This is a repository copy of *Determining Risk of Barrett's Esophagus and Esophageal Adenocarcinoma Based on Epidemiologic Factors and Genetic Variants*.

Determining Risk of Barrett's Esophagus and Esophageal Adenocarcinoma Based on

Epidemiologic Factors and Genetic Variants

Short title: Integrative risk models for Barrett's esophagus and esophageal adenocarcinoma

Jing Dong,[1] Matthew F. Buas,[2,3] Puya Gharahkhani,[4] Bradley J. Kendall,[5,6,7] Lynn Onstad,[2] Shanshan Zhao,[8] Lesley A. Anderson,[9] Anna H. Wu,[10] Weimin Ye,[11] Nigel C. Bird,[12] Leslie Bernstein,[13] Wong-Ho Chow,[14] Marilie D. Gammon,[15] Geoffrey Liu,[16] Carlos Caldas,[17] Paul D. Pharoah,[18,19] Harvey A. Risch,[20] Prasad G. Iyer,[21] Brian J. Reid,[2] Laura J. Hardie,[22] Jesper Lagergren,[23,24] Nicholas J. Shaheen,[25] Douglas A. Corley,[26,27] Rebecca C. Fitzgerald,[28] Stomach and Oesophageal Cancer Study (SOCS) consortium, David C. Whiteman,[5] Thomas L. Vaughan,[2] Aaron P. Thrift[1,29]

[1]Dan L Duncan Comprehensive Cancer Center, Baylor College of Medicine, Houston, TX, USA.

[2]Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA, USA.

[3]Department of Cancer Prevention and Control, Roswell Park Cancer Institute, Buffalo, NY, USA.

[4]Statistical Genetics, QIMR Berghofer Medical Research Institute, Brisbane, QLD, Australia.

[5]Cancer Control, QIMR Berghofer Medical Research Institute, Brisbane, QLD, Australia.

[6]School of Medicine, University of Queensland, Brisbane, QLD, Australia.

[7]Department of Gastroenterology and Hepatology, Princess Alexandra Hospital, Brisbane, QLD, Australia.

[8]Biostatistics and Computational Biology Branch, National Institute of Environmental Health Sciences, National Institutes of Health, Research Triangle Park, NC, USA.

[9]Centre for Public Health, Queen's University Belfast, Belfast, Northern Ireland.

[10]Department of Preventive Medicine, University of Southern California/Norris Comprehensive Cancer Center, Los Angeles, CA, USA.

[11]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden.

[12]Department of Oncology, Medical School, University of Sheffield, Sheffield, UK.

[13]Department of Population Sciences, Beckman Research Institute and City of Hope Comprehensive Cancer Center, Duarte, CA, USA.

[14]Department of Epidemiology, MD Anderson Cancer Center, Houston, TX, USA.

[15]Department of Epidemiology, University of North Carolina, Chapel Hill, NC, USA.

[16]Pharmacogenomic Epidemiology, Ontario Cancer Institute, Toronto, Ontario, Canada M5G 2M9.

[17]Cancer Research UK, Cambridge Institute, Cambridge, UK.

[18]Department of Oncology, University of Cambridge, Cambridge, UK.

[19]Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK.

[20]Department of Chronic Disease Epidemiology, Yale School of Public Health, New Haven, CT, USA.

[21]Division of Gastroenterology and Hepatology, Mayo Clinic, Rochester, MN, USA.

[22]Division of Epidemiology, LICAMM, School of Medicine, University of Leeds, Leeds, UK.

[23]Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm, Sweden.

[24]School of Cancer Sciences, King's College London, London, UK.

[25]Division of Gastroenterology and Hepatology, University of North Carolina School of Medicine, Chapel Hill, NC, USA.

[26]Division of Research, Kaiser Permanente Northern California, Oakland, CA, USA.

[27]San Francisco Medical Center, Kaiser Permanente Northern California, San Francisco, CA, USA.

[28]Medical Research Council (MRC) Cancer Unit, Hutchison-MRC Research Centre, University of Cambridge, Cambridge, UK.

[29]Section of Epidemiology and Population Sciences, Department of Medicine, Baylor College of Medicine, Houston, TX, USA.

Abbreviations: AUC, area under the receiver operating characteristic curve; BE, Barrett's esophagus; BEACON, Barrett's and Esophageal Adenocarcinoma Consortium; BMI, body mass index; CI, confidence interval; EAC, esophageal adenocarcinoma; GERD, gastroesophageal reflux disease; GWAS, genome-wide association study; NSAIDs, nonsteroidal anti-inflammatory drugs; OR, odds ratio; PRS, polygenic risk score.

Correspondence to: Aaron P. Thrift, Ph.D., Dan L Duncan Comprehensive Cancer Center, Baylor College of Medicine, One Baylor Plaza, MS: BCM305, Houston, Texas 77030-3498. Tel: 713-798-9107; Fax: 713-798-3658; E-mail: aaron.thrift@bcm.edu

Conflicts of interest: The authors disclose no potential conflicts of interest.

