Clinical consequences of nonadherence to Barrett's esophagus surveillance recommendations: a multicenter prospective cohort study

Roumans, C.A.M.; Bogt, R.D. van der; Nieboer, D.; Steyerberg, E.W.; Rizopoulos, D.; Lansdorp-Vogelaar, I.; ... ; ProBar Study Grp

Diseases of the Esophagus (2022), **00**, 1–10
https://doi.org/10.1093/dote/doac113

DISEASES OF THE ESOPHAGUS

ISDE The International Society for Diseases of the Esophagus

Original Article

Clinical consequences of nonadherence to Barrett's esophagus surveillance recommendations: a Multicenter prospective cohort study

Carlijn A.M. Roumans,[1,2] Ruben D. van der Bogt,[1] Daan Nieboer,[2] Ewout W. Steyerberg,[2,3] Dimitris Rizopoulos,[4] Iris Lansdorp-Vogelaar,[2] Katharina Biermann,[5] Marco J. Bruno,[1] Manon C.W. Spaander,[1] On behalf of the ProBar Study Group*

[1] *Department of Gastroenterology and Hepatology, Erasmus MC, University Medical Center Rotterdam, Rotterdam, The Netherlands,* [2] *Department of Public Health, Erasmus MC, University Medical Center Rotterdam, Rotterdam, The Netherlands,* [3] *Department of Biomedical Data Sciences, Leiden University Medical Center, Leiden, The Netherlands,* [4] *Department of Biostatistics, Erasmus MC, University Medical Center Rotterdam, Rotterdam, The Netherlands, and* [5] *Department of Pathology, Erasmus MC, University Medical Center Rotterdam, Rotterdam, The Netherlands*

SUMMARY. **Half of Barrett's esophagus (BE) surveillance endoscopies do not adhere to guideline recommendations. In this multicenter prospective cohort study, we assessed the clinical consequences of nonadherence to recommended surveillance intervals and biopsy protocol. Data from BE surveillance patients were collected from endoscopy and pathology reports; questionnaires were distributed among endoscopists. We estimated the association between (non)adherence and (i) endoscopic curability of esophageal adenocarcinoma (EAC), (ii) mortality, and (iii) misclassification of histological diagnosis according to a multistate hidden Markov model. Potential explanatory parameters (patient, facility, endoscopist variables) for nonadherence, related to clinical impact, were analyzed. In 726 BE patients, 3802 endoscopies were performed by 167 endoscopists. Adherence to surveillance interval was 16% for non-dysplastic (ND)BE, 55% for low-grade dysplasia (LGD), and 54% of endoscopies followed the Seattle protocol. There was no evidence to support the following statements: longer surveillance intervals or fewer biopsies than recommended affect endoscopic curability of EAC or cause-specific mortality ($P > 0.20$); insufficient biopsies affect the probability of NDBE (OR 1.0) or LGD (OR 2.3) being misclassified as high-grade dysplasia/EAC ($P > 0.05$). Better adherence was associated with older patients (OR 1.1), BE segments ≤ 2 cm (OR 8.3), visible abnormalities (OR 1.8, all $P \leq 0.05$), endoscopists with a subspecialty (OR 3.2), and endoscopists who deemed histological diagnosis an adequate marker (OR 2.0). Clinical consequences of nonadherence to guidelines appeared to be limited with respect to endoscopic curability of EAC and mortality. This indicates that BE surveillance recommendations should be optimized to minimize the burden of endoscopies.**

KEY WORDS: **Barrett's esophagus, surveillance, adherence, guideline.**

INTRODUCTION

In the last decades, the incidence of esophageal adenocarcinoma (EAC) has been rising and mortality rates because of advanced EAC are high. The recommended therapy for advanced EAC is neoadjuvant chemoradiotherapy followed by surgery, which is invasive and carries a considerable complication risk.[1] The only known precursor lesion for EAC is Barrett's esophagus (BE). Timely detection of neoplasia provides the opportunity to cure patients from EAC using a more favorable option, such as endoscopic eradication therapy. Therefore, BE patients undergo surveillance.

Surveillance is performed periodically by upper endoscopy. During these endoscopies, a visual appraisal of the esophageal mucosa is performed and routine biopsies are taken. If the histological diagnosis reveals any dysplasia, the interval until the subsequent endoscopy is shortened, or the Barrett's segment is eradicated by endoscopic therapy. In guidelines, surveillance recommendations have been formulated in three areas in particular: surveillance intervals, biopsy protocol, and landmark identification.[2,3]

Address correspondence to Manon C.W. Spaander, MD, PhD, Professor, Department of Gastroenterology and Hepatology (room Na-612), Erasmus MC, University Medical Center, Doctor Molewaterplein 40, 3015 CD Rotterdam, The Netherlands.
Email: v.spaander@erasmusmc.nl
*J. Alderliesten, G. van Ingen, F.T.M. Peters, A. Karrenbeld, F. ter Borg, M. Louwen, P.C.J. ter Borg, M.A. den Bakker, A. Alkhalaf, F.C.P. Moll, J.J. Kolkman, J. van Baarlen, J.P.W. Burger, J.W.R. Meijer, A.J.P. van Tilburg, H. van der Valk, M.J.L. Romberg-Camps, W. Vos, R.J.F. Felt, T.J. Tang.

