An Interactive Web-based Educational Tool Improves Detection and Delineation of Barrett's Esophagus Related Neoplasia

J.J. Bergman[1*], A.J. de Groof[1*], O. Pech[2], K. Ragunath[3], D. Armstrong[4], N. Mostafavi[5], L. Lundell[6], J. Dent[7], M. Vieth[8], G.N. Tytgat[1], P. Sharma[9], on behalf of the International Working Group for Classification of Oesophagitis (IWGCO).

[1] Department of Gastroenterology and Hepatology, Amsterdam UMC, University of Amsterdam, Amsterdam, the Netherlands.
[2] Gastroenterology and interventional Endoscopy, Krankenhaus Barmherzige Brüder, Regensburg, Germany.
[3] Nottingham Digestive Diseases Centre, University of Nottingham and NIHR Nottingham BRC, Nottingham University Hospitals NHS Trust, Nottingham, United Kingdom
[4] Division of Gastroenterology, McMaster University Medical Centre, Ontario, Canada
[5] Biostatistical unit, department of Gastroenterology and Hepatology, Amsterdam UMC, University of Amsterdam, Amsterdam, the Netherlands.
[6] Department of Surgery, CLINTEC, Karolinska Institutet, Stockholm, Sweden
[7] Department of Medicine, University of Adelaide and Royal Adelaide Hospital, Adelaide, South Australia
[8] Institute of Pathology, Otto-von-Guericke University, Magdeburg, Germany
[9] Department of Veterans Affairs Medical Center, University of Kansas School of Medicine, Kansas City, United States

*Authors contributed equally to this manuscript and share first authorship.

All authors declared to have no disclosures relevant to this manuscript.

Corresponding author:
J.J. Bergman, MD PhD
Professor of Gastrointestinal Endoscopy
Director of Endoscopy
Academic Medical Center Amsterdam
Meibergdreef 9
1105 AZ Amsterdam
The Netherlands
Phone number: +31 (0)20 5663556
Fax number: +31 (0)20 6917033

This study was supported by unrestricted grants from AstraZeneca and Medtronic, and by an educational activity grant from the United European Gastroenterology (UEG), which had no involvement in the design, recruitment, data collection, analysis or interpretation or writing of the manuscript.

Author contribution:

Study design: J.J. Bergman, O. Pech, K. Ragunath, D. Armstrong, L. Lundell, J. Dent, M. Vieth, G.N. Tytgat, P. Sharma;

Patient recruitment and video collection: J.J. Bergman, O. Pech, P. Sharma;

Video review: J.J. Bergman, O. Pech, K. Ragunath, G. Tytgat, P. Sharma;

Video delineation: J.J. Bergman, O. Pech, K. Ragunath, P. Sharma;

Data analysis and interpretation: J.J. Bergman, A.J. de Groof, N. Mostafavi

Writing of manuscript: J.J. Bergman, A.J. de Groof, N. Mostafavi, J. Dent, O. Pech, K. Ragunath, D. Armstrong, P. Sharma;

All authors reviewed the final manuscript for important intellectual content and agreed to submit.

ABSTRACT

Background & Aims: Endoscopic detection of early Barrett's esophagus-related neoplasia (BORN) is a challenge. We aimed to develop a web-based teaching tool for improving detection and delineation of BORN.

Methods: We made high-definition digital videos during endoscopies of patients with BORN and non-dysplastic Barrett's esophagus (NDBE). Three experts superimposed their delineations of BORN lesions on the videos using special tools. In phase 1, 68 general endoscopists from 4 countries assessed 4 batches of 20 videos. After each batch, mandatory feedback compared assessors interpretations with those from experts . These data informed selection of 25 videos for the phase 2 module, which was completed by 121 new assessors from 5 countries. A 5-video test batch was completed before and after scoring of the four 5-video training batches. Mandatory feedback was as in phase 1. Outcome measures were scores for detection, delineation, agreement delineation, and relative delineation of BORN.

Results: A linear mixed-effect model showed significant sequential improvement for all 4 outcomes over successive training batches in both phases. In phase 2, median detection rates of BORN in the test batch increased by 30% (P<.001) after training. From baseline to the end of the study, there were relative increases in scores of 46% for detection, 129% for delineation, 105% for agreement delineation, and 106% for relative delineation (all P<.001). Scores improved independent of assessors' country of origin or level of endoscopic experience.

Conclusions: We developed a web-based teaching tool for endoscopic recognition of BORN that is easily accessible, efficient, and increases detection and delineation of neoplastic lesions. Widespread use of this tool might improve management of Barrett's esophagus by general endoscopists.