Specific author contributions: APT 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: JD, MFB, DCW, TLV, APT. Statistical analysis and interpretation of data: JD, MFB, SZ, APT. Acquisition of data and obtained funding: LAA, AHW, WY, NCB, LB, WHC, MDG, GL, CC, PDP, HAR, PGI, BJR, LJH, JL, NJS, DAC, RCF, DCW, TLV. Data preparation: JD, PG, LO, CC, PDP, RCF. Drafting of the manuscript: JD, APT. Critical review of the manuscript for important intellectual content: JD, MFB, PG, BJK, LO, SZ, LAA, AHW, WY, NCB, LB, WHC, MDG, GL, CC, PDP, HAR, PGI, BJR, LJH, JL, NJS, DAC, RCF, DCW, TLV, APT. Study supervision: APT. All authors read and approved the final version for submission.

Word count: 5356

Manuscript Number: GASTRO 17-01611

Title: Determining Risk of Barrett's Esophagus and Esophageal Adenocarcinoma Based on Epidemiologic Factors and Genetic Variants

Abstract:

Background & Aims: We developed comprehensive models to determine risk of Barrett's esophagus (BE) or esophageal adenocarcinoma (EAC) based on genetic and non-genetic factors.

Methods: We used pooled data from 3288 patients with BE, 2511 patients with EAC, and 2177 individuals without either (controls) from participants in the international Barrett's and EAC consortium as well as the United Kingdom's BE gene study and stomach and esophageal cancer study. We collected data on 23 genetic variants associated with risk for BE or EAC, and constructed a polygenic risk score (PRS) for cases and controls by summing the risk allele counts for the variants weighted by their natural log-transformed effect estimates (odds ratios) extracted from genome-wide association studies. We also collected data on demographic and lifestyle factors (age, sex, smoking, body mass index, use of nonsteroidal anti-inflammatory drugs) and symptoms of gastroesophageal reflux disease (GERD). Risk models with various combinations of non-genetic factors and the PRS were compared for their accuracy in identifying patients with BE or EAC using the area under the receiver operating characteristic curve (AUC) analysis.

Results: Individuals in the highest quartile of risk, based on genetic factors (PRS), had a 2-fold higher risk of BE (odds ratio [OR], 2.22; 95% CI, 1.89–2.60) or EAC (OR, 2.46; 95% CI, 2.07–2.92) than individual in the lowest quartile of risk based on PRS. Risk models developed based on only demographic or lifestyle factors or GERD symptoms identified patients with BE or EAC with AUC values ranging from 0.637 to 0.667. Combining data on demographic or lifestyle factors with data on GERD symptoms identified patients with BE with an AUC of 0.793 and patients with EAC with an AUC of 0.745. Including PRSs with these data only minimally increased the AUC values for BE (to 0.799) and EAC (to 0.754). Including the PRSs in the model developed based on non-genetic factors resulted in a net reclassification improvement for BE of 3.0% and for EAC of 5.6%.

Conclusions: We used data from 3 large databases of patients from studies of BE or EAC to develop a risk prediction model based on genetic, clinical, and demographic/lifestyle factors. We identified a PRS that increases discrimination and net reclassification of individuals with vs. without BE and EAC. However, the absolute magnitude of improvement is not sufficient to justify its clinical use.

KEY WORDS: BEACON, SNP, neoplasm, environmental exposures

INTRODUCTION

The incidence of esophageal adenocarcinoma (EAC) has increased sharply in many Western populations during the last four decades.[1,2] Despite recent advances in early detection and the availability of endoscopic therapy for early neoplastic lesions, the estimated overall 5-year survival rate for all EAC patients remains poor at 18%.[1,3] Most cases of EAC are diagnosed at advanced stages, with a median survival of less than 12 months.[1]

Barrett's esophagus (BE) is the precursor lesion for EAC and is a metaplastic columnar epithelium of the distal esophagus.[4] Chronic and frequent reflux of duodenogastric contents into the distal esophagus is the primary risk factor for BE and EAC.[5,6] The current approach to the prevention of deaths from EAC relies on upper endoscopy with biopsy to identify BE among individuals with frequent symptoms of gastroesophageal reflux disease (GERD), followed by regular endoscopic surveillance of BE to identify patients with neoplastic progression before invasive EAC occurs. However, only 5-7% of patients diagnosed with EAC have a prior diagnosis of BE,[7] and a large proportion of patients with BE are asymptomatic and undetected,[8] greatly limiting the usefulness of this strategy.

An alternative approach is screening using a risk prediction model, based on demographic and clinical characteristics, genetic markers and other factors, for patient risk stratification. Providing clinicians with a tool that allows them to estimate a patient's risk may better aid them in deciding whom to screen and make future resource utilization more efficient. Risk models incorporating genetic and non-genetic factors have been developed for predicting risks of breast cancer,[9] epithelial ovarian cancer,[10] and lung cancer,[11,12] among others. Previously constructed models for EAC[13,14] and BE[15-19] have combined demographic, clinical and lifestyle information into single risk models and shown moderate discriminatory ability. However, no studies have examined the clinical utility of genetic-based risk prediction models for BE and EAC, alone or in combination with non-genetic risk factors.

