Post-Endoscopy Esophageal Neoplasia in Barrett's Esophagus: Consensus Statements from an International Expert Panel

Supplementary Material & Appendices

TABLE OF CONTENTS

SUPPLEMENTARY TABLE 1 ..... **2**

SUPPLEMENTARY TABLE 2 ..... **3**

SUPPLEMENTARY TABLE 3 ..... **4**

SUPPLEMENTARY TABLE 4 ..... **6**

SUPPLEMENTAL MATERIAL **7**

APPENDIX 1 INTAKE SURVEY **12**

APPENDIX 2 PEEC: ROUND 1 ..... **14**

APPENDIX 3 RANKING INSTRUCTIONS **19**

APPENDIX 4 PEEC: ROUND 2 ..... **20**

APPENDIX 5 DETECTION OF HGD-EAC DURING SCREENING/SURVEILLANCE ENDOSCOPY **25**

APPENDIX 6 RESPONSES RELATED TO STATEMENTS FOR FUTURE RESEARCH **26**

APPENDIX 7 ESTABLISHING AN INFRASTRUCTURE FOR FUTURE RESEARCH **28**

Supplementary Table 1: GRADE evidence profile comparing chromoendoscopy (dye-based or virtual) compared with high-definition white light endoscopy (HD-WLE) for detection of high-grade dysplasia or esophageal adenocarcinoma in patients with known Barrett's esophagus undergoing screening or surveillance.

Summary of the Evidence:

Chromoendoscopy COMPARED TO HD-WLE FOR detection of HGD/EAC in patients with known BE undergoing screening or surveillance

Outcomes	Study event rates (95% CI)		Relative effect (95% CI)	Absolute effect* (assumed Neoplasia Detection Rate 5%)	№ of participants (studies)	Certainty of the evidence (GRADE)
	Dysplasia detection with HD-WLE	Dysplasia detection with Chromoendoscopy				
HGD/EAC	51/504 (10.1%)	74/504 (14.7%)	RR 1.44 (1.05 to 1.98)	22 more per 1,000 (from 3 more to 49 more)	504 (7 RCTs)	⨁⨁◯◯[1,2] LOW

GRADE Working Group grades of evidence
High quality: We are very confident that the true effect lies close to that of the estimate of the effect
Moderate quality: We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different
Low quality: Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect
Very low quality: We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect

[1]Rated down for risk of bias with studies primarily from referral centers with an exceptionally high rates of HGD/EAC detection (10-15%) during screening/surveillance in patients with NDBE, compared with population-based studies suggesting rates <1%
[2]Rated down for imprecision due to optimal information size (event rate <200)

Supplementary Table 2: GRADE evidence profile comparing Seattle biopsy protocol compared with non-protocolized biopsy for detection of any dysplasia in patients with known Barrett's esophagus undergoing screening or surveillance.

Summary of the Evidence:

Seattle biopsy protocol COMPARED TO non-protocolized biopsy FOR detection of any dysplasia in patients with known BE undergoing screening or surveillance

Outcomes	Study event rates (95% CI)		Relative effect (95% CI)	Absolute effect* (assumed NDR 5%)	№ of participants (studies)	Certainty of the evidence (GRADE)
	Dysplasia detection with non-protocolized biopsy	Dysplasia detection with Seattle biopsy protocol				
Any dysplasia	6/234 (2.6%)	52/272 (19.1%)	RR 6.27 (2.75 to 14.33)	264 more per 1,000 (from 88 more to 667 more)	506 (2 observational studies)	⨁◯◯◯[1,2,3] VERY LOW

GRADE Working Group grades of evidence
High quality: We are very confident that the true effect lies close to that of the estimate of the effect
Moderate quality: We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different
Low quality: Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect
Very low quality: We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect

[1] Observational studies with intrinsic risk of bias
[2] Rated down for imprecision due to optimal information size (event rate <200)
[3] Rated down for indirectness since detection of any dysplasia is a surrogate for post-endoscopy esophageal neoplasia

Supplementary Table 3: Statements voted as uncertain or inappropriate using the RAND/University of California, Los Angeles Appropriateness Method with median score, and number of experts in each category range

Statements	Median Score	# of Experts 1-3 Range (Inappropriate) (n, %)	# of Experts 4-6 Range (Uncertain) (n, %)	# of Experts 7-9 Range (Appropriate) (n, %)	MAD-M Score
Terminology and Definitions Round 1					
1) Post-endoscopy esophageal cancer (PEEC) is the preferred term for low-grade dysplasia (LGD), high-grade dysplasia (HGD) or esophageal adenocarcinoma (EAC) detected before the next recommended surveillance endoscopy after an index upper endoscopy showed no evidence of HGD/EAC in a patient diagnosed with non-dysplastic Barrett's esophagus (NDBE)	3	17 (71)	5 (21)	2 (8)	1.4
2) PEEC is the preferred term for LGD or HGD or EAC detected before the next recommended surveillance endoscopy after an index endoscopy showed no evidence of HGD or EAC in a patient diagnosed with NDBE at screening or surveillance endoscopy	3	17 (71)	6 (25)	1 (4)	1.3
Quality Review of PEEN/PEEC cases Round 1					
3) Endoscopy practices should implement a formal process to identify and register PEEC cases to review for potential causative factors and perform a root cause analysis	7	0	8 (33)	16 (77)	1.3

