Model for Identifying Individuals at Risk for Esophageal Adenocarcinoma

Kunzmann, A. T., Thrift, A. P., Cardwell, C. R., Lagergren, J., Xie, S., Johnston, B. T., Anderson, L. A., Busby, J., McMenamin, U. C., Spence, A. D., & Coleman, H. G. (2018). Model for Identifying Individuals at Risk for Esophageal Adenocarcinoma. *Clin. Gastroenterol. Hepatol.* Advance online publication. https://doi.org/10.1016/j.cgh.2018.03.014

Published in:
Clin. Gastroenterol. Hepatol.

Document Version:
Peer reviewed version

Title: A Model for Identifying Individuals at Risk for EAC

Authors: Andrew T Kunzmann [1], Aaron P Thrift [2], Chris R Cardwell [1], Jesper Lagergren [3,4], Shaohua Xie [3], Brian T Johnston [5], Lesley A Anderson [1], John Busby [6], Úna C McMenamin [1], Andrew D Spence [1] & Helen G Coleman [1, 7]

Authors' affiliations:

[1] Cancer Epidemiology and Health Services Research Group, Centre for Public Health, Queen's University Belfast, Belfast, Northern Ireland, United Kingdom

[2] Section of Epidemiology and Population Sciences, Department of Medicine, Baylor College of Medicine, Houston, Texas

[3] Upper Gastrointestinal Surgery, Department of Molecular medicine and Surgery, Karolinska Institutet, Karolinska University Hospital, Stockholm, Sweden

[4] School of Cancer and Pharmaceutical Sciences, King's College London, London, United Kingdom

[5] Royal Victoria Hospital, Belfast Health & Social Care Trust, Belfast, N. Ireland, United Kingdom

[6] Centre for Experimental Medicine, Queen's University Belfast, Belfast, Northern Ireland, United Kingdom

[7] Centre for Cancer Research and Cell Biology, Queen's University Belfast, Belfast, Northern Ireland, United Kingdom

Grant support: The project was kindly funded by **Ochre charity (**Registered charity number: SC032343). The authors have no conflicts of interest to report.

Abbreviations: AUROC, area under the receiver operating characteristic curve; BE, Barrett's esophagus; BMI, body mass index; CI, confidence interval; EAC, EAC; GERD, gastroesophageal reflux disease; NSAIDs, nonsteroidal anti-inflammatory drugs; OR, odds ratio.

Correspondence: Dr Andrew Kunzmann, Institute of Clinical Sciences Block B, Queen's University Belfast, Royal Victoria Hospital, Belfast, Northern Ireland, BT12 6BA.

Phone: +44 (0) 28 9063 5009

Fax: +44 (0) 28 90235900

Email: a.kunzmann@qub.ac.uk

Disclosures: The authors disclose no potential conflicts of interest.

Author contributions: ATK had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. Study concept and design: ATK, CRC, APT, HGC. Statistical analysis and interpretation of data: ATK, CRC. Acquisition of data and obtained funding: ATK, UCM, HGC. Data preparation: ATK, UCM, ADS. Drafting of the manuscript: ATK. Critical review of the manuscript for important intellectual content: ATK, APT, CRC, JL, SX, BTJ, LAA, JB, UMC, ADS, HGC. Study supervision: HGC. All authors read and approved the final version for submission.

Abstract:

Background & Aims: Esophageal adenocarcinoma (EAC) prognosis is poor because patients often 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 (aged ≥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.

KEY WORDS: BMI, upper gastrointestinal cancer, risk-prediction, esophagus

Introduction

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] as the poor survival is largely attributable to late clinical presentation with advanced disease [2].

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

Clinical guidelines suggest that individuals with chronic symptomatic gastro-esophageal reflux disease (GERD) should be referred for endoscopy screening to identify Barrett's esophagus[7], a pre-cursor to EAC present in 1-2% of the adult population[8]. Individuals with Barrett's esophagus are often entered into endoscopy surveillance programmes[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 due to a lack of both specificity, as risk of EAC in individuals with GERD is low[10,11], and sensitivity, as 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 multi-stage 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 based on a combination of factors widely available to clinicians that may predict risk of EAC development within 5 years. Secondary analyses aim 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 one 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 one 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 (as upper gastrointestinal cancers are rare aged <50), without a history of cancer (excluding non-melanoma skin cancer) at or before baseline or within 6 months following 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 enquired 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 body mass index, 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 which 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 (ONS). Cancer data was provided up until 30th September 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 (due to potential diagnostic delays) and 5 years from baseline was 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 tumour 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 ($P<0.1$).

