Cohort study: Risk stratification of Barrett's oesophagus using a non-endoscopic sampling method coupled with a biomarker panel

Caryn S Ross-Innes[1]*, PhD, Hamza Chettouh[1]*, PhD, Achilleas Achilleos[1]*, PhD, Nuria Galeano-Dalmau[1], MSc, Irene Debiram-Beecham[1], MSc, Shona MacRae[1], PhD, Petros Fessas[1], BA, Elaine Walker[1], BSc, Sibu Varghese[1], PhD Theodore Evan[1], MBBChir, Pierre S Lao-Sirieix[1], PhD, Maria O'Donovan[2], MD FRCPath, Shalini Malhotra[2], MBBS FRCPath, Prof Marco Novelli[3], PhD FRCPath, Babett Disep[4], MBChB FRCPath, Phillip V Kaye[5], MBChB FRCPath, Prof Laurence B Lovat[3], MBBS PhD FRCP, Rehan Haidry[3], MD FRCP, Prof Michael Griffin[4], FRCS, Prof Krish Ragunath[5], DNB MPhil FRCP, Prof Pradeep Bhandari[6], MD FRCP, Adam Haycock[7], MD MRCP, Danielle Morris[8], MD MRCP, Prof Stephen Attwood[9,10], MBBCh FRCS, Anjan Dhar[9,11], MD FRCP, Prof Colin Rees[9,12], MBBS FRCP, Prof Matt D Rutter[9,13], MD FRCP, Richard Ostler[14], MSc, Benoit Aigret[14], MSc, Peter D Sasieni[14], PhD and Prof Rebecca C Fitzgerald[1], MD FMedSci, on behalf of the BEST2 study group.

*contributed equally to this work

1. MRC Cancer Unit, Hutchison/MRC Research Centre, University of Cambridge, Cambridge, UK
2. Dept. Histopathology, Addenbrooke's Hospital, Cambridge, UK
3. University College London Hospital, London, UK
4. Royal Victoria Infirmary, Newcastle upon Tyne, UK
5. NIHR Nottingham Digestive Disease Biomedical Research Unit, Queens Medical Centre, Nottingham University Hospital NHS Trust
6. Queen Alexandra Hospital, Portsmouth, UK

7. St Mark's Hospital, London, UK

8. East and North Hertfordshire NHS Trust - QEII and Lister Hospitals, UK

9. Northern Region Endoscopy Group (NREG), UK

10. North Tyneside General Hospital, UK

11. County Durham and Darlington NHS Foundation Trust, UK

12. South Tyneside NHS Foundation Trust, Tyne and Wear, UK

13. North Tees and Hartlepool NHS Foundation Trust, UK

14. UK Cancer Prevention Trials Unit, London, UK

15. Multi-centre BEST2 study group are listed at the end of the manuscript

Correspondence to:

Professor Rebecca Fitzgerald

MRC Cancer Unit, Hutchison-MRC Research Centre, University of Cambridge

Hills Road, Cambridge, CB2 0XZ

Tel: 01223 763292

Email: rcf29@MRC-CU.cam.ac.uk

Abstract

Background

Barrett's oesophagus predisposes to adenocarcinoma. However, most Barrett's patients will not progress and endoscopic surveillance is invasive, expensive, and fraught by issues of sampling bias and the subjective assessment of dysplasia. This study investigated whether a non-endoscopic device, the Cytosponge, could be coupled with clinical and molecular biomarkers to identify a low-risk group for non-endoscopic follow-up.

Methods

In this multicentre cohort study (BEST2), patients with Barrett's underwent the Cytosponge test prior to their surveillance endoscopy. Clinical and demographic data were collected and Cytosponge samples were tested for a molecular biomarker panel including three protein biomarkers (p53, c-Myc, and Aurora kinase A), two methylation markers (*MYOD1* and *RUNX3*), glandular atypia and *TP53* mutation status. A multivariable logistic regression model was used to compute the conditional probability of dysplasia status. A simple model with high classification accuracy was selected and applied to an independent validation cohort.

Findings

In a discovery cohort (n=468) a model with high classification accuracy consisted of: glandular atypia, p53 abnormality and Aurora kinase A positivity and the interaction of age, obesity and length of the Barrett's segment**.** 35% patients fell into the low-risk category and the probability of being a true non-dysplastic patient was 162/162 (Confidence Intervals at 0.01 significance level 96-99·99%). In the validation cohort (n=65) 25/65 patients (38%) were classified as low risk and the probability of being a true non-dysplastic was 96·0% (99% CI 73·8-99·99%). The moderate-risk group comprised 27 non-dysplastic and 8 high grade dysplasia (HGD) cases while the high risk group had 0 non-dysplastic and 5 HGD patients (8% cohort).

Interpretation

A combination of biomarker assays from a single Cytosponge sample can be used to determine a group of patients at low risk for whom endoscopy could be avoided. This strategy could help to avoid over-diagnosis and over-treatment in Barrett's patients.

Funding

The BEST2 study was funded by Cancer Research UK.

Research in context

Evidence before the study

We searched PubMed from database inception to September 1st, 2001 with the MESH terms: biomarkers, cancer progression, Barrett's oesophagus prior to the start of the BEST2 trial, in order to review the status of the literature. This important area has attracted a lot of attention and aside from the histopathological assessment of dysplasia the most promising biomarkers include p53 status, copy number alterations and recently methylation panels have been investigated. All of these studies have relied on endoscopic sampling which is invasive and resource intensive. Furthermore, most effort has focussed on identifying patients at high risk for cancer. However since the majority will not progress to cancer, strategies to risk-stratify patients and avoid over-diagnosis are also very important. Furthermore, given the heterogeneity in the molecular genetic patient profiles of those progressing to cancer, identifying very low risk patients may be a more achievable biomarker strategy. A non-endoscopic cell collection device (Cytosponge) has been shown to be effective to diagnose Barrett's when coupled with a biomarker TFF3. The aim of this study was to determine whether additional biomarkers could identify patients at very low risk in whom further monitoring could be performed via the Cytosponge. Endoscopy and treatment could then focus on the highest risk patients.

Added value of the study

We tested a clinical and molecular biomarker panel on Cytosponge samples (7 variables) and used multivariable logistic regression to identify a minimal panel that

could risk stratify patients. The optimal panel comprises: age, length of Barrett's segment, waist:hip ratio and three biomarkers scored in a binary fashion to indicate p53 status, Aurora kinase A expression and glandular atypia and divides patients into low, moderate and high risk. This risk score was validated in an independent patient cohort and shown to confidently identify approximately 30% patients with a low risk for malignant progression. The high and moderate risk patients would be triaged for endoscopy which could be prioritised accordingly. To the best of our knowledge this is the first multidimensional panel applied to non-endoscopic oesophageal cell samples (ie, one involving DNA mutation, protein expression and clinical variables) to determine the risk profile for a patient with Barrett's oesophagus.

Implications of the evidence

Currently monitoring of patients with diagnosed Barrett's relies on endoscopy which is invasive for patients and expensive for the health care system. The use of a Cytosponge-biomarker test has the potential to more objectively risk stratify patients and identify a low-risk group for monitoring within the primary care setting that could be spared endoscopy.

Introduction

Oesophageal adenocarcinoma (OAC) patients have a median survival of one year, despite advances in therapy.[1] The burden of OAC could be reduced by diagnosing more cases of the precursor lesion Barrett's oesophagus (BO) while identifying those at increased risk for cancer development before treating them endoscopically.[2] However, this is a formidable task since the incidence of gastro-oesophageal reflux is approximately 5 per 1000 person-years in the UK.[3] Furthermore, despite clinical guidelines for endoscopic referral, primary care practice varies and low endoscopy referral rates correlate with poor outcomes.[4] The scale of the problem and the psychological and fiscal costs of endoscopy necessitate a new strategy.

We have developed a non-endoscopic diagnostic test which involves a cell-collection device which, coupled with the biomarker trefoil factor 3 (TFF3), diagnoses BO.[5,6] The device, called Cytosponge, comprises a medical-grade foam sphere on a string compressed within a gelatine capsule that is swallowed whilst holding onto the string. After 5 minutes, this dissolves within the stomach allowing the foam sphere to expand, before being pulled from the stomach through the oesophagus to the mouth. Cells are collected along the entire oesophageal lining, minimising the sampling bias inevitable with endoscopic biopsies. The sample is transported to the laboratory in preservative at room temperature and processed to paraffin for TFF3 biomarker assessment. Data from studies so far, amounting to over 2,000 patients, have shown that this approach is safe, has favourable acceptability rates compared to endoscopy, a sensitivity of 79·5-87% (depending Barrett's segment length), and a specificity of 92·4% for diagnosing BO.[5,6]

For TFF3-positive patients, it is essential that additional biomarkers are tested to assess the presence of genetic or molecular abnormalities indicative of dysplasia, as positive tests in patients with very benign disease at low risk of progression are commonplace in early cancer detection.[7] Over-diagnosis and over-treatment have adverse consequences for both patient and healthcare providers.[8] Our primary goal is thus to identify Barrett's patients with a very low risk profile, so that these individuals can be reassured without endoscopy.