BE and EAC are highly polygenic, with a number of overlapping genetic variants contributing to disease susceptibility.[20,21] Taking advantage of the genetic variants identified by recent large genome-wide association studies (GWAS) of BE and EAC,[20,22-24] we aimed to derive a polygenic risk score (PRS) and develop a comprehensive risk prediction model based on the PRS and a panel of well-established risk factors for BE or EAC (i.e., GERD symptoms, smoking status, body mass index (BMI), and use of nonsteroidal anti-inflammatory drugs (NSAIDs)).

METHODS

Study population

The current analysis used a pooled dataset that included participants from the international Barrett's and Esophageal Adenocarcinoma Consortium (BEACON; http://beacon.tlvnet.net/) GWAS, additional BE cases from the United Kingdom (UK) Barrett's Esophagus Gene Study, and EAC cases from the UK Stomach and Oesophageal Cancer Study. The BEACON GWAS included 2,413 BE cases, 1,512 EAC cases and 2,185 controls of European ancestry from 14 epidemiologic studies conducted in North America, Western Europe, and Australia, as previously described.[23] Histological confirmation of BE and EAC was carried out in all participating studies. Among them, after GWAS data cleaning procedures (see full details below), 2,406 BE cases, 1,508 EAC cases, and 2,177 controls were available for the current analysis. We included an additional 882 BE cases and 1,003 EAC cases from the two UK-based studies.[20] These BE cases were UK Barrett's Esophagus Gene Study participants identified at endoscopy with confirmed histopathological diagnoses of intestinal metaplasia.[20] The EAC cases were selected from the UK Stomach and Oesophageal Cancer Study and had International Classification of Diseases coding of malignant neoplasm of the esophagus (C15) and a pathological diagnosis of adenocarcinoma (M8140-8575).[20] The final study sample included 3,288 BE cases, 2,511 EAC cases, and 2,177 controls. Each contributing study complied with their institutional review board requirements and all participants gave informed consent.

Genetic variant selection, genotyping and imputation

We selected genetic variants reported to be associated at genome-wide significance ($P < 5\times10^{-8}$) with risk of BE and/or EAC in published GWAS.[20,22-24] Only the variant with the lowest P value was selected when multiple variants were observed in high linkage disequilibrium (LD) ($r^2 > 0.8$). In total, 23 genetic variants were included to generate the PRS (Supplementary Table 1).

Genotyping of buffy coat or whole blood DNA from all participants was performed using the Illumina Omni1M Quad platform, in accordance with standard quality control procedures.[20,25] A detailed quality control of both BEACON and UK datasets has been previously described.[20] In brief, all genotyped samples and variants met the following inclusion criteria: per variant and per sample missingness ≤ 3%; single nucleotide polymorphisms (SNPs) with a minor allele frequency (MAF) > 1%; SNPs with P ≥ 0.0001 in controls and P ≥ 5 × 10^{-10} in BE and EAC cases for Hardy-Weinberg equilibrium, no familial relationships, extreme heterozygosity rate, or outliers. Imputation was performed in each study using IMPUTE2 based on the 1000 Genomes Phase 1 haplotypes (June, 2014 release) reference panel.[20] Post-imputation quality control was conducted by excluding SNPs with imputation quality score < 0.4 or MAF < 0.01.

Based on samples and variants passing standard GWAS quality control, imputed dosage data for the 23 variants were extracted. Among them, eight variants had been directly genotyped; we therefore used the available genotyping data for these eight variants instead of imputation data. Imputed dosage data for the remaining 15 variants were converted to genotype data using GTOOL (http://www.well.ox.ac.uk/~cfreeman/software/gwas/gtool.html), and then merged with the genotypes of the eight directly genotyped variants.

Polygenic risk score

We constructed the PRS for cases and controls by summing the risk allele counts (i.e., participants have 0, 1, or 2 risk alleles) for the 23 variants weighted by their natural log-transformed effect estimates (odds ratios) extracted from the corresponding GWAS.[20,22-24] For each participant, we summed the weighted risk allele counts and then divided the total by 23 (i.e., the total number of loci) to derive a mean weighted score. The mean weighted score was then transformed into a Z score using the controls as the reference. This method for calculating risk scores mirrors the approach used in the software package PLINK.[26]

Demographic, lifestyle and clinical factors

Responses to self-reported questionnaires or interviews were collected at or near the time of BE or EAC

diagnosis for cases and at the time of recruitment for controls. Heartburn symptoms were defined as a

burning or aching pain behind the breastbone/sternum not attributed to heart problems, and regurgitation

symptoms as a sour taste resulting from regurgitation of acid, bile or other stomach contents. Heartburn

and regurgitation symptoms were harmonized and modeled dichotomously as recurrent vs. not recurrent

based on a frequency of weekly or greater for 'recurrent'. The term GERD will be used to refer to the

exposure of heartburn or regurgitation symptoms or both. BMI was calculated as weight in kg divided by

the square of height in meters (kg/m^2). In this current analysis, we used usual adult weight (prior to any

disease-related weight loss). If usual adult weight was unavailable, we used weight at 1, 5, or 20 years

before interview, depending upon which data were available.[27,28] Tobacco smoking was categorized as

ever vs. never regular use of cigarettes. Use of NSAIDs was defined as ever vs. never regular use of any

aspirin or non-aspirin NSAIDs. Exposures were harmonized across studies and merged into a single

dataset for analysis.