4) In order to benchmark endoscopy services, the quality of index endoscopy in BE can be measured by calculating unadjusted PEEC proportion (number of PEEC cases divided by the number of PEEC cases plus detected HGD/EAC cases)	5.5	4 (17)	11 (46)	9 (37)	1.8
5) PEEC proportion should be initially reported at a service level rather than individual endoscopist level	7	1 (4)	9 (37)	14 (59)	1.5
6) In order to benchmark endoscopy services, the quality of index endoscopy in BE can be measured by calculating the neoplasia detection rate (NDR, rate of detection of HGD/EAC) in patients with BE undergoing screening endoscopy	6	4 (17)	11 (46)	9 (37)	1.4
7) In order to benchmark endoscopy services, the quality of index endoscopy in BE can be measured by calculating the neoplasia detection rate (NDR, rate of detection of HGD/EAC) in patients with BE undergoing screening endoscopy	6	4 (17)	9 (37)	11 (46)	1.5
8) Endoscopists should complete validated training courses that highlight best practices in performing endoscopy in patients with BE (e.g. BORN project, CORE-BE)	6.5	0	12 (50)	12 (50)	1.0
Best Practice Advice to Reduce PEEN/PEEC Round 1 and 2					
9) Endoscopists should use advanced sampling techniques such as wide-area transepithelial sampling (WATS3D) as an adjunct to the Seattle biopsy protocol to augment detection of neoplasia during screening and surveillance endoscopy for BE					
Round 1	5	8 (33)	11 (46)	5 (21)	1.6
Round 2	3	13 (52)	8 (32)	4 (16)	1.6

Supplementary Table 4: GRADE evidence profile comparing wide area transepithelial sampling plus forceps biopsy compared with forceps biopsy alone for detection of dysplasia in patients with known Barrett's esophagus undergoing screening or surveillance.

Summary of the Evidence:

WATS + forceps biopsy COMPARED TO forceps biopsy alone FOR detection of dysplasia in patients with known BE undergoing screening or surveillance

Outcomes	Study event rates (95% CI)		Relative effect (95% CI)	Absolute effect* (assumed NDR 5%)	№ of participants (studies)	Certainty of the evidence (GRADE)
	Dysplasia detection with forceps biopsy alone	Dysplasia detection with WATS+FB				
Any dysplasia	169/13950 (1.2%)	438/13590 (3.2%)	RR 2.13 (1.30 to 3.47)	56 more per 1,000 (from 15 more to 124 more)	13,950 (5 observational studies + 1 clinical trial observational)	⨁◯◯◯[1,2,3] VERY LOW

GRADE Working Group grades of evidence
High quality: We are very confident that the true effect lies close to that of the estimate of the effect
Moderate quality: We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different
Low quality: Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect
Very low quality: We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect

[1] Rated down for risk of bias since evidence largely derived from observational studies
[2] Rated down for inconsistency, both statistical (I^2=82%) as well as wide variability in NDR between different studies and differences in outcomes (1 study only reported risk of HGD/EAC)
[3] Rated down for indirectness, since the outcome was detection of dysplasia which is (sub-optimal) surrogate for post-endoscopy HGD/EAC

Supplemental Material

Methods:

Study Design

In this prospective study, the RAND/University of California, Los Angeles Appropriateness Methodology (RAM) was employed to develop recommendations for post-endoscopy HGD and EAC in BE patients. An appropriate indicator is one in which the expected health benefit exceeds the expected negative consequences by a sufficiently wide margin, exclusive of costs.[1]

We used a modified Delphi process that, unlike the original Delphi process, provides panelists with the opportunity to discuss their judgments in a face-to-face meeting between rating rounds. This methodology is especially useful when randomized controlled trials are not available or cannot provide evidence at a level of detail sufficient to apply to the wide range of patients seen in everyday clinical practice.[1] It is a well-described methodology that has been applied across a broad range of disease processes and procedures within gastroenterology.[2-4] A multidisciplinary team of international experts including gastroenterologists, epidemiologists, a pathologist, methodologists and a statistician was recruited. The personnel selection criteria included leadership in the field of BE and EAC, subject-matter expertise through relevant peer-reviewed publications in the area, and diversity of geography and practice settings. Expert panelists participated in a three-round process. A total of 25 international experts accepted the invitation and participated in both Rounds 1 and 2 [mean years in clinical practice: 17.7, standard deviation (SD) 10.1]. The vast majority of panelists practice at academic centers, all endoscopists routinely perform endoscopic surveillance in NDBE patients (mean monthly exams: 14 (SD 17.1), and 65.2% perform EET [mean monthly procedures: 6.2 (SD 4.4)]. For all statements, the base-case was assumed to be adults with NDBE with no prior diagnosis of BE-associated neoplasia and no prior EET or foregut surgery.

