



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

7

ADHERENCE TO PRE-SET BENCHMARK QUALITY CRITERIA TO QUALIFY AS EXPERT ASSESSOR OF DYSPLASIA IN BARRETT'S ESOPHAGUS BIOPSIES: TOWARDS DIGITAL REVIEW OF BARRETT'S ESOPHAGUS

M. J. van der Wel, E. Klaver, L. C. Duits, R. E. Pouw, C. A. Seldenrijk,
G. J. A. Offerhaus, M. Visser, F. J. W. ten Kate, K. Biermann,
L. A. A. Brosens, M. Doukas, C. Huysentruyt, A. Karrenbeld,
G. Kats-Ugurlu, J. S. van der Laan, G. van Lijnschoten, F. C. P. Moll,
A. H. A. G. Ooms, J. G. P. Tijssen, J. J. G. H. M. Bergman, S. L. Meijer

United European Gastroenterology Journal 2019; 7 (7); 889-896

ABSTRACT

Background

Dysplasia assessment of Barrett's esophagus (BE)-biopsies is associated with low observer agreement; guidelines advise expert review. We have developed a web-based review panel for dysplastic BE biopsies.

Objective

To test if ten gastro-intestinal (GI) pathologists working at Dutch BE expert centers met pre-set benchmark scores (BMS) for quality criteria.

Methods

Ten GI pathologists twice assessed 60 digitalized BE cases, enriched for dysplasia; then randomized (7520 assessments). We tested predefined benchmark quality criteria: 1) percentage of 'indefinite for dysplasia' diagnoses, BMS $\leq 14\%$ for all cases, $\leq 16\%$ for dysplastic subset, 2) intra-observer agreement; BMS ≥ 0.66 / ≥ 0.39 , 3) percentage agreement with 'gold standard diagnosis'; BMS $\geq 82\%$ / $\geq 73\%$, 4) proportion of cases with high-grade dysplasia (HGD) underdiagnosed as non-dysplastic BE (NDBE); BMS $\leq 1/78$ ($\leq 1.28\%$) assessments for dysplastic subset.

Results

GI pathologists had 7 years' BE-experience, handling 7 BE-cases weekly. Three met stringent BMS; all cases and dysplastic subset, three met extended BMS. Four pathologists lacked one quality criterion to meet BMS.

Conclusion

Predefined BMS for expert assessment of BE-dysplasia biopsies are stringent and met by some GI pathologists. The majority of assessors however, only showed limited deviation from BMS. We expect further training with group discussions leads to adherence of all GI pathologists to quality criteria, and therefore eligible to join the review panel.

INTRODUCTION

In Barrett's esophagus (BE), the normal stratified squamous epithelium of the distal esophagus has been replaced by columnar epithelium with or without goblet cells. Patients with BE have a risk to develop esophageal adenocarcinoma (EAC) and malignant transformation follows the metaplasia – dysplasia – carcinoma sequence. ¹ BE patients therefore undergo endoscopic surveillance. Low-grade dysplasia (LGD) in biopsies obtained during endoscopic surveillance is an accepted risk factor for progression, but diagnosis can be difficult and interobserver agreement is low. ^{2,3} Therefore, guidelines advise review of all dysplastic cases by a second, preferably expert pathologist. ⁴⁻¹¹

In the Netherlands we have initiated a national digital review panel for dysplastic BE, consisting of five core pathologists considered 'experts' in the field of BE. These five expert BE pathologists have been dedicated to the field of BE for a minimum of ten years, have a minimum caseload of five BE cases per week of which $\geq 25\%$ is dysplastic, have participated in multiple training programs (www.best-academia.eu), and have co-authored on >5 peer reviewed publications in this field. Moreover, it is the only BE expert pathologist group worldwide that has validated their BE diagnostic assessments in prospective clinical studies. ¹²⁻¹⁴ To optimize the throughput time of the Dutch digital review panel, to divide the workload and to gain nationwide coverage, we aim to expand the panel with ten gastrointestinal (GI) pathologists from the eight BE expert centers in the Netherlands. To maintain panel diagnostic quality we need to confirm that these pathologists' assessments correspond with the assessment standards of the current five core pathologists. In an earlier study, we defined benchmark quality criteria, based on the assessment of the core pathologists of a study set of 60 whole-endoscopy cases. ¹⁵ The aim of this study was to evaluate if the assessment scores of ten dedicated GI pathologists, for reviewing the same study set of 60 whole-endoscopy cases, fall within the predefined range for these benchmark quality criteria.

