



UvA-DARE (Digital Academic Repository)

What makes an expert Barrett's pathologist?

Concordance and pathologist expertise within a digital review panel

van der Wel, M.J.

Publication date

2019

Document Version

Other version

License

Other

[Link to publication](#)

Citation for published version (APA):

van der Wel, M. J. (2019). *What makes an expert Barrett's pathologist? Concordance and pathologist expertise within a digital review panel*. [Thesis, fully internal, Universiteit van Amsterdam].

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

CHAPTER

DIGITAL MICROSCOPY AS VALID ALTERNATIVE TO CONVENTIONAL MICROSCOPY FOR HISTOLOGICAL EVALUATION OF BARRETT'S ESOPHAGUS BIOPSIES

4

M. J. van der Wel, L. C. Duits, C. A. Seldenrijk,
G. J. Offerhaus, M. Visser, F. J. Ten Kate, O. J. de Boer,
J. G. P. Tijssen, J. J. Bergman, S. L. Meijer

Diseases of the Esophagus. 2017 Nov 1;30(11):1-7

ABSTRACT

Background & Aims

Management of Barrett's esophagus (BE) relies heavily on histopathological assessment of biopsies, associated with significant intra- and interobserver variability. Guidelines recommend biopsy review by an expert in case of dysplasia. Conventional review of biopsies, however, is impractical and does not allow for teleconferencing or annotations. An expert digital review platform might overcome these limitations. We compared diagnostic agreement of digital and conventional microscopy for diagnosing BE dysplasia.

Methods

Sixty BE biopsy glass slides (non-dysplastic BE (NDBE); n 25, low-grade dysplasia (LGD); n 20; high-grade dysplasia (HGD); n 15) were scanned at 20 magnification. The slides were assessed four times by five expert BE pathologists, all practicing histopathologists (range: 5–30 years), in 2 alternating rounds of digital and conventional microscopy, each in randomized order and sequence of slides. Intraobserver and pairwise interobserver agreement were calculated, using custom weighted Cohen's kappa, adjusted for the maximum possible kappa scores.

Results

Split into three categories (NDBE, IND, LGD HGD), the mean intraobserver agreement was 0.75 and 0.84 for digital and conventional assessment, respectively (p 0.35). Mean pairwise interobserver agreement was 0.80 for digital and 0.85 for conventional microscopy (p 0.17). In 47/60 (78%) of digital microscopy reviews a majority vote of 3 pathologists was reached before consensus meeting. After group discussion, a majority vote was achieved in all cases (60/60).

Conclusion

Diagnostic agreement of digital microscopy is comparable to that of conventional microscopy. These outcomes justify the use of digital slides in a nationwide, web-based BE revision platform in the Netherlands. This will overcome the practical issues associated with conventional histologic review by multiple pathologists.

INTRODUCTION

Esophageal adenocarcinoma (EAC) is the cancer with the fastest rising incidence in Western world. ¹ EAC develops from a precursor lesion, known as Barrett's esophagus (BE), which is defined as replacement of esophageal squamous epithelium by metaplastic columnar epithelium containing intestinal metaplasia. BE predisposes to EAC through a metaplasia-dysplasia-carcinoma sequence. ² BE patients are therefore offered regular endoscopic surveillance with biopsies to detect neoplasia at an early and curable stage. Current guidelines recommend a 3-5 year surveillance interval in patients with non-dysplastic BE (NDBE), whereas BE patients with HGD or EAC are offered endoscopic or surgical therapy. ³⁻⁵ Low-grade dysplasia (LGD) in BE is an accepted risk factor for progression ⁶ and these patients are offered intensified endoscopic surveillance or prophylactic ablation therapy. ^{3,4}

Despite the ongoing search for biomarkers to predict malignant progression, histopathologic identification of dysplasia through the modified Vienna classification ^{4,7-9} (see Table 1) is the most important factor in deciding on the surveillance interval and/or the indication for treatment. ^{10,11} The histological evaluation, however, is associated with substantial inter- and intraobserver variability, ¹²⁻¹⁵ especially when diagnosing LGD. ⁶ In a previous study by our group, all 147 LGD cases diagnosed between 2000 and 2006 in six community hospitals in the Amsterdam region were reviewed by two expert pathologists. Seventy-five percent of these cases were down-staged to NDBE, demonstrating a neoplastic rate of only 0.49% per year, whereas the cases in which the original LGD diagnosis was confirmed, progressed to HGD/EAC at a rate of 13.4% per year. ⁶ These results were confirmed in an independent cohort of 293 community hospital patients with a referral diagnosis of LGD using a different panel of expert pathologists. Again, the majority of cases (59%) was downstaged to NDBE with a progression rate of 0.6% per patient per year, while in cases with confirmed LGD the annual progression rate was 9.1%. ¹⁶

