Development of Evidence Based Surveillance Intervals following Radiofrequency Ablation of Barrett's Esophagus

Cary C Cotton[1*], Rehan Haidry[2,5*], Aaron P Thrift[3,4], Laurence Lovat[2,5†], Nicholas J Shaheen[1†]

* Co-first authors, † co-last authors.

Author affiliations: 1) University of North Carolina at Chapel Hill, Department of Medicine, Division of Gastroenterology and Hepatology, Center for Esophageal Diseases and Swallowing, Chapel Hill, NC, USA. 2) University College Hospital, Department of Gastroenterology, Fitzrovia, London, UK. 3) Section of Epidemiology and Population Sciences, Department of Medicine, Baylor College of Medicine, Houston, TX, USA. 4) Dan L Duncan Comprehensive Cancer Center, Baylor College of Medicine, Houston, TX, USA. 5) Division of Surgery & Interventional Science, University College London, London, UK

Abbreviations: RFA, radiofrequency ablation, CEIM, complete eradication of intestinal metaplasia, BE, Barrett's esophagus, US RFA, United States Radiofrequency Ablation Registry, EAC, invasive esophageal adenocarcinoma, AIM, ablation of intestinal metaplasia, NDBE, non-dysplastic Barrett's esophagus, UK NHR, United Kingdom National Halo Registry, NRI, net reclassification improvement, IDI, integrated discrimination improvement, CM, centimeter, AIC, Akaike information criterion, EMR, endoscopic mucosal resection.

Corresponding Author: Nicholas J. Shaheen, MD, MPH
        Professor of Medicine and Epidemiology
        University of North Carolina School of Medicine
        CB#7080
        Chapel Hill, NC 27599-7080
        Phone: (919) 966-2513
        Fax: (919) 843-2508

Abstract

Background and Aims: Barrett's esophagus (BE) recurs in 25% or more of patients treated successfully with radiofrequency ablation (RFA), so surveillance endoscopy is recommended after complete eradication of intestinal metaplasia (CEIM). The frequency of surveillance is informed only by expert opinion. We aimed to model the incidence of neoplastic recurrence, validate the model in an independent cohort, and propose evidence-based surveillance intervals.

Methods: We collected data from the United States Radiofrequency Ablation Registry (US RFA, 2004–2013) and the United Kingdom National Halo Registry (UK NHR, 2007–2015) to build and validate models to predict the incidence of neoplasia recurrence following initially successful RFA. We developed 3 categories of risk and modeled intervals to yield 0.1% risk of recurrence with invasive adenocarcinoma. We fit Cox proportional hazards models assessing discrimination by C statistic and 95% confidence limits (CL).

Results: The incidence of neoplastic recurrence was associated with most severe histologic grade prior to CEIM, age, endoscopic mucosal resection, sex, and baseline BE segment length. In multivariate analysis, a model based solely on most severe pre-CEIM histology predictied neoplastic recurrence with a C statistic 0.892 (95% CL, 0.863–0.921) in the US RFA registry. This model also performed well when we used data from the UK NHR. Our model divided patients into 3 risk groups based on baseline histologic grade: non-dysplastic BE or indefinite-for-dysplasia, low-grade dysplasia, and high-grade dysplasia or intramucosal adenocarcinoma. For patients with low-grade dysplasia, we propose surveillance endoscopy at 1 and 3 years after CEIM; for patients with high-grade dysplasia or intramucosal adenocarcinoma we propose surveillance endoscopy at 0.25, 0.5, and 1 year after CEIM, then annually.

Conclusion: In analyses of data from the US RFA and UK NHR for BE, a much-attenuated schedule of surveillance endoscopy would provide protection from invasive adenocarcinoma. Adherence to the recommended surveillance intervals could decrease the number of endoscopies performed yet identify unresectable cancers at rates less than 1/1000 endoscopies.

Keywords: esophageal cancer, risk of progression, NDBE, LGD, HGD

Background

Radiofrequency ablation (RFA) is a safe and effective therapy for the treatment of Barrett's esophagus (BE).[1, 2] While the majority of patients undergoing RFA for dysplastic BE achieve durable complete eradication of intestinal metaplasia (CEIM), 25% or more may have recurrence of intestinal metaplasia.[3-12] Fortunately, most of these recurrences are non-dysplastic, and responsive to further treatment with RFA or other treatment modalities.[13] However, a small proportion of these recurrences are neoplastic, and up to 1% of patients have a recurrence with invasive esophageal adenocarcinoma (EAC).

Because of the risk of recurrence, following CEIM, endoscopic surveillance is performed to identify and treat recurrence of neoplasia and to prevent progression to invasive EAC.[5, 14] Surveillance practices after therapy vary widely and are informed by expert opinion alone.[15] The most commonly recommended surveillance algorithm is based on intervals utilized in the AIM Dysplasia Trial, a multicenter randomized controlled trial demonstrating the utility of RFA in the treatment of dysplastic BE.[16] While these intervals appear in clinical practice to lead to low rates of unresectable EAC,[3] they are likely too aggressive, given that they are identical to recommendations for patients who do not undergo RFA. If RFA lowers cancer risk, it is also reasonable to hypothesize that the need for surveillance endoscopy should also be lessened.