KEY WORDS: The BORN Project, esophageal adenocarcinoma, Barrett Esophagus, Endoscopy

INTRODUCTION

Barrett's esophagus (BE) patients undergo regular endoscopic surveillance to detect curable lesions which have high risk for developing into invasive esophageal adenocarcinoma (EAC). We have labeled these lesions Barrett's Esophagus Related Neoplasia (BORN), which consist of both high-grade dysplasia (HGD) and/or EAC. The reliable endoscopic detection of BORN is, however, difficult because the endoscopic appearances of early lesions are often subtle. Overall, progression to neoplasm is relatively rare in BE (<1% annually)[1], so that general endoscopists performing BE surveillance encounter early BORN lesions infrequently, which limits their familiarity with their endoscopic appearances.

Studies have evaluated whether specialized endoscopic imaging techniques, such as optical chromoscopy or magnification, may improve the endoscopic detection and delineation of BORN by general endoscopists; outcomes have generally been disappointing. Because of this, all current international guidelines recommend high definition white light endoscopy (HD-WLE) as the best surveillance technique.[2, 3]

Most studies suggest that endoscopists who have referral practices that specialize in BE management detect BORN lesions more reliably with HD-WLE than general endoscopists.[4, 5] It therefore seems likely that the endoscopic detection and delineation of BORN by general endoscopists can be improved by the use of training tools that enhance the endoscopic recognition of BORN with HD-WLE endoscopy. The potential benefit from effective training on use of HD-WLE appears to exceed any possible gains from widespread (and expensive) use of currently available specialized imaging technologies outside of specialist BE centers.

Over the past two decades, the International Working Group for the Classification of Oesophagitis (IWGCO) has been engaged in research and educational activities to improve endoscopic assessment of gastro-oesophageal reflux disease and BE. The IWGCO has developed and validated the Los Angeles Classification of reflux esophagitis (1996) and the Prague C&M criteria for BE (2004).[6-9] More recently, the IWGCO has been working on an interactive web-based teaching tool to improve the endoscopic recognition of BORN. Herein we describe the step-wise development and validation of the BORN teaching tool.

METHODS

Setting and design

For the BORN project, a subgroup of the IWGCO was formed consisting of the authors of this manuscript. Members of this subgroup convened multiple times a year between 2005-2015. The first step was the development of an endoscopic pullback imaging protocol tailored to the needs of this project during several meetings.

Endoscopic video recordings

High quality videos containing BORN and non-dysplastic BE (NDBE) were then prospectively collected with the standardized endoscopic pullback procedure which was illustrated by an instruction video to ensure that videos were recorded as follows: first the BE segment was washed thoroughly, then the video recording was started in the proximal stomach, approximately 1 centimeter distal to the diaphragmatic pinch. The endoscope was then pulled back slowly with its position centered in the esophagus, pausing for several seconds every 1-2 centimeters. During the pullback the esophagus was kept inflated, thereby providing an overview of the entire circumference of the BE segment. The pullback was continued for 3 centimeters proximal to the upper extent of the BE segment. Thus, the video recorded the appearances of the entire BE segment without giving any clues about possible BORN lesions. Videos were made in three tertiary referral centers (Amsterdam, Wiesbaden, Kansas City) by expert BE endoscopists (JB, OP, and PS) using HD-WLE. All video output was digitally recorded direct onto the hard-drive of a computer equipped with special software, thus preventing any loss of resolution in the original recording and in subsequent copying and transmission via the internet to the many endoscopists involved in this project.

Patient selection

Patients were enrolled from those referred for surveillance of known BE or for the first endoscopic treatment of proven BORN (HGD or EAC). Figure 1 summarizes the video selection and review processes. To ensure that BORN videos used for training contained an early, endoscopically curable neoplastic lesion, the following criteria were required: (a) at least one prior biopsy-based diagnosis of HGD or EAC confirmed by a 2nd pathologist; (b) HD-WLE showed the presence of a lesion that was subsequently resected by EMR; (c) confirmation of presence of HGD and/or EAC in the EMR specimen by a pathologist expert in BE and (d) no signs of deep submucosal infiltration (>T1sm1) in the resection specimens.

The following criteria ensured that NDBE videos did not contain visible neoplastic lesions: (a) no low grade dysplasia (LGD) or HGD diagnosis in at least two prior surveillance endoscopies; (b) in addition

to (a), all biopsies obtained during the endoscopy at which the video recording was made were also free of LGD and HGD and (c) close examination of the video by 2 endoscopists expert in BE did not reveal any changes suggestive of BORN (see below).