Statistical analysis

We derived separate risk models for BE and EAC. The combined prediction model used information from

the PRS, age, sex, GERD symptoms, smoking, BMI, and use of NSAIDs to estimate disease risk. All

seven risk factors were entered into a logistic regression model to estimate odds ratios (OR) and

corresponding 95% confidence intervals (95% CI).

To evaluate model performance, we first used the Hosmer-Lemeshow statistic[29] to assess model

goodness-of-fit. Receiver operating characteristic curves analysis was performed and the area under the

curve (AUC) was calculated to evaluate discriminatory performance, with repeated 10-fold cross-

validation and bootstrapping techniques used for internal validation of the risk models. To be useful, the

predicted risks must discriminate well between those participants who do have BE/EAC (cases) and those

participants who do not have BE/EAC (i.e., controls). The AUC gives the probability that for any randomly selected pair of individuals, one case and one control, the model assigns a higher probability to the case. A value of 1 indicates the model has perfect discrimination, while a value of 0.5 indicates the model discriminates no better than chance. Likelihood ratio statistics were also computed to compare model fit with demographic/lifestyle factors only (age, sex, smoking, BMI, and use of NSAIDs), GERD symptoms only, the PRS only, and the combination of these predictors.

We further assessed the clinical impact of the PRS by reclassification using the net reclassification index (NRI).[30] The NRI, after inclusion of the PRS in the model with demographic/lifestyle factors and GERD symptoms, was computed as [proportion of all cases reclassified at higher risk – proportion reclassified at lower risk] – [proportion of all controls reclassified at higher risk – proportion reclassified at lower risk].

All data were analyzed using R software (version 3.3.3; The R Foundation for Statistical Computing, Vienna, Austria) and Stata version 14 (StataCorp, College Station, Texas, USA). Statistical significance was determined at $\alpha = 0.05$, and all p-values for statistical significance were two-sided.

RESULTS

Participants

Characteristics of the study sample are shown in Table 1. The mean ages of BE cases, EAC cases, and controls were 62.9, 64.5, and 61.7 years, respectively. As expected, a large proportion of cases (BE, 75.6%; EAC, 87.3%) and controls (78.6%) were male. BE and EAC cases had higher BMI, and were more likely to have recurrent GERD symptoms and to have smoked compared to controls.

Polygenic risk score

The PRS was approximately normally distributed for each of the BEACON and UK datasets (Supplementary Figure 1). The PRS was significantly higher for case groups than in controls ($P < 0.001$ for both BE vs. controls and EAC vs. controls), with mean (sd.) PRS of 0.31 (1.01) in BE, 0.36 (1.02) in EAC and 0.00 (1.00) in controls (Supplementary Figure 2). An increase in the PRS was associated with statistically significantly higher risks of BE and EAC (Table 2), adjusted for demographic/lifestyle factors and GERD symptoms. A risk gradient was also observed across quartiles of the PRS, such that individuals in the highest quartile of the PRS had over 2-fold higher risks of BE (OR, 2.22; 95% CI, 1.89-2.60) and EAC (OR, 2.46; 95% CI, 2.07-2.92) compared to those in the lowest quartile of the PRS (Table 2). The PRS showed poor-to-moderate predictive ability for discriminating between BE cases and controls (AUC, 0.588; 95% CI, 0.572-0.603) and between EAC cases and controls (AUC, 0.599; 95% CI, 0.583-0.615) (Table 3).

Combined risk prediction models

We evaluated the predictive performance of demographic/lifestyle factors (age, sex, smoking, BMI, and use of NSAIDs) and GERD symptoms for discriminating BE and EAC cases from controls with and without the inclusion of the PRS in the models (Figure 1 & Table 3). The demographic/lifestyle-only and GERD symptoms-only models discriminate BE and EAC cases from controls with moderate accuracy (AUCs ranged from 0.637-0.667). When combined into a single model, compared with the

12

demographic/lifestyle-only and GERD symptoms-only models, the AUC values for the models with demographic/lifestyle factors and GERD symptoms were statistically significantly higher (BE: AUC, 0.793; 95% CI, 0.776-0.810; EAC: AUC, 0.745; 95% CI, 0.721-0.769).