Meetings

Round 0 was conducted virtually on November 7, 2020. During this meeting, panelists were oriented to RAM and study objectives were discussed. All panelists completed a survey that assessed baseline characteristics (**Appendix 1**). Five members of the study team (S.W., R.Y., S.S, T.S., D.K.) were responsible for drafting the study protocol and initial statements under the following categories related to post-endoscopy HGD and EAC: (i) terminology and definitions, (ii) potential explanations, (iii) best practice advice to reduce post-endoscopy HGD and EAC, (iv) development of performance metrics (definition for rates, standardized methodology for calculation) to ultimately allow for benchmarking and comparison between services, and (v) creation of an infrastructure for future research. The best practice advice statements focused on the value of the following interventions *a priori*:

(i) Identification, classification and photo-documentation of landmarks, BE length and description of visible lesions,

(ii) Use of high-definition white light endoscopy and virtual chromoendoscopy,

(iii) Appropriate time spent for inspection of the BE segment, and,

(iv) Appropriate sampling of the BE segment using the Seattle biopsy protocol and additional sampling techniques such as wide area transepithelial sampling (WATS$^{3D®}$).

During *Round 1,* panelists independently ranked statements using an electronic survey using REDCap (University of California, San Diego) (**Appendix 2**) with specific instructions for ranking (**Appendix 3**). As per RAM protocol, panel members were instructed to ignore cost implications and feasibility issues inherent in implementing the measure. The panelists were advised to apply their ranking to the average patient presenting to the average physician at an average facility. Each statement was ranked on a 9-point interval scale with a score of 1 to 3 considered inappropriate, 4 to 6 of uncertain appropriateness and 7 to

9 appropriate. The panelists also had the opportunity to provide written comments regarding each statement and suggest modifications, which were reviewed by the core team members (S.W., R.Y., S.S, T.S., D.K.).

The *Round 2* meeting that was conducted virtually on April 24, 2021. Prior to this meeting, individual summary results of *Round 1* rankings, with overall aggregated results were provided to each panelist. In addition, a summary of the systematic review and evidence profiles for each suggested intervention, with details regarding the RAM, were provided (operational definitions of levels of appropriateness and methods to assess levels of disagreement). This meeting involved presentations by expert RAM and GRADE methodologists (RY, SS) and an expert statistician (MH), followed by a detailed discussion among the panelists with regards to each proposed statement. Panelists reviewed existing literature and areas of disagreement followed by rewording of statements when applicable. New statements could be proposed during this meeting. After the meeting, the panelists then re-ranked each statement for its perceived levels of appropriateness (**Appendix 4**).

Literature review and rating quality of evidence

For pre-defined focused questions on best practices for endoscopic assessment of BE, a comprehensive literature search of Ovid Medline (Ovid MEDLINE in-process and other non-indexed citations, OVID MEDLINE), Embase and Cochrane Database of Systematic Reviews/Cochrane Register of controlled trials was performed in December 2020. The focused questions were transformed into the PICO format: (P) population in questions; (I) intervention; (C) comparator; and (O) outcomes of interest. For all clinical questions, potential patient-relevant outcomes were identified *a priori* and rated from not important to critical through a consensus process. Development of HGD and/or EAC *after* an endoscopy showing NDBE was considered a critical outcome; detection of any dysplasia at the index endoscopy was considered a surrogate outcome since detection of any dysplasia during index endoscopy is likely to

be inversely related to the detection of HGD and/or EAC on endoscopy performed prior to next recommended endoscopy. For each PICO question, either existing systematic reviews and/or meta-analyses were identified and reviewed or a new systematic review and meta-analysis was conducted. The certainty of the evidence was determined using the Grading of Recommendations Assessment, Development and Evaluation (GRADE) framework, and classified as high, moderate, low or very low. [5] In this approach, direct evidence from randomized controlled trials (RCTs) starts at high quality and can be downgraded based on risk of bias in the body of evidence (or study quality), indirectness (addressing a different but related population, intervention, or outcome, from the one of interest), imprecision (of summary estimate and boundaries of 95% CI), inconsistency (or heterogeneity, both statistical in effect estimate, and conceptually in intervention or outcomes), and/or publication bias. In contrast to RCTs, evidence derived primarily from observational studies starts at low quality, and can be downgraded (similar to evidence from RCTs), or upgraded if the magnitude of benefit is large. The certainty of evidence favoring one intervention over another for the pre-specified focused questions was presented to the panel, and appropriateness of the specific intervention was assessed using the RAM approach. The GRADE evidence profiles were developed using GDTpro application (http://gdt.guidelinedevelopment.org/app).

Statistical Analysis

For *Rounds 1 and 2*, reports included histograms for each question demonstrating the distribution of ratings and also included (a) the median response, (b) measures of spread of the responses; and (c) three measures of appropriateness. The measures of spread included the count of responses in each 3 point region (1-3, 4-6, and 7-9), and the mean absolute deviation from the median (MAD-M). Appropriateness of a metric (typically classified as appropriate, uncertain, or inappropriate) was based on (a) the median rating, and (b) whether or not panelists agreed, as measured by the amount of

dispersion of the ratings. Since there is no consensus on the best approach to measuring dispersion, and different methods can result in different conclusions, the participants were provided three different versions: BIOMED Classical, p-value, and Interpercentile Range Adjusted for Symmetry (IPRAS). The BIOMED Classical definition is an extension of the RAND classical definition beyond nine panel members, and defines for different panel sizes, the maximum number of responses that are allowed to fall outside the 3-point region that contains the median, in order to conclude agreement. The p-value method's definition of agreement is the result of a binomial hypothesis test that 80% of the ratings are within the 3-point region containing the median. Finally, the IPRAS is based on the non-parametric measure of spread, the interpercentile range (IPR= 70^{th} percentile - 30^{th} percentile). After adjusting for the lack of symmetry in the responses, if the resulting IPRAS is less than the IPR, there is disagreement. A measure was considered appropriate if the metric met the definition of appropriateness using all defined statistical methods. All statistical analyses were performed using SAS v. 9.4 (SAS Institute, Cary, NC).