Remarkably, in only half of surveillance endoscopies, endoscopists adhere to these guideline recommendations.[4,5] Little is known about the clinical consequences of nonadherence. The detection rate of dysplasia is known to reduce if fewer biopsies are taken than the recommended Seattle protocol.[4] However, the effect of nonadherence on endoscopic curability of EAC, mortality, and misclassification risk of histological diagnosis is unknown. Also, it is unclear why this adherence is so low. In previous studies, data were restricted by a lack of endoscopist variables (e.g. opinion about the length of surveillance interval) or the use of self-report.[6]

In this study, we aimed to estimate the clinical consequences of using longer surveillance intervals than recommended by BE guidelines and sampling less biopsies than the (Seattle) biopsy protocol dictates.[2,3] Also, we aimed to identify explanatory parameters to explain the gap between policy and practice and its potential clinical impact, by combining both endoscopist questionnaires and endoscopy and pathology reports.

PATIENTS AND METHODS

Study design

We performed a large multicenter prospective cohort study that has been described previously.[7] In summary, consecutive BE patients from 15 Dutch hospitals (3 university and 12 general hospitals) were included between September 2003 and December 2004; data were collected until November 2018. At index endoscopy, demographic information was collected. Before every follow-up (FU) endoscopy patients completed a questionnaire, concerning weight, length, symptoms of gastroesophageal reflux disease (GERD), use of medication, smoking, and alcohol use. During endoscopies, landmarks and the presence of visible abnormalities or esophagitis were identified. Instructions for taking biopsies were according to the Seattle protocol.[2] To determine the presence of intestinal metaplasia and the histological diagnosis, hematoxylin eosin slides were examined consecutively by a local pathologist and an expert pathologist. If discordant, a second expert pathologist reviewed the slides. In case of lasting disagreement, slides were examined by an extra pathologist until consensus was reached. Instructions for surveillance were according to the guideline of the American College of Gastroenterology (ACG).[2,8,9] The endpoint of the study was detection of high-grade dysplasia (HGD)/EAC. Information about the stage of EAC and treatment was collected.

The municipal registry was consulted to check mortality status of patients who had already dropped out of the study and, if applicable, time of death. If patients had passed away, the cause of death was searched in the electronic patient file of centers included where patients were having surveillance; otherwise their general practitioner was contacted.

Besides, all endoscopists who had ever performed an endoscopy in these patients were sent a postal questionnaire, covering demographic characteristics of the endoscopist, the organizational structure of surveillance at the endoscopy center, knowledge of the guideline, neoplastic progression risk estimations, and their opinion of Barrett's surveillance (Appendix 1). In an accompanying letter, endoscopists were informed that the questionnaire was anonymous, and that a personal code would be used to link the answers to their endoscopies performed. They were sent a reminder 4 weeks later.

Study population

Inclusion criteria were histologically confirmed intestinal metaplasia in biopsies obtained from columnar-lined epithelium in the esophagus, Barrett segment $\geq$2 cm, and absence of a history of HGD/EAC. To exclude prevalent cases of neoplasia at baseline, only BE patients with $\geq$6 months of FU in the study without detection of HGD/EAC were selected.

Adherence to guideline recommendations

To calculate clinical consequences and explanatory parameters for nonadherence, the actual adherence rate to ACG guideline recommendations was determined first. As this guideline was issued in 2002, and updated in 2008 and 2015, we used the recommendations of the guideline prevalent at the date that endoscopies were performed.[2,8,9] A total of 12 recommendations were selected within three domains: surveillance interval, biopsy protocol, and landmark identification (Fig. 1). Adherence rates to guideline recommendations were assessed as a binary variable per performed upper endoscopy: adherent or nonadherent. All endoscopies were assessed for adherence to all 12 recommendations separately. Adherence was defined as the proportion of the number of endoscopies in line with the guideline divided by the total number of endoscopies; corresponding 95% confidence intervals were calculated. Surveillance interval, Seattle protocol, and Prague classification were assigned to be 'primary recommendations' of the guideline. Surveillance intervals were assumed to be adherent to guideline recommendations within an accepted time range, prevalent at the time of the endoscopy (Supplementary Table 1). This indicated roughly an accepted range of 6 months shorter or longer than the guideline-recommended interval (3–5 years) in cases of non-dysplastic BE (NDBE); a range of 3 months shorter or longer than the guideline-recommended interval (3 months to 1 year) was accepted in cases of low-grade dysplasia (LGD)

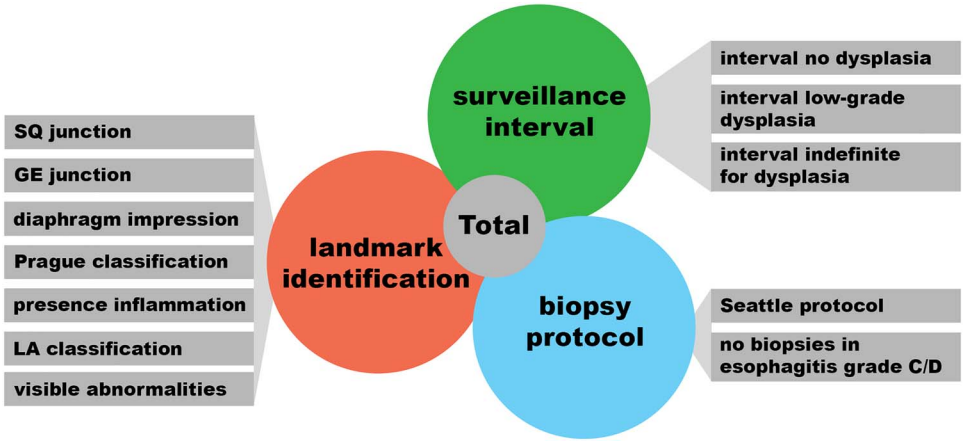

Fig. 1 Three domains and 12 guideline recommendations.

or indefinite for dysplasia (IND). The Seattle protocol was assumed to be executed appropriately if at least 4 biopsies were sampled per 2 centimeters (cm) of the maximum BE length. However, since it is not always necessary to sample 4 biopsies in a Barrett's tongue, an adjusted adherence rate was calculated: per additional two cm on top of the circular segment at least one (instead of 2) biopsies had to be sampled. As the Prague classification was not implemented in the guideline until 2008, adherence to this recommendation was assessed only for endoscopies performed afterwards.[9,10] Adherence to the Prague classification was classified as a report of the length of the Barrett's segment with respect to the circular part (C) and the maximum length (M). Total adherence was defined as the proportion of endoscopies adherent to all primary recommendations.