We have shown that *TP53* mutations qualify for a risk stratification biomarker and are detectable using the Cytosponge.[9] We have also shown that the Cytosponge-sample is representative of the multiple clones within a diverse genetic content.[10] However, for a risk-stratification tool to be clinically applicable, it is important not to miss patients with high grade dysplasia (HGD) and intramucosal adenocarcinoma (IMC). Although *TP53* mutation has a high specificity, it will not identify all patients with HGD or early cancer, since the prevalence is 70-80%.[9,11] To increase our sensitivity for detecting high-risk Barrett's patients, we propose including additional molecular biomarkers combined with clinical factors.

Histopathological grade of dysplasia is a good predictor when confirmed by independent pathologists,[12] so we included a measure of glandular atypia present within the Cytosponge sample. c-Myc was included as it is recurrently amplified in OAC,[15] while many of its target genes have been identified as overexpressed in Barrett's with dysplasia.[15] Copy number change is a strong predictor of progression and changes dramatically in the transition to invasive disease.[10,16,17] Due to the infeasibility, expense and low sensitivity of performing cell cycle analysis or SNP arrays on FFPE Cytosponge material, we selected immunohistochemical expression of Aurora kinase A (AurKA) as a surrogate aneuploidy marker.[18] AurKA expression has also been shown to be significantly upregulated in Barrett's with HGD and OAC compared to non-dysplastic.[19]

In addition to *TP53* sequencing, p53 staining was included as this has been shown to be associated with dysplasia,[20] although the p53 staining-absent pattern cannot be reliably scored on the Cytosponge samples and was excluded. Recent evidence has shown that methylation is a good predictor of progression.[22] Five genes shown to be methylated with increasing grade of dysplasia were tested: p16, *ESR1*, *MYOD1*, *HPP1*, and *RUNX3.*[21] After reviewing pilot data, only *MYOD1* and *RUNX3* were taken forward, as these were the most promising (See Appendix, p3).

In summary, the tissue biomarkers taken forward into this study were therefore: atypia, AurKA, p53 abnormality (mutations and protein over-expression), c-Myc overexpression, *MYOD1* and *RUNX3* methylation (Appendix, p3). Clinical information

including demographics and symptoms were also assessed as additional biomarkers of risk.

Hence, the study aims were: to identify a clinically applicable risk-stratification biomarker panel which could be performed on the Cytosponge samples; - to apply this panel to a large cohort of patients with BO and a TFF3-positive Cytosponge sample to confidently identify a low-risk group of at least 30% of cases who could be spared endoscopy; and - to test the Cytosponge risk-stratification model on an independent cohort of patients.

Methods:

Study design and participants

The prospective, multi-centre BEST2 case-control cohort was designed to examine risk stratification in the Barrett's (case) arm as a secondary objective.[6] Ethics approval was obtained from the East of England–Cambridge Central Research Ethics Committee (No: 10/H0308/71) and registered in the UK Clinical Research Network Study Portfolio (9461). Patients in the validation cohort were selected from newly registered BEST2 patients and from the CASE1 study (see Appendix, p4).[22] Earliest date of enrolment for participants was 07/07/2011. Written informed consent was obtained for each patient. Data was collected on demographics, clinical exposures (alcohol, tobacco, drugs), and symptoms and anonymised. We collected W:H ratio and BMI in the BEST2 study but found that BMI is much more frequently measured in clinical practice and is less prone to error. For the validation cohort we therefore used BMI only. There was no minimum segment length requirement for BO provided they had a least one TFF3 positive cell. Cytosponge specimens were processed to paraffin blocks, as previously published.[6] The gastroscopies were performed within an hour of Cytosponge collection. Biopsy samples were taken from any visible lesions and from each quadrant, every 2 cm as recommended by the Seattle surveillance protocol. Diagnostic biopsies were reviewed locally. Biopsies with a diagnosis of dysplasia were reviewed in a consensus meeting by experienced pathologists from four institutions (M.O., M.N., B.D., and P.K.) that were blinded to the Cytosponge test result.

Procedures

The processed TFF3 positive sample was cut into consecutive sections and the first slide containing two sections was stained with H&E. No enrichment was made for TFF3 positive or histopathological atypical areas prior to assessment for molecular biomarkers – the entire Cytosponge section was evaluated for each biomarker.

Any glandular atypia present on this slide was assessed by two expert pathologists (M.O. and S.M.) blinded to the clinical diagnosis and a consensus reached (Kappa

0.66). For protein biomarker immunohistochemistry (IHC), slide 4 was used for p53, slide 8 for c-Myc and slide 10 for AurKA. Slides were stained using the BondMax autostainer with the Leica Bond Polymer Detection kit. The conditions and antibodies used can be found in the Appendix, p5.

p53 staining with an intensity of 3 was considered significant, as previously published.[20] c-Myc intensity was scored as 0-3 with 2 and 3 being considered significant staining.[23] As p53 is most commonly scored in a binary fashion and also due to the varying amount of Barrett's tissue in the Cytosponge samples, we decided against using a % scoring system. AurKA scoring was determined with a pilot study to evaluate different cut-offs and bearing in mind clinical feasibility. Thus AurKA was scored as non-significant or significant staining, with non-significant being fewer than 5 positive-staining cells, in the whole section.

Genomic DNA was extracted from 8x10µm sections of the processed Cytosponge FFPE clot using Deparaffinization Buffer (Qiagen) and the QIAamp FFPE DNA Tissue Kit (Qiagen). The manufacturer's protocol was followed with the exception that samples were incubated at 56^{o}C for 24 hours instead of 1 hour, and 10 µl of extra Proteinase K were added to the samples halfway through the incubation. FFPE-extracted DNA was quantified by PCR using primers specific to ALU115 repetitive elements (See Appendix, p1). 10-25ng quantified DNA was used for library preparation using *TP53* Accel-Amplicon comprehensive panel (Swift Biosciences) according to manufacturer's guidelines and sequenced on the Illumina Miseq machine using 150bp paired-end sequencing to achieve an average of 10,000-fold coverage per sample.

Outcomes

Sequencing quality was checked using the FastQC program. Trimmed reads were aligned to GRCh37 human reference genome (hg19) using BWA aligner, then sorted by genomic positions and processed using the GATK pipeline. The GRCh38 version was not available when we started this work. We used LoFreq to detect low allele-frequency variants from the Cytosponge samples and annotated the results using VEP command from ensembl database. Libraries were prepared in duplicate for each

sample and mutations called in both duplicates with an allele frequency above 0·5% were considered as true positives.

DNA was bisulphite-converted using the EZ DNA Methylation-Gold™ kit (as described by Zymo Research). Samples were eluted in 25 µl water and 2 µl was used per MethyLight reaction, as previously described.[24] β-actin was used to normalise for amount of input DNA. Universally methylated and bisulphite-converted DNA (D5010-1, Zymo Research) was used to derive standard curves for each primer and probe set and a calibrator was used in all experiments to allow absolute quantification of methylation levels. Amplification conditions used for all reactions were: 95°C for 10 mins followed by 50 cycles of 95°C for 15 seconds and 60°C for 1 minute. Percentage methylation of each gene was calculated as previously published,[17] as was the combined methylation value for *RUNX3* plus *MYOD1*. The data underlying our choice of these 2 biomarkers is shown in the Appendix, p3.

Statistical analysis

The histopathological diagnosis from the endoscopy biopsy taken at the same time as the Cytosponge was used as the gold standard for comparison. Data analysis was performed using R statistical software version 3.0.2,[31] and missing values were imputed using iterative regression imputation.[32] A logistic regression model was used to compute the conditional probability to have Barrett's with high grade dysplasia/intramucosal adenocarcinoma (HGD/IMC), given the covariates. Consider a simple decision theory problem. Rather than modelling the dichotomous response variable Y directly (Best2diagnosis; 1: HGD/IMC, 0: Negative), we modelled the probability that Y belongs to a particular category, given the value of p covariates, say $X=\{X_1,\ldots,X_p\}$. We classified samples based on estimating the conditional probabilities $p(Y|X)$ and then predicted the risk-profile of a patient as follows:

Low-risk if $p(Y=1|X) <= c_1$

Moderate-risk if $c_1 < p(Y=1|X) < c_2$

High-risk if $p(Y=1|X) >= c_2$

where c_1 and c_2 are estimated from the data (see Appendix, p1). A model consisted of glandular atypia, p53 abnormality, Aurora kinase A IHC and the interaction of BMI

(or waist:hip ratio), Barrett's oesophagus maximum length and age was selected. The model with BMI is preferable clinically. Estimated model coefficients and confidence intervals are provided in the Appendix, page 1. Clopper-Pearson Binomial confidence intervals for proportions are provided.