The Dutch BE guideline and the recent ESGE BE guideline, therefore, advise that all diagnoses of LGD should be revised by an expert GI-pathologist before decisions are made on surveillance interval or prophylactic ablation therapy. In our country, the Barrett's Advisory Committee (BAC) has been accommodating review requests for BE since 1998. ^{17,18} Current practice for the committee is to use conventional

microscopy and glass slides for this review process. This requires that glass slides are physically transported before review or consultation can take place. This procedure is cumbersome and carries the risk of glass slides getting damaged or lost in the process, especially if they have to be reviewed by multiple expert pathologists at different institutions. Other researchers have encountered these problems in their validation process as well.¹⁹

We anticipate that the implementation of new guidelines will lead to a significant increase in the number of review requests in the near future. The use of digital slides would significantly improve the practicality of this review process compared to using glass slides. Evaluation of digital slides allows multiple experts to review cases in parallel without the need to transfer glass slides and enables consensus meetings through teleconferencing without the need to share the glass slides during face-to-face meetings. Furthermore, the reports and digital annotations generated by the review panel might be used for feedback to the referring center and for future teaching tools. In the process of creating a national, digital BE advisory platform, an important prerequisite is that evaluation of digital and glass slides is not significantly different. The aim of this study, therefore, is to compare the diagnostic performance of the five expert GI pathologists involved in the BAC over the last years, when using digital and conventional microscopy slides of BE biopsy specimens.

METHODS

Slide selection and scanning

For this proof of concept study, a total of 60 single slides were selected from 60 individual BE biopsy cases that were referred to the BAC earlier for pathology review and from the BE surveillance program at the Amsterdam Academic Medical Center between 2007 and 2013. All glass slides were formalin fixed, paraffin embedded (FFPE), and hematoxylin and eosin (H&E) stained. The referring diagnosis was LGD in 20 cases and HGD or EAC in 15 cases. These cases were supplemented with 25 NDBE reference cases. From each case, a single, representative slide was selected by the study coordinator, based on detailed reviewing of the pathology report. The selected single slides were fully digitalized

using a scanner with a x20 microscope objective (.Slide, Olympus, Tokyo, Japan) and checked for focus and acuity by the study coordinator. Subsequently, the slides were stored on a secure server and made available online using the virtual slide system 'Digital Slidebox 4.5' (<https://dsb.amc.nl/dsb/login.php>, Slidepath, Leica Microsystems, Dublin, Ireland).

Histologic assessment

The central expert pathology panel consisted of five pathologists (FJWtK, CAS, SLM, MV, GJAO). The expert pathologists who participated in this study were considered as such by their (international) peers. They have been dedicated to the field of Barrett's for a minimum of 5 years (range: 5–30 years) and have a minimum caseload of 5–10 cases per week of which 75% is dysplastic. All pathologists have participated in the Dutch Barrett advisory committee for many years¹⁶⁻¹⁸ and are actively practicing histopathologists. All pathologists participated in multiple training programs for endoscopists and pathologists (<http://www.best-academia.eu>) and each has co-authored more than 10 peer reviewed publications in this field.^{6 16 18 20-32} All pathologists assessed each slide four times: twice by conventional microscopy and twice by digital microscopy. A minimum interval of 2 weeks was maintained between the rounds to minimize recognition bias, as recommended by the College of American Pathologists. (10) In order to further minimize recognition bias, the complete slide set was randomly divided in two subsets of 30 cases each. Subsequently, each subset was randomized 20 times to create four uniquely ordered versions of each subset for each pathologist. During each round of assessment, pathologists viewed one randomly ordered subset with one modality and the other randomly ordered subset with the other modality. The five participating pathologists were blinded for any clinical information or identifying slide features at all times and no correspondence about the dataset was allowed during the study. The digital slides were viewed on a calibrated full HD monitor (Eizo EcoView 2436 WFS-BK, using calibration software 'Datacolor Spyder 4 Pro'), which was recalibrated at every meeting. The glass slides were transported to the pathologists by the study coordinator and were assessed using their own microscopes. Glass slides were viewed at a maximum of x20 magnification, similar to the scanning resolution. The study coordinator was present during all assessments and completed a case record form (CRF) for each slide assessment, which contained questions about the quality of the slide and its diagnostic characteristics. The pathologists diagnosed the cases according

to the Vienna classification for gastrointestinal neoplasia: 'NDBE,' 'IND,' 'LGD,' 'HGD or (suspicion of) invasive carcinoma'.⁷ The categories of HGD and (suspicion of) invasive carcinoma were grouped together. Once the pathologists had completed a slide assessment and signed off the CRF, they could not re- evaluate their diagnosis.