The United States Radiofrequency Ablation (US RFA) Registry is the largest existing study of patients undergoing RFA for BE[17] in the United States, and the United Kingdom National Halo Registry (UK NHR) is the largest existing study of such patients outside the United States.[18] These registries present a unique opportunity to assess patterns of recurrence to produce evidence-based surveillance intervals, and then validate them in an independent population. Such an approach should allow for risk stratification for dysplastic recurrence and then provide independent validation in a contemporaneous, geographically distinct population that such risk-stratification is robust and generalizable.[19]

Our aims were to 1) model the rate of neoplastic recurrence (low-grade dysplasia [LGD], high-grade dysplasia [HGD] or EAC) following RFA for BE using data from the US RFA Registry based on characteristics known at the time CEIM is established, 2) to validate this model in the UK NHR, 3) to simplify the model into categories of surveillance risk, and 4) to propose evidence-based surveillance intervals following RFA for BE.

Methods

Studies and Inclusion in the Surveillance Cohorts

The US RFA Registry and the UK NHR are the largest studies of patients with BE treated with RFA.[17, 18] Enrollment in the US RFA began in 2004 and surveillance data collected until 2013. Enrollment in the UK NHR began in 2007, with data collection ongoing. A data lock of the UK NHR in 2015 was used for this study, which allowed at least 18 months of follow-up for all included subjects. While the US RFA Registry is set in academic-affiliated and independent practices and includes patients with both non-dysplastic and dysplastic BE at their initial RFA, the UK NHR is set in academic centers and includes only patients with dysplastic BE at their initial RFA. Patients that achieved CEIM (defined as one post-treatment endoscopy showing no histological or endoscopic evidence of intestinal metaplasia or associated neoplasia) and entered endoscopic surveillance (defined as having at least one additional surveillance endoscopy with histologic sampling following CEIM), were included for this analysis. Patients that had any history of invasive esophageal adenocarcinoma (EAC) or esophagectomy prior to entering surveillance were excluded. Patients began to accrue person-time in surveillance at the endoscopy demonstrating CEIM, and continued until censoring for retreatment, censoring for non-dysplastic recurrence, or the development of a neoplastic recurrence.

Definition of the Primary and Secondary Outcomes

Neoplastic recurrence was the primary outcome and was defined as the first finding on histologic examination of LGD, HGD or EAC in the esophagus or cardia after CEIM confirmed in a single endoscopy. In secondary analyses, we also examined models of recurrence with HGD or higher, and intramucosal adenocarcinoma or higher.

Definition and Description of Clinical Characteristics to Predict Neoplastic Recurrence

Variables included in forward model selection were those which would be known to the clinician at the time of CEIM and with clinically or biologically plausible effects on the rate of neoplastic recurrence.

These included: age at first RFA treatment, sex, initial BE segment length in centimeters, presence of prior endoscopically resected nodular disease, performance of any endoscopic mucosal resection of nodular disease during endoscopic eradication treatment, and the patients' most severe histologic grade prior to CEIM. The number and type of treatments given, race and ethnicity, and previous Nissen fundoplication were also examined as potential predictors, but did not reach the *a priori* statistical significance thresholds for inclusion in the model (p<0.05). Baseline characteristics were described with mean and standard deviation for continuous variables and with number and percent for categorical variables.

Model Form and Procedure for Selection of Variables in the Predictive Models

We modeled multiple approaches to parameterization of continuous and categorical variables. For age at first treatment, we fit models of age as a continuous variable, age in two categories with a boundary at 65, age in three categories without boundaries at 50 and 75, age in deciles from 50 to 80, age in deciles from 55 to 85, and age in five year strata from 55 to 85. For baseline Barrett's segment length, we fit models of length as a continuous variable, length in two categories with a boundary of 4 centimeters (cm), length in two categories with a boundary of 2 cm, and length in three categories with boundaries at 4 and 8 cm. For most severe histology prior to entering surveillance, we fit models of all five categories of histologic grade, models combining indefinite for dysplasia with NDBE, and models combining indefinite for dysplasia with LGD. For continuous variables, numbers equal to the boundary went with the higher stratum and categories extended below and above the outer boundaries to the minimum and maximum observed values. The parameterization that had the lowest Akaike information criterion (AIC), a model fit statistic for which lower values compared to a nested model indicate better fit, was used for the variable in model selection.

The authors performed model selection by an *a priori* model selection process. We fit Cox proportional hazards models of the cumulative incidence of neoplastic recurrence after CEIM. Model building was performed by forward selection with a significance threshold for parameter entry of *p* less

than 0.05 and with the parameter with the lowest AIC entered first. We also performed a sensitivity analysis selecting the model among only those patients with dysplasia.

Assessment of Discrimination and Calibration and Validation of the Predictive Model

We calculated the C statistic, net reclassification improvement (NRI), and integrated discrimination improvement (IDI) comparing each candidate model in the selection process to a nested model without addition of the new variable.[20] The C statistic is a measure of how well the model separates patients with higher and lower risk. It ranges from 0 to 1, with 1 representing perfect discrimination by the model. We also calculated the NRI and IDI of the selected model to a referent.[21] Values of NRI and IDI near the null suggest that the model with an additional parameter does not outperform the referent model, and values less than zero denote a model that does worse than the referent model.[22] The authors assessed calibration by examination of stratified Kaplan-Meier plots overlying curves from the US RFA and UK NHR datasets. For validation of the model in the UK NHR, the subset of US RFA patients with low-grade dysplasia or more severe histologic grade were selected as the referent group.