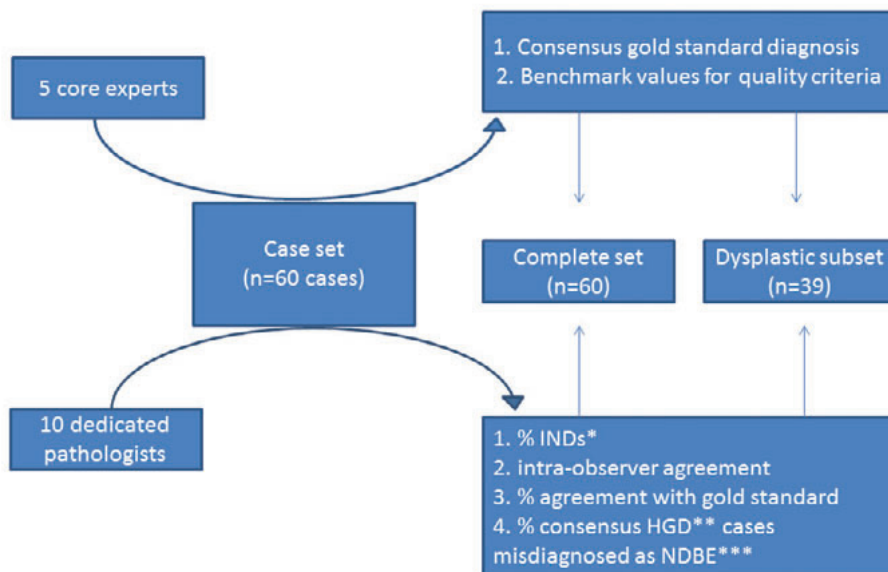
MATERIALS AND METHODS

Since the materials used in this study were anonymized, the medical ethical committee of the AMC waived the need for approval.

Case selection and slide scanning

For 60 patients who had an endoscopy for BE surveillance, we selected all formalin fixed, paraffin embedded tissue blocks and/or slides of the biopsies obtained during the endoscopy. The case set consisted of 39 cases with an original diagnosis of LGD (n=20) or high-grade dysplasia (HGD; n=19) that had been sent to our centre for consultation, between 2012 and 2014. These 39 dysplastic cases were supplemented with 21 consecutive non-dysplastic BE (NDBE) cases from a community hospital in the Amsterdam region. The five core expert BE pathologists had assessed this case set twice individually at an earlier stage followed by consensus meetings to create a gold standard diagnosis for all cases.¹⁵ Their scores were used to create benchmark values for the quality criteria.

Figure 1: Flowchart of study set-up



*indefinite for dysplasia, **high-grade dysplasia, ***non-dysplastic Barrett's esophagus

Assessors

The assessors were ten dedicated GI pathologists working in the 8 BE expert centers in the Netherlands. In accordance with Dutch guidelines, work-up and treatment of dysplastic BE is centralized in these specialized centers. All pathologists had been dedicated to the field of BE for a median of seven years (range 5-30 years) and had a median case load of seven BE cases per week (range 5-17), of which $\geq 25\%$ was dysplastic. All were actively practicing pathologists, considered experts by their peers, and had already participated in a joint training program consisting of evaluating and discussing 60 single-slide BE cases (35 dysplastic), of which results were published separately.¹⁶

Histological assessment of samples

For the current study, the pathologists independently assessed all 60 cases twice in random order, with a wash-out time of at least one month, scoring them according to the modified Vienna criteria for GI neoplasms.^{2,17} The p53 immunohistochemistry (IHC) was used as a diagnostic adjunct and scored according to the p53 decision rule developed earlier.¹⁶ They individually logged onto the virtual slide system to assess the cases and record the highest diagnostic grade per case. After the two assessment rounds, all cases that did not have a majority diagnosis were discussed in a face-to-face group discussion with all participants, for education purposes.

Definition of benchmark values for quality criteria

For each of the four quality criteria, a benchmark range had been calculated based on the scores of the five core pathologists in the aforementioned earlier study (see **Table 1**).¹⁵ The flow chart of the study can be appreciated in **Figure 1**.

Table 1: Values for benchmark quality criteria based on 95% prediction interval of 5 core pathologists ¹⁶

Quality criterium	95% PI [†] core pathologists all cases (n=60)	Benchmark value	95% PI core pathologists dysplastic cases (n=39)	Benchmark value
Percentage of IND* cases (%)	3-14%	≤14%	-2-16%	≤16%
Intra-observer agreement in 3 categories (K)	0.66-1.02	≥0.66	0.39-0.73	≥0.39
Agreement with consensus gold standard diagnosis (%)	82-98%	≥82%	73-104%	≥73%
Consensus HGD [‡] cases misdiagnosed as NDBE [§] (%; fraction)	0.8% (1/120 assessments)	≤0.8% (1/120 assessments)	≤1.3% (1/78 assessments)	≤1.3% (1/78 assessments)

*IND = indefinite for dysplasia; [†]PI = prediction interval; [‡]HGD = high-grade dysplasia; [§]NDBE = non-dysplastic Barrett's esophagus

Figure 2: Example of 4x4 cross table of pathologist against consensus gold standard diagnosis, showing the position of agreement, overdiagnosis and underdiagnosis

		Consensus gold standard diagnosis			
		NDBE	IND	LGD	HGD
Pathologist	NDBE		underdiagnosis		*
	IND				
	LGD	overdiagnosis		agreement with gold	
	HGD			standard diagnosis	

*number of consensus high-grade dysplasia cases misdiagnosed as non-dysplastic Barrett's esophagus

Outcome measurements

Per pathologist, we established whether the scores met all benchmark quality criteria, for the complete case set as well as for the dysplastic subset. The different outcome measurements were calculated as per our previous study.¹⁵ **Figure 2** illustrates the spectrum of diagnostic agreement, over- and underdiagnoses.

RESULTS

Baseline characteristics of samples in case set

Median age of patients at diagnosis was 66 years (IQR 13) and 73% was male. Cases contained a total of 151 sets of quadrant biopsies with a median of 5 slides (IQR 3-9), from a median of 2 levels (IQR 1-4) with 4 biopsies per level (IQR 3-4.5), with a total of 376 slides to be assessed.