Points-based models were created from the coefficient-based model by dividing the coefficient of 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 cut-offs[28]. For example, the coefficient for men was 1.64 and the smallest coefficient in the model was 0.40 (BMI of 25-<30kg/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 using the area under the receiver operating characteristic (AUROC) curve with 95% confidence interval (CI). An internally validated AUROC was calculated 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 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 cut-off threshold.

To assess if a more complex model increased model performance (AUROC), additional factors, not widely available to clinicians or which may require additional examination, were added to the initial model in turn (excluding individuals with missing data) then with all factors significant at P<0.1 added to a stepwise selection model.

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

- including individuals with a prior history of non-upper GI cancers;
- conducting multiple imputation for missing variables (using ten imputations and combined using Rubin's rules);
- using different cancer follow-up periods;
- using different age periods;
- using different categories for smoking, BMI and separate variables for each oesophageal 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 as 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, as the tool could be used as triage to 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 ($P<0.1$) and assessed the discriminative ability of a new points- based model.

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

Results

Participants

There were 502,640 participants in the UK Biobank, of whom 117,891 (23.5%) were excluded as they were aged under 50 years, 30,665 (6.1%) were excluded due to a history of cancer (or cancer within 6 months of baseline), and 4,060 were excluded due to 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 NSAIDs, statins or asthma inhalers (Table 1). Mean follow-up time was 4.8 (standard deviation 0.6) years.

Non-participants

Individuals excluded due to 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.

EAC 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 an AUROC of 0.80 (95% CI 0.77-0.82, see Figure 1A). 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, chi^2 statistic=6.58, p=0.58), and the calibration curve (Figure 1B).

EAC risk-prediction model: points-based model

A points-based model assigned additional points based on age (55-60 years: 1.5; 60-65 years: 2.5; 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 cut-off threshold are provided in Table 3. A cut-off threshold of 8+ points with the highest Youden's index (0.48), had a sensitivity of 77.5%, a specificity of 70.5%, a positive predictive value of 0.16% (Figure 1C) and would mean 612 referrals for 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 cut-off threshold, 76.4% (79,970 individuals) had not reported a prior 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 oesophageal conditions by type and after multiple imputation (Supplementary table 3-4).

Upper gastrointestinal cancers (gastric and esophageal cancers)

The factors selected in the model for upper gastrointestinal cancer (gastric or esophageal cancers) within 5 years was 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 histological type within 5 years was 0.76 (95% CI 0.73-0.79) (Supplementary Table 3).

Due to similarity in structure, for simplicity, we report results using the EAC points system. The 8-point cut-off 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 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 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 due to 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, I.e. NSAIDs were not confirmed to have utility in risk-prediction, though 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 a 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, I.e. 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 five-year risk of EAC in individuals above the 8-point cut-off 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 five-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 modelling would be required to assess the cost-effectiveness of changes to screening practices.

Strengths & 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 generalisability of the UK Biobank to the general population has been

criticised due to the healthy participant effect[22]. Further studies could validate the findings of the current study using electronic clinical record databases, where symptom history may be better captured, as 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, as 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 due to 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, though the points attributed and positive predictive values for specific cut offs require external validation.

Table and Figure legends

Table 1. Characteristics of study population in terms of candidate predictor variables at the baseline assessment visit, who did or did not develop oEAC within 5 years.

Table 2. Risk factors associated with oEAC in a stepwise logistic regression and the points assigned for the points based model.

Table 3. Statistics for the performance of a points-based oEAC risk-prediction model at different points based cut-offs.

Figure 1. (A) Receiver operating characteristic (ROC) curve for the based model for predicting risk of EAC within 5 years; (B) Hosmer-Lemeshow calibration plot: Observed versus predicted proportion, based upon risk score, of individuals with EAC within 5 years, and; (C) Cumulative proportion developing EAC from 6 months to five years following baseline based upon risk score (<8 points versus 8+ points).

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

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 2017;In press.

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;27:757–63.

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 2017;Epub ahead.

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;7:749–56.

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;14:325–36.

29. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. Epidemiology 2010;21:128–38.

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;14:5254–60.

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.