Ethics

The institutional review board of Erasmus MC University Medical Center (Rotterdam, The Netherlands) and the boards of each participating center approved the study protocol. Written informed consent was acquired from patients included before their first endoscopy.

Statistical analysis

Clinical consequences of nonadherence were assessed for endoscopic curability, mortality, and risk of misclassification of histological diagnosis. Of the 12 recommendations investigated for adherence in this study (Fig. 1), only surveillance interval (NDBE, IND, and LGD combined) and Seattle protocol were used in this part of the analysis, since the use of landmarks in reporting surveillance was particularly recommended to gain uniformity in reports. The 'endoscopic curability' of EAC (stage T1a versus ≥ T1b) and the number of patients who either died from EAC ('cause-specific mortality') and the patients who did not die from EAC were assessed separately for endoscopies adherent and endoscopies nonadherent to the recommended surveillance interval (particularly longer intervals) or Seattle protocol. For both outcome measurements, the statistical significance of a potential difference was estimated by using a Fisher's exact test. The risk of 'misclassification of histological diagnosis' was estimated in a multistate hidden Markov model. This model has been described before and was used in a modified design (Supplementary Fig. 1).[7] Histological diagnosis was subdivided in NDBE, LGD, and HGD/EAC. One state could be misclassified as another: true states versus observed states. For example, LGD could be the true state, but it might be misclassified as NDBE because of sampling error or because of misdiagnosis of the pathologist. In this analysis, the probability of misclassification was assessed first. In cases of nonadherence to the Seattle protocol, the probability of misclassification might be higher. Therefore, the odds of an increment of this probability were estimated afterwards. The data presented in this study are available on request from the corresponding author.

To explain nonadherence and its potential clinical consequences, the association between adherence rates to surveillance interval and Seattle protocol and potential explanatory parameters was determined. These parameters were subdivided into three categories: patient, facility, and endoscopist variables. The risk of nonadherence for patient and facility variables was modeled in mixed-effects logistic regression models, using a random intercept per patient. Patient variables included age, gender, BE length (<3 versus ≥3 cm), esophagitis (present or absent), and visible abnormalities (present or absent). The facility variable was binary (university hospital or general hospital). These results were tested for robustness with other definitions of BE length. Missing values of covariates included in the model, patient or facility variables, were imputed if >3% was missing. The imputation model took into account the multilevel structure of our data. Missing outcomes ([non]adherence) were

Table 1 Baseline characteristics of patients included and of the survey among endoscopists

Baseline characteristics of patients included

Characteristic		Median (IQR)/proportion ($n = 726$)	Missing
FU time (years)		8.2 (5.3–10)	0
n° of FU		4.0 (3.0–5.0)	0
Age (years)		61 (53–69)	0
Male gender		529 (73%)	0
GERD		221 (30%)	9
PPI use		654 (90%)	2
NSAID use		34 (4.7%)	2
Aspirin use		102 (14%)	1
Statin use		209 (29%)	161
Smoking	Current	146 (20%)	12
	Ever	329 (45%)	
	Never	239 (33%)	
Alcohol	Current	552 (76%)	12
	Ever	66 (9.1%)	
	Never	96 (13%)	
BMI (kg/m^2)		27 (25–29)	88
Length of BE	Continuous	4.0 (2.0–5.0)	0
	≥3 cm	537 (74%)	0
Esophagitis present		74 (10%)	1
Nodularity present		32 (4.4%)	0

Baseline characteristics of survey among endoscopists

Characteristic		Median (IQR)/Proportion ($n = 57$)
GENERAL INFORMATION		
Age (years)		45 (40–56)
Male gender		40 (70%)
Specialism	Gastroenterology	54 (95%)
	Internal medicine	3 (5%)
	Surgery	0 (0%)
	Other	0 (0%)
Subspecialism[†]	General	25 (44%)
	Upper digestive tract	11 (19%)
	Lower digestive tract	6 (11%)
	Biliairy pathology	16 (28%)
	Hepatology	14 (25%)
	IBD	12 (21%)
	Oncology	13 (23%)
	Other	6 (11%)
Status of training	Resident	16 (28%)
	Specialist	37 (65%)
Number of years working in the field		9.0 (5.0–19)
Type of practice[†]	University hospital	19 (33%)
	Teaching hospital	32 (56%)
	General hospital	19 (33%)
	Barrett expert center	2 (3.5%)
ENDOSCOPY		
Time for surveillance endoscopy	10 minutes	2 (3.5%)
	15 minutes	36 (63%)
	20 minutes	10 (18%)
	25 minutes	2 (3.5%)
	Other	5 (8.8%)
BE surveillance endoscopy per year	≤50	41 (72%)
	51–100	12 (21%)
	101–150	2 (3.5%)
	≥151	0 (0%)
Upper endoscopies in general per month	≤20	6 (11%)
	21–35	13 (23%)
	36–49	22 (39%)
	≥50	15 (26%)
Years of experience in surveillance of BE	0–5 years	15 (26%)
	6–10 years	15 (26%)
	11–15 years	7 (12%)
	≥16 years	19 (33%)
Knowledge guideline	Surveillance interval	36 (63%)
	Seattle protocol	49 (86%)
	Histopathology	39 (68%)