Performance of 10 pathologists for the complete case series (n=60)

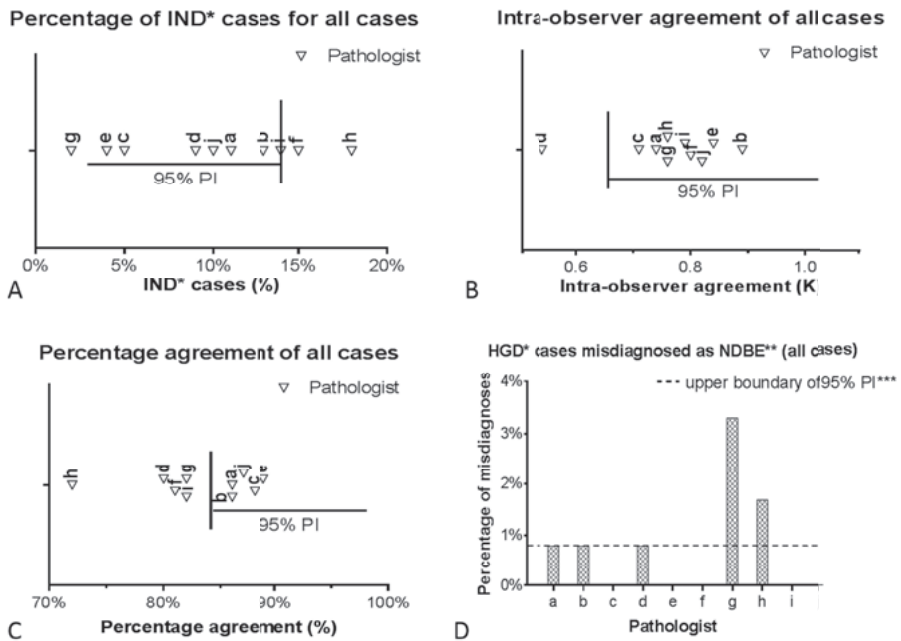
The pathologists generated a total of 1200 case diagnoses over 7520 assessed slides. For the percentage of IND cases, eight out of 10 pathologists met the benchmark value (see **Figure 3A**). For the intra-observer agreement, nine out of 10 pathologists fell within the benchmark value (see **Figure 3B**). For the percentage agreement with the consensus gold standard diagnosis, five out of 10 pathologists fell within the benchmark value (see **Figure 3C**). For the consensus HGD cases misdiagnosed as NDBE, eight pathologists fell within the benchmark value (see **Figure 3D**). In **Supplementary Table 1**, these results are visualized in cross tables per pathologist compared to the consensus gold standard diagnosis. For the complete case set, five out of 10 pathologists met the benchmark values for all four criteria.

Performance of 10 pathologists for subset of dysplastic cases (n=39)

For the percentage of IND cases, all pathologists fell within the benchmark value (see **Figure 3E**). For the intra-observer agreement, six out of 10 pathologists fell within the benchmark value (see **Figure 3F**). For the percentage agreement with the consensus gold standard diagnosis, nine out of 10 pathologists fell within the benchmark value (see **Figure 3G**). For the consensus HGD cases misdiagnosed as NDBE, eight out of 10 pathologists fell within the benchmark value (see **Figure 3H**). In **Supplementary Table 2** these results are visualized in cross tables per pathologist compared to the

consensus gold standard diagnosis. For the dysplastic subset, six out of 10 pathologists met the benchmark values for all four criteria.

Figure 3: (A-D) Performance of 10 gastrointestinal pathologists relative to benchmark criteria, for the complete case set

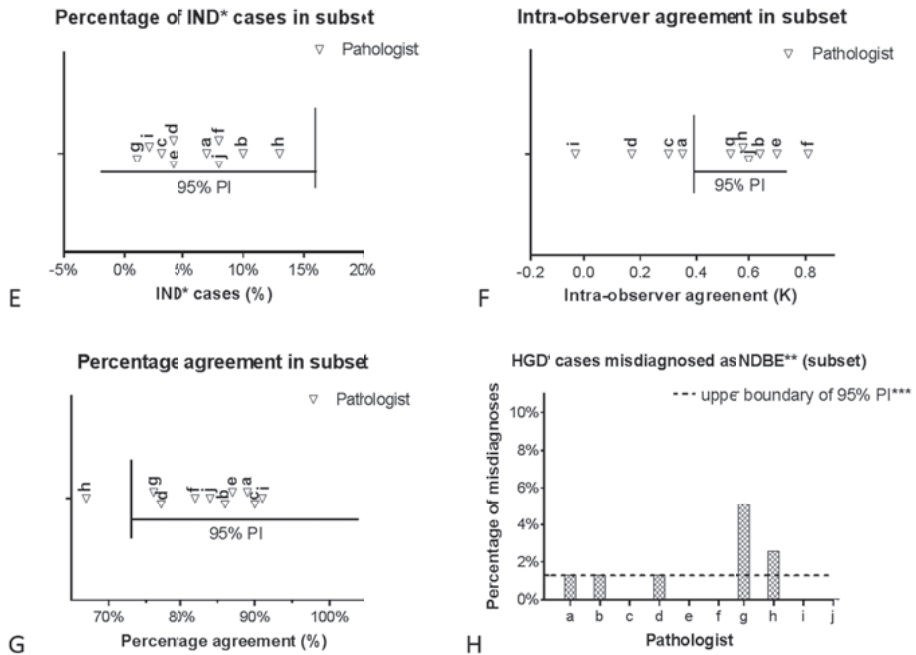


Vertical line; benchmark value, horizontal line; 95% prediction interval (PI)