Development of Proposed Surveillance Intervals from the Predictive Model

The selected predictive model was simplified into categories by collapsing levels of categorical variables that had similar estimates into a single level. Stratified by these categories, the changing incidence of neoplastic recurrence over time was estimated as a baseline hazard function with smoothing by a cubic spline with four degrees of freedom.[23] This allowed the authors to generate surveillance intervals sensitive to the changing instantaneous rate of recurrence observed in other studies.[24]

Generating logical surveillance intervals requires the identification of a tolerable degree of risk of neoplastic progression at each exam. In an appropriate surveillance program, the risk of progression at each endoscopic surveillance session would approach, but not exceed, the tolerable risk, and this risk should be approximately equal at each examination, i.e. it would be illogical to tolerate twice as much risk at one surveillance exam than the one either before or after it. Intervals were chosen that were expected to yield 0.1% incidence of EAC at each visit, which roughly equates to the expected proportion of serious

adverse events from esophagogastroduodenoscopy in a patient population such as that of the US RFA Registry.[25-27] For the purposes of this calculation, subjects that had EAC at the visit following neoplastic recurrence, even if the initial recurrence was of lower histologic grade, were considered recurrence with EAC. The tolerance threshold for recurrence with dysplasia or more severe histology was calculated as 0.1% divided by the proportion of visits with dysplasia or more severe histology. For patients at elevated risk of endoscopy complications, intervals were also modeled to yield 0.2% incidence of EAC. In this way, clinicians caring for patients with significant comorbidities, which in their estimate at least doubled the risk of complication of endoscopy, could alter surveillance intervals to account for this comorbidity. For proposed surveillance intervals, the times were rounded to the nearest 0.25-0.5 fraction of a year that approximated the observed shape of neoplastic recurrence.

Sensitivity analyses

The authors performed flexible estimation of the baseline hazard function to model the changing rate of recurrence over time. When modeling the shape of recurrence over time, it is possible that an arbitrary decision in parameterization of model of the baseline hazard function could produce an estimate of shape that is biased and not robust. The authors examined alternative spline parameterizations of the baseline hazard function with three rather than four degrees of freedom and with piecewise estimates across three and four equally spaced intervals. We overlaid these estimates for the highest category of surveillance risk, in which the rate of neoplastic recurrence was found to decrease over time.

In both the US RFA Registry and the UK NHR, some patients had non-dysplastic recurrence of BE, and underwent treatment of this lesion. This could bias our analysis, if censoring these patients removed patients from our pool who would otherwise be disproportionately at higher risk to meet our primary endpoint of neoplastic recurrence. In order to address this concern, we performed imputation analyses, imputing a hypothetical twofold and fourfold higher rate of neoplastic recurrence among patients who were censored for recurrence with NDBE. We performed 1,000 iterations of the model with imputation of neoplastic recurrence for patients who were censored for treatment of non-dysplastic recurrence. These

hypothetical recurrences were modeled with exponentially distributed times of recurrence with neoplasia. The survival time distribution was derived from the estimated rate of neoplastic recurrence in the year following censoring for recurrence with NDBE under the model resulting from the model selection process with the same spline estimate of the baseline hazard function as in the primary analysis. Statistical analysis was performed in SAS version 9.4.

Results

Definition and baseline characteristics of the surveillance cohorts

In the US RFA Registry cohort, there were an initial 5,521 subjects, from which 117 subjects were

excluded due to invasive adenocarcinoma before or during treatment and 29 subjects were excluded due

to esophagectomy before or during treatment. Of 5,444 included at baseline, 4,087 (75%) achieved CEIM

and 3,105 had at least one subsequent visit (Figure 1a). Baseline characteristics were similar between the

overall Registry participants and patients meeting criteria for inclusion in this analysis (Supplemental

Table 1). There were 7,984 surveillance visits included for analysis with a mean of 2.57 visits per patient

(standard deviation 1.88). The median time to first surveillance visit was 1.0 years (standard deviation

0.76) for non-dysplastic BE, 0.9 years (standard deviation 0.61) for BE with LGD, 0.6 years (standard

deviation 0.49) for BE with HGD, and 0.5 years (standard deviation 0.35) for intramucosal EAC.

In the UK NHR, 391 subjects met these criteria and were included, from which 12 were excluded for

missing values of predictor variables, and 373 remained (Figure 1b). Baseline characteristics were

generally similar between the studies, though, importantly, the UK NHR did not include patients with

NDBE or BE indeterminate for dysplasia. Therefore, participants in the UK Registry were generally

older, with longer segments of disease (Table 1).

Predicting the Incidence of Neoplastic Recurrence

Unadjusted Associations of Characteristics at Entry into Surveillance with Neoplastic Recurrence

In bivariate analyses, we observed statistically significant differences in the incidence of dysplastic

recurrence using 5 predictor variables: 1) the most severe histologic grade prior to CEIM, 2) age at the

first treatment, 3) performance of endoscopic mucosal resection (EMR) before or during treatment 4) sex,

and 5) baseline Barrett's segment length (Figure 2, Supplemental Figures 1-4). Patients who were older,

male, those with more severe baseline histology, those who underwent EMR before or during RFA, and

those with long segment length recurred at a higher rate than patients that were younger, female, with less severe histology, who did not undergo EMR, and those with shorter segment length, respectively.

Multivariable Model Building and Assessment of Predictive Performance

Most severe histologic grade before entering surveillance had the lowest AIC among first terms considered for addition to the model and was entered into the model first, then baseline age. No further variables met significance criteria for entry into the model. Model building selected the same predictor variables among the subgroup of patients with low-grade dysplasia or more severe histologic grade.