(*Continued*)

Table 1 Continued

Baseline characteristics of survey among endoscopists		
Characteristic		**Median (IQR)/Proportion ($n = 57$)**
RISK		
Risk estimation neoplastic progression NDBE	0.1–0.2%	21 (37%)
	0.3–1%	33 (58%)
	2–5%	0 (0%)
	6–10%	0 (0%)
OPINION		
Surveillance is cost-effective		10 (18%)
Evidence underpinning guideline		34 (60%)
Survival benefit because of surveillance		39 (68%)
Agreement Seattle protocol	Agree	25 (44%)
	Disagree	8 (14%)
	Do not know	22 (39%)
Histological diagnosis is an adequate marker	Agree	7 (12%)
	Disagree	47 (82%)
	Do not know	2 (3.5%)
Surveillance interval of 3 years in NDBE	Too short	21 (37%)
	Adequate	32 (56%)
	Too long	0 (0%)
Surveillance interval of 1 year in LGD	Too short	8 (14%)
	Adequate	33 (58%)
	Too long	14 (25%)

[†]Multiple answers were allowed.

not imputed. Endoscopist variables were continuous, binary, or categorical, following the answering possibilities in the postal questionnaire. The risk of nonadherence for endoscopist variables was estimated using multivariable Lasso regression. All variables of which ORs were reported in Table 4 were associated with better (OR > 1) or worse (OR < 1) adherence. If not reported, nonadherence could not be explained by these particular parameters. Because of the statistical analysis used, confidence intervals were not calculated. Because not all endoscopists responded to the questionnaire, only those endoscopies performed by respondents were included in the analysis. Consequently, not for every variable sufficient data were available to include all endoscopist variables investigated in the questionnaire. To evaluate a potential impact on the results of excluding those endoscopies performed by nonresponders from the analysis, a nonresponder analysis was performed.

RESULTS

Baseline characteristics

A total of 726 patients were included, with a median FU time of 8.2 years (IQR 5.3–10) (Table 1). The median age was 61 years (IQR 53–69) and the cohort predominantly consisted of males (73%). The median BE length was 4.0 cm (IQR 3.0–5.0) with 74% patients having a long segment BE (≥3 cm); 30% of patients had symptoms of GERD.

In these patients, 3802 endoscopies were performed by 167 endoscopists. Questionnaires were sent to 155 endoscopists; of 12 endoscopists we were unable to obtain contact information. Sixty-three (41%) endoscopists returned the questionnaire; six were not filled out for various reasons (e.g. the recipient was no longer employed at the contacted institution). Consequently, 57 questionnaires (37%) were used in the analysis (Table 1). The median age of the endoscopists was 45 years (IQR 40–56), mostly male (70%) and gastroenterologist (95%); 19% had the upper digestive tract as a subspecialty, and 23% oncology.

Adherence to guideline recommendations

Adherence ranged from 16 to 99%, depending on the recommendation investigated (Supplementary Fig. 2 and Supplementary Table 2). Total adherence to all 'primary recommendations' was 5.5% (161/2944).

Surveillance interval

The interval until the next surveillance endoscopy was according to guideline recommendations in 16% (363/2344) of endoscopies with a histological diagnosis of NDBE. With 82% (1921/2344) the majority of the nonadherent endoscopies were performed at a shorter interval; 2.6% (60/2344) were performed at a longer interval. If LGD was detected, the interval until the next surveillance endoscopy was adherent in 55% (298/545) of endoscopies. Most of these nonadherent endoscopies (32% [174/545]) were performed at a longer interval than recommended; 13% (73/545) at a shorter interval.

Seattle protocol

The Seattle protocol for taking biopsies was followed appropriately in 54% (1665/3105); in all other endoscopies, fewer biopsies were taken than