Role of the funding source

The BEST2 study was funded by Cancer Research UK. (Grant ref; C14478/A12088, http://www.cancerresearchuk.org/). The funders had no role in study design, data collection, analysis and interpretation, decision to publish, or writing of the report. The study received infrastructure support from the Cambridge Human Research Tissue Bank, which is supported by the National Institute for Health Research (NIHR) Cambridge Biomedical Research Centre, from Addenbrooke's Hospital. R.C.F. is funded by an NIHR Professorship and receives core funding from the Medical Research Council and infrastructure support from the Biomedical Research Centre and the Experimental Cancer Medicine Centre. R.C.F. and C.S.R. had full access to all the data in the study. The corresponding author had the final responsibility to submit for publication.

Results

Patient characteristics

TFF3-positive Cytosponge samples from 468 Barrett's patients with intestinal metaplasia were assessed. Of these, 376 had no dysplasia and 92 had HGD/IMC (Figure 1). Patients with HGD/IMC were older (Mann-Whitney test, $p<0.0001$), had longer Barrett's segments ($p<0.0001$) and had a higher waist:hip ratio ($p=0.008$) (Table 1).

Biomarker panel performance

Seven biomarkers, including three protein biomarkers for p53, c-Myc, and AurKA (Figure 2), two methylation markers (*MYOD1* and *RUNX3*), glandular atypia and *TP53* mutation were assessed on all 468 Cytosponge samples with very few assays failing

and resulting in missing data (Table 2). All biomarkers were analysed separately to estimate their sensitivity and specificity for detecting HGD/IMC compared to no dysplasia on the Cytosponge specimen. A positive value for p53 abnormality resulted from either significant p53 staining or TP53 mutation. The combined methylation value for *MYOD1* and *RUNX3* was also analysed (Table 2). AurKA IHC was the most sensitive biomarker (78%; 65-88%) for detecting HGD/IMC compared to no dysplasia. p53 IHC was the most specific biomarker (96%; 92-98%) p53 IHC and *TP53* mutation alone both had sensitivities of 58% (44-70%), but combining these increased sensitivity to 72% (58-83%) (Table 2).

No biomarkers were sensitive and specific enough individually. As our main aim was to identify a low-risk group who would be spared endoscopy, we performed a multivariable logistic regression analysis considering all biomarkers and a list of demographic information, clinical exposures (smoking, alcohol, medication), symptoms, and endoscopic findings (see Appendix, p6) to identify such a group. Using a 3-risk model approach (low-, moderate-, and high-risk), we selected the smallest possible number of predictors that attained the best possible specificity. The model with the best classification performance consisted of four predictors, with coefficient estimate shown in brackets: glandular atypia (2.435), p53 abnormality (1.802), AurKA IHC (0.876), and the interaction of age, waist:hip ratio, and maximum Barrett's segment length (0.0000634). Atypia and a positive AurKA score were associated with a longer segment length (Wilcox text p=<0.0001) and p53 status to a lesser extent (p=0.002), (Appendix, p10).

Testing this model on the full cohort resulted in all HGD/IMC Barrett's patients being categorised as moderate- or high-risk (Figure 3 and Table 3). For the low-risk category, the probability of being truly non-dysplastic was 162/162 (Confidence Interval at 0.01 significance level 96-100%) and the probability of being a HGD/IMC patient was 0/162 (0·01-4%). Furthermore, given a patient has HGD/IMC, the probability of this sample being classified as low-risk was less than 5·6%. For the moderate-risk group, the probability of being truly non-dysplastic was 86% (79-91%) and the probability of being a true HGD/IMC was 14% (9-21%). For high-risk patients, the probability of being a true HGD/IMC was 87% (73-95%) and the probability of

having non-dysplastic endoscopic biopsies was 13% (5-27%). Analysis of clinical data on the non-dysplastic patients misclassified as high-risk determined that one patient was diagnosed with metastatic adenocarcinoma, two had HGD/IMC requiring endoscopic therapy, two had an endoscopic diagnosis of indefinite for dysplasia and remain under follow-up, while the remaining 5 have not yet developed dysplasia.

Next, we checked the model in a cohort of 65 patients, with baseline characteristics similar to the discovery cohort (Table 1). 25/65 patients (38%) were classified as low-risk and these included 24 non-dysplastic cases and 1 with HGD (Table 3). Hence, if categorised as low-risk, the probability of being truly non-dysplastic was 96·0% (99%;73·8-99·99%). There were 27 non-dysplastic and 8 HGD cases classified as moderate-risk which made up 54% of the cohort and the high-risk group comprised 5 HGD and no non-dysplastic patients (8% cohort).

Low grade dysplasia (LGD) is a difficult histopathological diagnosis with low inter intra-observer variability and variable outcomes[12], so including them in the model from the start could cause confounding. We therefore analysed them separately. Out of 50 patients with LGD diagnosed from endoscopic biopsies taken on the same day as the Cytosponge, 16 were categorised as high-risk, 29 as moderate-risk and 5 as low-risk. Since a diagnosis of LGD is now a clinical indication for endoscopic ablation therapy, following evidence for the superiority of ablation therapy compared with surveillance,[13] it is impossible to assess the patients' natural history. However, we also examined how the diagnosis related to the highest grade of dysplasia ever recorded in these patients with LGD at the time of the Cytosponge and found that 7/16 (44%) patients in the high-risk category had a diagnosis of HGD/IMC on endoscopic biopsies taken before or after the Cytosponge test. This was more likely than in LGD patients assigned to the moderate or low-risk groups – 7/29 (24%) and 1/5 (20%), respectively (Appendix, p7). We also examined the treatment offered to these patients as a surrogate for the severity of disease: patients categorised as high risk were more likely to require surgical or endoscopic resection in addition to ablation therapy (Figure 4).

Discussion

This study demonstrates the possibility of performing a combination of biomarker assays from a single Cytosponge sample. Combining the biomarkers (p53 abnormality, glandular atypia, and AurKA staining) with clinical variables (age, length of Barrett's, and obesity), and using logistic regression enables us to risk-stratify patients into three risk groups. The primary objective was to determine with high confidence a group of patients at low risk for whom endoscopy could be avoided. The validation cohort suggests that the algorithm provides a high level of confidence (74-99% confidence intervals) for determining patients who are dysplasia-free. In both discovery and validation, approximately one third of patients fell into the low-risk category which could have a substantial impact on the financial and psychological burden for this group. High-risk patients should be prioritised for endoscopy and endoscopic therapy if dysplasia is confirmed. For moderate-risk patients an endoscopy would be recommended, though in the future a repeat Cytosponge might be an alternative given that the natural history of BO is generally slow and we have previously shown that repeating the Cytosponge improves sensitivity for biomarkers.[6]

Current practice relies on subjective histopathological assessment of dysplasia.[20] The latest UK BSG guidelines suggest using p53 IHC to help improve diagnostic confidence, and further studies have strengthened the evidence for this,[20] although this practice is not yet widely adopted. In the risk score developed here we use several objective measures, but also include pathologists' assessment of glandular atypia. The p53 status is based on both whole-gene sequencing as well as an immunoscore. We sequenced most coding exons (E2-11) and did not rely on *TP53* hotspot analysis, since sequencing has demonstrated that mutations can occur anywhere along the gene.[9] The AurKA staining is used as a surrogate for copy number and the scoring is binarised to make it as straightforward as possible. Currently, clinical features are excluded from clinical assessment of patients undergoing surveillance, although Barrett's segment length is a well-recognised risk factor.[27] The relevant contribution of individual biomarkers to the risk prediction varies, with glandular atypia and p53 status being the most important (see Appendix, p8).

We restricted the main analyses in the discovery and validation cohorts to distinguish between non-dysplastic Barrett's and HGD, because LGD is a diagnostic conundrum. Recent studies have shown that the outcome of a LGD diagnosis can vary considerably depending on whether or not the diagnosis was confirmed by multiple pathologists at different institutions.[29] When we performed the risk stratification panel on patients with LGD, we found that patients fell into all three risk categories. These cases had all been reviewed by at least two pathologists. Unsurprisingly, these cases therefore mainly (88%) fell into the moderate and high-risk categories, in keeping with previous studies in which a consensus diagnosis of LGD has a higher risk of malignant progression.[12] It is interesting that in our study LGD patients in the high-risk category were more likely to have a diagnosis of HGD/IMC at some time during their endoscopic history and to harbour a visible lesion, suggesting that the risk category is meaningful.