With the addition of the PRS, the AUCs showed small improvement: demographic/lifestyle-only model increased from 0.637 to 0.656 for BE, and from 0.665 to 0.697 for EAC; and the AUC for the GERD symptoms-only model increased from 0.667 to 0.709 for BE, and from 0.638 to 0.684 for EAC. Compared with the demographic/lifestyle-GERD symptoms combined models, adding the PRS into the model slightly improved the prediction performance, with an AUC of 0.799 (95% CI, 0.782-0.816) for BE and 0.754 (95% CI, 0.729-0.778) for EAC (Figure 1, Table 3 and Supplementary Tables 2 and 3). Likelihood-ratio statistics were then calculated comparing the model with demographic/lifestyle-only, GERD symptoms-only and demographic/lifestyle-GERD symptoms combined with the models that included the PRS (Supplementary Table 2). While all comparisons were statistically significant (all P < 0.001), the absolute improvement in AUCs afforded by adding the PRS was small. Upon examining the reclassification properties of adding the PRS to the demographic/lifestyle-GERD symptoms combined models for BE and EAC, we observed an overall improvement in the net risk stratification of 3.0% (number of participants reclassified of overall sample was 461) for BE risk prediction and 5.6% (number of participants reclassified of overall sample was 285) for EAC risk prediction (Table 4).

Each model showed good internal validity and we found little evidence of overfitting in the repeated 10-fold cross-validation analysis (combined model for BE: AUC, 0.797; combined model for EAC: AUC, 0.748) (Table 3). Similar results were observed when we used bootstrapping methods for internal validation (combined model for BE: AUC, 0.795; combined model for EAC: AUC, 0.746). The models also demonstrated strong calibration in the study population (Hosmer-Lemeshow statistic, P = 0.55 for BE and P = 0.15 for EAC), showing good agreement between the predicted probabilities and actual BE/EAC risks across the observed range of risk.

DISCUSSION

To the best of our knowledge, this is the first study to develop a risk prediction model for BE and EAC combining non-genetic (demographics, lifestyle factors, and GERD symptoms) and genetic data. The results of this risk assessment study demonstrate that our combined risk prediction model may have limited clinical utility for discriminating BE and EAC cases from controls. Risk models including age, sex, smoking status, BMI, use of NSAIDs, and frequency of GERD symptoms showed moderate discriminatory ability and were well calibrated. Adding the PRS provided significantly but largely incremental improvements in the AUC. These data provided 3-6% improvement in the net reclassification index, suggesting that genetic predictors in combination with information on demographic and lifestyle factors and history of GERD symptoms may improve risk stratification efforts for BE and EAC, but the absolute magnitude of improvement may not be sufficient to justify its clinical use.

Given the strong association between GERD symptoms and the risks of BE and EAC, most existing strategies for preventing deaths from EAC have suggested screening individuals with chronic and frequent GERD symptoms by endoscopy to detect BE, and monitoring those with BE by endoscopic biopsy surveillance to detect dysplasia or early cancers, which are amenable to endoscopic ablation and/or surgical resection.[4,31] However, a critical limitation of these approaches, as implemented, is that most individuals who ultimately develop EAC (~93%) are never referred for endoscopic or other screening tests until they present with late-stage cancer.[32] Assuming these patients had access to medical care, they either had asymptomatic BE (without GERD symptoms), or they had GERD symptoms not sufficiently severe to justify either presentation to their medical provider or endoscopy in the judgment of their medical provider.[32] To enable more effective prevention and early detection of EAC, a key challenge lies in identifying the "at risk" population with greater precision. To that end, our combined prediction model included demographic/lifestyle (age, sex, smoking, BMI, and use of NSAIDs) factors, frequency of GERD symptoms and genetic factors (the PRS) to estimate relative disease risk. This is a BE/EAC-specific model incorporating BE/EAC-specific risk factors such as GERD symptoms and the PRS. None

14

of the factors included in our combined model were used in the clinical diagnosis of BE and EAC, or used as part of study population recruitment for any involved study in this risk prediction model. This allows us to minimize overestimating the predictive ability of the integrative risk models.

Previous studies have developed prediction models to estimate the absolute 5-year risk of EAC.[13,14] For example, a study from a nation-wide population-based case-control study in Sweden used a set of readily available (non-genetic) factors and reported an AUC of 0.84 among 189 incident EAC cases and 820 controls.[13] An Australian study combined BMI, smoking, GERD symptoms, use of acid-suppressant medications and NSAIDs, and attained educational level to estimate the absolute risk of EAC from 364 incident EAC cases and 1,580 controls, and obtained an AUC of 0.75.[14] These models have not been validated in independent populations. Similarly, past risk prediction models for BE based on GERD symptoms and other non-genetic risk factors have yielded low to moderate predictive accuracy.[15-19] These studies observed that using multiple factors improved discrimination compared with using only GERD symptoms (frequency and duration), but that history of GERD symptoms was a particularly important predictor.[16,17] Consistent with these findings, GERD symptoms was the most important predictor of BE and EAC risk in our model, and explained the largest fraction of the phenotypic variance for BE (8.9%) and EAC (6.3%). However, approximately 40% of patients with EAC do not report a history of GERD symptoms at cancer diagnosis,[33] and we showed that inclusion of additional factors in risk stratification (e.g., the AUC for EAC was 0.745 for the demographic/lifestyle-GERD symptoms combined model vs. 0.638 for the GERD symptoms-only model; $P < 0.001$) is necessary to improve performance for clinical application.