REFERENCES:

1. Fitch K AM, Burnand B, et al. The RAND/UCLA appropriateness method user's manual. Santa Monica: RAND 2001.
2. Wani S, Muthusamy VR, Shaheen NJ, et al. Development of Quality Indicators for Endoscopic Eradication Therapies in Barrett's Esophagus: The TREAT-BE (Treatment With Resection and Endoscopic Ablation Techniques for Barrett's Esophagus) Consortium. Am J Gastroenterol 2017;112:1032-1048.
3. Yadlapati R, Vaezi MF, Vela MF, et al. Management options for patients with GERD and persistent symptoms on proton pump inhibitors: recommendations from an expert panel. Am J Gastroenterol 2018;113:980-986.
4. Gawron AJ, Bell R, Abu Dayyeh BK, et al. Surgical and endoscopic management options for patients with GERD based on proton pump inhibitor symptom response: recommendations from an expert U.S. panel. Gastrointest Endosc 2020;92:78-87 e2.
5. Guyatt GH, Oxman AD, Vist GE, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ 2008;336:924-6.

Appendix 1

Intake Survey

First, please answer a few questions about your practice	
Please enter your name (last, first)	_____
What is your age?	◯ 30 to 40 years ◯ 41 to 50 years ◯ 51 to 60 years ◯ 61-70 years
How many years have you been in practice?	_____
What is your institutional affiliation?	_____
What type of practice do you work in?	☐ Academic (University) ☐ Academic (VA affiliated) ☐ Community ☐ Other
Please describe other	_____
Please answer these questions about your BE practice	
Do you perform surveillance for non-dysplastic BE?	◯ Yes ◯ No
How many in a month, on average?	_____
Do you perform endoscopic eradication therapy (EET) for BE?	◯ Yes ◯ No
How many new cases of BE related neoplasia do you see in a month, on average, for EET?	_____
Do you care for patients with esophageal adenocarcinoma?	◯ Yes ◯ No
How many in a month, on average?	_____
On average, how much time would you spend inspecting a BE segment for a patient with C2M2 (minutes)	_____
On average, how much time would you spend inspecting a BE segment for a patient with C10M11 (minutes)	_____

This initiative is to standardize the terminology and definitions related to post-endoscopy esophageal adenocarcinoma (PEEC), the concept of missed HGD/EAC following endoscopy. We would like to understand your perspectives about this concept.
What do you believe is the rate of missed HGD/EAC within one year of an index endoscopy in a patient with non-dysplastic BE (NDBE)? _____
What do you believe is the rate of any missed dysplasia/EAC within one year of an index endoscopy in a patient with NDBE? _____
The definition of post-endoscopy esophageal adenocarcinoma should include the following ◯ EAC alone ◯ EAC + HGD ◯ EAC + HGD +LGD ◯ EAC + HGD + LGD + Indefinite dysplasia

Appendix 2

PEEC: ROUND 1

Consensus Statements on Post-Endoscopy Esophageal Adenocarcinoma - Round 1

Please enter your name (Last, First)

_____

Round 1 will assess the group's initial agreement for which statements regarding PEEC should be recommended, should be recommended against, or are not appropriate for a recommendation.

You will rank each statement for your perceived appropriateness and necessity in clinical practice. Ranking is on a scale of 1 to 9 where 1 means that the expected harms greatly outweigh the expected benefits, and 9 means that the expected benefits greatly outweigh the expected harms.
Generally scores of:
1, 2 or 3 = inappropriate measure;
4, 5, or 6 = equivocal/uncertain appropriateness;
7, 8, or 9 = appropriate measure.

Please use the following instructions when ranking the appropriateness of these statements:

1. The statements do not necessarily have to apply to any one specific patient, but rather, they may pertain to the overall care of BE patients.

2. A statement/measure (when applicable) is considered "valid" if adherence with this measure is critical to provide quality care to patients with BE exclusive of costs or feasibility. Do not consider cost implications or the feasibility of implementing the measure in your rankings.

3. Base your rankings on your own personal judgment, and not what you believe other experts or the panel might say.

4. Consider these measures for the average patient presenting to the average physician at an average hospital.

The ultimate purpose of this process is to develop recommendations to assist practitioners with quality improvement.

Section 1: To standardize terminology and definitions related to PEEC and interval EAC

High-grade dysplasia (HGD); Esophageal adenocarcinoma (EAC); Non-dysplastic Barrett's esophagus (NDBE); Low-grade dysplasia (LGD)

Please rank the appropriateness of the following definitions of PEEC. Note that the type of dysplasia and/or the indication for the index upper endoscopy vary in the proposed definitions.