The question arises as to how clinically applicable the risk score would be. We did not perform any enrichment of the Barrett's cells within the Cytosponge samples and we did not normalise any of the biomarkers according to the number of TFF3+ cells in order to make this laboratory panel as clinically applicable as possible. p53 immunostaining is routine in histopathology for various conditions and automation of nuclear scoring is well-described. Sequencing of a single gene, such as p53, is becoming common-place in oncology.[30] The AurKA stain is cytoplasmic and less robust: we have used a binary score to make it as practical as possible and, since this contributes the least to the final score, further studies should evaluate whether this could be omitted. Our analysis used BMI or waist:hip ratio and we have found that these can be used interchangeably: BMI would be preferable clinically, being less prone to measurement error. The risk category can be easily calculated (see Appendix, p9, and separate Excel file) and this is becoming common in management of several conditions, including cardiovascular disease.[31] The economics of our approach would need to be determined, but avoiding endoscopy and histopathological assessment of multiple biopsies could save money and reduce the burden of an invasive procedure for patients.

There are a number of limitations to our study. Most centres involved are tertiary referral centres and so the population studied may not represent a surveillance cohort in a general hospital i.e. at higher risk and also may have a different standard of care. The number of patients is reasonable for a risk-stratification study but testing in larger prospective cohorts is required. Another difficulty is with regards to the choice of a gold standard for comparison. We have chosen the grade of dysplasia at the endoscopy performed alongside the Cytosponge. Due to sampling error and subjective histopathological assessment, this will lead to miscategorisation. Consensus pathology review of biopsies taken at the same time as the Cytosponge was performed on all cases with dysplastic Barrett's however this was not feasible for the large number of cases without dysplasia which could lead to a false-negative gold standard diagnosis. For example, when we examined the patients categorised as high-risk by the Cytosponge who had no dysplasia (discovery cohort) or LGD on endoscopic biopsies, we found that some of them had had a diagnosis of HGD in an endoscopy performed before or soon after the Cytosponge. If you take non dysplastic patients at the time of Cytosponge but who had previously had high grade dysplasia, i.e. 12 patients - then 4 were high risk, 5 moderate risk and 3 low risk. The number of cases with missing biomarker data was very small. However, the sections cut for p53 sequencing were done last and in some cases there wasn't sufficient to perform the assay. We hope that collection of long-term follow-up data on these patients will provide further confirmation of the value of the Cytosponge risk score.

We expect that the risk stratification algorithm presented here can be further improved by alternative biomarkers. Before incorporating the Cytosponge-biomarker test into clinical practice for surveillance, a large randomised clinical trial powered according to the data presented here will be required in which patients are allocated to a follow-up or treatment strategy based on their Cytosponge result. In our view, this innovative strategy combining a non-endoscopic device with a multi-dimensional biomarker panel has the potential to improve management of patients with pre-malignant BO in whom risk stratification is essential to avoid over-diagnosis and over-treatment.

Author contributions:

RCF and PDS designed the study. CSR-I, HC, and RCF performed the data analysis and data interpretation. CSR-I, HC, N G-D, CEW, SV, TE performed wet lab assays. AA performed the statistical analysis. ID-B co-ordinated the clinical study. CSR-I, ID-B, MOD, CEW, SV, LL, MG, KR, RH, SSS, PB, AH, DM, SA, AD, CR, PF, PLS and MDDR were involved in data collection. MOD, PK, MN, BD, and SM were involved with pathology review. RCF, CSR-I, and PF wrote the manuscript. CSR-I, AA and PF constructed the figures. RO built the study database. BA and PDS had oversight of the clinical trial.

Competing interests:

Since this study was conducted the Cytosponge™-TFF3 technology has been licensed to Covidien GI solutions (now owned by Medtronic) by the Medical Research Council. Rebecca Fitzgerald, Maria O'Donovan and Pierre Lao-Sirieix are named inventors on patents pertaining to the Cytosponge™. Covidien Solutions and Medtronic have not been privy to this manuscript or the data therein. All other authors declare no conflicts of interest.

Ethics committee approval:

Ethics approval was obtained from the East of England–Cambridge Central Research Ethics Committee (No: 10/H0308/71) and registered in the UK Clinical Research Network Study Portfolio (9461).

Acknowledgements:

R. C. F. and C. S. R. had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. We would like to thank Nitzan Rosenfeld and Francesco Marass for their helpful advice on p53

sequencing methods with optimal sensitivity to detect a low allele fraction. We would like to thank all the patients who took part in the study. We would also like to thank the Genomics Core Facility at the CRUK Cambridge Institute for their sequencing service.

References

1. Oesophageal cancer survival statistics. Cancer Research UK. 2015. http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/oesophageal-cancer/survival (accessed 19 July 2016).
2. Vaughan T, Fitzgerald R. Precision prevention of oesophageal adenocarcinoma. Nature Reviews Gastroenterology & Hepatology 2015; 12: 243-248.
3. El-Serag H, Sweet S, Winchester C, Dent J. Update on the epidemiology of gastro-oesophageal reflux disease: a systematic review. Gut 2013; 63: 871-880.
4. Shawihdi M, Thompson E, Kapoor N et al. Variation in gastroscopy rate in English general practice and outcome for oesophagogastric cancer: retrospective analysis of Hospital Episode Statistics. Gut 2013; 63: 250-261.
5. Kadri S, Lao-Sirieix P, O'Donovan M et al. Acceptability and accuracy of a non-endoscopic screening test for Barrett's oesophagus in primary care: cohort study. BMJ 2010; 341: c4372-c4372.
6. Ross-Innes C, Debiram-Beecham I, O'Donovan M et al. Evaluation of a Minimally Invasive Cell Sampling Device Coupled with Assessment of Trefoil Factor 3 Expression for Diagnosing Barrett's Esophagus: A Multi-Center Case–Control Study. PLoS Med 2015; 12: e1001780.
7. van Luijt P, Heijnsdijk E, Fracheboud J et al. The distribution of ductal carcinoma in situ (DCIS) grade in 4232 women and its impact on overdiagnosis in breast cancer screening. Breast Cancer Research 2016; 18. DOI:10.1186/s13058-016-0705-5.
8. Moynihan R, Doust J, Henry D. Preventing overdiagnosis: how to stop harming the healthy. BMJ 2012; 344: e3502-e3502.
9. Weaver J, Ross-Innes C, Shannon N et al. Ordering of mutations in preinvasive disease stages of esophageal carcinogenesis. Nature Genetics2014; 46: 837-843.
10. Ross-Innes C, Becq J, Warren A et al. Whole-genome sequencing provides new insights into the clonal architecture of Barrett's esophagus and esophageal adenocarcinoma. Nature Genetics 2015; 47: 1038-1046.
11. Dulak A, Stojanov P, Peng S et al. Exome and whole-genome sequencing of esophageal adenocarcinoma identifies recurrent driver events and mutational complexity. Nature Genetics 2013; 45: 478-486.
12. Duits L, Phoa K, Curvers W et al. Barrett's oesophagus patients with low-grade dysplasia can be accurately risk-stratified after histological review by an expert pathology panel. Gut 2014; 64: 700-706.
13. Phoa K, van Vilsteren F, Weusten B et al. Radiofrequency Ablation vs Endoscopic Surveillance for Patients With Barrett Esophagus and Low-Grade Dysplasia. JAMA 2014; 311: 1209.
14. Bandla S, Pennathur A, Luketich J et al. Comparative Genomics of Esophageal Adenocarcinoma and Squamous Cell Carcinoma. The Annals of Thoracic Surgery 2012; 93: 1101-1106.

15. Varghese S, Newton R, Ross-Innes C et al. Analysis of Dysplasia in Patients With Barrett's Esophagus Based on Expression Pattern of 90 Genes.Gastroenterology 2015; 149: 1511-1518.e5.
16. Stachler M, Taylor-Weiner A, Peng S et al. Paired exome analysis of Barrett's esophagus and adenocarcinoma. Nature Genetics 2015; 47: 1047-1055.
17. Nones K, Waddell N, Wayte N et al. Genomic catastrophes frequently arise in esophageal adenocarcinoma and drive tumorigenesis. Nature Communications 2014; 5: 5224.
18. Zhou H, Kuang J, Zhong L et al. Tumour amplified kinase STK15/BTAK induces centrosome amplification, aneuploidy and transformation. Nature Genetics 1998; 20: 189-193.
19. Rugge M, Fassan M, Zaninotto G et al. Aurora kinase A in Barrett's carcinogenesis. Human Pathology 2010; 41: 1380-1386.
20. Kaye P. p53 Immunohistochemistry as a biomarker of dysplasia and neoplastic progression in Barrett's oesophagus. Diagnostic Histopathology 2015; 21: 89-98.
21. Schulmann K, Sterian A, Berki A et al. Inactivation of p16, RUNX3, and HPP1 occurs early in Barrett's-associated neoplastic progression and predicts progression risk. Oncogene 2005; 24: 4138-4148.
22. Lao-Sirieix P, Debiram -Beecham I, Sarah K et al. 54 Evaluation of a Minimally-Invasive Cytosponge Esophageal Cell Collection System in Patients With Barrett's Esophagus. Gastroenterology 2015; 148: S-16.
23. Boult J, Tanière P, Hallissey M, Campbell M, Tselepis C. Oesophageal adenocarcinoma is associated with a deregulation in the MYC/MAX/MAD network.Br J Cancer 2008; 98: 1985-1992.
24. Eads C, Danenberg K, Kawakami K et al. MethyLight: a high-throughput assay to measure DNA methylation. Nucleic Acids Research 2000; 28: 32e-0.
25. Gelman A, Hill J. Data Analysis Using Regression and Multilevel/Hierarchical Models. Leiden: Cambridge University Press, 2006.
26. Venables W, Ripley B. Modern applied statistics with S., Springer New York, 2002
27. Coleman H, Bhat S, Murray L et al. Symptoms and Endoscopic Features at Barrett's Esophagus Diagnosis: Implications for Neoplastic Progression Risk.Am J Gastroenterol 2014; 109: 527-534.
28. Greene W. Econometric analysis. Upper Saddle River, NJ: Pearson, 2008.
29. Kaye P, Haider S, Ilyas M et al. Barrett's dysplasia and the Vienna classification: reproducibility, prediction of progression and impact of consensus reporting and p53 immunohistochemistry. Histopathology 2009; 54: 699-712.
30. Irshad H, Veillard A, Roux L, Racoceanu D. Methods for Nuclei Detection, Segmentation, and Classification in Digital Histopathology: A Review—Current Status and Future Potential.IEEE Rev Biomed Eng 2014; 7: 97-114.
31. Dalton ABull R. Risk stratification could reduce costs in primary prevention of cardiovascular disease. BMJ 2011; 343: d4913-d4913.