Performance of pathologists relative to benchmark scores

Overall, three out of 10 pathologists met all benchmark values for the complete case set as well as for the dysplastic subset. When we extended our benchmark quality criteria by using a wider range, the 99% PI scores of the five core pathologists, an extra three pathologists met the benchmark range (results not shown). Four pathologists did not meet the 99% PI benchmark range on one quality criterion, namely the intra-observer agreement of the dysplastic subset, or the percentage of HGD gold standard cases misdiagnosed as NDBE.

Figure 3: (E-H) Performance of 10 gastrointestinal pathologists relative to benchmark criteria, for the dysplastic subset



Vertical line; benchmark value, horizontal line; 95% prediction interval (PI)

DISCUSSION

The aim of this study was to test if ten GI pathologists working at the eight BE expert centers in the Netherlands met pre-set benchmark scores of pre-defined quality criteria for evaluating BE biopsies (see **Table 1**), by assessing a case set of 60 BE cases consisting of 376 slides. To our knowledge, this is the first time histopathological expertise is quantified in the assessment of dysplastic BE biopsies. These criteria and benchmark values were established in an earlier study by using the assessments of five core expert BE pathologists as reference. These five pathologists are considered experts in BE diagnostics according to previously defined criteria: they have been dedicated to the field of BE for at least 15 years (range 15-45 years); have a median case load of 7 cases per week (range 5-15), of which $\geq 25\%$ is dysplastic; they participated in the Dutch Barrett advisory committee for 5-13 years¹⁸⁻²⁰ and have co-authored more than ten peer reviewed publications in this field.^{12-16 18 20-30} To create a consensus gold

standard diagnosis for this case set, these five core pathologists assessed the same dataset as used in this study, two times independently, followed by group discussions.

In the current study, the boundaries of the 95% PI and 99% PI of their individual scores were used as benchmark ranges to assess the performance of ten dedicated GI pathologists working at the eight Dutch BE expert centers. These ten pathologists have been dedicated to the field of BE with varying levels of experience (median of 7 years (range 5-30), minimum case load of 7 BE biopsies per week (range 5-17) of which $\geq 25\%$ is dysplastic, however they did not have the intensive collaboration that the five core pathologists had. When comparing their assessments to the benchmark scores, we found that three out of ten GI pathologists met all pre-set benchmark ranges for the quality criteria, for the complete case set as well as for the subset of dysplastic cases. Performance according to the benchmark range of the dysplastic subset is of key importance, since this subset is the main patient population for review requests by the national digital review panel. Their adherence to the benchmark ranges implies that these three pathologists perform similar to the five core members in their diagnostic assessment. Expanding the digital revision panel with these three pathologists would therefore not compromise the current assessment homogeneity.

The results of our study need to be interpreted with caution. First of all, we have used the intra-observer agreement (weighted kappa) as an indirect measure of expertise, because it underscores the individual reproducibility of the pathologist. However, calculating a kappa score can be less reliable when marginal totals are skewed, leading to a high chance agreement and therefore a low kappa score. This is of particular relevance for the subanalysis of the dysplastic cases (where it is amplified by the low numbers in the subanalysis). Taking these aspects of the kappa calculation into account, we feel that the intra-observer agreement of the subanalysis is less reliable as a benchmark score than measuring diversions from the consensus gold standard diagnoses, i.e. the percentage agreement per pathologist. The percentage agreement of nine out of 10 pathologists falls within the 95% PI benchmark range for this criterion. This outcome signifies correct detection of dysplastic cases (**Figure 3G**). If we did not take the intra-observer agreement into account, one additional pathologist meets the predefined benchmark values. Second, the 95% PI benchmark range of percentage of cases 'indefinite for dysplasia' is inflated compared to clinical practice. This is explained by the fact that our case set was strongly enriched for difficult dysplastic cases, as

encountered in a review panel setting, and by the fact that we used a p53 decision rule in the interpretation of p53 IHC as a diagnostic adjunct.¹⁶ Moreover, the current study is part of a structured training program and after the individual assessments presented here, the 10 GI pathologists participated in face-to-face plenary group meetings, discussing cases that were discrepant with the consensus gold standard diagnosis. Importantly, after completion of this study set, all pathologists have assessed a case set of 62 endoscopic resection cases, and are currently reviewing 40 cases sent to the national digital review panel. These assessments were again combined with face-to-face plenary group meetings, to discuss difficult and discrepant cases. This will further improve the experience and homogeneity of panel members. We aim to reevaluate their performance in the near future, and consequently expect more pathologists to meet the benchmark quality criteria presented here.

This study has some limitations. The benchmark quality criteria used in this study depend on the distribution of diagnoses in this dataset and the individual scores of the five core pathologists. The benchmark scores only apply to this specific digital study set and the number and scores of the core pathologists. However, because there is no standardized way to define expertise in BE diagnostics, we feel that these benchmark quality criteria are currently the best choice to quantify expertise when diagnosing BE dysplasia in biopsy samples in the Netherlands.