Review of videos by IWGCO members for inclusion in the draft phase 1 module

The videos that met the above requirements were uploaded to a secure server, with their standardized case record forms (CRFs) for review by expert endoscopist IWGCO members. Factors evaluated were the need for extensive video editing, image quality of the recording, and satisfactory technical quality of the pullback. This was judged by the level of air inflation, amount of residual mucus and fluids, any major adverse effects of motility on image quality and adequate demonstration of the diaphragmatic pinch, gastroesophageal junction and circumferential and maximum BE extent. In the case of videos from BORN patients, the adequacy of imaging of neoplastic lesions was additionally assessed. Videos considered of inadequate quality by any reviewer for any of these items were excluded. NDBE cases were evaluated by OP and KR and the BORN videos by GT and KR, neither of whom had made the original recordings.

Delineation of BORN lesions by IWGCO: After culling of unsuitable videos by OP, KR and GT, the lesions in all remaining BORN videos were delineated online on selected image frames (see below) by at least three of four IWGCO BE endoscopic experts (JB, OP, KR and PS) using the online software module (Meducati AB, Göteborg, Sweden, www.meducati.com) developed specifically for this project. The aim of this process was to create a ground truth for videos considered potentially suitable for inclusion in the phase 1 module.

First, the endoscopist who had recorded the BORN video (the content provider) marked the time during which the BORN lesion was visible. Within this time span, one video frame was selected for every second as having best image quality. The content provider and two of the remaining three experts then delineated the lesion on each of the selected second-by second image frames, without consultation with the other experts. The three partially overlapping delineations for all selected frames were then superimposed on the image (Figures 2 and 3). The “sweet spot” was the area delineated by all three experts (expert 1 AND expert 2 AND expert 3): this was defined as the most easily recognized part of the BORN lesion.

Comparisons were made for delineations of each selected BORN video frame according to 3 pairs of experts (expert1/expert2; expert 1/expert3; expert2/expert3). For each pair of delineations on each of the selected images, an ‘AND/OR’ ratio was calculated by dividing the area where the 2 delineations overlapped by the total area of the 2 delineations. The ‘AND/OR’ ratio was taken as the

measure of level of agreement on delineation between two assessors. An 'AND/OR' ratio of <25% was taken as inadequate agreement for that video frame. For the time span during which the BORN lesion was visible, the video was judged suitable for inclusion in the phase 1 materials if the 'AND/OR ratio' was ≥25% on at least three quarters of all selected second-by-second video frames. For all such videos, each of the remaining video frames with an 'AND/OR ratio' of <25% were re-delineated by expert 2 and expert 3. These re-delineations were used in the final version of that video, as expert 1 (the content provider) was considered 'correct' since he had the most extensive knowledge of the lesion.

For videos with inadequate agreement, defined as an 'AND/OR' ratio ≥25% in less than three-quarters of video frames, a fourth expert delineated the lesion on all of the selected second-by-second video frames, to replace the 'worst' expert (the expert with the lowest mean 'AND/OR' score; either expert 2 or expert 3). If there was still no resolution, the video was reviewed during a face-to-face consensus meeting between experts. After reviewing and discussing a video, the experts repeated their delineations independently, after which either the agreement threshold was reached or the video was not included in the phase 1 module.

The above process led to selection of 48 delineated BORN videos and 32 NDBE videos.

IWGCO development of the phase 1 draft training module:

This phase 1 'draft' training module included 4 batches of 20 videos each, containing 11-12 BORN videos and 8-9 NDBE videos. To ensure that every batch was equal in terms of difficulty, the BORN videos were ranked in 4 classes of difficulty, based on their mean IWGCO expert 'AND/OR' scores (the lower the mean 'AND/OR' score, the more difficult the video). Sorting the videos into batches of equal difficulty was important because it ensured that differences of performance among different training batches were a true reflection of learning (or its lack), rather than due to differences in difficulty among batches. The randomization process was driven by computer-generated numbers, and maintained a comparable mix of difficulty among batches: within this constraint though, individual videos were randomized to different batches among different assessors, so that a particular video was evaluated in a different order from batches 1 to 4 by different assessors. Thus, the composition of batches of comparable difficulty varied for different assessors. The randomization of a particular video to appear in different places across the training materials made it possible to determine whether its assessment improved as the training proceeded, indicative of a learning effect.

Two IWGCO members finally checked the software and the general functioning of this training module before general endoscopist assessments started.