Recent large-scale GWAS of BE and EAC have provided an opportunity to develop risk models incorporating genetic variants. PRS using expanded sets of genetic variants has been reported to improve discrimination for diseases such as coronary heart disease, breast, lung and prostate cancer.[11,34-36] However, empirical evidence for the effectiveness of genetic risk stratification in BE and EAC is lacking.

15

In the present study, the PRS was associated with risks of BE and EAC, and accounted for 1.7% and 2.2% phenotypic variance of BE and EAC, respectively. Those in the highest quartile of the PRS had 122% and 146% increased risk for BE and EAC, respectively, as compared with those with lowest quartile. How to best assess and quantify the improvement in a risk prediction model offered by adding new markers has long been a challenge. Calculating the improvement in the AUC is a traditional way to evaluate the prediction increment of a new marker. However, for prediction models already possessing relatively good discrimination, it is difficult for new promising markers to produce large increases in the AUC. For example, in the current study, the model including GERD symptoms and demographic/lifestyle factors provides reasonably good discrimination between BE/EAC cases and controls; thus, adding the PRS by itself does not meaningfully increase the AUC. New metrics, such as NRI, have been proposed for quantifying the prediction increment of a new marker,[37] and have become widely used. The NRI focuses on reclassification tables constructed separately for cases and controls, and quantifies the correct movement in categories-upwards for cases and downwards for controls. In the current study, we found that after including the PRS in the model, 121 BE cases were reclassified as greater risk and 128 controls were reclassified as lower risk; while for EAC, 65 cases were corrected upwards and 100 controls downwards. Although the PRS assigned more individuals to the tails of the risk distributions for BE and EAC, the improvements were limited (3.0% and 5.6% for BE and EAC, respectively). These data suggested that compared with the model combining demographic/lifestyle factors and GERD symptoms, the additional prediction value provided by the PRS is not sufficient for clinical use. However, actively monitoring the individuals at the tails of this distribution may still help to identify more high risk patients for cancer prevention and avoid overtreatment for those individuals at lowest risk.

Strengths of the current study include its large sample size of both BE and EAC cases, the high-quality case-control and cohort parent studies, and the available non-genetic and genetic data. The combined model provides moderate predictiveness for BE and EAC, with estimated AUCs of 0.799 and 0.754, respectively. Because internal validation indicates an upper limit of the expected performance in new

16

settings, the repeated 10-fold cross-validation and bootstrapping analysis suggested minimal overfitting of the models. Reclassification analysis showed small improvement of adding the PRS into the prediction model, suggesting potential impact of the PRS to improve the identification of individuals at higher risk for BE and EAC. Resequencing studies of genetic loci that are now underway might help to further refine the PRS of our combined models, and the identification of other, consistently associated risk factors for BE and EAC may also improve discrimination.

Our study also has some limitations. First, only genome-wide significant GWAS variants were selected to generate the PRS; additional loci yet to be identified, rare variants with larger magnitudes of association and copy number variants should be included in future PRS. Because we selected SNPs and the corresponding regression coefficients for the PRS based on those SNPs identified in a meta-analysis of BE/EAC GWAS[20] that includes a subset of participants used in the current analysis, we may have overestimated the predictive ability of the PRS due to overfitting. The "true" incremental improvement in discrimination afforded by a PRS may therefore be smaller than reported here. Although external validation is the gold standard because it establishes the model's transportability and generalizability,[38] we were unable to externally validate our risk prediction models. Amassing sufficient study power for risk prediction analysis is a challenge, particularly given that paired clinical and GWAS data are often not available. Given the large worldwide consortial sample of patients participating in this work, few additional EA and BE patients with non-genetic and genetic data are currently available for replication, thus such work may require additional time for study patients to accrue. We elected to pool all studies in the analysis to gain the greatest statistical power and used 10-fold cross-validation and bootstrapping techniques for internal validation rather than a split sample approach which can lead to unstable models with a suboptimal predictive ability, and observed similar results.[39] Although we do not have an external replication set, the consistency of associations of BE/EAC risk factors across BEACON studies and internal validation analyses suggest that our models would have adequate transportability and generalizability. Since we used case-control data to estimate relative risks, some level of recall bias,

misclassification and missing data are inevitable for exposure variables such as smoking, GERD symptoms, BMI, and NSAID use. Recently, central obesity (e.g., waist-to-hip ratio) has been found to be more strongly associated with the risks of BE and EAC than BMI; however, waist and hip measurements were not collected in the majority of the included studies. It is likely that adding such measures to the model, rather than BMI, may improve predictive ability. In addition, our model was trained in a study sample comprised of individuals of European ancestry; it may not be applicable to other racial or ethnic groups. Because our pooled dataset included some cohort studies with no controls (i.e., studies provided only BE or EAC cases), we were not able adjust for study site. Heterogeneity between cases and controls from different centers may have influenced our results and requires additional work to establish the transportability and generalizability of the models. Finally, our prediction models were constructed based on regression approaches, and used a small number of variables to predict the probability of BE or EAC. These traditional regression models are based on theory and assumptions and have a priori hypotheses. Additional advanced machine learning techniques may be able to produce a more flexible relationship among the predictors and the outcome and help confront issues of multiple and correlated predictors, non-linear relationships, and interactions between predictors and endpoints, in large datasets. However, the output from these methods may not be as user friendly and have limited clinical utility.