This study is unique because of a number of features. First, it is part of a structured approach to guarantee quality and uniformity of histological diagnosis of BE biopsies in the Netherlands. Over the past five years, our group has set up a national digital review panel for BE after conducting five preliminary studies.^{15 16 30} This is the first time worldwide an expert pathology review panel has been set-up conducting such quantifiable preliminary work in such a meticulous way. Second, the case set used for this study consisted of all slides from all biopsy levels of a single endoscopy (376 slides in total), was fully digitalized and only contained review cases from clinical practice. There were two assessment rounds with an adequate wash-out time. In order to improve homogeneity of the group the pathologists held a group discussion afterwards to discuss cases that did not have a majority diagnosis. This digital case set of dysplastic BE cases will be made available to allow pathologists in- and outside the Netherlands to evaluate if they meet the aforementioned benchmark ranges for quality criteria.

Our goal for the future remains to improve the knowledge of BE-related diagnostic pathology among GI pathologists in the Netherlands; and to include all GI pathologists working at the BE expert centers in the Netherlands in our review panel. For this, we first need to ensure quality and homogeneity of the panel as outlined above. Subsequently, we need a prediction model that allows us to efficiently select the number of pathologists needed for reviewing cases and to divide the workload equally among panel members. We aim to improve and expand training in BE pathology both nationally as well as internationally by constructing a freely available, accredited training module incorporating the information gathered from all study sets and group discussions thus far.^{15 16 30} In this way, pathologists with an interest in BE can train themselves and reflect on their performance relative to the benchmark scores of the training set.

REFERENCES

1. 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.
2. 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.
3. 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
4. Fitzgerald RC, di Pietro M, Ragnunath 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
5. Spechler SJ, Sharma P, Souza RF, et al. American Gastroenterological Association medical position statement on the management of Barrett's esophagus. *Gastroenterology* 2011;140(3):1084-91. doi: 10.1053/j.gastro.2011.01.030
6. Fock KM, Talley N, Goh KL, et al. Asia-Pacific consensus on the management of gastro-oesophageal reflux disease: an update focusing on refractory reflux disease and Barrett's oesophagus. *Gut* 2016;65(9):1402-15. doi: 10.1136/gutjnl-2016-311715
7. Shaheen NJ, Falk GW, Iyer PG, et al. ACG Clinical Guideline: Diagnosis and Management of Barrett's Esophagus. *The American journal of gastroenterology* 2016;111(1):30-50. doi: 10.1038/ajg.2015.322
8. Wang KK, Sampliner RE, Practice Parameters Committee of the American College of G. Updated guidelines 2008 for the diagnosis, surveillance and therapy of Barrett's esophagus. *The American journal of gastroenterology* 2008;103(3):788-97. doi: 10.1111/j.1572-0241.2008.01835.x
9. 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
10. Whiteman DC, Appleyard M, Bahin FF, et al. Australian clinical practice guidelines for the diagnosis and management of Barrett's Esophagus and Early Esophageal Adenocarcinoma. *Journal of gastroenterology and hepatology* 2015 doi: 10.1111/jgh.12913
11. Maag-Darm-Leverartsen Nvv. Richtlijn Barrett-oesofagus. www.richtlijndatabase.nl: IKNL, 2018:1-71.
12. Duits LC, van der Wel MJ, Cotton CC, et al. Patients With Barrett's Esophagus and Confirmed Persistent Low-Grade Dysplasia Are at Increased Risk for Progression to Neoplasia. *Gastroenterology* 2017;152(5):993-1001 e1. doi: 10.1053/j.gastro.2016.12.008
13. 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

14. 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
15. van der Wel MJ, Duits LC, Klaver E, et al. Development of benchmark quality criteria for assessing whole-endoscopy Barrett's esophagus biopsy cases. *United European gastroenterology journal* 2018;6(6):830-37. doi: 10.1177/2050640618764710 [published Online First: 2018/07/20]
16. van der Wel MJ, Duits LC, Pouw RE, et al. Improved diagnostic stratification of digitised Barrett's oesophagus biopsies by TP53 immunohistochemical staining. *Histopathology* 2018;72(6):1015-23. doi: 10.1111/his.13462 [published Online First: 2018 Feb 28]
17. 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
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. 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.
20. 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
21. 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
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. van der Wel MJ, Duits LC, Seldenrijk CA, et al. Digital microscopy as valid alternative to conventional microscopy for histological evaluation of Barrett's esophagus biopsies. *Diseases of the esophagus : official journal of the International Society for Diseases of the Esophagus / ISDE* 2017;30(11):1-7. doi: 10.1093/dote/dox078

Supplementary Table 1: Agreement of 10 GI* pathologists with consensus gold standard diagnosis (mean over 2 assessment rounds) for the complete case set (n=60) in 4 categories (4x4 cross tables)