Evaluation of the phase 1 draft training module by general endoscopist assessors

The suitability of individual videos for training general endoscopists was assessed from scoring of the draft module by 68 general endoscopist assessors from 4 countries. Assessors were grouped according to three categories of experience (Table 1). None of the assessors were considered to be experts in BE management, on the basis of assessors' responses to the question on whether their practice had a focus on this and a review of the list of assessor names by IWGCO board members. The general endoscopist assessors were provided with a personal login account to access the online training module. They had to review an instruction video and complete a questionnaire, after which they were granted access to their first batch of videos in the module.

Assessors were asked to review the *entire* video, typically less than 90 seconds and then indicate if a BORN lesion was present. They could pause, rewind and forward the video before making this judgement. If BORN was diagnosed, assessors selected the frame in which they felt the lesion was clearest. Then, if the BORN diagnosis was correct, a fade-through-black function automatically shifted to the nearest image frame (free of any BE expert assessments) that had been selected by the experts as the best image of the lesion within that second. Assessors were required to mark their preferred biopsy spot (Figure 3) and then to delineate the entire lesion with the software tools. The video was then locked to prevent repeat assessments.

When the assessor selected a frame from a NDBE video or from a BORN video outside the time that the actual BORN lesion was visible, after the fade-though-black function, the frame that had been selected by the assessor was redisplayed, since there were no delineated images available for that time period chosen by the assessor.

Assessors had to finish each video in a single session and complete each batch of 20 videos within 2 weeks. After completion of each batch, assessors were guided through a mandatory, tailored feedback session on all 20 videos in that batch. This feedback allowed re-run of all videos, review of all scores (see endpoints below) and comparison of their delineations and scores with those of the experts for the same image that they had delineated. Also, assessors could 'follow the experts', by viewing all other selected second-by-second video frames with BORN lesions delineated by the experts. For each of these frames, the assessor could add or remove any of the 3 experts' delineations to better reassess the image for extent of the lesion. Only after completion of this mandatory feedback session, did the software allow progression to the next batch.

Outcome measurements for the phase 1 draft training module

The following 4 primary outcome measurements were evaluated.

Detection score: Division of the number of correctly identified NDBE videos plus the number of BORN videos in which the biopsy was positioned within the sweet spot, by the total number of NDBE and BORN videos

BORN delineation score: Percentage of the sweet spot of expert delineations marked by the assessor.

BORN agreement delineation score*:* The mean 'AND/OR' score of the assessors' delineation with the individual expert delineation of the three experts was calculated, using the methodology described above. This mean 'AND/OR' score was considered as the 'agreement delineation score' of the assessor.

BORN relative delineation score: The agreement delineation score of the assessor divided by the mean agreement delineation scores amongst experts, thereby correcting for disagreement among experts.

Development of the phase 2 condensed training module by the IWGCO

The data from phase 1, and written feedback from assessors were reviewed by IWGCO members: these inputs led to the adjustment of the training module in several ways to improve learning efficiency, as described in Results.

The 20 BORN videos used in the training batches of the condensed version of the training module were already classified into 4 levels of difficulty, as described above for the phase 1 draft module. The order of appearance and assessment of each video was again randomized across the batches delivered to different assessors in the same way as described for phase 1.

Assessment and validation of the phase 2 condensed training module by general endoscopists

The phase 2 evaluation of the condensed training module was then carried out using the software tools and processes outlined for phase 1, with a new group of 121 general endoscopists from the Netherlands, United Kingdom, Germany, United States, and Canada, classed in the same three categories of experience (Table 1). Assessors were not aware that all videos contained early BORN lesions.

The outcome measures were as for phase 1, except that in phase 2, the detection score was defined as the number of BORN lesions correctly identified by positioning the biopsy in the sweet spot, since there were no NDBE videos.

The phase 2 training module started with a "test" batch of five videos, all of which contained BORN, to evaluate the performance of the assessor in scoring absence/presence and position of BORN lesions prior to the start of the training batches. Importantly, no feedback was given at this stage on the assessor's interpretations of these test videos.

Assessors then evaluated the four training batches in the same way as in the phase 1 evaluations, including mandatory, tailored, structured feedback after completion of each batch. None of the videos in the test batch were included in the training batches. After completion of the fourth training batch, including its feedback session, the module required the assessor to repeat the evaluation of the test batch for the second time, to provide 'before and after training' measures of assessor performance for recognition and delineation of BORN lesions.