A model of the incidence of neoplastic recurrence (recurrence with LGD, HGD, IMC or invasive EAC) including most severe histologic grade prior to CEIM offered excellent discrimination (C statistic 0.892, 95% confidence limit 0.863 to 0.921) when applied to the histologically diverse US RFA cohort, and good discrimination (C statistic 0.746, 95% confidence limit 0.680 to 0.812) when limited to patients in the US RFA cohort with baseline dysplastic Barrett's esophagus (Supplemental Table 2).

The model that resulted from forward selection to predict any neoplastic recurrence included most severe histologic grade prior to CEIM and baseline age. This model also had very good discrimination (C statistic 0.837, 95% confidence limit 0.807 to 0.868) when applied to the full US RFA cohort, and acceptable discrimination (C statistic 0.685, 95% confidence limit 0.639 to 0.732) when limited to patients in the US RFA cohort with baseline dysplastic Barrett's esophagus, the comparator group for validation in the UK NHR.

Modeling Overall Recurrence and Recurrence with Higher Grade Neoplasia (HGD and EAC)

Similar to models for any neoplastic recurrence, models for higher-grade recurrence (HGD and EAC) that resulted from the model selection procedure included most severe histologic grade before entering surveillance and baseline age. Models for recurrence with high-grade dysplasia or for adenocarcinoma offered slightly better discrimination (C statistic 0.870 to 0.917) than the primary model for neoplastic recurrence (Supplemental Table 3). Too few events occurred to fit models of EAC. Models for overall

recurrence (including non-dysplastic recurrence) found a large effect of initial Barrett's segment length and had comparatively poor discrimination (C statistic 0.630, 95% confidence limit 0.609 to 0.650).

Validation in the United Kingdom National Halo Registry

The model with histologic grade alone had good discrimination in the UK NHR (C statistic 0.728, 95% confidence limit 0.584 to 0.871), performing in an almost identical fashion to that of the subset of patients with baseline dysplastic BE in the US RFA Registry (C statistic 0.746, 95% confidence limit 0.680 to 0.812). Models that included age in addition to most severe histologic grade prior to surveillance did poorly in external validation in the UK NHR (C statistic 0.581, 95% confidence limit 0.503 to 0.659). The addition of age to histologic grade in the UK NHR produce a negative net reclassification index and an integrated discrimination improvement of near null.

Development of Three Surveillance Risk Groups

The model selection process produced a model with most severe histologic grade prior to CEIM and age as a continuous variable. Though age was a statistically significant model parameter, including it in the models 1) decreased overall measures of model discrimination, 2) did not perform well in the validation dataset, and 3) failed to move subjects between three simple categories of histologic grade. Age was therefore not used for developing categories of surveillance risk. In the chosen model with histologic grade alone, HGD and intramucosal adenocarcinoma widely overlapped in estimated risk of neoplastic recurrence. Similarly, NDBE and indeterminate for dysplasia were also modeled to have the same surveillance risk. For the aforementioned reasons, the authors chose three groups classified by their most severe histology prior to CEIM: 1) NDBE or indeterminate for dysplasia, 2) LGD, and 3) HGD or IMC.

Modeling Surveillance Intervals in the United States Radiofrequency Ablation Registry

The annual rate of recurrence with neoplasia (LGD, HGD or EAC) was 0.19% (95% confidence limit 0.09 to 0.40%) in risk group one (patients with pre-CEIM NDBE/indeterminate for dysplasia), 1.98% (95% confidence limit 1.34 to 2.93%) in risk group two (patients with pre-CEIM LGD), and 5.93% (95% confidence limit 4.77 to 7.36%) in risk group three (patients with pre-CEIM HGD/IMC). In the higher

risk groups, neoplastic recurrence occurred at a higher rate in the first year, but at a constant estimated rate thereafter (Supplemental Figure 5). Among 114 initial cases of neoplastic recurrence, 2 (1.8%) held EAC, and an additional 2 (1.8%) had EAC within six months. We chose 2.9% as the rate of neoplastic recurrence per visit to yield an estimated rate of invasive adenocarcinoma of 0.1%. This level of risk tolerance was chosen so that the risk of complications from surveillance endoscopy (approximately 1/1000 in this patient population) would roughly approximate the risk of invasive carcinoma discovered at the exam. In a secondary analysis for subjects at higher risk of endoscopic complications, we chose a 5.7% rate of neoplastic recurrence to yield 0.2% risk of invasive cancer. This analysis allowed us to estimate surveillance intervals for patients at higher risk of endoscopic complications (Supplemental table 4). Using our model, for each of our 3 risk categories, we estimated surveillance intervals predicted to yield these rates of neoplastic recurrence. As would be expected, the higher the risk tolerance, the longer the period between endoscopic surveillance intervals.

The large majority of recurrences with dysplasia in the US RFA Registry were of histologic grade amenable to endoscopic retreatment (Figure 3). The newly proposed surveillance intervals are presented in Table 2, with recommendations for patients at increased risk of endoscopic complications in supplemental table 4. For patients with pre-CEIM NDBE or BE indefinite for dysplasia, we defer proposal of a specific surveillance interval because endoscopic eradication therapy is not recommended for this group. However, among patients who were treated for NDBE in this cohort, our data suggest the yield of surveillance endoscopy is very low compared to patients treated for dysplasia or intramucosal adenocarcinoma. In fact, at 7 years post-ablation, such patients had not crossed the risk threshold for surveillance endoscopy in our model, suggesting that surveillance endoscopy prior to that period would be extremely low yield. For patients with pre-CEIM LGD, we propose surveillance endoscopy at 1 and 3 years following the establishment of CEIM. For patients with pre-CEIM HGD or IMC, we propose surveillance endoscopy at 0.25, 0.5, 1, 2, 3, 4, and 5 years following the establishment of CEIM. Recommendations beyond five years would require extrapolation beyond the present data.