	(Highly inappropriate) 1	2	3	4	(Equivocal) 5	6	7	8	(Highly appropriate) 9

PEEC is the preferred term for HGD or EAC detected before the next recommended surveillance endoscopy after an index upper endoscopy showed no evidence of HGD/EAC in a patient diagnosed with NDBE at screening or surveillance endoscopy	○	○	○	○	○	○	○	○	○
PEEC is the preferred term for HGD or EAC detected before the next recommended surveillance endoscopy after an index upper endoscopy showed no evidence of HGD/EAC in a patient diagnosed with NDBE at screening endoscopy	○	○	○	○	○	○	○	○	○
PEEC is the preferred term for LGD or HGD or EAC detected before the next recommended surveillance endoscopy after an index upper endoscopy showed no evidence of HGD/EAC in a patient diagnosed with NDBE at screening endoscopy	○	○	○	○	○	○	○	○	○
PEEC is the preferred term for LGD or HGD or EAC detected before the next recommended surveillance endoscopy after an index upper endoscopy showed no evidence of HGD/EAC in a patient diagnosed with NDBE at screening or surveillance endoscopy	○	○	○	○	○	○	○	○	○

Please rank the appropriateness of the following time interval for which PEEC applies.									
	(Highly inappropriate) 1	2	3	4	(Equivocal) 5	6	7	8	(Highly appropriate) 9
The time interval for which the occurrence of PEEC applies is ≤ 1 year following screening or surveillance endoscopy	○	○	○	○	○	○	○	○	○

The time interval for which the occurrence of PEEC applies is < 3 years following screening or surveillance endoscopy	○	○	○	○	○	○	○	○	○
The time interval for which the occurrence of PEEC applies is ≤ 1 year following screening endoscopy and < 3 years following surveillance endoscopy	○	○	○	○	○	○	○	○	○

Please provide comments, if any, about section 1.

_____

Section 2: To standardize potential explanations for PEEC

Please rank the appropriateness of the following explanations for PEEC

	(Highly inappropriate) 1	2	3	4	(Equivocal) 5	6	7	8	(Highly appropriate) 9
The potential explanations for PEEC include missed HGD/EAC, rapidly progressive EAC or incomplete endoscopic eradication therapy (ablation and/or resection)	○	○	○	○	○	○	○	○	○
Missed HGD/EAC contributes to the majority of PEEC cases	○	○	○	○	○	○	○	○	○

Please provide comments, if any, about section 2.

_____

Section 3: To provide best-practice advice on reducing rates of PEEC in clinical practice

Please rank the appropriateness of the following statements in terms of their impact on reducing PEEC

	(Highly inappropriate) 1	2	3	4	(Equivocal) 5	6	7	8	(Highly appropriate) 9

Endoscopists should define the extent of BE using a standardized grading system documenting the circumferential and maximal extent of the columnar lined esophagus (Prague classification) with a clear description of landmarks and all (including subtle) visible lesions using the Paris classification	○	○	○	○	○	○	○	○	○
Screening and surveillance endoscopy for BE should be performed using high-definition white light endoscopy (HD-WLE) and chromoendoscopy (traditional or virtual)	○	○	○	○	○	○	○	○	○
Endoscopists should spend adequate time inspecting the BE segment at a suggested rate of 1 minute per centimeter of circumferential BE	○	○	○	○	○	○	○	○	○
In patients undergoing screening or surveillance endoscopy for BE, endoscopists should obtain biopsies using the Seattle biopsy protocol	○	○	○	○	○	○	○	○	○
Endoscopists should use advanced sampling techniques such as wide-area transepithelial sampling (WATS-3D) as an adjunct to the Seattle biopsy protocol to augment detection of neoplasia during screening and surveillance endoscopy for BE	○	○	○	○	○	○	○	○	○

Please provide comments, if any, about section 3.

_____

Section 4: To develop PEEC as a performance measure (definition of PEEC rate, standardized methodology for calculation) to ultimately allow for benchmarking and comparison between services

The ultimate purpose of these statements is to assist practitioners with quality improvement. All measures are intended to be calculated and reported at the practice level and need not have a direct benefit to an individual patient. Practices will be able to compare themselves to one another in hopes that practices will feed back their own data to institute quality improvement initiatives when appropriate.