Statistical analysis

Statistical analysis was performed using SPSS Statistical software package for Windows (version 24, SPSS Inc., Chicago) and R version 3.4.0. Since this was the first study to develop and validate a large endoscopic training program in this field, no formal sample size calculation was feasible.
For descriptive statistics, normally distributed data were shown as mean (± standard deviation) and variables with skewed distribution were shown as median (interquartile range [IQR]). To test differences in outcome parameters, paired T-tests, Wilcoxon Tests and Wilcoxon Signed-rank tests were performed. Linear mixed-effect models were performed for each outcome parameter to assess learning effects over the training batches. A random intercept was set for each subject to capture the correlation among measurements within the same subject. To control for potential confounding, models were adjusted for the effect of country of origin and endoscopic experience.

RESULTS

Assessment of the phase 1 draft training module

A total of 68 assessors from the Netherlands (46), United Kingdom (8), Germany (10), and United States (4) completed the draft training module in full. They were classed in three categories of endoscopic experience: trainees (fellows in training), junior general gastroenterologists (board certified ≤2 years of practice) and senior general gastroenterologists (≥5 years in practice) (Table 1).

Outcomes: Scores for median detection, delineation, agreement delineation, and relative delineation showed a gradual improvement over the 4 batches (Table 2, and Figure 4 in supplementary materials). In a linear mixed-effect model to assess this trend, batch number was an independent statistically significant factor associated with an increase in all of the four performance measures, thus indicating a learning effect across all 4 batches. (Table 3, and Figure 4 in supplementary materials).

The relative improvement in scores between batch 1 and batch 4 was 21% for detection (95% CI 6-40, P=0.01), 64% for delineation (95% CI 36-101, P<0.001), 55% for agreement delineation (95% CI 27-89, P<0.001), and 55% for relative delineation (95% CI 29-93, P<0.001).

Development of the final condensed phase 2 module from the draft module

All NDBE videos (n=32) and twenty-three BORN videos were removed, since these contributed little or nothing to the learning process in phase 1, on the basis of minimal or no improvement of their relative delineation scores from batches 1 to 4. Removed BORN videos included those with a baseline relative delineation score >85% indicating these were “too obvious” and those with a relative delineation score <25% across all batches which demonstrated these were “too difficult”. A post-hoc analysis showed better median performance scores after exclusion of these videos (data not shown).

The final version of the phase 2 final condensed training module therefore consisted of the twice-assessed test batch (see Methods) and four training batches, each of 5 videos, making a total 25 BORN videos.

Assessment and validation of the final condensed phase 2 training module

The phase 2 module was completed in full by a new group of 121 general endoscopic assessors from the Netherlands (33), United Kingdom (28), Germany (20), United States (21), and Canada (19), who were classed in three experience categories (see above and Table 1). None had been involved in phase 1 or had a highly developed special focus on management of BE (see Methods).

Outcomes: There were sequential improvements in the scores for detection, delineation, agreement delineation, and relative delineation of BORN in videos from the first to the fourth training batches. The improvements of all measures were significant and superior to those in phase 1, with a relative score increase of 46% (95% CI 33-50, p<0.001) for detection, 129% (95% CI 106-160, P<0.001) for delineation, 105% (95% CI 83-130, p<0.001) for agreement delineation, and 106% (95% CI 85-132, p<0.001) for relative delineation (Table 4, and Figure 5 in supplementary materials).

Batch number was the only significant predictive factor for score improvement in the linear mixed effect model and, as in phase 1, the learning effect was independent of endoscopic expertise and country of origin (Table 3).

Comparison of the test batch assessments before and then after completion of the four training batches showed significant increases for all performance scores (Table 4). Most notably, the median detection score rose by 30% (p<0.001).

DISCUSSION

This paper first describes the demanding processes involved in the development of an educational module for the endoscopic diagnosis of early BORN. The module was designed for automated on-line delivery of high-definition videos, their scoring, and provision of feedback on the endoscopic judgements made by training participants. Secondly, we report on validation studies that led to refinement of the phase 1 draft training program to the final phase 2 condensed educational module. This module has been shown to substantially improve the recognition and delineation of early BORN lesions by general endoscopists, regardless of their level of experience. This outcome is consistent with earlier studies that have concluded that early BORN lesions are detected less reliably by endoscopists who practice outside highly specialized BE referral centers.[4]

The condensed, final phase 2 BORN training module is now ready for widespread use. It is CME-accredited and available at no cost via either at www.iwgco.net, www.ueg.eu, or www.best-academia.eu. To our knowledge, this is the first validated online, interactive endoscopic training program in our field. Crucial to the project was its use of highresolution endoscopic video recordings, with replay and methods of delivery to assessors that fully preserved the resolution of the live images. The project depended on the highly disciplined commitment of 189 volunteer general endoscopist assessors from 5 countries. These assessors generated the efficacy data of the phase 1 draft and the final, much shorter phase 2 training programs. The authors are most grateful to these contributors who are listed in Supplementary Materials. We believe that the scale of the evaluation is unprecedented in the literature on endoscopic training and diagnosis.