Observed		Observed (%)									
		GS					GS				
		NDBE	IND	LGD	HGD	Total	NDBE	IND	LGD	HGD	Total
P1	NDBE	17.0	1.0	2.5	0.5	21.0	28.4%	1.7%	4.2%	0.9%	35.0%
	IND	1.0	2.0	2.5	1.0	6.5	1.7%	3.3%	4.2%	1.7%	10.9%
	LGD	0.0	0.0	9.5	5.0	14.5	0.0%	0.0%	15.9%	8.4%	24.2%
	HGD	0.0	0.0	4.5	13.5	18.0	0.0%	0.0%	7.5%	22.5%	30.0%
	Total	18.0	3.0	19.0	20.0	60.0	30.0%	5.0%	31.7%	33.3%	100.0%
Observed		Observed (%)									
		GS					GS				
		NDBE	IND	LGD	HGD	Total	NDBE	IND	LGD	HGD	Total
P2	NDBE	18.0	1.5	0.0	0.5	20.0	30.0%	2.5%	0.0%	0.9%	33.3%
	IND	0.0	1.5	6.5	0.0	8.0	0.0%	2.5%	10.9%	0.0%	13.4%
	LGD	0.0	0.0	11.0	3.5	14.5	0.0%	0.0%	18.3%	5.9%	24.2%
	HGD	0.0	0.0	1.5	16.0	17.5	0.0%	0.0%	2.5%	26.7%	29.2%
	Total	18.0	3.0	19.0	20.0	60.0	30.0%	5.0%	31.7%	33.3%	100.0%
Observed		Observed (%)									
		GS					GS				
		NDBE	IND	LGD	HGD	Total	NDBE	IND	LGD	HGD	Total
P3	NDBE	15.0	0.5	0.0	0.0	15.5	25.0%	1%	0.0%	0.0%	25.8%
	IND	1.0	0.5	1.0	0.5	3.0	1.7%	0.8%	1.7%	0.8%	5.0%
	LGD	1.5	2.0	16.0	5.0	24.5	2.5%	3.3%	26.7%	8.3%	40.8%
	HGD	0.5	0.0	2.0	14.5	17.0	0.9%	0.0%	3.4%	24.2%	28.3%
	Total	18.0	3.0	19.0	20.0	60.0	30.0%	5.0%	31.7%	33.3%	100.0%
Observed		Observed (%)									
		GS					GS				
		NDBE	IND	LGD	HGD	Total	NDBE	IND	LGD	HGD	Total
P4	NDBE	16.5	1.0	4.0	0.5	22.0	27.5%	1.7%	6.7%	0.9%	36.7%
	IND	1.5	0.5	3.0	0.5	5.5	2.5%	0.9%	5.0%	0.9%	9.2%
	LGD	0.0	1.5	10.5	8.5	20.5	0.0%	2.5%	17.5%	14.2%	34.2%
	HGD	0.0	0.0	1.5	10.5	12.0	0.0%	0.0%	2.5%	17.5%	20.0%
	Total	18.0	3.0	19.0	20.0	60.0	30.0%	5.0%	31.7%	33.3%	100.0%

Observed		Observed (%)									
		GS					GS				
		NDBE	IND	LGD	HGD	Total	NDBE	IND	LGD	HGD	Total
P5	NDBE	18.0	2.5	1.0	0.0	21.5	30.0%	4.2%	1.7%	0.0%	35.9%
	IND	0.0	0.0	2.0	0.5	2.5	0.0%	0.0%	3.4%	0.9%	4.2%
	LGD	0.0	0.5	13.5	9.5	23.5	0.0%	0.9%	22.5%	15.8%	39.2%
	HGD	0.0	0.0	2.	10.0	12.5	0.0%	0.0%	4.2%	16.7%	20.8%
	Total	18.0	3.0	19.0	20.0	60.0	30.0%	5.0%	31.7%	33.3%	100.0%
Observed		Observed (%)									
		GS					GS				
		NDBE	IND	LGD	HGD	Total	NDBE	IND	LGD	HGD	Total
P6	NDBE	13.5	0.0	0.0	0.0	13.5	22.5%	0.0%	0.0%	0.0%	22.5%
	IND	4.0	0.5	4.0	0.5	9.0	6.7%	0.9%	6.7%	0.9%	15.0%
	LGD	0.5	2.5	14.5	6.0	23.5	0.9%	4.2%	24.2%	10.0%	39.2%
	HGD	0.0	0.0	0.5	13.5	14.0	0.0%	0.0%	0.9%	22.5%	23.4%
	Total	18.0	3.0	19.0	20.0	60.0	30.0%	5.0%	31.7%	33.3%	100.0%
Observed		Observed (%)									
		GS					GS				
		NDBE	IND	LGD	HGD	Total	NDBE	IND	LGD	HGD	Total
P7	NDBE	17.0	2.0	5.0	2.0	26.0	28.4%	3.3%	8.4%	3.3%	43.3%
	IND	1.0	0.0	0.0	0.0	1.0	1.7%	0.0%	0.0%	0.0%	1.7%
	LGD	0.0	1.0	13.5	13.5	28.0	0.0%	1.7%	22.5%	22.5%	46.7%
	HGD	0.0	0.0	0.5	4.5	5.0	0.0%	0.0%	0.9%	7.5%	8.3%
	Total	18.0	3.0	19.0	20.0	60.0	30.0%	5.0%	31.7%	33.3%	100.0%
Observed		Observed (%)									
		GS					GS				
		NDBE	IND	LGD	HGD	Total	NDBE	IND	LGD	HGD	Total
P8	NDBE	16.5	2.0	3.0	1.0	22.5	27.5%	3.3%	5.0%	1.7%	37.5%
	IND	1.5	0.5	7.5	1.5	11.0	2.5%	0.9%	12.5%	2.5%	18.0%
	LGD	0.0	0.5	5.5	7.0	13.0	0.0%	0.9%	9.2%	11.7%	21.7%
	HGD	0.0	0.0	3.0	10.5	13.5	0.0%	0.0%	5.0%	17.5%	22.5%
	Total	18.0	3.0	19.0	20.0	60.0	30.0%	5.0%	31.7%	33.3%	100.0%