Please rank the appropriateness of the following statements									
	(Highly inappropriate) 1	2	3	4	(Equivocal) 5	6	7	8	(Highly appropriate) 9
Endoscopy services should implement a formal process to identify and register PEEC cases to review for potential causative factors and perform a root-cause analysis	○	○	○	○	○	○	○	○	○
In order to benchmark endoscopy services, the quality of index endoscopy in BE can be measured by calculating unadjusted PEEC proportion (number of PEEC cases divided by the number of PEEC cases plus detected HGD/EAC cases)	○	○	○	○	○	○	○	○	○
PEEC proportion should be initially reported at a service level rather than individual endoscopist level	○	○	○	○	○	○	○	○	○
In order to benchmark endoscopy services, the quality of index endoscopy in BE can be measured by calculating neoplasia detection rate (NDR, rate of detection of HGD/EAC) in BE patients undergoing screening endoscopy	○	○	○	○	○	○	○	○	○
In order to benchmark endoscopy services, the quality of endoscopy in BE can be measured by calculating neoplasia detection rate (NDR, rate of detection of HGD/EAC) in BE patients undergoing screening or surveillance endoscopy	○	○	○	○	○	○	○	○	○
Endoscopists should complete validated training courses that highlight best practices in performing endoscopy in BE patients (e.g. BORN project, CORE-BE)	○	○	○	○	○	○	○	○	○

Appendix 3

Ranking instructions provided to all panel members

Dear Colleagues,

Thank you for participating in our project titled "**Consensus Statements on Post-Endoscopy Esophageal Adenocarcinoma."**

The following link will take you to an online list of proposed statements. You will be prompted to rate the proposed statements according to your perceived appropriateness and necessity in clinical practice. Please use the following instructions when ranking the appropriateness of these statements (with the most important considerations in bold):

1. The statements do not necessarily have to apply to any one specific patient, but rather, they may pertain to the overall care of BE patients.

2. A statement/measure (when applicable) is considered "valid" if adherence with this measure is critical to provide quality care to patients with BE **exclusive of costs or feasibility. Do not consider cost implications or the feasibility of implementing the measure in your rankings.**

3. Base your rankings on your own personal judgment, and not what you believe other experts or the panel might say.

4. Consider these measures for the average patient presenting to the average physician at an average hospital.

5. The purpose of these statements is to assist practitioners with quality improvement. All measures are intended to be calculated and reported at the practice level and need not have a direct benefit to an individual patient. Practices will be able to compare themselves to one another in hopes that practices will feed back their own data to institute quality improvement initiatives when appropriate.

6. Please complete the online questionnaire by **xx**.

7. We will contact you to revise your rankings if the rankings do not follow the instructions above. If you feel a measure is unreasonable, not useful, or dangerous, please rank it a 1 instead of leaving it blank.

8. Once all of your responses are received, we will analyze the rankings and presentthe blinded, aggregate results prior to and during our Round 2 meeting.

Thank you and please feel free to contact us with any questions.

Appendix 4

PEEC: ROUND 2

Round 2 Voting

Please enter your name (Last, First)

_____

Round 2 will assess the group's agreement for which statements should be recommended following our recent group meeting.

You will rank each statement for your perceived appropriateness and necessity in clinical practice. Ranking is on a scale of 1 to 9 where 1 means that the expected harms greatly outweigh the expected benefits, and 9 means that the expected benefits greatly outweigh the expected harms.
Generally scores of:
1, 2 or 3 = inappropriate measure;
4, 5, or 6 = equivocal/uncertain appropriateness;
7, 8, or 9 = appropriate measure.

Please use the following instructions when ranking the appropriateness of these statements:

1. The statements do not necessarily have to apply to any one specific patient, but rather, they may pertain to the overall care of BE patients.

2. A statement/measure (when applicable) is considered "valid" if adherence with this measure is critical to provide quality care to patients with BE exclusive of costs or feasibility. Do not consider cost implications or the feasibility of implementing the measure in your rankings.

3. Base your rankings on your own personal judgment, and not what you believe other experts or the panel might say.

4. Consider these measures for the average patient presenting to the average physician at an average hospital.

The ultimate purpose of this process is to develop recommendations to assist practitioners with quality improvement.

Section 1: To standardize terminology and definitions related to PEEC/PEEN

Discussion points in the manuscript will clarify that the patient population has non-dysplastic BE (NDBE) without any concerning features for neoplasia which could include erosive esophagitis, indeterminate for dysplasia, low-grade dysplasia (LGD), high-grade dysplasia (HGD), or esophageal adenocarcinoma (EAC).

Discussion will also provide details on why 6 months was selected to allow for patients to be brought back to check for healing of erosive esophagitis or to schedule patients for resection of visible lesions or appropriate staging of suspected malignancy.

	(Highly inappropriate) 1	2	3	4	(Equivocal) 5	6	7	8	(Highly appropriate) 9

Post-endoscopy esophageal adenocarcinoma (PEEC) is the preferred term for esophageal adenocarcinoma (EAC) detected before the next recommended surveillance endoscopy in a patient with non-dysplastic Barrett's esophagus (NDBE)	○	○	○	○	○	○	○	○	○
The time interval for which the occurrence of PEEC applies is between 6 months and 3 years following screening or surveillance endoscopy	○	○	○	○	○	○	○	○	○
Post-endoscopy esophageal neoplasia (PEEN) is the preferred term for high-grade dysplasia (HGD) or esophageal adenocarcinoma (EAC) detected before the next recommended surveillance endoscopy in a patient with non-dysplastic Barrett's esophagus (NDBE)	○	○	○	○	○	○	○	○	○
The time interval for which the occurrence of PEEN applies is between 6 months and 3 years following screening or surveillance endoscopy	○	○	○	○	○	○	○	○	○

Please provide comments, if any, about section 1.