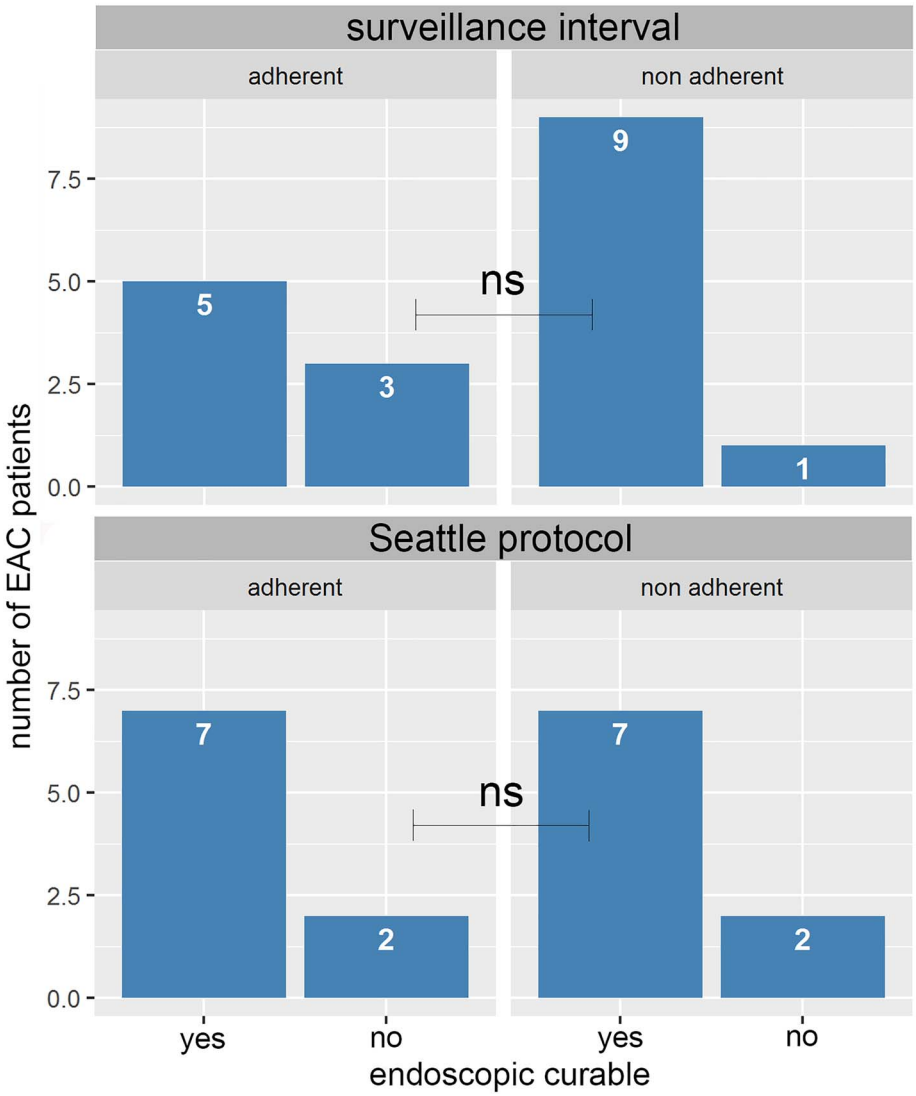

Fig. 2 Association of nonadherence to primary guideline recommendations and endoscopic curability in patients who developed EAC.

recommended. The mean number of biopsies per cm was 1.9 (SD 0.9).

Prague classification

Length of BE was reported according to Prague classification in 61% (1121/1850) of endoscopies.

Clinical consequences of nonadherence

Endoscopic curability of EAC

EAC was detected in 18 patients. Fourteen patients were cured endoscopically; four patients needed a more invasive treatment (Fig. 2 and Supplementary Table 3).

Out of these 18 EAC patients, 10 endoscopies were nonadherent to surveillance interval. Eight were performed too early, two were performed too late; out of those performed too late, one EAC was endoscopically curable and one was not endoscopically curable. There was no statistically significant difference between adherence to surveillance interval and endoscopic curability of EAC ($P = 0.27$).

The same proportion of EACs was endoscopically curable if biopsies were taken as recommended by the Seattle protocol and if fewer biopsies were taken ($P = 1.0$).

Mortality because of EAC

In our cohort, 164 (23%) patients died, of which six because of EAC.

Of the six patients who died from EAC, none of the endoscopies were performed at intervals longer than recommended, only earlier or in time (Table 2 and Supplementary Tables 4 and 5).

The cause-specific mortality was not higher in patients whose endoscopy before EAC detection was nonadherent to the Seattle protocol compared with those adherent ($p = 0.68$).

Table 2 Probability of mortality because of EAC in cases of (non)adherence. Only the last interval or endoscopy prior to detection of EAC was used

		Mortality because of EAC	Mortality from other causes
Surveillance interval	Adherent	83% (5/6)	25% (40/158)
	Too early	17% (1/6)	61% (97/158)
	Too late	0% (0/0)	13% (21/158)
			P = n.a.
Seattle protocol	Adherent	67% (4/6)	39% (64/158)
	Nonadherent	33% (2/6)	40% (66/158)
	Missing	0% (0/6)	17% (28/158)
			$P = 0.68$

It is not applicable (n.a.) to calculate the difference in cause-specific mortality for surveillance interval. The question of interest is whether a longer interval than recommended increased the risk of cause-specific mortality. However, among patients who have died because of EAC, there were no intervals too long.

Table 3 Odds of misclassification of histological diagnosis in cases of nonadherence to the Seattle protocol.

OBSERVED histological diagnosis	TRUE histological diagnosis	Adherence to Seattle protocol	OR (95% CI)
Any histological diagnosis	Any other histological diagnosis	Adherence	Ref.
		Nonadherence	**1.30 (1.02; 1.67)**
ND	LGD	Adherence	Ref.
		Nonadherence	**3.44 (1.19; 9.95)**
LGD	ND	Adherence	Ref.
		Nonadherence	**1.30 (1.03; 1.66)**
HGD/EAC	LGD	Adherence	Ref.
		Nonadherence	2.31 (0.21; 26.0)
HGD/EAC	ND	Adherence	Ref.
		Nonadherence	1.04 (0.52; 2.06)

Statistically significant results are presented in bold.

Misclassification of histological diagnosis

The probability of misclassification of any histological diagnosis increased by 30% if biopsies were not taken according to Seattle protocol (OR 1.3, 95% CI 1.0; 1.7) (Table 3). If the odds of every potential misclassification were determined separately, this increased the risk of misclassification turned out to be particularly caused by the following: NDBE being observed as LGD (OR 3.4, 95% CI 1.2; 10), and LGD being observed as NDBE (OR 1.3, 95% CI 1.0; 1.7). There was no statistically significant association between nonadherence to the Seattle protocol and the misclassification in histological diagnosis of LGD or NDBE being observed as HGD/EAC (OR 2.3, 95% CI 0.2; 26, OR 1.0, 95% CI 0.5; 2.1, respectively), which could have caused overtreatment.

Explanatory parameters for nonadherence

The association between adherence rates to 'primary recommendations' and potential explanatory parameters is shown in Table 4.