Table 1: Modified Vienna classification for gastrointestinal neoplasms*

Category	Diagnosis
1	Negative for dysplasia
2	Indefinite for dysplasia
3	Low grade dysplasia
4	High grade dysplasia, non-invasive carcinoma, suspicion of invasive carcinoma, intramucosal carcinoma
5	Submucosal carcinoma or beyond

*Adapted from Schlemper et al. 2000⁷

Consensus meeting

After completion of the four rounds, a consensus meeting was held in which the five participating pathologists discussed discrepant cases and attempted to reach a consensus diagnosis for all digital slides. Cases were considered discrepant if three pathologists or less agreed in the second digital round of assessment. Discrepant cases were assessed as a group, digitally, using the same screen and viewing software as in the initial assessment. Discrepant cases were discussed in random order with each of the five pathologists alternating in presenting the case. Three minutes discussion time per discrepant case (slide) was allowed. If a unanimous consensus diagnosis could not be reached, a majority vote was agreed upon.

Statistical analyses

No formal sample size calculations were conducted for the purpose of this study. We anticipated a significant background variability in the histological interpretation because of differences between pathologists and the inherently subjective interpretation of morphological changes that are associated with reactive changes and dysplasia. We therefore selected a realistic number of cases and case mix that would allow the participating pathologists to perform the required four

assessments rounds within a reasonable timeframe. Intra- and interobserver agreements were calculated using a custom weighted Cohen's kappa (K).^{33 34} A diagnosis of IND is ranked '2' in the Vienna classification, but this diagnosis is not necessarily ranked ordinally between NDBE (1) and LGD (3). Therefore, custom weights were assigned to (dis)agreements. Complete agreements were assigned a weight of 1. Disagreements between NDBE and HGD were assigned a weight of 0, all other disagreements were assigned a weight of 0.5. Moreover, the maximum possible kappa score calculated for pairwise assessments does not always equal 1.³⁵ Therefore, strength of agreement was corrected for the differences in maximum kappa and traditionally categorized as: a value of zero or less indicates agreement no better than chance alone; 0.00–0.20, slight; 0.21–0.40, fair; 0.41–0.60, moderate; 0.61–0.80, substantial; 0.81–1.00, almost perfect.³⁶ All kappa's were calculated using the clinically significant grouping of 'no dysplasia' versus 'dysplasia' (3 categories: NDBE-IND-LGD + HGD). The intraobserver agreement between the first and second rating with the same methodology (digital or conventional) was assessed for each pathologist. This is called the intramethod (test-retest) agreement. The interobserver agreement was calculated between each possible pair of pathologists. From these duos, the average intramethod agreement was calculated for both methods. Statistical significance between the mean kappas was calculated using the paired t-test, considering the 2 assessment modalities as paired observations and the five pathologists as one sample. We hypothesized that there was no difference between the two modalities (H0 hypothesis). A *p*-value of 0.05 was considered statistically significant and a reason to reject H0. Statistical analyses were performed using the Statistical Package for the Social Sciences (SPSS 22.0, IBM Corp., Armonk, New York, USA). The custom weighted kappa was developed using the self-automated program Agreestat (version 2013.2, Advanced Analytics, LCC, Gaithersburg, USA).

RESULTS

Intraobserver agreement

When split into three categories (i.e. NDBE, IND, and LGD + HGD) and corrected for maximum possible kappa scores, the mean intraobserver (intramethod) agreement was 0.75 for digital microscopy and 0.84 for conventional microscopy (p -value 0.34, H_0 accepted). Intraobserver agreements between individual pathologists varied between 0.53 and 0.88 (**Table 2**). When data were split into four categories (i.e. presenting LGD and HGD as separate categories), data were highly comparable (results not shown).