The BEST2 Study group full authorship list:

BEST2 is a multicenter UK collaboration. In addition to the listed authors, BEST2 members involved in this study were: Karen Coker, Wanfeng Zhao, Kathryn Brown, Beverley Haynes (Cambridge Tissue Bank); Tara Nuckcheddy Grant, Massimiliano di Pietro, Eleanor Dewhurst, Bincy Alias, Leanne Mills, Caroline Wilson, Elizabeth Bird-Lieberman, Jan Bornschein, Yean Lim, Kareem Shariff (Cambridge MRC Cancer Unit); Roberto Cayado Lopez, Myrna Udarbe (NIHR Clinical Investigation Ward: Research Endoscopy Unit, Cambridge); Claire Shaw, Glynis Rose (County Durham and Darlington NHS Foundation Trust); Ian Sargeant, M Al-Izzi, Roisin Schimmel, Elizabeth Green (East and North Herts QEII and Lister Hospitals); Morgan Moorghen, Reshma Kanani, Mariann Baulf, Jayne Butcher (St Mark's Hospital, London); Adil Butt, Steve Bown, Vinay Sehgal, John Louis-Auguste, Darina Kohoutova, Sarah Kerr, Victor Eneh, Nigel Butter, Haroon Miah, David Graham, Rommel Butawan, Manuel Rodriguez-Justo, Alison Winstanley, Grace Adesina, Sabrina Holohan (University College London Hospital); Joan Idris (Cancer Prevention Trials Unit, London, UK); Nick Hayes, Shajahan Wahed, Nelson Kath Houghton, Marc Hopton (Royal Victoria Hospital, Newcastle Upon Tyne); Anne Eastick, Debasis Majumdar, Kassem Manuf (North Tees and Hartlepool NHS Foundation Trust); Lyndsey Fieldson, Helen Bailey (North Tyneside); Jacobo Fernandez-Sordo Ortiz, Mina Patel, Suzanne Henry, Samantha Warburton, Jonathan White (Nottingham Queen's Medical Centre, Nottingham); Lisa Gadeke, Beverley Longhurst, Richmond Abeseabe, Peter Basford, Rupam Bhattacharyya, Scott Elliot, (Queen Alexandra Hospital, Portsmouth); Roisin Bevan, Carly Brown, Philippa Laverick, Gayle Clifford (South Tyneside NHS

Foundation Trust, Tyne and Wear); Anita Gibbons, Julie Ingmire (Hinchingbrooke); Abdullah Mawas (West Suffolk); Jacquelyn Harvey, Sharon Cave (Bedford Hospital).

Table 1: Demographics of the discovery and validation cohorts

	Discovery cohort		Validation cohort	
	Non-dysplastic Barrett's oesophagus	Barrett's oesophagus with HGD or IMC	Non-dysplastic Barrett's oesophagus	Barrett's oesophagus with HGD
Number	376	92	51	14
Age (years)	64 (56-71)	69 (63-74)	63 (53-69)	62 (49-67)
Ethnicity number (%)	366 (97·3%)	91 (98·9%)	49 (96·1%)	14 (100·0%)
White	9 (2·4%)	1 (1·1%)	2 (3·9%)	0 (0·0%)
Other	1 (0·3%)	0 (0·0%)	0 (0·0%)	0 (0·0%)
Refused				
Sex (M:F)	3·8 : 1	7·4 : 1	4·1 : 1	13·0 : 1
BMI	28·1 (25·5-30·8)	28.8 (26·1-31·1)	28·1 (25·6-30·4)	27.8 (25·3-31.9)
Waist:hip ratio	0·94 (0·90-0·98)	0.95 (0·93-1·01)	Not recorded*	Not recorded*
Hiatus hernia (%)	78·4%	82·0%	86·3%	92·9%
Barrett's oesophagus maximum length (cm)	4 (3-7)	7 (5-10)	4 (3-6)	5 (4-7)

Data shown reflect median (IQR) for age, BMI, waist:hip ratio and Barrett's oesophagus maximum length. Male: female ratio rounded to the nearest tenth. HGD = high-grade dysplasia, IMC = intramucosal adenocarcinoma.

*Waist:hip ratio was seldom recorded, so only BMI was used for the validation cohort.

Table 2: Sensitivity and specificities estimates for the individual biomarkers comparing non-dysplastic Barrett's oesophagus and Barrett's oesophagus with high grade dysplasia. p53 abnormality was computed using the combined data for p53 IHC and *TP53* mutation (i.e. either significant p53 staining or *TP53* mutation would give a positive value for p53 abnormality). Clopper and Pearson Binomial Confidence Intervals are provided for dichotomous variables. Logistic regression classification was performed using each of the continuous variables. Number of samples with missing data is shown.

Biomarker	Sensitivity (%)	Specificity (%)	Missing data (%)
p53 IHC	58 (44-70)	96 (92-98)	1 (0.2)
TP53 mutation	58 (44-70)	85 (80-90)	13 (2.8)
p53 abnormality	72 (58-83)	83 (77-88)	13 (2.8)
Glandular atypia	64 (50-77)	94 (90-97)	1 (0.2)
cMyc IHC	63 (49-75)	72 (66-78)	3 (0.6)
Aurora kinase A IHC	78 (65-88)	70 (64-77)	4 (0.9)
MYOD1 methylation	67 (61-74)	64 (50-77)	7 (1.5)
RUNX3 methylation	74 (67-79)	60 (46-73)	7 (1.5)
Combined MYOD1 and RUNX3 methylation	70 (63-76)	62 (48-75)	7 (1.5)

Table 3: Risk classification results for the discovery and validation cohorts. NDBO= Non-dysplastic Barrett's oesophagus, HGD= Barrett's oesophagus with high grade dysplasia, IMC=intramucosal adenocarcinoma

Discovery cohort (n=468)			Validation cohort (n=65)		
Risk classification	NDBO	HGD/IMC	Risk classification	NDBO	HGD
Low (n=162)	162 (43%)	0 (0%)	Low (n=25)	24 (47%)	1 (7%)
Moderate (n=238)	205 (55%)	33 (36%)	Moderate (n=35)	27 (53%)	8 (57%)
High (n=68)	9 (2%)	59 (64%)	High (n=5)	0 (0%)	5 (36%)