For patients with dysplastic BE, proposed intervals result in a marked reduction of the number of surveillance endoscopies when compared to current practice and societal guidelines. Current guidelines[2] adopt the surveillance protocol of the AIM Dysplasia trial,[16] which utilized surveillance for BE with LGD every 6 months in year 1, and annually, and for BE HGD every three months in year 1, every 6 months in year 2 then annually. Using these intervals for comparison, over five years, patients with baseline LGD would have 2 rather than 6 exams over 5 years, a 66% reduction, and patients with baseline HGD or IMC would have 7 rather than 9 exams, a 22% reduction. The practical difference in endoscopic surveillance visits over the first five years in each registry is presented in Table 3. Adherence to these recommended attenuated surveillance intervals would result in a 38% reduction in the numbers of surveillance endoscopy over 5 years in patients with dysplasia in the US RFA cohort, if endoscopists were following the current guidelines.

Proposed Intervals in the United Kingdom National Halo Registry

In the proposed risk groups, the incidence of neoplastic recurrence over time was similar between the UK NHR and the US RFA Registry (Figure 4). Applying the proposed surveillance intervals to outcomes data from the UK NHR demonstrated that the mean estimated yield of dysplastic recurrence for each endoscopic examination was similar to that in the US RFA Registry (3.7% for high risk and 4.7% for low risk in the UK NHR vs. 2.9% per exam in US RFA Registry). Several individual exams appeared to have much higher yield in the UK NHR than in the US data. However, given the relatively small numbers of overall neoplastic recurrences in UK NHR, the individual exam data are markedly impacted by as few as 1-2 recurrences (Supplemental Table 5). As with the US RFA data, most recurrences in the UK NHR were of histologic grade that is amenable to endoscopic treatment.

Sensitivity analyses

Alternative parameterizations of the baseline hazard generally reproduced the higher rate of recurrence of neoplasia early after CEIM among the highest risk group (Supplemental Figure 6). The close overlap in estimates suggests the decrease in rate of neoplastic recurrence after year one in the high-

risk group is robust to modeling assumptions. Sensitivity analysis found that imputed rates of neoplastic

recurrence after subjects were censored for recurrence without dysplasia did not alter our findings without

extreme assumptions. For example, even if the assumed rate of recurrence with neoplasia were fourfold

higher among censored patients compared to subjects not censored, the overall rate of neoplastic

recurrence was not significantly altered (Supplemental Figure 7).

Discussion

Our analysis used the two largest prospective studies of patients treated with RFA for BE to develop an evidence-based recommendation for surveillance intervals after CEIM. We fit models of the incidence of neoplastic recurrence, defined as recurrence with LGD, HGD, or EAC, in the US RFA Registry, then externally validated the models in the UK NHR. We developed three categories of surveillance risk based on the best performing model, and then estimated the yield for neoplastic recurrence in these three risk groups over time. For a given tolerance for recurrence of neoplasia, time intervals were estimated to produce a constant yield of neoplastic recurrence at each visit. Our findings suggest the frequency of surveillance endoscopies after CEIM should differ broadly from what is currently recommended and rely only on most severe histologic grade before CEIM is achieved.

Our model selection process produced a model that included most severe pre-CEIM histologic grade and age, which was compared to a model with most severe pre-CEIM histologic grade alone. The model including age was statistically significant, but performed poorly in discrimination statistics relative to the performance of the histologic grade only model, was not well validated in the UK NHR, and was of limited utility given age did not move subjects across bounds of similar histologic grade. Increasing age, though a risk factor for neoplastic recurrence, is also a strong indicator of the risk of adverse effects of surveillance endoscopy. As such, and with an eye for reducing complexity of surveillance recommendations, we selected the histologic grade only model for timing of surveillance visits.

We then estimated the changing rate of neoplastic recurrence over time using flexible time-to-event models to facilitate the ideal timing of surveillance visits. The shape of recurrence over time was not sensitive to examined modelling assumptions, and tended to match the shape in a prior randomized controlled trial with the rate of events highest in the year after CEIM, then slowing down to a lower, constant rate.[24] NDBE and indeterminate for dysplasia had an essentially constant, low rate of recurrence with neoplasia after CEIM in the US RFA Registry; such patients were not available for analysis in the

prior trial or the UK NHR. We selected surveillance intervals for a tolerance for recurrence of dysplasia of 2.9% per visit, a rate that is consistent with a one per thousand incidence of invasive adenocarcinoma. This risk tolerance yielded surveillance endoscopies at one and three years after CEIM for patients with baseline LGD and endoscopies at three months, six months, one year, two years, three years, four years, and five years for patients with baseline HGD or IMC. These visits produced similar mean yield for recurrence of dysplasia when externally validated in the UK NHR.