Interobserver agreement

The mean interobserver agreement for digital and conventional microscopy demonstrated kappa values of 0.80 and 0.85, respectively (p -value 0.17, H_0 accepted). The agreement between individual pairs of pathologists varied between 0.71 and 1.00 (**Table 3**). When data were split into four categories (i.e. presenting LGD and HGD as separate categories), data were again highly comparable (results not shown).

Consensus

In the last round of digital microscopy assessments, 13 out of 60 cases (22%) had a unanimous agreement of the five expert pathologists; 16 (26%) had an agreement of 4 out of 5 pathologists, and 18 (30%) had an agreement of 3 pathologists, bringing the total number of cases with a majority vote to 47 out of 60 cases (78%). Thirteen cases (22%) had an agreement of two pathologists, and these were discussed in a consensus meeting, together with the 18 cases on which 3 pathologists agreed, adding up to a total of 31 discussed cases. The discrepancies between diagnoses of these cases mainly concerned IND versus NDBE or LGD (14/31, 45%) or LGD versus HGD (10/31, 32%). After group discussion, 15 additional cases (25%) attained a unanimous agreement of the five expert pathologists. In the remaining 16 cases, the pathologists reached a majority vote of 3 or 4 pathologists, bringing the total amount of cases with unanimous agreement to 28 (47%) and the total amount of cases with a majority vote of 3 or 4 pathologists to 60 (100%). These results can be appreciated in **Table 4**. Pictures taken of an NDBE, LGD, and

HGD case with full panel consensus using digital microscopy can be appreciated in **Figure 1**.

Table 2: Intraobserver agreement of 5 expert BE pathologists for digital and conventional microscopy

Path*	Digital microscopy		Conventional microscopy			p-value [‡]	
	Weighted K [†] (95% CI)	Max K	Weighted / max K	Weighted K (95% CI)	Max K		Weighted / max K
3 categories: NDBE – IND – LGD + HGD[§]							
1	0.85 (0.74-0.95)	0.91	0.92	0.79 (0.68-0.91)	0.90	0.88	
2	0.43 (0.26-0.61)	0.61	0.71	0.78 (0.64-0.91)	0.89	0.87	
3	0.89 (0.81-0.98)	0.97	0.93	0.75 (0.62-0.89)	1.00	0.75	
4	0.54 (0.34-0.74)	0.83	0.65	0.63 (0.44-0.82)	0.71	0.88	
5	0.51 (0.33-0.69)	0.97	0.53	0.77 (0.65-0.90)	0.95	0.82	
Mean	0.64	0.86	0.75	0.74	0.89	0.84	0.35

*Path=pathologist, [†]K = kappa, for weighted / max K, [‡]when ≤ 0.05, [§]NDBE = non-dysplastic Barrett's esophagus, IND = indefinite for dysplasia, LGD = low-grade dysplasia, HGD = high-grade dysplasia

Figure 1: Pictures of digitalized cases of non-dysplastic Barrett's esophagus (A), low-grade dysplasia (B), and high-grade dysplasia (C) with full panel consensus.

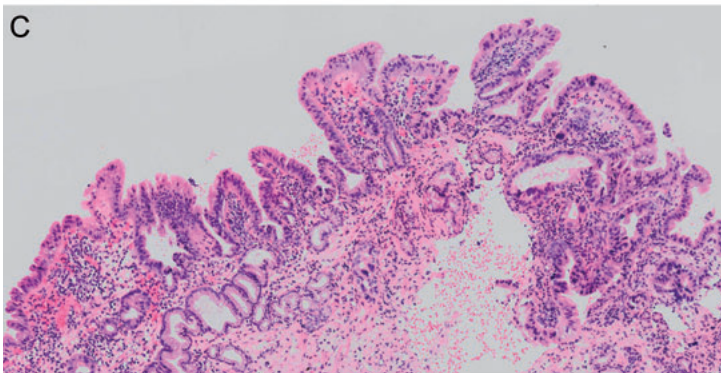
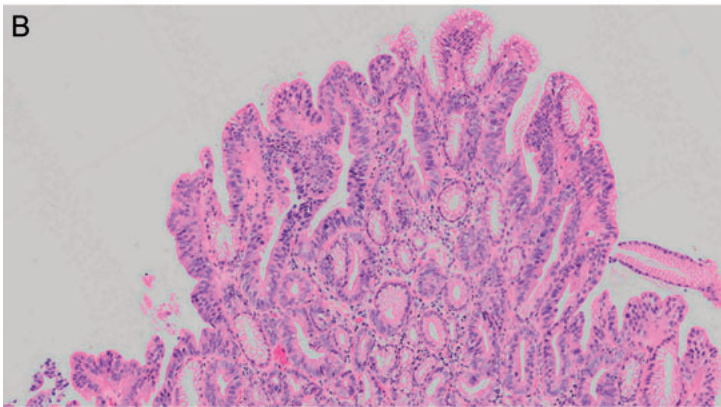
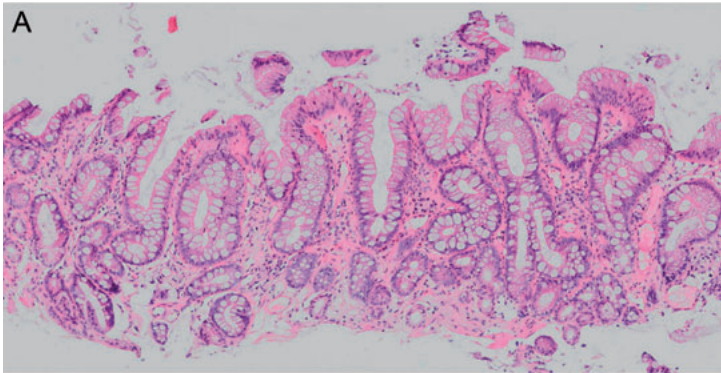