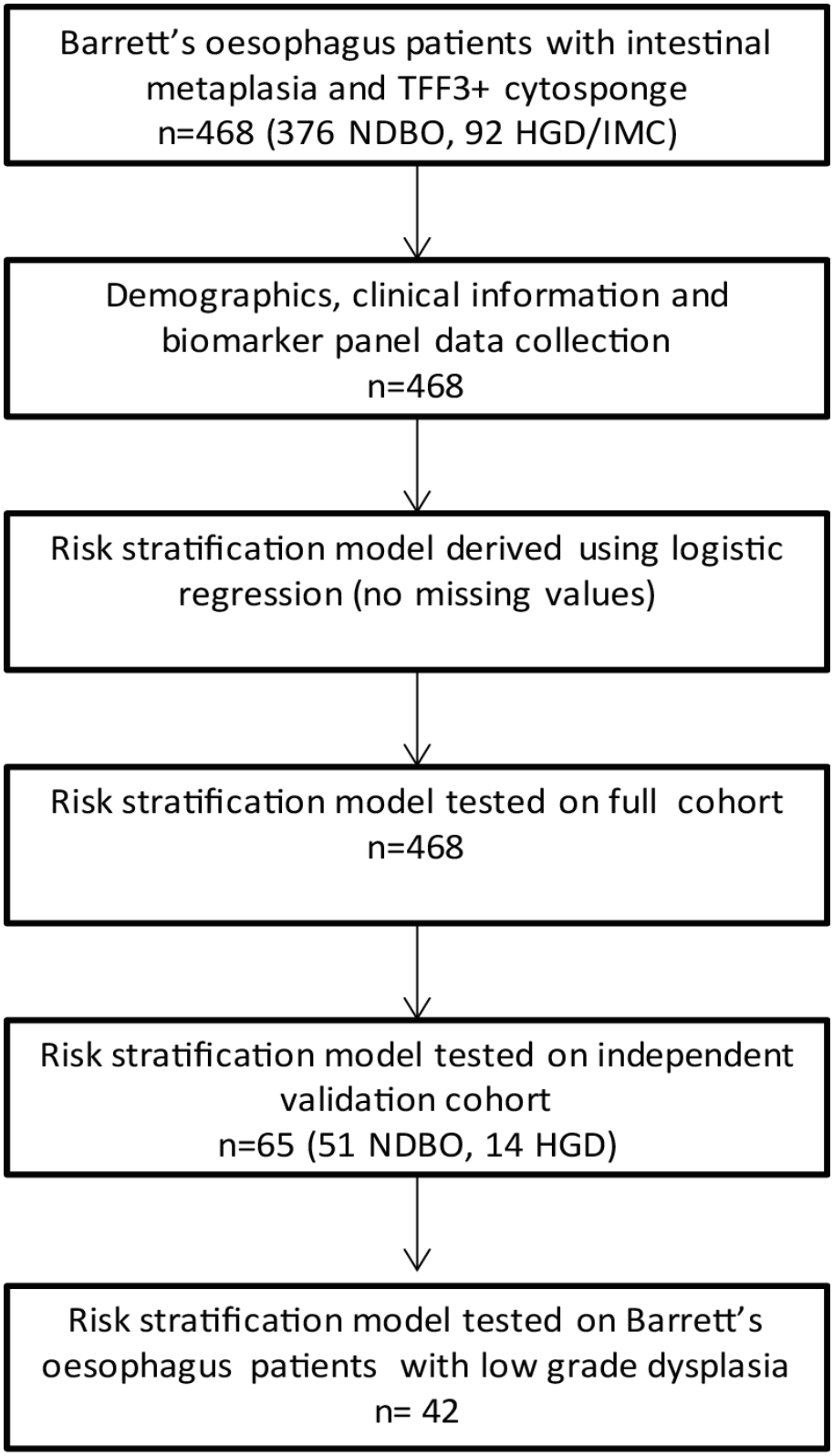


Figure 1: Study flow diagram

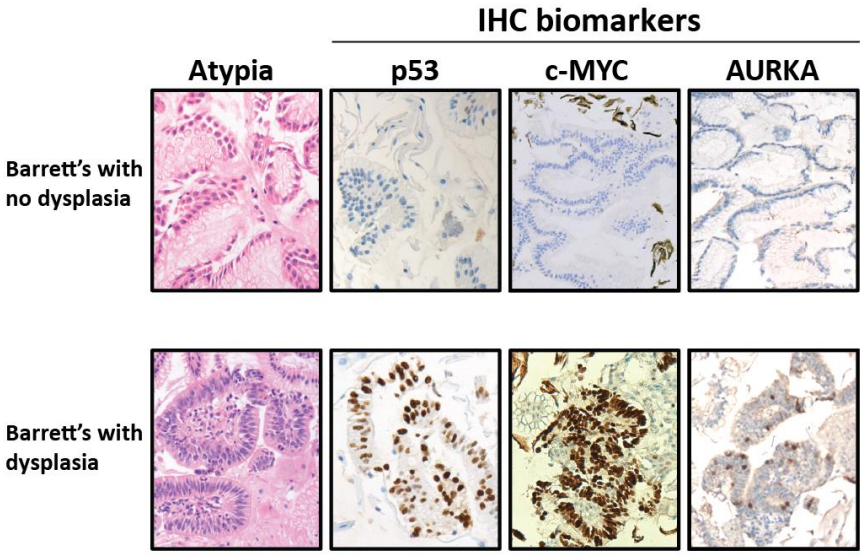
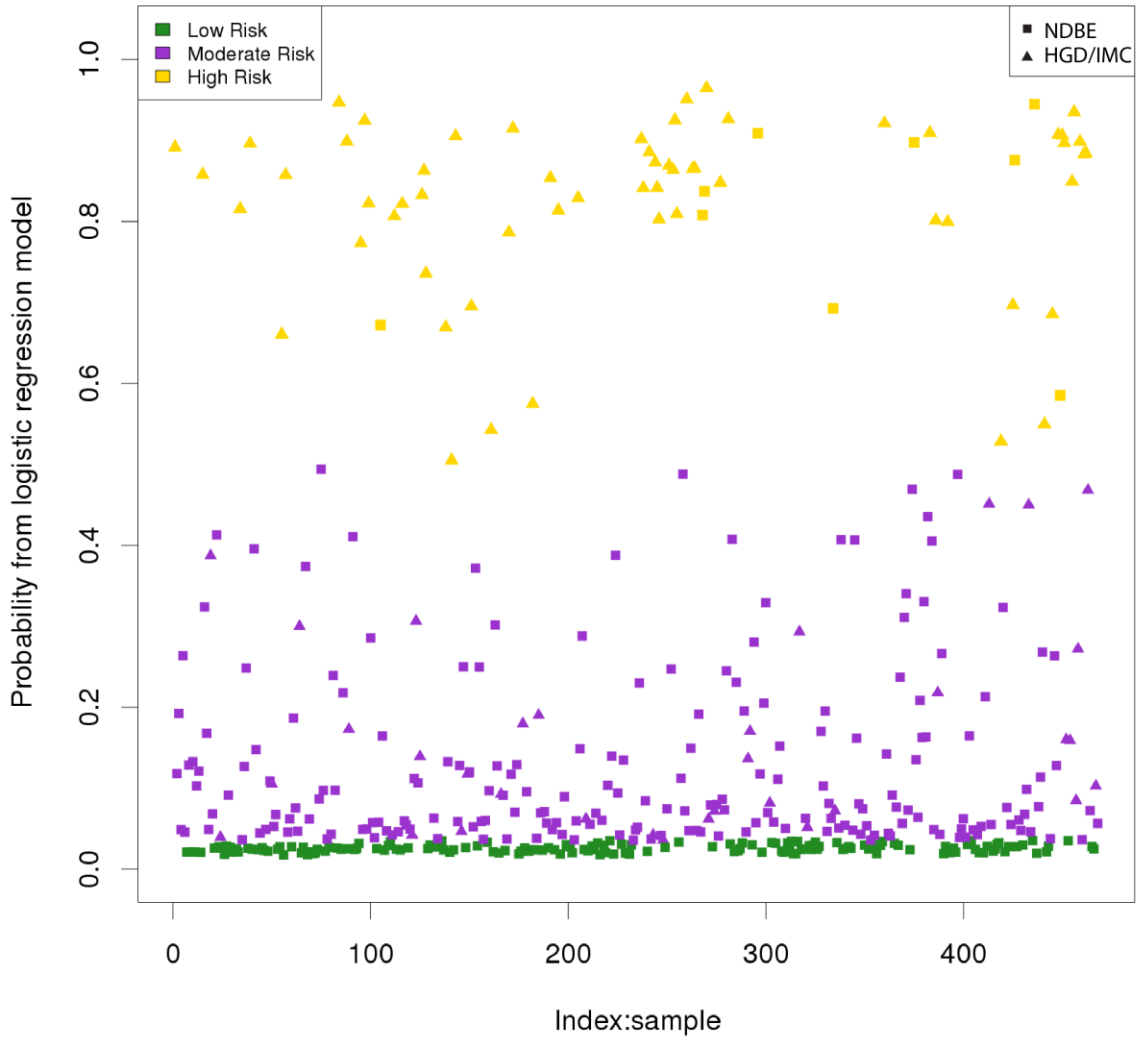


Figure 2: Examples of significant and non-significant staining for the three protein biomarkers as well as an example of glandular atypia as seen on a cytosponge sample.

Figure 3: Risk classification results for the 468 discovery cohort patients. Barrett's patients with no dysplasia are depicted by a square and patients with high-grade dysplasia or intramucosal adenocarcinoma are depicted by triangles.