Societal recommendations do not endorse the performance of endoscopic eradication therapy in NDBE; the models of post-ablation NDBE and the proposed surveillance intervals above are provided because data from the US RFA Registry show that a large proportion of patients undergoing RFA in the US received it for non-dysplastic disease or indefinite for dysplasia. Putting aside the wisdom of endoscopic therapy in this patient population, they also require informed endoscopic surveillance. Based on our analysis, such patients can be surveyed on a long interval of at least 7 years, while still maintaining acceptably low levels of risk of neoplastic progression. Though it would require untenable extrapolation of a constant rate beyond the seven years of surveillance follow-up in these patients, the rate of neoplastic recurrence among the lowest risk group was so small that a single surveillance endoscopy at 36 years after CEIM was estimated to produce the same yield of recurrence with neoplasia as the first recommended surveillance endoscopy for the middle and high-risk groups.

These findings should be taken within the context of prior literature describing the incidence of neoplastic recurrence after CEIM. Because recurrence of NDBE is usually scant in amount and easily treated with low rates of progression to unresectable EAC, the clinical relevance of recurrence rates of non-dysplastic intestinal metaplasia is uncertain. For this reason, we performed this analysis with an outcome of neoplastic recurrence, an endpoint likely to be regarded by endoscopists as being clinically important. A number of studies have reported the incidence of recurrence of intestinal metaplasia after CEIM, but have not generally reported the rate of recurrence with neoplasia.[5, 8] In the AIM Dysplasia

Trial the rate of neoplastic recurrence was 3.3 and 7.3 per 100 person-years among patients with LGD and HGD, respectively.[24] This is similar to the rates for LGD of 2.0 and for HGD of 5.5 in the present study.

This work has limitations. Our models assume subjects are censored at recurrence without dysplasia, because such patients would be logical candidates for endoscopic retreatment. Thus, our conclusions and suggested surveillance intervals cannot be applied to patients after they are retreated and regain CEIM, since such patients may represent a group at increased risk of a second recurrence. Also, because patients with NDBE were not included in the UK NHR, our validation exercise could only be performed on data from patients with baseline LGD, HGD, or intramucosal IMC, and these are the only patients in whom we have recommended new surveillance intervals. Additionally, though the shape of recurrence over time in this study matches a prior study, its decreasing rate of recurrence with time is not statistically significant in this study. It is also important to note that we performed this analysis using an indirect outcome, neoplastic recurrence, to make inferences about the clinically important outcome, interval development of invasive cancer during surveillance. Even in these largest two registry populations, initial recurrence with invasive cancer is too uncommon for robust modeling. Finally, the smaller size and differences in baseline characteristics in the UK NHR may explain the observed differences in surveillance yield for individual endoscopies from the US RFA Registry.

There were only two cases of initial recurrence with invasive cancer and two additional cases that had invasive cancer progression at follow-up visits within six months after recurring with lower-grade disease. These latter two cases could represent true progression of disease between 3 and 6 months, but in the interest of erring toward a more conservative surveillance regimen, they were counted as invasive cancer recurrences due to the possibility they were simply missed on the initial visit with lower-grade recurrence.

This study also has important strengths. Ours are the largest two existing studies of patients with BE treated with RFA and the largest two surveillance populations. To our knowledge, this is the only study that has developed surveillance intervals after RFA for BE using an evidence-based process, with a stated

degree of expected recurrence of disease, as well as an *a priori* suggested tolerable risk of progression. The US RFA Registry offers a unique opportunity to study surveillance intervals because the study did not mandate any one surveillance schedule, and functions as a sort of "natural experiment." This work allows current opinion-based recommendations to instead be governed by data, and in this respect, represents a marked improvement over guidance currently found in our societal guidelines. To generate proposed surveillance intervals for patients after the attainment of CEIM, we selected a tolerance for neoplastic recurrence consistent with a risk of invasive adenocarcinoma of approximately 1/1000. With this degree of risk tolerance, the risk of invasive adenocarcinoma comes close to the risk of the endoscopic procedure itself, making it illogical to survey more aggressively.

Of note, this approach also allows clinicians and patients to set surveillance practices based on their degree of risk tolerance. For instance, if patients were comfortable with higher risks of interval progression to cancer, the surveillance intervals proposed for patients at increased risk of endoscopic complications could be recommended for them as well. Analyses of this type make explicit the implied trade-off in decisions about endoscopic surveillance intervals – the longer the interval between surveillance exams, the less risk of adverse events, costs, and inconvenience to the patient to mount the program, but the higher the risk of recurrent, and potentially unresectable, disease. With this degree of risk tolerance, and by better spacing of the examinations, we were able to markedly decrease the number of necessary examinations while maintaining a low and acceptable risk of neoplastic recurrence.

In summary, the recurrence of neoplasia after RFA for BE is best predicted by the most severe histologic grade identified before CEIM. Patients with NDBE have similar rates of recurrence of dysplasia to patients with indeterminate for dysplasia and patients with HGD have similar rates of recurrence of dysplasia to patients with IMC, yielding three groups of risk based on the patient's most severe histology prior to CEIM: 1) NDBE/Indefinite for dysplasia, 2) LGD and 3) HGD/IMC. Using these risk prediction models and assuming a reasonable and conservative tolerance for invasive carcinoma, we suggest surveillance intervals based on the recurrence patterns seen in these two large

studies. Patients with LGD and lesser histologic grade could undergo surveillance much less frequently than guidelines currently recommend, and patients with HGD or worse histology should also undergo fewer visits over the five years following CEIM, but with a more gradual taper, to improve detection of neoplastic recurrence. Implementation of our proposed surveillance intervals may reduce the risk of progression to cancer, decrease harms and costs associated with over-surveillance, and among endoscopists following current guidelines, substantially reduce the overall number of surveillance endoscopies in cohorts undergoing RFA for BE. This approach is also the first to provide a direct basis in evidence for surveillance practices in BE.