Table 3: Pairwise interobserver agreement of 5 expert BE pathologists for digital and conventional microscopy

Path*	Digital microscopy				Conventional microscopy			
	Weighted K [†] (95% CI)	Max K	Weighted / max K	Max K	Weighted K (95% CI)	Max K	Weighted / max K	p-value [‡]
3 categories: NDBE – IND – LGD + HGD[§]								
1-2	0.39 (0.25-0.53)	0.44	0.89	0.57	0.53 (0.38-0.69)	0.95	0.95	
1-3	0.78 (0.66-0.90)	0.92	0.85	0.93	0.77 (0.65-0.89)	0.83	0.83	
1-4	0.53 (0.35-0.70)	0.65	0.80	0.63	0.56 (0.40-0.71)	0.89	0.89	
1-5	0.65 (0.50-0.80)	0.89	0.73	0.91	0.60 (0.43-0.76)	0.66	0.66	
2-3	0.35 (0.21-0.49)	0.38	0.91	0.53	0.47 (0.31-0.63)	0.88	0.88	
2-4	0.23 (0.11-0.34)	0.23	1.00	0.30	0.28 (0.15-0.42)	0.94	0.94	
2-5	0.32 (0.18-0.47)	0.48	0.68	0.61	0.49 (0.32-0.66)	0.81	0.81	
3-4	0.57 (0.39-0.75)	0.72	0.79	0.65	0.60 (0.43-0.76)	0.92	0.92	
3-5	0.60 (0.44-0.76)	0.84	0.71	0.90	0.64 (0.49-0.80)	0.71	0.71	
4-5	0.41 (0.22-0.59)	0.58	0.69	0.57	0.52 (0.35-0.68)	0.92	0.92	
Mean	0.48	0.61	0.80	0.66	0.55	0.85	0.85	0.17

*Path=pathologist, [†]K = kappa, [‡]for weighted / max K, significant when ≤ 0.05 , [§]NDBE = non-dysplastic Barrett's esophagus, IND = indefinite for dysplasia, LGD = low-grade dysplasia, HGD = high-grade dysplasia

Table 4: Consensus of 5 expert BE pathologists before and after group discussion

Number of pathologists	Agreement	
	Before discussion (%)	After discussion (%)
Initial majority vote (≥ 3)	47 (78)	60 (100)
5	13 (22)	28 (47)
4	16 (27)	24 (40)
3	18 (30)	8 (13)
2	13 (22)	-
Total	60	60
Discrepancies when ≤ 3 agreement	Before discussion (%)	After discussion (%)
LGD vs HGD*	10 (32)	3 (38)
NDBE vs LGD/HGD	7 (23)	-
IND vs NDBE / LGD	14 (45)	5 (62)

*NDBE = non-dysplastic Barrett's esophagus, IND = indefinite for dysplasia, LGD = low-grade dysplasia, HGD = high-grade dysplasia