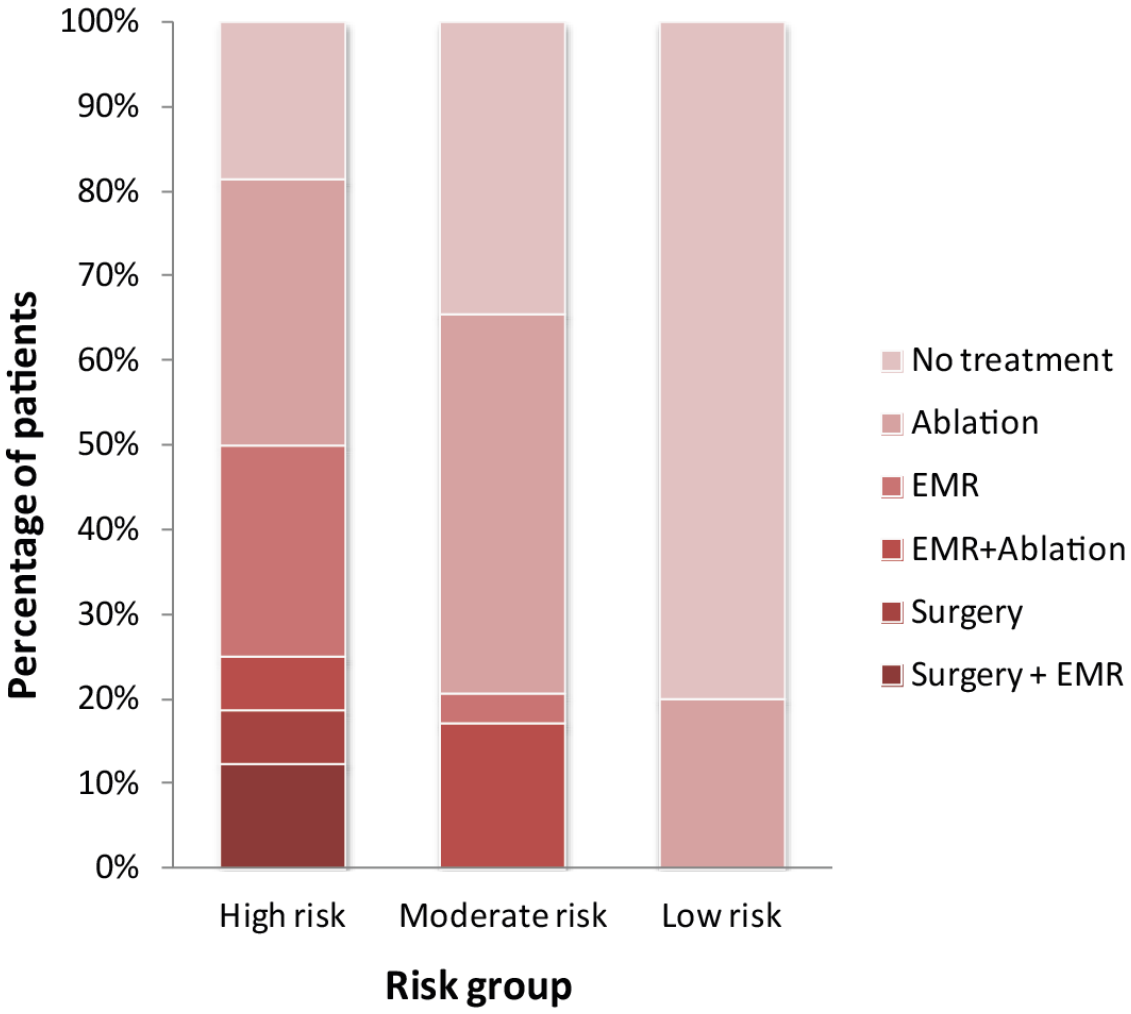


Figure 4: Intervention follow up of patients with low grade dysplasia assigned to high, moderate and low risks groups by the Cytosponge risk score. EMR = Endoscopic mucosal resection; Ablation = Radiofrequency ablation or Argon plasma coagulation.

Appendix: Risk stratification of Barrett's oesophagus using a non-endoscopic sampling method coupled with a biomarker panel

Materials and methods

FFPE DNA extraction and *TP53* sequencing on Cytosponge samples

The primer sequences specific to ALU115 repetitive elements used for quantification of FFPE extracted DNA by PCR were: (forward) 5'-CCTGAGGTCAGGAGTTCGAG-3' and (reverse) 5'-CCCGAGTAGCTGGGATTACA-3'.

Methylation analysis on the Cytosponge

The sequences of the primers and probes used were: MYOD1 forward primer: 5'-GAGCGCGCGTAGTTAGCG-3', MYOD1 reverse primer: 5'-TCCGACACGCCCTTTCC-3', MYOD1 probe: 5'-6FAM-CTCCAACACCCGACTACTATATCCGCGAAA-TAMRA-3', ACTB forward primer: 5'-TGGTGATGGAGGAGGTTTAGTAAGT-3', ACTB reverse primer: 5'-AACCAATAAAACCTACTCCTCCCTTAA-3', ACTB probe: 5'-6FAM-ACCACCACCCAACACACAATAACAAACACA-TAMRA-3' (from Eads, Danenberg et al. 2000), RUNX3 forward primer: 5'-GGCTTTTGGCGAGTAGTGGTC-3', RUNX3 reverse primer: 5'-ACGACCGACGCGAACG-3', RUNX3 protein: 5'-6FAM-CGTTTTGAGGTTCGGGTTTCGTCGTT-TAMRA-3' from the Meltzer laboratory.

Statistical analysis

Let Y be the response variable (Best2diagnosis; 1: HGD/IMC, 0: Negative) and $X=\{X_1,\ldots,X_p\}$ represent p covariates. Recall that we classified samples based on estimating the conditional probabilities p(Y|X) using a logistic regression model and then predicted the risk-profile of a patient as described in main text, where c_1 and c_2 are estimated from the data as described below. First, a biomarkers logistic regression model (with glandular atypia, p53 abnormality and Aurora kinase A IHC as covariates) was selected using a backward model selection approach[32]. More precisely, we started with a model consisting of all biomarkers then the least significant variable was dropped, so long as it was not significant at the 0.05 critical level. We proceeded by removing non-significant variables in a sequential manner (applying the same rule) until all remaining variables were statistically significant. The fitted conditional probabilities of being HGD/IMC were computed using the selected logistic regression model and then they were classified in three risk-groups, i.e. low, moderate, and high risk, as described above. The lower threshold c_1 was selected to maximise (minimise) the number of negative (HGD/IMC) samples classified in the low-risk group. The other threshold c_2 was selected to maximise (minimise) the number of HGD/IMC (negative) samples classified in the high-risk group. In an attempt to increase classification accuracy (as defined below), we fitted logistic regression models consisting of the selected biomarkers plus an additional covariate representing any of all available clinical variables. We have also fitted models including the main effects of clinical features and/or second and/or third order interactions between the clinical variables. The classification procedure described above was subsequently applied on each fitted model. The predictive performance of each model was assessed using the following simple approach. All models were ranked (i) in an ascending order according to the total number of misclassified patients (HGD/IMC in low-risk group and negative patients in high-risk group) and, subsequently, (ii) in a descending order based on the total number of correctly classified patients (negative patients falling into the low-risk group and HGD/IMC patients falling into the high-risk group). In this fashion, the higher the ranking of a model is, the better predictive performance is supposed to have. A model consisted of glandular atypia, p53 abnormality, Aurora kinase A IHC and the interaction of BMI (or waist:hip ratio), Barrett's oesophagus maximum length and age was selected. We are aware of the increased chance of overfitting, thus we emphasized on enrolling a validation cohort to assess the "out-of-sample" performance and get more trustworthy results. The predictive performance of the selected model on the validation cohort seems promising.

Estimated model coefficients, standard errors, z-statistics and p-values are provided below:

Coefficient	**Estimate**	**Std. Error**	**Z-value**	**P-value**
Intercept	-4.019	0.384	-10.471	<0.00000000000000002
Atypia	2.435	0.362	6.732	0.000000000017
p53_Status	1.802	0.325	5.544	0.00000003
AURKA	0.876	0.340	2.574	0.01005
AgexMxBMI	0.0000634	0.000022	2.885	0.00392

Tables

Sensitivity and specificity assessment of methylated regions to detect dysplasia on the Cytosponge.

In a pilot experiment consisting of 113 Cytosponge samples (15 controls, 54 Barrett's with no dysplasia, 20 Barrett's with LGD and 24 Barrett's with HGD), five methylated regions (p16, HPP1, RUNX3, ESR1 and MYOD1) were assessed to see which subset of methylated regions performed the best and had the best sensitivity and specificity to detect dysplasia on the Cytosponge. Together RUNX3 and MYOD1 gave the best area under the curve when comparing any dysplasia with no dysplasia and were therefore taken forward to evaluate further on the Cytosponge samples.