The training videos used a standardized pullback, without specific attention on the area containing a BORN lesion, so that the videos could be used to teach detection of these lesions. The quality of the pullbacks also provides a reference to general endoscopists for how a Barrett's segment should be optimized for endoscopic inspection with respect to the amount of mucus, bubbles, and the amount of insufflation. The draft phase 1 and condensed final phase 2 BORN modules encompassed a variety of proven early BORN lesions, which were all treated endoscopically. Histology confirmed all of these to contain HGD or EAC, without deep submucosal infiltration.

The culling of content from the phase 1 draft module was informed by evaluation of the learning curves of all videos separately. Videos shown to contribute insignificantly to learning were removed. As described below, these assessments were made possible through the design of the module's software support of the training process. The final version of the BORN module contains videos of

relatively subtle early BORN lesions, which were all shown by the phase 1 data to contribute to the learning.

The results of the phase 2 evaluations show a significant learning process that was superior to outcomes from the draft phase 1 module, despite its relative brevity. Each of the subcategories of endoscopic experience and country of origin showed significant skill gains, suggesting a high extrinsic validity of our findings (Tables 2, 3, 4). Because of this external validation, users of the module for training purposes can not only relate their video assessments to the expert delineations, but can also benchmark their progress throughout the different batches relative to the 121 phase 2 assessors, with reference to country of origin and level of endoscopic experience.

The primary focus of this study has been on training in recognition of lesions using white light endoscopy in overview, without the use of magnification or optical chromoscopy techniques. In our opinion, these latter techniques are characterization tools, useful once lesions have been primarily recognized with white light endoscopy (see Introduction).

All four outcome measures of the BORN module are of clinical relevance. The "detection score" provides feedback on the number of lesions correctly identified and appropriately targeted for biopsy; in the phase 2 evaluation, only 60% of lesions were identified in the first assessment of the test batch prior to completion of the training batches, whereas a high proportion of lesions were identified in training batches 2-4 and in the second evaluation of the test batch. The other three performance measures relate to a more detailed assessment of BORN lesions and its subtle borders. It is important to note that it was not our primary aim to train general endoscopists in the delineation of BORN lesions, as a prelude to resection. This is usually done in the setting of a highly specialized center with the addition of optical chromoscopy and magnified endoscopy, with the aim of achieving the highest possible cure rate with the first therapy. Yet, by requiring users of this educational module to delineate lesions, our aim was to enhance their recognition of the most subtle appearances of BORN lesions, which are usually around the periphery and still highly relevant to cure.

The "delineation score" expresses how much of the "sweet spot" (i.e. the area delineated by all 3 experts) was delineated by the assessor. This is likely to be the most abnormal part of the lesion that generally requires endoscopic resection rather than ablation.[5, 10] The "agreement delineation score" displays the ratio between the assessor correctly identifying the most involved part of the lesion versus delineating areas that were not considered neoplastic (i.e. no HGD or EAC) by any of the IWGCO specialists, as defined by their mapping of the entire lesion. This assessment is useful to

express the risk for over- or under-treatment. Since the IWGCO experts disagreed to some extent with each other, we created the "relative delineation score", which basically measures how close the assessor is to being "as good" as one of the three experts. These three delineation parameters all showed a clear learning effect, with an especially impressive increase in the phase 2 module (relative improvements from baseline of 129%, 105%, and 106%, respectively; table 4).

A first reaction to the exclusion of NDBE videos from the final phase 2 module might be that this is a weakness of this study, but it is in fact a strength. Assessors were not informed that all videos in the final module contained BORN lesions and the accuracy of recognition of BORN lesions was tested by requiring assessors to make an accurate identification of the position of the lesion by marking it on a video frame, so guessing this would be apparent. Thirdly, the convincingly superior results obtained with the phase 2 module of just 25 videos, compared to the phase 1 draft module of 80 videos, support the data-driven exclusion of the NDBE videos, on the basis these contributed nothing to learning in phase 1.

The greatly reduced size of the phase 2 module, compared to the draft phase 1 module, made its completion much less daunting. The high learning efficiency of the phase 2 module should enhance its level of uptake by both trained and trainee endoscopists.