Patient variables

Increasing age was associated with a marginal increment in the probability of adherence to surveillance interval (OR 1.1, $P = 0.05$). Adherence to the Seattle protocol was better in cases of visible abnormalities (OR 1.8, $P = 0.01$). However, fewer biopsies than recommended were taken in long BE segments (OR 0.1, $P < 0.01$). Adherence to the Prague classification was lower in older patients (OR 0.9, $P < 0.01$).

Facility variables

Adherence to the Seattle protocol was better in university than in general hospitals (OR 2.1, $P < 0.01$). Facility variables did not impact adherence to surveillance interval or the Prague classification.

Endoscopist variables

In all, 60% of the endoscopists believed that the evidence underpinning the guideline is sufficient, 44% agreed with the Seattle protocol, 12% deemed histological diagnosis to be an adequate marker, and, respectively, 56 and 58% thought that the surveillance interval for NDBE and LGD is appropriate (Table 1).

The complete results of variables that increased the risk of nonadherence are reported in Supplementary Table 7. Those parameters affecting adherence more than twice were the following: (i) performing more BE surveillance endoscopies annually was associated with higher adherence to surveillance interval guidelines (51–100 per year OR 1.8, 101–150 per year OR 2.6). (ii) Endoscopies performed by endoscopists who deemed histological diagnosis to be an adequate marker were associated with higher adherence to surveillance intervals (OR 2.0). (iii) Board-certified specialists were associated with higher adherence to the Seattle protocol than gastroenterologists

Table 4 Explanatory parameters for improved adherence to primary guideline recommendations. Endoscopist-related variables affecting adherence more than twice for at least one primary recommendation were reported.

Domain		Surveillance interval	Biopsy protocol	Landmark identification
Recommendation guideline		NDBE and LGD and IND	Seattle protocol	Prague classification
		OR (95% CI)	OR (95% CI)	OR (95% CI)
Patient related				
Age (10 years older)		**1.10 (1.00; 1.22)**	0.93 (0.81; 1.04)	**0.88 (0.77; 0.95)**
Gender (female)		0.85 (0.67; 1.08)	0.86 (0.65; 1.15)	1.08 (0.87; 1.35)
BE length (LSBE)		1.03 (0.82; 1.30)	**0.12 (0.09; 0.16)**	1.17 (0.94; 1.47)
Inflammation (present)		0.86 (0.54; 1.39)	0.91 (0.60; 1.39)	1.22 (0.76; 1.96)
Visible abnormality (present)		0.73 (0.47; 1.16)	**1.77 (1.16; 2.69)**	0.69 (0.45; 1.06)
Facility related				
Type of practice (university)		1.19 (0.91; 1.53)	**2.14 (1.57; 2.92)**	1.20 (0.94; 1.53)
		OR	OR	OR
Endoscopist related				
Subspecialty upper digestive tract or oncology		1.65	0.62	3.20
Status of training (specialist)		—	3.90	—
BE surveillance endoscopies per year	≤50	Ref.	Ref.	Ref.
	51–100	1.77	—	—
	101–150	2.59	—	—
Upper endoscopies in general per month	≤20	Ref.	Ref.	Ref.
	21–35	—	0.36	—
	36–49	—	0.29	—
	≥50	—	0.49	—
Years of experience in surveillance of BE	0–10	Ref.	Ref.	Ref.
	11–15	1.05	—	0.53
	≥16	1.01	—	1.65
Knowledge guideline	Bad/OK	Ref.	Ref.	Ref.
	Good	—	0.45	0.76
Histological diagnosis is an adequate marker (yes)		—	1.97	1.71

Statistically significant results are presented in bold.

in training (OR 3.9). (iv) Those endoscopists whose answers in the theoretical assessment were in line with guideline recommendations were associated with reduced adherence to the Seattle protocol (OR 0.5). (v) Endoscopies of endoscopists who performed more upper endoscopies per month, not necessarily with BE surveillance as an indication, were associated with less adherence (21–35 OR 0.4, 36–49 OR 0.3, ≥50 OR 0.5). (vi) Use of the Prague classification was associated with higher adherence if endoscopists had the upper digestive tract or oncology as subspecialty (OR 3.2). (vii) More years of experience with BE surveillance was associated with less adherence to the Prague classification (11–15 years OR 0.5), but with increasing experience (up to more ≥16 years) the association with adherence to Prague classification was reversed (OR 1.7).

The results were robust for other definitions of BE length (Supplementary Table 8). Adherence to surveillance interval and landmark identification was higher among endoscopies performed by respondents than performed by nonrespondents (both $P < 0.01$) (Supplementary Table 9).

DISCUSSION

In this study, based on our cohort, including a limited number of cases of neoplastic progression, we were not able to collect evidence that longer surveillance intervals and sampling fewer biopsies than recommended affect endoscopic curability of EAC, cause-specific mortality, and overtreatment of BE because of misclassification of histological diagnosis. Given the limited power of this part of our study, these findings should not be interpreted as ground for adjusting current surveillance intervals. It should, however, be a signal to re-evaluate the effectiveness of the guideline, including evidence underpinning recommendations as well as the strategy used to predict neoplastic progression risk. We found an adherence rate to surveillance interval, Seattle protocol, and Prague classification of only 5.5%; this was particularly caused by shorter intervals for NDBE, longer intervals for LGD, and sampling of fewer biopsies than recommended by the Seattle protocol. The most prominent variables associated with better adherence were shorter BE segments, surveillance performed in a university hospital, more experience in performing BE surveillance endoscopies, and if endoscopists deemed histological diagnosis to be an adequate marker. Endoscopists' opinion had a minor influence on adherence.