DISCUSSION

In this study, we found conventional and digital microscopic assessment to have comparable performance scores when used by five expert BE pathologists assessing BE biopsy slides: the two modalities had a comparable intraobserver agreement and interobserver agreement. These results suggest that pathologists are equally consistent in reassessing digital and glass slides. The observed differences between digital and conventional microscopic assessments were small compared to the observed differences between the individual pathologists for their intraobserver agreement or compared to the observed differences between pairs of pathologists for their interobserver agreement. This indicates that the variability in diagnostic viewpoints of pathologists is of greater importance than any differences induced by observing slides either through the microscope or digitally on a computer screen. It is important to note that none of the five pathologists in this study used digital microscopy in daily practice and their experience in BE diagnostics therefore was fully based on conventional microscopy. In our opinion, this biases the comparison between conventional microscopy and digital microscopy in favor of the former. We feel that the results of this study justify the use of digital microscopy in reviewing BE biopsies. Digital

microscopy allows pathologists to assess cases in parallel and to discuss cases in teleconferencing sessions, without slides being broken or lost in transportation between different facilities. Furthermore, the reports and digital annotations generated by the review panel might be used for feedback to the referring centers and for future teaching tools.

The kappa scores generated by our panel should be interpreted with caution. In order to achieve the study aim, raters were restricted to single slide H&E assessments. This does not reflect the normal microscopic assessment of BE biopsies, where the pathologist has access to multilevel biopsies and slides, depending on the length of the BE segment. Moreover, in clinical practice pathologists often use additional p53 immunohistochemical staining to aid in discrimination between histological categories. In preparation for our national digital review platform we are, therefore, currently assessing the interobserver agreement of our review platform by having our participating pathologists assess digital slides of complete BE endoscopies (i.e. all slides from all tissue blocks produced from one endoscopy, including p53 immunohistochemistry). In addition, there are a number of methodological issues that hamper the comparison of kappa scores between studies. First, many studies do not use a weighted kappa, which means that a discrepancy between NDBE and IND is valued the same way as a discrepancy between NDBE and HGD or cancer. Second, most studies do not take the maximum possible kappa per agreement comparison into account. Disagreements between two observers are statistically visible as divergent marginal totals and interchanging of categories (usually in borderline cases), leading to a maximum possible kappa of less than one. We therefore calculated a custom weighted kappa and depicted the kappa as fraction of the maximum possible kappa, thus increasing the validity of the kappa measure in this study. An additional implication that is true for every agreement study which makes use of the kappa statistic is that the results are not directly comparable across studies due to selection issues, which concern the number and experience of the raters and the number and complexity of the samples. This study has some additional strengths. First, our pathologists are renowned experts in the field of BE diagnostics and have proven this in many earlier studies.^{6 16 18 20-32} We are aware that no universal definition of an expert BE pathologist exists, but we feel that the extensive scientific record, diagnostic experience in the field of BE and

recognition as experts by their peers justifies this qualification. Second, we have tried to prevent systematic bias by randomizing the order of the samples as well as the assessment modality for every pathologist, by supervising all assessments, and by scoring all cases twice for every assessment modality. Third, the dysplastic cases in the study all had been sent for revision earlier and therefore reflect the type of cases our review platform will be dealing with in the near future. This study has some limitations. First, the panel pathologists were heterogeneous concerning their computer skills and subsequently in operating the viewing program for the digital cases. This in turn affected their fluidity of digital assessment. Second, their extensive experience in the assessment of BE cases was restricted to conventional microscopy. As mentioned, we feel that these factors may have biased our results in favor of conventional microscopy, which strengthens our conclusion.

In conclusion, this study demonstrated that digital microscopy is a valid alternative to conventional microscopy in the histological assessment of BE biopsies. Implementing digital microscopy in the setting of a national review platform will overcome many of the limitations associated with conventional microscopic review of glass slides. A digital platform will allow review of all slides from all biopsies taken per endoscopy, but studies testing the feasibility to implement this in actual patient care are needed. Therefore, we are currently investigating the agreement of these pathologists and nine other expert BE pathologists using another case set where they review all slides from all biopsy levels per patient endoscopy. The digital slide set of this study and corresponding consensus diagnoses will be used in future studies to develop an algorithm for agreement. Furthermore, the use of a digital review platform is not only limited to the interpretation of BE biopsies. It could also be employed in the future for the interpretation and risk stratification of endoscopic resection specimens, which encounters similar diagnostic problems. The set-up of this panel for BE biopsies and resection specimens, as well as improved consensus within the group of pathologists could prove to be a useful tool to streamline patient care at both ends of the diagnostic spectrum.