		Observed					Observed (%)				
		GS					GS				
		NDBE	IND	LGD	HGD	Total	NDBE	IND	LGD	HGD	Total
P9	NDBE	10.0	0.0	0.0	0.0	10.0	16.7%	0.0%	0.0%	0.0%	16.7%
	IND	6.5	1.0	1.0	0.0	8.5	10.9%	1.7%	1.7%	0.0%	14.2%
	LGD	1.5	2.0	15.0	3.5	22.0	2.5%	3.3%	25.0%	5.8%	36.7%
	HGD	0.0	0.0	3.0	16.5	19.5	0.0%	0.0%	5.0%	27.5%	32.5%
	Total	18.0	3.0	19.0	20.0	60.0	30.0%	5.0%	31.7%	33.3%	100.0%
		Observed					Observed (%)				
		GS					GS				
		NDBE	IND	LGD	HGD	Total	NDBE	IND	LGD	HGD	Total
P10	NDBE	17.5	1.5	1.0	0.0	20.0	29.2%	2.5%	1.7%	0.0%	33.3%
	IND	0.5	1.0	4.5	0.0	6.0	0.9%	1.7%	7.5%	0.0%	10.0%
	LGD	0.0	0.5	11.0	3.0	14.5	0.0%	0.9%	18.3%	5.0%	24.2%
	HGD	0.0	0.0	2.5	17.0	19.5	0.0%	0.0%	4.2%	28.3%	32.5%
	Total	18.0	3.0	19.0	20.0	60.0	30.0%	5.0%	31.7%	33.3%	100.0%

*GI = gastrointestinal; †NDBE = non-dysplastic Barrett's esophagus; ‡IND = indefinite for dysplasia; §LGD = low-grade dysplasia; ¶HGD = high-grade dysplasia

Supplementary Table 2: Agreement of 10 GI* pathologists with consensus gold standard diagnosis (mean over 2 assessment rounds) for the dysplastic subset (n=39) in 4 categories (4x4 cross tables)

		Observed					Observed (%)				
		GS					GS				
		NDBE	IND	LGD	HGD	Total	NDBE	IND	LGD	HGD	Total
P1	NDBE	1.0	0.0	2.0	0.5	3.5	2.6%	0.0%	5.1%	1.3%	9.0%
	IND	0.0	2.0	1.0	1.0	4.0	0.0%	5.1%	2.6%	2.6%	10.3%
	LGD	0.0	0.0	9.5	5.0	14.5	0.0%	0.0%	24.4%	12.8%	37.2%
	HGD	0.0	0.0	3.5	13.5	17.0	0.0%	0.0%	9.0%	34.7%	43.6%
	Total	1.0	2.0	16.0	20.0	39.0	2.6%	5.1%	41.0%	51.3%	100.0%

Observed		Observed (%)									
		GS					GS				
		NDBE	IND	LGD	HGD	Total	NDBE	IND	LGD	HGD	Total
P2	NDBE	1.0	0.5	0.0	0.5	2.0	2.6%	1.3%	0.0%	1.3%	5.1%
	IND	0.0	1.5	4.5	0.0	6.0	0.0%	3.9%	11.6%	0.0%	15.4%
	LGD	0.0	0.0	10.0	3.5	13.5	0.0%	0.0%	25.6%	9.0%	34.6%
	HGD	0.0	0.0	1.5	16.0	17.5	0.0%	0.0%	3.9%	41.0%	44.9%
	Total	1.0	2.0	16.0	20.0	39.0	2.6%	5.1%	41.0%	51.3%	100.0%
Observed		Observed (%)									
		GS					GS				
		NDBE	IND	LGD	HGD	Total	NDBE	IND	LGD	HGD	Total
P3	NDBE	0.5	0.0	0.0	0.0	0.5	1.3%	0.0%	0.0%	0.0%	1.3%
	IND	0.5	0.0	1.0	0.5	2.0	1.3%	0.0%	2.6%	1.3%	5.1%
	LGD	0.0	2.0	13.5	5.0	20.5	0.0%	5.1%	34.7%	12.8%	52.6%
	HGD	0.0	0.0	1.5	14.5	16.0	0.0%	0.0%	3.9%	37.2%	41.1%
	Total	1.0	2.0	16.0	20.0	39.0	2.6%	5.1%	41.0%	51.3%	100.0%
Observed		Observed (%)									
		GS					GS				
		NDBE	IND	LGD	HGD	Total	NDBE	IND	LGD	HGD	Total
P4	NDBE	1.0	0.5	4.0	0.5	6.0	2.6%	1.3%	10.3%	1.3%	15.4%
	IND	0.0	0.0	2.0	0.5	2.5	0.0%	0.0%	5.2%	1.3%	6.4%
	LGD	0.0	1.5	8.5	8.5	18.5	0.0%	3.9%	21.8%	21.8%	47.4%
	HGD	0.0	0.0	1.5	10.5	12.0	0.0%	0.0%	3.9%	27.0%	30.8%
	Total	1.0	2.0	16.0	20.0	39.0	2.6%	5.1%	41.0%	51.3%	100.0%
Observed		Observed (%)									
		GS					GS				
		NDBE	IND	LGD	HGD	Total	NDBE	IND	LGD	HGD	Total
P5	NDBE	1.0	2.0	0.5	0.0	3.5	2.6%	5%	1.3%	0.0%	9.0%
	IND	0.0	0.0	2.0	0.5	2.5	0.0%	0.0%	5.2%	1.3%	6.4%
	LGD	0.0	0.0	11.0	9.5	20.5	0.0%	0.0%	28.2%	24.4%	52.6%
	HGD	0.0	0.0	2.5	10.0	12.5	0.0%	0.0%	6.4%	25.7%	32.1%
	Total	1.0	2.0	16.0	20.0	39.0	2.6%	5.1%	41.0%	51.3%	100.0%