One of the special features of the BORN module is that its online assessment of each video is fully driven by the trainees: they select the preferred video frame to position a biopsy mark and to delineate the lesion. The module software gives tailored, interactive feedback on the selected video frame. During the feedback sessions at completion of each training batch of videos, the assessor can add and remove experts' delineations as well as their own delineations and thus fully appreciate the subtle appearance of the lesion on the selected time frame. Since expert delineations are available for every second for the time the BORN lesion is visible in the video, assessors can expand feedback to themselves from other parts of the video, allowing for multiple "seeing-recognition" iterations throughout the video. The learning opportunity provided by these viewing and feedback features far exceeds the learning opportunity provided by observation of a live endoscopy in a BORN patient.

What are the potential clinical implications of the BORN module? The availability of a validated teaching tool for recognition of BORN enhances the materials available to endoscopists in training and undergoing recertification. Specific BORN performance scores (see above) could function as a quality requirement. Although this was not our primary aim, the phase 2 BORN module should also be a useful training and assessment tool for endoscopists who undertake the treatment of early BORN lesions.[11] For example, one of the ground rules for safe and effective application of endoscopic

therapy is that endoscopic resection, rather than mucosal ablation should be used to remove all visibly abnormal mucosa, even areas with the most subtle abnormalities.[5, 10] The relative delineation score might be a suitable parameter to measure how close an assessor is to being an expert in detection and delineation of BORN. Numerical thresholds could be established for BORN module scores that define competence in the recognition and spatial assessment of BORN lesions.

In conclusion, we have developed and validated a powerful and efficient interactive web-based teaching tool. The results of this study demonstrate that by completing the BORN training module, general endoscopists with a wide range of experience and from different countries of origin can substantially and conveniently increase their skills for detection and delineation of early BORN lesions. Therefore the module could provide training in an essential upper GI endoscopic skill that is not otherwise readily available.

REFERENCES

1. Hvid-Jensen F, Pedersen L, Drewes AM, et al. Incidence of adenocarcinoma among patients with Barrett's esophagus. N Engl J Med 2011;365:1375-83.
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; quiz 51.
3. Weusten B, Bisschops R, Coron E, et al. Endoscopic management of Barrett's esophagus: European Society of Gastrointestinal Endoscopy (ESGE) Position Statement. Endoscopy 2017;49:191-198.
4. Scholvinck DW, van der Meulen K, Bergman JJ, et al. Detection of lesions in dysplastic Barrett's esophagus by community and expert endoscopists. Endoscopy 2017;49:113-120.
5. Boerwinkel DF, Swager A, Curvers WL, et al. The clinical consequences of advanced imaging techniques in Barrett's esophagus. Gastroenterology 2014;146:622-629 e4.
6. Armstrong D, Bennett JR, Blum AL, et al. The endoscopic assessment of esophagitis: a progress report on observer agreement. Gastroenterology 1996;111:85-92.
7. Lundell LR, Dent J, Bennett JR, et al. Endoscopic assessment of oesophagitis: clinical and functional correlates and further validation of the Los Angeles classification. Gut 1999;45:172-80.
8. 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;131:1392-9.
9. Update on the Paris classification of superficial neoplastic lesions in the digestive tract. Endoscopy 2005;37:570-8.
10. Phoa KN, Pouw RE, Bisschops R, et al. Multimodality endoscopic eradication for neoplastic Barrett oesophagus: results of an European multicentre study (EURO-II). Gut 2016;65:555-62.
11. 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. Gastrointest Endosc 2017;86:1-17.e3.

TABLES AND FIGURES

Figure 1. Flow diagram of video selection for assessment phase 1 and phase 2.

Figure 2. Example of a video frame showing a BORN lesion (A) with 3 individual expert delineations (B-D), all three expert delineations (E) and the creation of the sweet spot (F).

Figure 3. Example of a video frame showing a BORN lesion (A), with a correct biopsy within the sweet spot (B), a delineation overlapping the sweet spot (C) and a delineation overlapping with individual expert delineations (D-F).

Table 1. Characteristics of assessors in phase 1 and 2.

	Phase 1				Phase 2			
Country	**Trainee**	**Junior GE**	**Senior GE**	**Total**	**Trainee**	**Junior GE**	**Senior GE**	**Total**
Netherlands	18	16	12	46	11	12	10	33
Germany	1	4	5	10	7	8	5	20
USA	1	2	1	4	6	10	5	21
Canada	-	-	-	-	5	6	8	19
UK	2	2	4	8	9	4	15	28
Total	22	24	22	68	38	40	43	121

Table 2. Median scores of outcome parameters per batch in phase 1.