REFERENCES

1. Thrift AP, Whiteman DC. The incidence of esophageal adenocarcinoma continues to rise: analysis of period and birth cohort effects on recent trends. *Ann Oncol* 2012;23(12):3155-62. doi: 10.1093/annonc/mds181
2. Haggitt RC, Tryzelaar J, Ellis FH, et al. Adenocarcinoma complicating columnar epithelium-lined (Barrett's) esophagus. *Am J Clin Pathol* 1978;70(1):1-5.
3. Fitzgerald RC, di Pietro M, Ragunath K, et al. British Society of Gastroenterology guidelines on the diagnosis and management of Barrett's oesophagus. *Gut* 2014;63(1):7-42. doi: 10.1136/gutjnl-2013-305372
4. Spechler SJ, Sharma P, Souza RF, et al. American Gastroenterological Association technical review on the management of Barrett's esophagus. *Gastroenterology* 2011;140(3):e18-52; quiz e13. doi: 10.1053/j.gastro.2011.01.031
5. 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(2):191-98. doi: 10.1055/s-0042-122140
6. Curvers WL, ten Kate FJ, Krishnadath KK, et al. Low-grade dysplasia in Barrett's esophagus: overdiagnosed and underestimated. *The American journal of gastroenterology* 2010;105(7):1523-30. doi: 10.1038/ajg.2010.171
7. Schlemper R, Riddell R, Kato Y, et al. The Vienna classification of gastrointestinal epithelial neoplasia. *Gut* 2000;47(2):251-5. doi: 10.1136/gut.47.2.251
8. Reid BJ, Haggitt RC, Rubin CE, et al. Observer variation in the diagnosis of dysplasia in Barrett's esophagus. *Human pathology* 1988;19(2):166-78.
9. Schlemper RJ, Kato Y, Stolte M. Diagnostic criteria for gastrointestinal carcinomas in Japan and Western countries: proposal for a new classification system of gastrointestinal epithelial neoplasia. *Journal of gastroenterology and hepatology* 2000;15 Suppl:G49-57.
10. Pantanowitz L, Sinard JH, Henricks WH, et al. Validating whole slide imaging for diagnostic purposes in pathology: guideline from the College of American Pathologists Pathology and Laboratory Quality Center. *Arch Pathol Lab Med* 2013;137(12):1710-22. doi: 10.5858/arpa.2013-0093-CP
11. Vologgi L, Montgomery E, Lam-Himlin D. A clinical and histopathologic focus on Barrett esophagus and Barrett-related dysplasia. *Arch Pathol Lab Med* 2011;135:1249-60. doi: 10.5858/arpa.2011-0019-RA
12. Montgomery E, Bronner MP, Goldblum JR, et al. Reproducibility of the diagnosis of dysplasia in Barrett esophagus: a reaffirmation. *Human pathology* 2001;32(4):368-78. doi: 10.1053/hupa.2001.23510
13. Kerkhof M, van Dekken H, Steyerberg EW, et al. Grading of dysplasia in Barrett's oesophagus: substantial interobserver variation between general and gastrointestinal pathologists. *Histopathology* 2007;50(7):920-7. doi: 10.1111/j.1365-2559.2007.02706.x