In conclusion, this was the first attempt to quantify the discriminatory ability of genetic factors associated with BE and EAC, alone and in combination with non-genetic factors. The PRS was strongly associated with risks of BE and EAC; however, the combined risk prediction model provided here is complex and the absolute difference in discriminatory ability of the combined model compared to risk models based on only non-genetic factors (i.e., demographic and lifestyle factors and history of GERD symptoms) was not sufficient to influence clinical decision-making. A better understanding and use of genetic information for risk prediction is needed before such integrated models could be recommended for clinical use. Future studies are also necessary to refine and evaluate prediction models based on data from prospective studies, which allow model calibration with temporally developed information.

REFERENCES

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2016. CA Cancer J Clin 2016;66:7-30.

2. 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.

3. Thrift AP. The epidemic of oesophageal carcinoma: Where are we now? Cancer Epidemiol 2016;41:88-95.

4. Reid BJ, Li X, Galipeau PC, et al. Barrett's oesophagus and oesophageal adenocarcinoma: time for a new synthesis. Nat Rev Cancer 2010;10:87-101.

5. Shaheen NJ, Richter JE. Barrett's oesophagus. Lancet 2009;373:850-61.

6. Thrift AP, Kramer JR, Qureshi Z, et al. Age at onset of GERD symptoms predicts risk of Barrett's esophagus. Am J Gastroenterol 2013;108:915-22.

7. Bhat SK, McManus DT, Coleman HG, et al. Oesophageal adenocarcinoma and prior diagnosis of Barrett's oesophagus: a population-based study. Gut 2015;64:20-5.

8. Dulai GS, Guha S, Kahn KL, et al. Preoperative prevalence of Barrett's esophagus in esophageal adenocarcinoma: a systematic review. Gastroenterology 2002;122:26-33.

9. Shieh Y, Eklund M, Madlensky L, et al. Breast Cancer Screening in the Precision Medicine Era: Risk-Based Screening in a Population-Based Trial. J Natl Cancer Inst 2017;109.

10. Clyde MA, Palmieri Weber R, Iversen ES, et al. Risk Prediction for Epithelial Ovarian Cancer in 11 United States-Based Case-Control Studies: Incorporation of Epidemiologic Risk Factors and 17 Confirmed Genetic Loci. Am J Epidemiol 2016;184:579-589.

11. Cheng Y, Jiang T, Zhu M, et al. Risk assessment models for genetic risk predictors of lung cancer using two-stage replication for Asian and European populations. Oncotarget 2016.

12. Spitz MR, Amos CI, D'Amelio A, Jr., et al. Re: Discriminatory accuracy from single-nucleotide polymorphisms in models to predict breast cancer risk. J Natl Cancer Inst 2009;101:1731-2; author reply 1732.

13.   Xie SH, Lagergren J. A model for predicting individuals' absolute risk of esophageal adenocarcinoma: Moving toward tailored screening and prevention. Int J Cancer 2016;138:2813-9.

14.   Thrift AP, Kendall BJ, Pandeya N, et al. A model to determine absolute risk for esophageal adenocarcinoma. Clin Gastroenterol Hepatol 2013;11:138-44.

15.   Thrift AP, Kendall BJ, Pandeya N, et al. A clinical risk prediction model for Barrett esophagus. Cancer Prev Res (Phila) 2012;5:1115-23.

16.   Thrift AP, Garcia JM, El-Serag HB. A multibiomarker risk score helps predict risk for Barrett's esophagus. Clin Gastroenterol Hepatol 2014;12:1267-71.

17.   Rubenstein JH, Morgenstern H, Appelman H, et al. Prediction of Barrett's esophagus among men. Am J Gastroenterol 2013;108:353-62.

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.   Liu X, Wong A, Kadri SR, et al. Gastro-esophageal reflux disease symptoms and demographic factors as a pre-screening tool for Barrett's esophagus. PLoS One 2014;9:e94163.

20.   Gharahkhani P, Fitzgerald RC, Vaughan TL, et al. Genome-wide association studies in oesophageal adenocarcinoma and Barrett's oesophagus: a large-scale meta-analysis. Lancet Oncol 2016;17:1363-1373.

21.   Ek WE, Levine DM, D'Amato M, et al. Germline genetic contributions to risk for esophageal adenocarcinoma, Barrett's esophagus, and gastroesophageal reflux. J Natl Cancer Inst 2013;105:1711-8.