	Training Batch 1 (%)	Training Batch 2 (%)	Training Batch 3 (%)	Training Batch 4 (%)	Median absolute increase batch 1-batch 4 (%)*	P-value	Median relative increase batch 1-batch 4 (%)**	P-value
Median detection score	64 (IQR 54-82)	69 (IQR 54-81)	69 (IQR 54-82)	73 (IQR 54-91)	8 (95% CI 0-16)	0.07	21 (95% CI 6-40)	0.01
Median delineation score	41 (IQR 23-56)	52 (IQR 38-68)	59 (IQR 43-68)	63 (IQR 48-78)	22 (95% CI 14-30)	<0.001	64 (95% CI 36-101)	<0.001
Median agreement delineation score	32 (IQR 18-41)	39 (IQR 27-49)	42 (IQR 29-50)	44 (IQR 32-52)	13 (95% CI 8-19)	<0.001	55 (95% CI 27-89)	<0.001
Median relative delineation score	45 (IQR 25-60)	57 (IQR 40-71)	61 (IQR 43-72)	65 (IQR 47-77)	19 (95% CI 11-28)	<0.001	55 (95% CI 29-93)	<0.001

*** Wilcoxon Signed-Rank Tests.**

**** Wilcoxon Tests.**

Table 3. Linear mixed-effects model of outcome measurements in phase 1 and phase 2.

		Phase 1		Phase 2	
		Estimate	**P-value**	**Estimate**	**P-value**
Detection scores	Experience junior	0.086	0.002	0.041	0.039
	Experience senior	0.075	0.009	0.021	0.281
	Country Germany	-0.013	0.691	0.007	0.779
	Country USA	-0.001	0.983	0.004	0.868
	Country Canada	-	-	0.013	0.602
	Country UK	-0.013	0.706	0.003	0.901
	Batch number	0.022	0.027	0.035	<0.001
Delineation scores	Experience junior	0.100	0.001	0.023	0.324
	Experience senior	0.080	0.009	0.035	0.124
	Country Germany	-0.034	0.338	-0.008	0.764
	Country USA	0.053	0.311	-0.001	0.968
	Country Canada	-	-	0.021	0.471
	Country UK	-0.004	0.916	0.017	0.513
	Batch number	0.068	<0.001	0.059	<0.001

Agreement delineation scores	Experience junior	0.072	<0.001	0.024	0.121
	Experience senior	0.060	0.004	0.028	0.062
	Country Germany	-0.023	0.337	-0.020	0.282
	Country USA	0.034	0.335	-0.008	0.681
	Country Canada	-	-	0.015	0.442
	Country UK	0.008	0.756	-0.009	0.609
	Batch number	0.040	<0.001	0.032	<0.001
Relative delineation scores	Experience junior	0.108	<0.001	0.031	0.177
	Experience senior	0.088	0.005	0.046	0.041
	Country Germany	-0.033	0.362	-0.014	0.622
	Country USA	0.063	0.230	0.001	0.968
	Country Canada	-	-	0.026	0.356
	Country UK	0.011	0.779	-0.006	0.831
	Batch number	0.059	<0.001	0.044	<0.001

Table 4. Median scores of outcome parameters per batch in phase 2.

	Evaluation of test batch at start (%)	Training Batch 1 (%)	Training Batch 2 (%)	Training Batch 3 (%)	Training Batch 4 (%)	Evaluation of test batch at end (%)	Median absolute increase start – end test batch (%)*	P-value	Median relative increase start – end test batch (%)**	P-value
Median detection score	60 (IQR 60-80)	60 (IQR 60-80)	100 (IQR 80-100)	100 (IQR 80-100)	100 (IQR 80 – 100)	80 (IQR 80-100)	30 (95% CI 20-30)	<0.001	46 (95% CI 33-50)	<0.001
Median delineation score	32 (IQR 21-46)	44 (IQR 31-55)	69 (IQR 53-81)	73 (IQR 63-81)	78 (IQR 65-85)	70 (IQR 61-77)	35 (95% CI 31-38)	<0.001	129 (95% CI 106-160)	<0.001
Median agreement delineation score	25 (IQR 17-35)	35 (IQR 24-42)	48 (IQR 40-57)	51 (IQR 45-58)	55 (IQR 46-61)	48 (IQR 41-54)	21 (95% CI 18-24)	<0.001	105 (95% CI 83-130)	<0.001
Median relative delineation score	35 (IQR 22-48)	49 (IQR 34-61)	73 (IQR 59-85)	76 (IQR 66-85)	82 (IQR 69-91)	68 (IQR 58-74)	30 (95% CI 26-34)	<0.001	106 (95% CI 85-132)	<0.001

*** Wilcoxon Signed-Rank Tests.**

**** Wilcoxon Tests.**