14. Coco DP, Goldblum JR, Hornick JL, et al. Interobserver variability in the diagnosis of crypt dysplasia in Barrett esophagus. *The American journal of surgical pathology* 2011;35(1):45-54. doi: 10.1097/PAS.0b013e3181ffdd14
15. Goldblum JR. Barrett's esophagus and Barrett's-related dysplasia. *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc* 2003;16(4):316-24. doi: 10.1097/01.MP.0000062996.66432.12
16. Duits LC, Phoa KN, Curvers WL, et al. Barrett's oesophagus patients with low-grade dysplasia can be accurately risk-stratified after histological review by an expert pathology panel. *Gut* 2014 doi: 10.1136/gutjnl-2014-307278
17. Hulscher JB, Haringsma J, Benraadt J, et al. Comprehensive Cancer Centre Amsterdam Barrett Advisory Committee: first results. *Neth J Med* 2001;58(1):3-8.
18. Offerhaus GJ, Correa P, van Eeden S, et al. Report of an Amsterdam working group on Barrett esophagus. *Virchows Archiv : an international journal of pathology* 2003;443(5):602-8. doi: 10.1007/s00428-003-0906-z
19. Sanders DS, Grabsch H, Harrison R, et al. Comparing virtual with conventional microscopy for the consensus diagnosis of Barrett's neoplasia in the AspECT Barrett's chemoprevention trial pathology audit. *Histopathology* 2012;61(5):795-800. doi: 10.1111/j.1365-2559.2012.04288.x
20. Curvers WL, van Vilsteren FG, Baak LC, et al. Endoscopic trimodal imaging versus standard video endoscopy for detection of early Barrett's neoplasia: a multicenter, randomized, crossover study in general practice. *Gastrointestinal endoscopy* 2011;73(2):195-203. doi: 10.1016/j.gie.2010.10.014
21. Phoa KN, van Vilsteren FG, Weusten BL, et al. Radiofrequency ablation vs endoscopic surveillance for patients with Barrett esophagus and low-grade dysplasia: a randomized clinical trial. *JAMA : the journal of the American Medical Association* 2014;311(12):1209-17. doi: 10.1001/jama.2014.2511
22. Polkowski W, Baak JP, van Lanschot JJ, et al. Clinical decision making in Barrett's oesophagus can be supported by computerized immunoquantitation and morphometry of features associated with proliferation and differentiation. *The Journal of pathology* 1998;184(2):161-8. doi: 10.1002/(SICI)1096-9896(199802)184:2<161::AID-PATH971>3.0.CO;2-2
23. van Sandick JW, Baak JP, van Lanschot JJ, et al. Computerized quantitative pathology for the grading of dysplasia in surveillance biopsies of Barrett's oesophagus. *The Journal of pathology* 2000;190(2):177-83. doi: 10.1002/(SICI)1096-9896(200002)190:2<177::AID-PATH508>3.0.CO;2-X
24. van Sandick JW, van Lanschot JJ, Kuiken BW, et al. Impact of endoscopic biopsy surveillance of Barrett's oesophagus on pathological stage and clinical outcome of Barrett's carcinoma. *Gut* 1998;43(2):216-22.
25. 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(4):555-62. doi: 10.1136/gutjnl-2015-309298

26. Alvarez Herrero L, van Vilsteren FG, Pouw RE, et al. Endoscopic radiofrequency ablation combined with endoscopic resection for early neoplasia in Barrett's esophagus longer than 10 cm. *Gastrointestinal endoscopy* 2011;73(4):682-90. doi: 10.1016/j.gie.2010.11.016
27. van Vilsteren FG, Pouw RE, Seewald S, et al. Stepwise radical endoscopic resection versus radiofrequency ablation for Barrett's oesophagus with high-grade dysplasia or early cancer: a multicentre randomised trial. *Gut* 2011;60(6):765-73. doi: 10.1136/gut.2010.229310
28. Phoa KN, Pouw RE, van Vilsteren FG, et al. Remission of Barrett's esophagus with early neoplasia 5 years after radiofrequency ablation with endoscopic resection: a Netherlands cohort study. *Gastroenterology* 2013;145(1):96-104. doi: 10.1053/j.gastro.2013.03.046
29. Peters FP, Brakenhoff KP, Curvers WL, et al. Histologic evaluation of resection specimens obtained at 293 endoscopic resections in Barrett's esophagus. *Gastrointestinal endoscopy* 2008;67(4):604-9. doi: 10.1016/j.gie.2007.08.039
30. Pouw RE, Gondrie JJ, Sondermeijer CM, et al. Eradication of Barrett esophagus with early neoplasia by radiofrequency ablation, with or without endoscopic resection. *Journal of gastrointestinal surgery : official journal of the Society for Surgery of the Alimentary Tract* 2008;12(10):1627-36; discussion 36-7. doi: 10.1007/s11605-008-0629-1
31. Swager A, Boerwinkel DF, de Bruin DM, et al. Volumetric laser endomicroscopy in Barrett's esophagus: a feasibility study on histological correlation. *Diseases of the esophagus : official journal of the International Society for Diseases of the Esophagus / ISDE* 2016;29(6):505-12. doi: 10.1111/dote.12371
32. Davelaar AL, Calpe S, Lau L, et al. Aberrant TP53 detected by combining immunohistochemistry and DNA-FISH improves Barrett's esophagus progression prediction: a prospective follow-up study. *Genes Chromosomes Cancer* 2015;54(2):82-90. doi: 10.1002/gcc.22220
33. Cohen J. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. *Psychological Bulletin* 1968;70(4):213-19.
34. Cohen J. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* 1960;20(1):37-45.
35. Feinstein AR. High agreement but low kappa: I the problems of two paradoxes. *J Clin Epidemiol* 1990;43(6):543-49.
36. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33(1):159-74.