As BE surveillance aims to detect EAC at an early stage, one would expect that if guideline recommendations are not followed appropriately, EAC would be detected in an endoscopically non-curable stage, or patients may even die because of esophageal cancer. This hypothesis was not supported by the results or our study: there was no difference between adherence and nonadherence (i.e. longer surveillance intervals,

fewer endoscopies) in these outcome measurements. Although the numbers of EAC detected in our cohort are limited and the number of cause-specific deaths is small, they have been observed after a considerable FU time, which we expect to reflect clinical practice. Since EAC was not more often endoscopically incurable in case too long intervals were used, and patients having longer intervals did not die because of EAC, one could contemplate that it may be safe to lengthen surveillance intervals. The consideration to lengthen intervals is supported by economic considerations.[7,11] Also, the level of evidence for the currently recommended intervals in the guideline is moderate at best.[2]

Besides, we observed that endoscopic curability of EAC or mortality was not related to nonadherence to the Seattle protocol. Contrarily, we did observe that sampling fewer biopsies than recommended was associated with a higher misclassification rate of histological diagnosis in general, which is particularly caused by misclassification of NDBE as LGD and vice versa. Abrams *et al*.[4] pointed out in a large retrospective analysis that less dysplasia was detected in cases of nonadherence to the Seattle protocol. These results suggest that the Seattle protocol may be effective in the detection of dysplasia, but our results add that its additional value may be limited with respect to the diagnosis of HGD/EAC, since nonadherence does not affect that risk of misclassification in our series. Given the limited clinical consequences of nonadherence, a less stringent or alternative biopsy protocol may be considered. Simultaneously, this may improve adherence to the biopsy protocol, as more than half of endoscopists did not agree with the Seattle protocol to be effective.

Explanatory factors for nonadherence that have already been reported in the literature were confirmed in our study: adherence is better in university hospitals and shorter BE length.[6,12,13] This study is the first that investigates patient, endoscopist, and facility variables as explanatory parameters for nonadherence, in which adherence rates are collected without self-report, but with the possibility to relate endoscopist variables (e.g. opinion and organizational structure of surveillance) to the endoscopies they performed. Previously, two studies have also combined both a survey, endoscopy, and pathology reports, but the survey was not among the endoscopists who performed the endoscopies and could therefore not be directly correlated.[14,15] Also, in the literature, studies were most often performed in univariable analysis; we used multivariable analysis.[5]

There are several limitations to our study. Because of the innovative design concerning the analysis investigating the association of adherence to primary recommendations and endoscopist variables it was difficult to include mediators and colliders. Therefore, the results should not be interpreted as a definitive causal relation.[16] It could be used as a proof-of-principle and a starting point for further research exploring these associations. In this part of the analysis, we have also observed a low response rate of the survey among endoscopists. Only endoscopies performed by endoscopists who responded were included. Differences were observed between endoscopies included and endoscopies excluded. Because of this low response rate, the analysis concerning explanatory parameters was separately for endoscopist variables and patient and facility variables. Therefore, it was not possible to adjust for all parameters simultaneously. Additionally, since endoscopies were included from 2003, recall bias may play a role in this study: endoscopists who have been performing study endoscopies may have changed their opinion over time. Taken together, this underscores the fact that these analyses are exploratory and results should be interpreted with caution. Another limitation is that ideally we would have adjusted for all potential clustering effects in our model investigating the association between adherence and both patient and facility variables. However, to prevent the model from overfitting, we have particularly focused on the most important potential clustering effect, which we expected to be patients, as included in the random effects part of the model. Given our aims and the large variation in endoscopists over the years of this observational study, we do think we can use this model to address this aim of the study. Important to notify is that when analyzing the association between explanatory parameters and adherence to surveillance interval, nonadherence was considered as performing endoscopies both at a shorter and longer interval than recommended by the guideline. Ideally, we would have performed the analysis separately for endoscopies with either NDBE or LGD being adherent or nonadherent, in which nonadherent would be subdivided in too early and too late. However, in order to maintain a considerable amount of power in the analysis and to be uniform with the analysis for the other domains, we combined them. Besides, the number of patients who have developed or who passed away because of EAC is limited because of the known low BE progression rate. This was particularly relevant for the part of the analysis in which surveillance interval was related to endoscopic curability and mortality. Only few endoscopies of these EAC patients were performed at a longer surveillance interval than recommended; consequently, results should be interpreted with caution. Strong points are the extensive methodology in evaluating the association between endoscopists' opinion and objective data collection from endoscopy and pathology reports. Also, the multistate hidden Markov model based on 726 BE patients is a solid method to provide evidence for the effect of nonadherence to the Seattle protocol on the misclassification risk of histological diagnosis.

In conclusion, the disadvantageous effect of longer surveillance intervals and fewer biopsies than recommended by BE guidelines may be limited with respect to endoscopic curability of EAC and mortality; nonadherence does not appear to affect the probability of misclassification of histological diagnosis. As this is an exploratory analysis given the low number of EACs included, this should be further investigated. The results of our study could, however, be interpreted as a signal that not optimal adherence to guideline recommendations itself is the goal, but the improvement of the methodology of surveillance. For example, the implementation of other biomarkers that contribute to a better risk estimation of neoplastic progression and corresponding (longer) risk-based intervals could be considered. Besides, the effectiveness of the recommended biopsy protocol may be re-evaluated, given it is time-consuming and error-prone because of nonadherence. Ultimately, improving the evidence underpinning the guideline would contribute most to improve the surveillance practice for BE.