_____

Section 2: To standardize potential explanations for PEEC/PEEN

Please rank the appropriateness of the following explanations for PEEC/PEEN

Discussion section will clarify some of the best practices in endoscopic eradication therapy (EET) and why incomplete EET was not included in this statement.

	(Highly inappropriate) 1	2	3	4	(Equivocal) 5	6	7	8	(Highly appropriate) 9

The potential explanations for PEEC/PEEN include missed HGD/EAC and rapidly progressive EAC	○	○	○	○	○	○	○	○	○

Please provide comments, if any, about section 2.

_____

Section 3: To provide best-practice advice on reducing rates of PEEC/PEEN in clinical practice

Please rank the appropriateness of the following statements in terms of their impact on reducing PEEC/PEEN

The discussion section will include reference to tools for grading systems and goals of grading. The discussion section will also include a description of what constitutes a high-quality endoscopic examination.

	(Highly inappropriate) 1	2	3	4	(Equivocal) 5	6	7	8	(Highly appropriate) 9
Endoscopists should define the extent of BE using a standardized grading system documenting the circumferential and maximal extent of the columnar lined esophagus (Prague classification) with a clear description of landmarks and characteristics of visible lesions, when present.	○	○	○	○	○	○	○	○	○
Screening and surveillance endoscopy for BE should be performed using high-definition white light endoscopy (HD-WLE) and chromoendoscopy (traditional or virtual).	○	○	○	○	○	○	○	○	○
Endoscopists should spend adequate time inspecting the BE segment	○	○	○	○	○	○	○	○	○

In patients undergoing screening or surveillance endoscopy for BE, endoscopists should obtain biopsies using the Seattle biopsy protocol (4-quadrant biopsies every 2 cm and obtaining target biopsies or resection or outlining a plan for resection for any visible lesions).	○	○	○	○	○	○	○	○	○
Endoscopists should use advanced sampling techniques such as wide-area transepithelial sampling (WATS-3D) as an adjunct to the Seattle biopsy protocol to augment detection of neoplasia during screening and surveillance endoscopy for BE.	○	○	○	○	○	○	○	○	○

Please provide comments, if any, about section 3.

Section 4: To develop PEEC/PEEN as a performance measure to ultimately allow for benchmarking and comparison between services

	(Highly inappropriate) 1	2	3	4	(Equivocal) 5	6	7	8	(Highly appropriate) 9
Endoscopy practices can consider reviewing PEEC cases to understand contributing factors and areas of improvement.	○	○	○	○	○	○	○	○	○

Based on the discussion during the meeting we propose this new statement:

To facilitate the use of a common language when categorizing PEEC/PEEN cases according to their most plausible explanations, we suggest that the following categories be used:

a. Possible missed visible lesion, prior examination adequate*
b. Possible missed visible lesion, prior examination inadequate
c. Detected visible lesion, no or inadequate sampling with target biopsies
d. Detected visible lesion, incomplete resection of previously identified lesion
e. Prior examination adequate and clinically indicated follow-up not recommended
f. Prior examination inadequate and clinically indicated follow-up not recommended
g. Prior examination adequate and failure of patient to follow-up on a recommended surveillance endoscopy interval

***Adequate examination will be defined by at least documenting landmarks, use of HD-WLE and chromoendoscopy (traditional or virtual) and sampling using Seattle biopsy protocol**

	(Highly inappropriate) 1	2	3	4	(Equivocal) 5	6	7	8	(Highly appropriate) 9
Please rank the appropriateness of this new statement	○	○	○	○	○	○	○	○	○

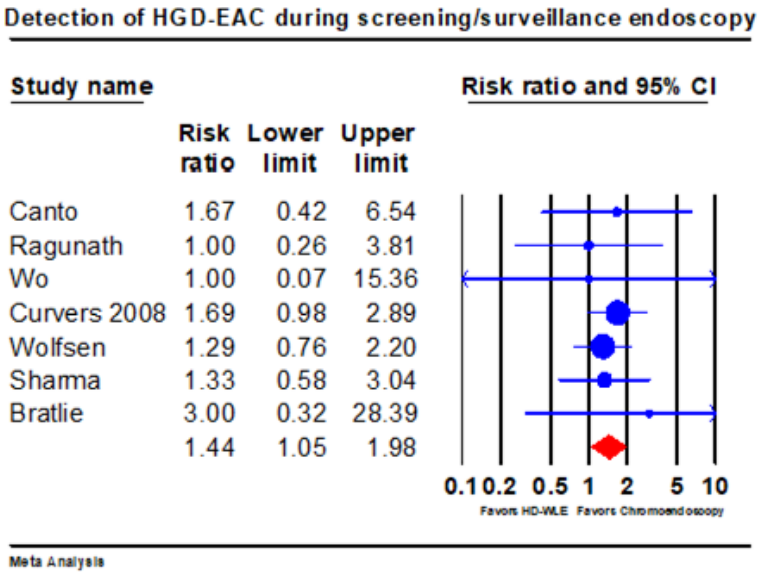
Section 5: Future Directions and Research:

The following statements will be moved to a future directions and research section and will not be voted on for appropriateness. Please indicate whether you agree or disagree with inclusion of these topics for future directions/research.