References

1.   Spechler SJ, Sharma P, et al. American Gastroenterological Association medical position statement on the management of Barrett's esophagus. Gastroenterology 2011;140:1084-91.
2.   Shaheen NJ, Falk GW, Iyer PG, et al. ACG Clinical Guideline: Diagnosis and Management of Barrett's Esophagus. Am J Gastroenterol 2016;111:30-50.
3.   Cotton CC, Wolf WA, Pasricha S, et al. Recurrent intestinal metaplasia after radiofrequency ablation for Barrett's esophagus: endoscopic findings and anatomic location. Gastrointest Endosc 2015;81:1362-9.
4.   Johnson CS, Louie BE, Wille A, et al. The Durability of Endoscopic Therapy for Treatment of Barrett's Metaplasia, Dysplasia, and Mucosal Cancer After Nissen Fundoplication. J Gastrointest Surg 2015;19:799-805.
5.   Pasricha S, Bulsiewicz WJ, Hathorn KE, et al. Durability and predictors of successful radiofrequency ablation for Barrett's esophagus. Clin Gastroenterol Hepatol 2014;12:1840-7 e1.
6.   Perry KA, Walker JP, Salazar M, et al. Endoscopic management of high-grade dysplasia and intramucosal carcinoma: experience in a large academic medical center. Surg Endosc 2014;28:777-82.
7.   Lee JK, Cameron RG, Binmoeller KF, et al. Recurrence of subsquamous dysplasia and carcinoma after successful endoscopic and radiofrequency ablation therapy for dysplastic Barrett's esophagus. Endoscopy 2013;45:571-4.
8.   Phoa KN, Pouw RE, van Vilsteren FG, et al. Remission of Barrett's esophagus with early neoplasia 5 years after radiofrequency ablation with endoscopic resection: a Netherlands cohort study. Gastroenterology 2013;145:96-104.
9.   Gupta M, Iyer PG, Lutzke L, et al. Recurrence of esophageal intestinal metaplasia after endoscopic mucosal resection and radiofrequency ablation of Barrett's esophagus: results from a US Multicenter Consortium. Gastroenterology 2013;145:79-86 e1.
10.   Guarner-Argente C, Buoncristiano T, Furth EE, et al. Long-term outcomes of patients with Barrett's esophagus and high-grade dysplasia or early cancer treated with endoluminal therapies with intention to complete eradication. Gastrointest Endosc 2013;77:190-9.
11.   Dulai PS, Pohl H, Levenick JM, et al. Radiofrequency ablation for long- and ultralong-segment Barrett's esophagus: a comparative long-term follow-up study. Gastrointest Endosc 2013;77:534-41.
12.   Orman ES, Kim HP, Bulsiewicz WJ, et al. Intestinal metaplasia recurs infrequently in patients successfully treated for Barrett's esophagus with radiofrequency ablation. Am J Gastroenterol 2013;108:187-95.
13.   Guthikonda A, Cotton CC, Madanick RD, et al. Clinical Outcomes Following Recurrence of Intestinal Metaplasia After Successful Treatment of Barrett's Esophagus With Radiofrequency Ablation. Am J Gastroenterol 2017;112:87-94.
14.   Haidry RJ, Dunn JM, Butt MA, et al. Radiofrequency ablation and endoscopic mucosal resection for dysplastic barrett's esophagus and early esophageal adenocarcinoma: outcomes of the UK National Halo RFA Registry. Gastroenterology 2013;145:87-95.
15.   Bedi AO, Kwon RS, Rubenstein JH, et al. A survey of expert follow-up practices after successful endoscopic eradication therapy for Barrett's esophagus with high-grade dysplasia and intramucosal adenocarcinoma. Gastrointest Endosc 2013;78:696-701.
16.   Shaheen NJ, Sharma P, Overholt BF, et al. Radiofrequency Ablation in Barrett's Esophagus with Dysplasia. The New England Journal of Medicine 2009;360:2277-88.
17.   Shaheen NJ, Kim HP, Bulsiewicz WJ, et al. Prior fundoplication does not improve safety or efficacy outcomes of radiofrequency ablation: results from the U.S. RFA Registry. J Gastrointest Surg 2013;17:21-8.