22.   Su Z, Gay LJ, Strange A, et al. Common variants at the MHC locus and at chromosome 16q24.1 predispose to Barrett's esophagus. Nat Genet 2012;44:1131-6.

23.   Levine DM, Ek WE, Zhang R, et al. A genome-wide association study identifies new susceptibility loci for esophageal adenocarcinoma and Barrett's esophagus. Nat Genet 2013;45:1487-93.

24.   Palles C, Chegwidden L, Li X, et al. Polymorphisms near TBX5 and GDF7 are associated with increased risk for Barrett's esophagus. Gastroenterology 2015;148:367-78.

25.   Laurie CC, Doheny KF, Mirel DB, et al. Quality control and quality assurance in genotypic data for genome-wide association studies. Genet Epidemiol 2010;34:591-602.

26.   Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007;81:559-75.

27.   Hoyo C, Cook MB, Kamangar F, et al. Body mass index in relation to oesophageal and oesophagogastric junction adenocarcinomas: a pooled analysis from the International BEACON Consortium. Int J Epidemiol 2012;41:1706-18.

28.   Thrift AP, Shaheen NJ, Gammon MD, et al. Obesity and risk of esophageal adenocarcinoma and Barrett's esophagus: a Mendelian randomization study. J Natl Cancer Inst 2014;106.

29.   Lemeshow S, Hosmer DW, Jr. A review of goodness of fit statistics for use in the development of logistic regression models. Am J Epidemiol 1982;115:92-106.

30.   Leening MJ, Vedder MM, Witteman JC, et al. Net reclassification improvement: computation, interpretation, and controversies: a literature review and clinician's guide. Ann Intern Med 2014;160:122-31.

31.   Sampliner RE. Practice guidelines on the diagnosis, surveillance, and therapy of Barrett's esophagus. The Practice Parameters Committee of the American College of Gastroenterology. Am J Gastroenterol 1998;93:1028-32.

32.   Vaughan TL, Fitzgerald RC. Precision prevention of oesophageal adenocarcinoma. Nat Rev Gastroenterol Hepatol 2015;12:243-8.

33.   Lagergren J, Bergstrom R, Lindgren A, et al. Symptomatic gastroesophageal reflux as a risk factor for esophageal adenocarcinoma. N Engl J Med 1999;340:825-31.

34.   Abraham G, Havulinna AS, Bhalala OG, et al. Genomic prediction of coronary heart disease. Eur Heart J 2016;37:3267-3278.

35. Mavaddat N, Pharoah PD, Michailidou K, et al. Prediction of breast cancer risk based on profiling with common genetic variants. J Natl Cancer Inst 2015;107.

36. Chen H, Liu X, Brendler CB, et al. Adding genetic risk score to family history identifies twice as many high-risk men for prostate cancer: Results from the prostate cancer prevention trial. Prostate 2016;76:1120-9.

37. Pencina MJ, D'Agostino RB, Sr., D'Agostino RB, Jr., et al. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. Stat Med 2008;27:157-172.

38. Thrift AP, Kanwal F, El-Serag HB. Prediction models for gastrointestinal and liver diseases: too many developed, too few validated. Clin Gastroenterol Hepatol 2016;14:1678-1680.

39. Steyerberg EW, Harrell FE, Jr. Prediction models need appropriate internal, internal-external, and external validation. J Clin Epidemiol 2016;69:245-247.

Figure Legend

Figure 1. The receiver operating characteristic curves associated with risk prediction models for A. Barrett's esophagus (BE) and B. esophageal adenocarcinoma (EAC). Demographic/lifestyle model includes age, sex, smoking, BMI and use of NSAIDs.

Sussex County Hospital, Freeman Hospital, Royal Victoria Infirmary, Victoria Hospital Blackpool, Weston Park Hospital, Royal Hampshire County Hospital, Conquest Hospital, Royal Bournemouth General Hospital, Mount Vernon Hospital, Lister Hospital, William Harvey Hospital, Kent and Canterbury Hospital, Great Western Hospital, Dumfries and Galloway Royal Infirmary, Poole General Hospital, St Hellier Hospital, North Devon District Hospital, Salisbury District Hospital, Weston General Hospital, University Hospital Coventry, Warwick Hospital, George Eliot Hospital, Alexandra Hospital, Nottingham University Hospital, Royal Chesterfield Hospital, Yeovil District Hospital, Darlington Memorial Hospital, University Hospital of North Durham, Bishop Auckland General Hospital, Musgrove Park Hospital, Rochdale Infirmary, North Manchester General, Altnagelvin Area Hospital, Dorset County Hospital, James Paget Hospital, Derriford Hospital, Newham General Hospital, Ealing Hospital, Pinderfields General Hospital, Clayton Hospital, Dewsbury & District Hospital, Pontefract General Infirmary, Worthing Hospital, Macclesfield Hospital, University Hospital of North Staffordshire, Salford Royal Hospital, Royal Shrewsbury Hospital, Manchester Royal Infirmary.