Observed		Observed (%)									
		GS					GS				
		NDBE	IND	LGD	HGD	Total	NDBE	IND	LGD	HGD	Total
P6	NDBE	0.5	0.0	0.0	0.0	0.5	1.3%	0.0%	0.0%	0.0%	1.3%
	IND	0.5	0.0	4.0	0.5	5.0	1.3%	0.0%	10.3%	1.3%	12.9%
	LGD	0.0	2.0	11.5	6.0	19.5	0.0%	5.1%	29.5%	15.4%	50.0%
	HGD	0.0	0.0	0.5	13.5	14.0	0.0%	0.0%	1.3%	34.6%	35.9%
	Total	1.0	2.0	16.0	20.0	39.0	2.6%	5.1%	41.0%	51.3%	100.0%
Observed		Observed (%)									
		GS					GS				
		NDBE	IND	LGD	HGD	Total	NDBE	IND	LGD	HGD	Total
P7	NDBE	0.5	1.0	5.0	2.0	8.5	1.3%	2.6%	12.9%	5.1%	21.8%
	IND	0.5	0.0	0.0	0.0	0.5	1.3%	0.0%	0.0%	0.0%	1.3%
	LGD	0.0	1.0	10.5	13.5	25.0	0.0%	2.6%	26.9%	34.6%	64.1%
	HGD	0.0	0.0	0.5	4.5	5.0	0.0%	0.0%	1.3%	11.6%	12.8%
	Total	1.0	2.0	16.0	20.0	39.0	2.6%	5.1%	41.0%	51.3%	100.0%
Observed		Observed (%)									
		GS					GS				
		NDBE	IND	LGD	HGD	Total	NDBE	IND	LGD	HGD	Total
P8	NDBE	1.0	1.0	3.0	1.0	6.0	2.6%	2.6%	7.7%	2.6%	15.4%
	IND	0.0	0.5	6.0	1.5	8.0	0.0%	1.3%	15.4%	3.9%	20.5%
	LGD	0.0	0.5	5.0	7.0	12.5	0.0%	1.3%	12.8%	18.0%	32.1%
	HGD	0.0	0.0	2.0	10.5	12.5	0.0%	0.0%	5.1%	26.9%	32.1%
	Total	1.0	2.0	16.0	20.0	39.0	2.6%	5.1%	41.0%	51.3%	100.0%
Observed		Observed (%)									
		GS					GS				
		NDBE	IND	LGD	HGD	Total	NDBE	IND	LGD	HGD	Total
P9	NDBE	0.5	0.0	0.0	0.0	0.5	1.3%	0.0%	0.0%	0.0%	1.3%
	IND	0.0	0.0	1.0	0.0	1.0	0.0%	0.0%	2.6%	0.0%	2.6%
	LGD	0.5	2.0	12.5	3.5	18.5	1.3%	5.1%	32.1%	9.0%	47.5%
	HGD	0.0	0.0	2.5	16.5	19.0	0.0%	0.0%	6.5%	42.4%	48.7%
	Total	1.0	2.0	16.0	20.0	39.0	2.6%	5.1%	41.0%	51.3%	100.0%

Observed		Observed (%)									
		GS					GS				
		NDBE	IND	LGD	HGD	Total	NDBE	IND	LGD	HGD	Total
P10	NDBE	1.0	1.0	1.0	0.0	0.5	2.6%	2.6%	2.6%	0.0%	7.7%
	IND	0.0	0.5	4.0	0.0	1.0	0.0%	1.3%	10.3%	0.0%	11.6%
	LGD	0.0	0.5	9.0	3.0	18.5	0.0%	1.3%	23.1%	7.7%	32.1%
	HGD	0.0	0.0	2.0	17.0	19.0	0.0%	0.0%	5.2%	43.6%	48.8%
	Total	1.0	2.0	16.0	20.0	39.0	2.6%	5.1%	41.0%	51.3%	100.0%

*GI = gastrointestinal; †NDBE = non-dysplastic Barrett's esophagus; ‡IND = indefinite for dysplasia;
§LGD = low-grade dysplasia; ¶HGD = high-grade dysplasia