physical activity[e]	0.78 (0.74–0.81)	.57	x
Fruit and vegetable intake[f]	0.78 (0.74–0.81)	.25	x
Alternative factors			
Smoking status by pack-years[g]	0.78 (0.75–0.81)	<.01	Yes
Abdominal obesity[b,h]	0.77 (0.74–0.81)	.75	Yes

AUROC, area under the receiver operating characteristic curve; CI, confidence interval.
[a]If no data missing for additional/alternative risk factors.
[b]International Diabetes Federation guidelines for abdominal obesity (Yes, >94.cm in men and >80.cm in women, vs no).
[c]Any or none. Self-reported.
[d]Never, light-moderate current (>0–14 units per week), heavy current (>14 units per week), or former drinkers.
[e]Low, medium, or high physical activity measured by using validated short-form International Physical Activity Questionnaire.
[f]<5 or 5+ pieces per day.
[g]<20 versus ≥20 pack-years instead of smoking status.
[h]Instead of body mass index categories.

Supplementary Table 3. Sensitivity Analyses Using the Risk-Prediction Model Identified in the Original Sample and Assessing the Variables Included or Excluded Using Stepwise Selection Within the New Sample

	Total n	Cases	AUROC (main model)	Factors from original model not selected in new model	Additional factors selected in new model
Main analysis	355,034	220	0.80 (0.77–0.82)	NA	NA
Additional individuals with cancer history[a]	379,585	237	0.80 (0.77–0.82)	None	None
Multiple imputation[b]	358,513	223	0.79 (0.76–0.82)	None	None
Follow-up time					
6 mo–3 y	355,034	124	0.82 (0.78–0.85)	None	None
6 mo–8 y	355,034	265	0.79 (0.76–0.81)	None	NSAID use
Smoking					
Ever smoking	355,034	220	0.79 (0.76–0.82)	None	None
Current smoking	355,034	220	0.79 (0.76–0.82)	None	None
Body mass index					
3 categories[c]	355,034	220	0.80 (0.77–0.82)	None	None
2 categories[d]	355,034	220	0.79 (0.76–0.82)	None	None
Age (*y*)					
50–60	157,644	55	0.83 (0.77–0.88)	None	Statin use, asthma inhaler use
Over 55	282,519	204	0.78 (0.75–0.81)	None	None
Over 60	197,390	165	0.77 (0.73–0.80)	None	None
Sex					
Men	164,153	185	0.73 (0.70–0.76)	None	Asthma inhaler use
Women	190,881	35	0.71 (0.63–0.79)	BMI	None
Esophageal condition					
No	306,377	161	0.80 (0.77–0.83)	None	None
Yes	48,657	59	0.76 (0.70–0.81)	BMI	None
Exclude prior Barrett's/esophagitis	353,809	214	0.80 (0.77–0.83)	None	None
Separating esophageal conditions[e]	355,034	220	0.80 (0.77–0.83)	None	None
BMI (kg/m^2)					
<25	111,426	35	0.83 (0.77–0.90)	None	Diabetes
25+	243,573	185	0.79 (0.75–0.82)	None	Hypertension
Smoking					
Never	188,080	56	0.78 (0.72–0.84)	None	Asthma inhaler use
Ever	166,954	164	0.76 (0.73–0.79)	None	Statin use, NSAID use
Cancer types					
Any upper gastrointestinal cancer	355,034	495	0.74 (0.72–0.76)	None	None
Gastric cancer	355,034	196	0.72 (0.69–0.76)	None	None
Esophageal cancer	355,034	284	0.76 (0.73–0.79)	None	None
Esophageal squamous cell carcinoma	355,034	64	0.71 (0.66–0.78)	Sex, BMI, esophageal preconditions	Asthma inhaler use

AUROC, area under the receiver operating characteristic curve; BMI, body mass index; NSAID, nonsteroidal anti-inflammatory drug.

[a]Excluding individuals with history of upper gastrointestinal cancers.

[b]Multiple imputation for BMI and smoking status.

[c]BMI categories (<25, 25–<30, 30+).

[d]BMI categories (<25, 25+).

[e]Esophageal conditions separated by type of condition: (1) gastroesophageal reflux disorder or anti-reflux medication (adjusted hazard ratio, 1.72; 95% confidence interval, 1.25–2.35); (2) Barrett's esophagus or esophagitis (adjusted hazard ratio, 4.19; 95% confidence interval, 1.79–9.77), and (3) other esophageal conditions including hiatus hernia, esophageal varices or stricture, and esophageal surgery (adjusted hazard ratio, 1.64; 95% confidence interval, 0.40–6.66).

Supplementary Table 4. Association Between Individual Esophageal Conditions at Baseline and Odds of Esophageal Adenocarcinoma When Conducting Sensitivity Analysis Separating Esophageal Conditions

	No esophageal adenocarcinoma	Esophageal adenocarcinoma	Adjusted odds ratio[a] (95% confidence interval)
Barrett's esophagus or esophagitis			
No	353,595	214	1.00 (reference)
Yes	1219	6	4.19 (1.79–9.77)
Gastroesophageal reflux disorder[b]			
No	306,963	162	1.00 (reference)
Yes	47,851	58	1.72 (1.25–2.35)
Other esophageal conditions[c]			
No	352,295	212	1.00 (reference)
Yes	2419	8	1.64 (0.40–6.66)

[a]Adjusted for age (50–55, 55–60, 60-65, 65+), sex, body mass index (<25, 25–<30, 30–<35, 35+), and smoking status (never, former, current).
[b]Includes individuals prescribed anti-reflux medications.
[c]Esophageal stricture, varices, disorder (unspecified), and hiatal hernia.

Supplementary Table 5. Statistics for Performance of Points-based Model for Upper Gastrointestinal Cancers at Different Points-based Cutoffs

Points cutoff	Patients deemed high-risk (%)	True cancers predicted (out of 495)	Sensitivity	Specificity	Youden's index	Absolute 5-year risk of upper gastrointestinal cancer[a] per 100,000	Referrals for additional screens per upper gastrointestinal cancer[a] predicted[b]	Odds ratio (95% confidence interval)
5+	222,686 (62.7)	434	87.7	37.3	0.25	194.89	513.1	4.23 (3.24–5.54)
6+	181,069 (51.0)	405	81.8	49.0	0.31	223.67	447.1	4.33 (3.44–5.44)
7+	143,675 (40.5)	374	75.6	59.6	0.35	260.31	384.2	4.56 (3.71–5.59)
8+	104,723 (29.5)	328	66.3	70.6	0.37	313.20	319.3	4.71 (3.91–5.67)
9+	72,840 (20.5)	270	54.5	79.5	0.34	370.68	269.8	4.66 (3.91–5.57)
10+	47,230 (13.3)	210	42.4	86.7	0.29	444.63	224.9	4.82 (4.03–5.76)
11+	22,982 (6.5)	116	23.4	93.6	0.17	504.74	198.1	4.44 (3.60–5.47)
12+	8821 (2.5)	52	10.5	97.5	0.08	589.50	169.6	4.63 (3.47–6.17)

[a]Within 5 years of baseline assessment.
[b]Not all cancers predicted will be detected during follow-up screens.