18. Haidry RJ, Butt MA, Dunn JM, et al. Improvement over time in outcomes for patients undergoing endoscopic therapy for Barrett's oesophagus-related neoplasia: 6-year experience from the first 500 patients treated in the UK patient registry. Gut 2015;64:1192-9.
19. Harrell FE, Jr., Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med 1996;15:361-87.
20. Kremers W. Concordance for survival time data including time-dependent covariates accounting for ties in predictor and time, Mayo Clinic, Technical Report Series, 2007.
21. Pencina MJ, Larson MG, D'Agostino RB. Choice of time scale and its effect on significance of predictors in longitudinal studies. Stat Med 2007;26:1343-59.
22. 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;27:157-72.
23. Royston P, Parmar MK. Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. Stat Med 2002;21:2175-97.
24. **Cary C. Cotton and W. Asher Wolf,** Bergein F. Overholt, Nan Li, Charles J. Lightdale, Herbert C. Wolfsen, Sarina Pasricha, Kenneth K. Wang, Nicholas J. Shaheen for the AIM Dysplasia Trial Group. Late recurrence of BE after complete eradication of IM is rare: Final report from ablation in intestinal metaplasia containing dysplasia trial.. Gastroenterology 2017;153:681-688.
25. Sieg A, Hachmoeller-Eisenbach U, Eisenbach T. Prospective evaluation of complications in outpatient GI endoscopy: a survey among German gastroenterologists. Gastrointest Endosc 2001;53:620-7.
26. Quine MA, Bell GD, McCloy RF, et al. Prospective audit of upper gastrointestinal endoscopy in two regions of England: safety, staffing, and sedation methods. Gut 1995;36:462-7.
27. Silvis SE, Nebel O, Rogers G, et al. Endoscopic complications. Results of the 1974 American Society for Gastrointestinal Endoscopy Survey. JAMA 1976;235:928-30.

Author names in bold designate shared co-first authorship.

Tables

Table 1. Baseline Demographic, Endoscopic, and Histologic Characteristics of Subjects in the United Kingdom National Halo Registry and United States Radiofrequency Ablation Cohort.

	United Kingdom National Halo Registry	United States Radiofrequency Ablation Registry with Dysplasia	Wilcoxon/ Fisher p compared to UK NHR	United States Radiofrequency Ablation Registry	Wilcoxon/ Fisher p compared to UK NHR
N	373	1,425		3,105	
Baseline age in years - mean (SD)	67.0 (9.7)	64.5 (10.3)	< 0.0001	61.5 (10.9)	< 0.0001
Male sex - N (percent)	303 (81.2)	1,174 (82.4)	0.60	2,258 (72.7)	0.0004
Non-dysplastic Barrett's esophagus - N (percent)	0 (0.0)	0 (0.0)	< 0.0001	1,441 (46.4)	< 0.0001
Indeterminate for dysplasia - N (percent)	0 (0.0)	0 (0.0)		242 (7.8)	
Low-grade dysplasia - N (percent)	83 (22.3)	658 (46.2)		643 (20.7)	
High-grade dysplasia - N (percent)	265 (71.1)	630 (44.2)		628 (20.2)	
Intramucosal adenocarcinoma - N (percent)	25 (6.7)	137 (9.6)		125 (4.0)	
Baseline segment length in CM - mean (SD)	5.0 (3.5)	4.4 (3.2)	0.0003	3.7 (3.0)	< 0.0001
Endoscopic resection before treatment - N (percent)	204 (54.7)	284 (19.9)	< 0.0001	289 (9.3)	< 0.0001
Endoscopic resection during treatment - N (percent)	15 (4.0)	92 (6.5)	0.085	95 (3.1)	0.35

SD, standard deviation; N, number; CM, centimeters.

Table 2. Recommended Time after Complete Eradication Intestinal Metaplasia of Surveillance Visits to Yield 2.9% Neoplastic Recurrence per Visit or 0.1% Invasive Adenocarcinoma for Patients at Average Risk of Endoscopic Complications.

Risk Category:	Visit 1	Visit 2	Visit 3	Visit 4	Visit 5	Visit 6	Visit 7	Visit 8
Low-grade dysplasia	1 years	3 years	> 5 years*	*	*	*	*	*
High-grade dysplasia or adenocarcinoma in situ	3 months	6 months	1 year	2 years	3 years	4 years	5 years	> 5 Years*

*Surveillance times were estimated to a limit of five years for the higher two risk categories and seven years for the lower to avoid extrapolation beyond the data.

Table 3. Comparing the Number of Surveillance Endoscopies that would be Performed in the United States Radiofrequency Ablation Registry and the United Kingdom National Halo Registry under Current and Proposed Surveillance Regimens.

Surveillance risk group:	Patients in surveillance	Endoscopies under current recommendations	Endoscopies under proposed recommendations	Actual reduction in stratum	Total reduction for population
United States Radiofrequency Ablation Registry					
2: Low-grade dysplasia	658	3,948	1,316	67%	38%
3: High-grade dysplasia or intramucosal adenocarcinoma	767	6,903	5,369	22%	
United Kingdom National Halo Registry					
2: Low-grade dysplasia	83	498	166	67%	29%
3: High-grade dysplasia or intramucosal adenocarcinoma	290	2,610	2,030	22%	

Figure Legend

Figure #	Title
1a and 1b	A) Inclusion of 3,105 Subjects in the Surveillance Cohort at Risk from 5,521 United States Radiofrequency Ablation Registry Subjects. B) Inclusion of 373 Subjects in the Surveillance Cohort at Risk from 577 United Kingdom National HALO Registry Subjects.
2	Kaplan-Meier Estimates of the Proportion of Subjects in the US RFA Registry without Recurrence of Neoplasia in Five Years after Complete Eradication of Intestinal Metaplasia by Most Severe Prior Histologic Grade.
3	The Rate of First Recurrence of Neoplasia with Low-grade Dysplasia, High-grade Dysplasia, Intramucosal Adenocarcinoma, and Invasive Adenocarcinoma among Simplified Categories of Surveillance Risk.
4	Kaplan-Meier Estimates of the Proportion of Subjects in the US RFA Registry and the UK National Halo Registry without Recurrence of Neoplasia in Five Years after Complete Eradication of Intestinal Metaplasia by Proposed Surveillance Risk Groups.