ACKNOWLEDGMENTS

The authors would like to thank the members of the ProBar Study Group for their role in this study: J. Alderliesten, G. van Ingen, F.T.M. Peters, A. Karrenbeld, F. ter Borg, M. Louwen, P.C.J. ter Borg, M.A. den Bakker, A. Alkhalaf, F.C.P. Moll, J.J. Kolkman, J. van Baarlen, J.P.W. Burger, J.W.R. Meijer, A.J.P. van Tilburg, H. van der Valk, M.J.L. Romberg-Camps, W. Vos, R.J.F. Felt, and T.J. Tang.

ABBREVIATIONS

ACG, American College of Gastroenterology; BE, Barrett's esophagus; EAC, esophageal adenocarcinoma; FU, follow-up; GE junction, gastroesophageal junction; GERD, gastroesophageal reflux disease; HGD, high-grade dysplasia; IND, indefinite for dysplasia; IQR, interquartile range; LGD, low-grade dysplasia; NDBE, non-dysplastic Barrett's esophagus; OR, odds ratio; SQ junction, squamocolumnar junction.

CONFLICT OF INTEREST

Nothing to disclosure.

AUTHOR CONTRIBUTIONS

C.A.M. R.: concept and design, study planning, interpretation data, collection and interpretation data, drafting manuscript. R.D.B.: collection and interpretation data, critical revision of the manuscript for intellectual content. D.N.: interpretation of data, critical revision of the manuscript for intellectual content. E.W.S.: interpretation of data, critical revision of the manuscript for intellectual content. D.R.: interpretation of data, critical revision of the manuscript for intellectual content. I.L.-V.: interpretation of data, critical revision of the manuscript for intellectual content. K.B.: interpretation of data, critical revision of the manuscript for intellectual content. M.J.B.: interpretation of data, critical revision of the manuscript for intellectual content. M.C.W.S.: concept and design, interpretation data, critical revision of the manuscript for intellectual content.

References

1. Werf L R van der, Busweiler L A D, van Sandick J W, van Berge Henegouwen M I, Wijnhoven B P L, Upper D, GICAg. Reporting National Outcomes after esophagectomy and gastrectomy according to the Esophageal Complications Consensus Group (ECCG). Ann Surg 2020; 271(6): 1095–101.
2. Shaheen N J, Falk G W, Iyer P G, Gerson L B, American College of Gastroenterology. ACG clinical guideline: diagnosis and management of Barrett's esophagus. Am J Gastroenterol 2016; 111(1): 30, quiz 1–50.
3. Siersema P D, Bergman J J G H M, Van Berge Henegouwen M I *et al.* Richtlijn Barrett-oesofagus. Dutch Society of Gastroenterology, Haarlem, The Netherlands. 2017.
4. Abrams J A, Kapel R C, Lindberg G M *et al.* Adherence to biopsy guidelines for Barrett's esophagus surveillance in the community setting in the United States. Clin Gastroenterol Hepatol 2009; 7(7): 736, quiz 10–42.
5. Roumans C A M, van der Bogt R D, Steyerberg E W *et al.* Adherence to recommendations of Barrett's esophagus surveillance guidelines: a systematic review and meta-analysis. Endoscopy 2020; 52(1): 17–28.
6. Curvers W L, Peters F P, Elzer B *et al.* Quality of Barrett's surveillance in the Netherlands: a standardized review of endoscopy and pathology reports. Eur J Gastroenterol Hepatol 2008; 20(7): 601–7.
7. Kastelein F, van Olphen S, Steyerberg E W *et al.* Surveillance in patients with long-segment Barrett's oesophagus: a cost-effectiveness analysis. Gut 2015; 64(6): 864–71.
8. Sampliner R E, Practice Parameters Committee of the American College of Gastroenterology. Updated guidelines for the diagnosis, surveillance, and therapy of Barrett's esophagus. Am J Gastroenterol 2002; 97(8): 1888–95.
9. Wang K K, Sampliner R E, Practice Parameters Committee of the American College of Gastroenterology. Updated guidelines 2008 for the diagnosis, surveillance and therapy of Barrett's esophagus. Am J Gastroenterol 2008; 103(3): 788–97.
10. Sharma P, Dent J, Armstrong D *et al.* The development and validation of an endoscopic grading system for Barrett's esophagus: the Prague C & M criteria. Gastroenterology 2006; 131(5): 1392–9.
11. Inadomi J M, Somsouk M, Madanick R D, Thomas J P, Shaheen N J. A cost-utility analysis of ablative therapy for Barrett's esophagus. Gastroenterology 2009; 136(7): 2101–14e1-6.
12. Menezes A, Tierney A, Yang Y X *et al.* Adherence to the 2011 American Gastroenterological Association medical position statement for the diagnosis and management of Barrett's esophagus. Dis Esophagus 2015; 28(6): 538–46.
13. Gordon-Cooke J, Hillman L C. Barrett's oesophagus: are we really compliant with a standard biopsy protocol for surveillance? J Gastroenterol Hepatol 2015; 30: 43.
14. Shen E F, Gladstone S, Milne G, Paterson-Brown S, Penman I D. Endoscopic surveillance practice for Barrett's oesophagus in Scotland and early experience in implementing local guidelines. Scott Med J 2003; 48(2): 43–5.
15. Das D, Ishaq S, Harrison R *et al.* Management of Barrett's esophagus in the UK: overtreated and underbiopsied but improved by the introduction of a national randomized trial. Am J Gastroenterol 2008; 103(5): 1079–89.
16. Westreich D, Greenland S. The table 2 fallacy: presenting and interpreting confounder and modifier coefficients. Am J Epidemiol 2013; 177(4): 292–8.