Methylation regions	**AUC**
ESR1	0.739
HPP1	0.754
MYOD1	0.771
P16	0.673
RUNX3	0.754
P16 HPP1	0.757
P16 RUNX3	0.727
P16 ESRI	0.741
P16 MYOD1	0.762
HPP1 RUNX3	0.770
HPP1 ESR1	0.755
HPP1 MYOD1	0.773
RUNX3 ESR1	0.754
RUNX3 MYOD1	0.786
ESR1 MYOD1	0.762
P16 HPP1 RUNX3	0.764
P16 HPP1 ESR1	0.753
P16 HPP1 MYOD1	0.770
P16 RUNX3 ESR1	0.754
P16 RUNX3 MYOD1	0.771
P16 ESR1 MYOD1	0.758
HPP1 RUNX3 ESR1	0.761
HPP1 RUNX3 MYOD1	0.776
HPP1 ESR1 MYOD1	0.763
RUNX3 ESR1 MYOD1	0.769
P16 HPP1 RUNX3 ESR1	0.759
P16 HPP1 RUNX3 MYOD1	0.773
P16 HPP1 ESR1 MYOD1	0.761
P16 RUNX3 ESR1 MYOD1	0.761
HPP1 RUNX3 ESR1 MYOD1	0.765
P16 HPP1 RUNX3 ESR1 MYOD1	0.763

Sites of patient recruitment

Recruitment site	Location	Principle investigator	Number of patients
Addenbrooke's Hospital	Cambridge	Rebecca Fitzgerald	195
University College London Hospital	London	Laurence Lovat	145
Royal Victoria Infirmary	Newcastle	Michael Griffin	118
Nottingham Queen's Medical Centre	Nottingham	Krish Ragunath	76
Queen Alexandra	Portsmouth	Pradeep Bhandari	35
South Tyneside NHS Foundation Trust	South Tyneside	Colin Rees	15
University Hospital of North Tees	Stockton-On-Tees	Mathew Rutter	12
University Hospital of North Durham	County Durham and Darlington	Anjan Dhar	12
Queen Elizabeth II Hospital	Welwyn Garden	Danielle Morris	8
St Mark's Hospital	London, Northwick Park	Adam Haycock	6
North Tyneside General Hospital	North Shields	Stephen Attwood	5
Hinchingbrooke Hospital	Huntingdon	Anita Gibbons	4
Bedford Hospital	Bedford	Jacqueline Harvey	2

IHC staining conditions and antibodies used.

Antigen	Protocol	Antigen retrieval	Antibody	Antibody dilution
p53	Protocol F	H1(30)	Novocastra™ Mouse Monoclonal Antibody p53 Protein (DO-7) Product Code: NCL-p53-DO7	1:50
c-Myc	MRC+E*	H2(20)	Epitomics c-MYC antibody, clone Y69, Rabbit monoclonal Cat #: 1472-1	1:50
Aurora kinase A	MRC+E	H2(30)	Millipore Anti-Aurora-A (C-term), clone EP1008Y, Rabbit Monoclonal Cat #: 04-1037	Lot specific dilution as the concentration is not normalised between lots

* For c-Myc staining, the primary antibody was incubated with 60 minutes

Demographics, clinical history and molecular biomarkers included in the model selection. Selected predictor and interaction variables are highlighted in bold.

	Model input	Options
Demographics		
	Age	
	Gender	M/F
	BMI	
	Waist:hip ratio	
	Family history of EAC	Y/N
Smoking history		
	Current smoker	Y/N
	Total years smoked	
	Time since stopped smoking (years)	
	Number of pack-years	
Alcohol history		
	Number of alcohol units per week	
Medication history		
	Ever taken NSAIDs	Y/N
	Current NSAID status	Currently taking, not taking, NA
	NSAID dose frequency	Never, on demand, at least daily, NA
	NSAID duration (years)	
	Currently on H pylori medication	Y/N, don't know
Symptoms		
	Chest pain	Never, sometimes, often/daily
	Burning chest	Never, sometimes, often/daily
	Acid taste in mouth	Never, sometimes, often/daily
	Number of years since heartburn started	Never, <10 yrs, 10-20 yrs, >20 yrs
Endoscopic findings		
	Hiatus hernia	Y/N
	Barrett's oesophagus maximum length (cm)	
	Oesophagitis	Y/N
	Presence of oesophageal nodules or ulcers	Y/N
	Barrett's oesophagus surveillance length (years)	
Biomarker panel on Cytosponge specimens		
	Glandular atypia	0/1
	p53 IHC	0/1
	TP53 mutation	0/1
	p53 abnormality	0/1
	cMYC IHC	0/1
	Aurora Kinase A IHC	0/1
	MYOD1 methylation (%)	
	RUNX3 methylation (%)	
	MYOD1+RUNX3 methylation (%)	

Diagnosis follow-up of patients in the high, moderate, and low risk groups diagnosed with low-grade dysplasia at baseline.

Risk group	Number of patients diagnosed with high grade dysplasia	Total number of patients	Percentage diagnosed with high grade dysplasia
High risk	7	16	43.75%
Moderate risk	7	29	24.10%
Low risk	1	5	20%

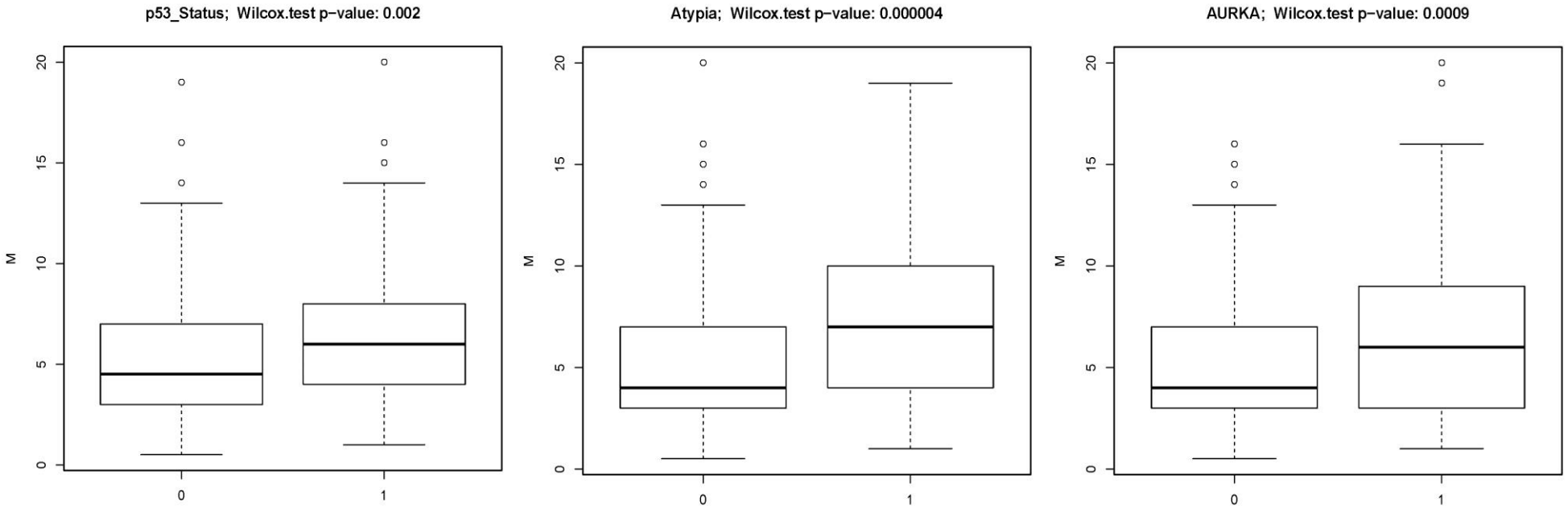
Marginal effects of individual predictors.

A marginal effect is a measure of the direct effect that a change in a particular biomarker, e.g. from negative to positive score, has on the predicted probability of having a high grade dysplasia, when the other predictors are kept fixed. All marginal effects were computed as in Greene[27] using the R function "logitmfx" from R-cran library "mfx".

Predictor	**Marginal effect**	**Range**
Atypia	40%	19-61%
p53 status	24%	10-38%
Aurora kinase A	9%	1-17%
log of interaction between age, maximal Barrett's length, and obesity	8%	2-14%

Risk stratification prediction tool

Predictor	Type	Vale	Probability (P)	Classification
p53 abnormality (combination of IHC and mutation)	Binary	a	=1/(EXP(4.019+a*(-1.802187)+b*(-2.434713)+c*(-0.8756866)+d*e*f*(-0.00006342102))+1)	Low risk P<0.0345 Moderate risk 0.0345<P<0.48 High risk P>0.48 In MS Excel: =IF(P<0.0345,"Low Risk",IF(P>0.48,"High-Risk","Moderate-Risk"))
Glandular atypia	Binary	b		
AurKA staining	Binary	c		
Age	Continuous	d		
BMI	Continuous	e		
Maximal Barrett's oesophagus length	Continuous	f		



Biomarker profiles associated with Barrett's segment length. Y–axis (M) refers to the maximum length of the Barrett's segment. X-axis 0 or 1 refers to p53 normality or abnormality respectively.