In order to benchmark endoscopy services, the quality of index endoscopy in BE can be measured by calculating unadjusted PEEN proportion (number of PEEN cases divided by the number of PEEN cases plus detected HGD/EAC cases). Unadjusted PEEC proportion = number of PEEN cases / total number of HGD/EAC cases	○ Agree ○ Disagree
PEEC/PEEN proportion should be initially reported at a service level rather than individual endoscopist level.	○ Agree ○ Disagree
In order to benchmark endoscopy services, the quality of index endoscopy in BE can be measured by calculating neoplasia detection rate (NDR, rate of detection of HGD/EAC) in BE patients undergoing screening endoscopy.	○ Agree ○ Disagree
In order to benchmark endoscopy services, the quality of endoscopy in BE can be measured by calculating neoplasia detection rate (NDR, rate of detection of HGD/EAC) in BE patients undergoing screening or surveillance endoscopy.	○ Agree ○ Disagree
Endoscopists should complete validated training courses that highlight best practices in performing endoscopy in BE patients (e.g. BORN project, CORE-BE, www.iwgco.net).	○ Agree ○ Disagree
The panel will propose a checklist for manuscripts that address PEEC/PEEN as a study outcome (adapted from Rutter M, Beintaris I et al, Gastroenterology 2018).	○ Agree ○ Disagree
Please provide comments, if any, about section 5.	_____

Appendix 5

Forest plot of studies included in the analysis comparing high-definition white light endoscopy plus chromoendoscopy with high-definition white light endoscopy alone in the detection of Barrett's associated high-grade dysplasia and esophageal adenocarcinoma



Appendix 6

Responses related to statements for future research

Statement	n (%)
In order to benchmark endoscopy services, the quality of index endoscopy in BE can be measured by calculating unadjusted PEEN proportion (number of PEEN cases divided by the number of PEEN cases plus detected HGD/EAC cases)	
Agree	17 (68)
Disagree	8 (32)
PEEN/PEEC proportion should be initially reported at a service level rather than individual endoscopist level	
Agree	18 (72)
Disagree	7 (28)
In order to benchmark endoscopy services, the quality of index endoscopy in BE can be measured by calculating neoplasia detection rate (NDR, rate of detection of HGD/EAC) in BE patients undergoing screening endoscopy	
Agree	17 (68)
Disagree	8 (32)
In order to benchmark endoscopy services, the quality of endoscopy in BE can be measured by calculating neoplasia detection rate (NDR, rate of detection of HGD/EAC) in BE patients undergoing screening or surveillance endoscopy	
Agree	16 (64)
Disagree	9 (36)

Endoscopists should complete validated training courses that highlight best practices in performing endoscopy in BE patients (e.g. BORN project, CORE-BE, www.iwgco.net) Agree Disagree	 24 (96) 1 (4)
The panel will propose a checklist for manuscripts that addresses PEEN/PEEC as a study outcome (adapted from Rutter M, Beintaris I et al., Gastroenterology 2018) Agree Disagree	 23 (92) 2 (8)

Appendix 7

Establishing an infrastructure for future Post-Endoscopy Esophageal Neoplasia (PEEN) and Post-Endoscopy Esophageal Adenocarcinoma (PEEC) research: proposed checklist for manuscripts that address PEEN/PEEC as a study outcome

(adapted from Rutter M, Beintaris I et al, Gastroenterology 2018).

	Recommendation
Title and Abstract	The project's design is indicated with the use of the term PEEN/PEEC in the title or the abstract The abstract includes a balanced summary of the results
Introduction and Study Aims	The scientific background and rationale for the project is described A clear description of the study hypothesis and study aims is included
Methods **Study Design**	Details regarding the study design are explicitly stated in the manuscript (observational study – retrospective, prospective; systematic review/meta-analysis)
Setting	Setting, locations and relevant timeframes including periods of recruitment/analysis, follow-up, and data collection are described. Analysis setting (e.g. endoscopist or service/unit/s level) is clearly defined
Participants	Description of eligibility criteria and sources and methods of selection of participants are given. Methods of follow-up are described. The rationale for patients in different categories is included.
Definitions and Variables	Definitions used to categorize patients should be included. For e.g. the definition used to categorize a patient with prevalent, incident EAC and PEEN/PEEC should be included.

	Variables collected including patient and endoscopy characteristics along with data source should be included
Study size, statistical analysis and sensitivity analysis	The rationale for sample/study size should be provided Statistical methods are described. Methods to examine subgroups, interactions and predictors should be included. How missing data, if applicable, were addressed should be provided.
Results	
Participants and descriptive data	Overall number of participants included in the analysis and those with study outcomes (PEEN/PEEC and others such as prevalent or incident HGD/EAC). A flow diagram is included. Descriptive data on characteristics of included and excluded participants Follow-up, if applicable, is included (e.g. mean, median, total amount)
Outcome data	PEEN/PEEC rates (adjusted and unadjusted with 95% confidence intervals)
Other analyses	Other analyses completed such as predictors of PEEN/PEEC, sensitivity analyses are reported
Discussion **Main results and interpretation**	Main results with results to study aims are summarized and placed in context with other published studies
Limitations	Description of potential bias, imprecision and generalizability
Funding source	Source of funding should be included