



OPEN ACCESS

# Performance of gastrointestinal pathologists within a national digital review panel for Barrett's oesophagus in the Netherlands: results of 80 prospective biopsy reviews

Esther Klaver ,<sup>1</sup> Myrtle van der Wel ,<sup>1,2</sup> Lucas Duits,<sup>1</sup> Roos Pouw,<sup>1</sup> Kees Seldenrijk,<sup>3</sup> Johan Offerhaus,<sup>4</sup> Mike Visser,<sup>5</sup> Fiebo ten Kate,<sup>4</sup> Katharina Biermann,<sup>6</sup> Lodewijk Brosens ,<sup>4</sup> Michael Doukas,<sup>6</sup> Clément Huysentruyt,<sup>7</sup> Arend Karrenbeld ,<sup>8</sup> Gursah Kats-Ugurlu,<sup>8</sup> Jaap van der Laan,<sup>9</sup> Ineke van Lijnschoten,<sup>7</sup> Freek Moll,<sup>10</sup> Ariadne Ooms,<sup>11</sup> Jan Tijssen,<sup>12</sup> Sybren Meijer,<sup>2</sup> Jacques Bergman<sup>1</sup>

For numbered affiliations see end of article.

## Correspondence to

Esther Klaver, Gastroenterology and Hepatology, Amsterdam UMC - Locatie AMC, Amsterdam 1105 AZ, Netherlands; e.klaver@amsterdamumc.nl

EK and MvdW contributed equally.

Received 13 February 2020  
Revised 17 April 2020  
Accepted 1 May 2020

## ABSTRACT

**Aims** The histopathological diagnosis of low-grade dysplasia (LGD) in Barrett's oesophagus (BO) is associated with poor interobserver agreement and guidelines dictate expert review. To facilitate nationwide expert review in the Netherlands, a web-based digital review panel has been set up, which currently consists of eight 'core' pathologists. The aim of this study was to evaluate if other pathologists from the Dutch BO expert centres qualify for the expert panel by assessing their performance in 80 consecutive LGD reviews submitted to the panel.

**Methods** Pathologists independently assessed digital slides in two phases. Both phases consisted of 40 cases, with a group discussion after phase I. For all cases, a previous consensus diagnosis made by five core pathologists was available, which was used as reference. The following criteria were used: (1) percentage of 'indefinite for dysplasia' diagnoses, (2) percentage agreement with consensus diagnosis and (3) proportion of cases with a consensus diagnosis of dysplasia underdiagnosed as non-dysplastic. Benchmarks were based on scores of the core pathologists.

**Results** After phase I, 1/7 pathologists met the benchmark score for all quality criteria, yet three pathologists only marginally failed the agreement with consensus diagnosis (score 68.3%, benchmark 69%). After a group discussion and phase II, 5/6 remaining aspirant panel members qualified with all scores within the benchmark range.

**Conclusions** The Dutch BO review panel now consists of 14 pathologists, who—after structured assessments and group discussions—can be considered homogeneous in their review of biopsies with LGD.

## INTRODUCTION

Low-grade dysplasia (LGD) is an important histological risk factor for progression to oesophageal adenocarcinoma in patients with Barrett's oesophagus (BO). The histological diagnosis of LGD is however challenging, because early morphological changes of dysplasia are difficult to distinguish from

reactive atypia of reflux-induced inflammation. As a result, the interobserver agreement for the diagnosis of LGD is poor, and LGD is widely overdiagnosed in community practice. Two studies from our group have shown that 73%–85% of the community LGD cases are downstaged to non-dysplastic BO (NDBO) by BO expert pathologists and that these downstaged cases have a low progression risk during follow-up. However, if the diagnosis LGD is confirmed by an expert BO pathologist, the risk of neoplastic progression is approximately 10% per patient-year.<sup>1,2</sup> This risk increases significantly if multiple expert pathologists agree on this diagnosis.<sup>3</sup>

New guidelines dictate review of all dysplastic BO cases by an expert pathologist,<sup>4–7</sup> but the concept of 'expert pathologist' is ill-defined, and access to expert review is not widely available. To facilitate expert review in the Netherlands, a national, web-based digital histology review platform has been set-up by the eight BO expert centres in the Netherlands. All diagnostic work-up and treatment of early BO neoplasia is centralised in these eight centres. The histology review panel was built around five 'core pathologists'. These 'core' BO expert pathologists had a track record in the field of BO of >10 years (range 10–30 years), had participated in multiple teaching programmes (ie, [www.best-academia.eu](http://www.best-academia.eu)) and had each coauthored on >10 peer reviewed publications in this field.<sup>8–12</sup>

In a prior study, we evaluated if we could expand the number of pathologists in the panel to reduce the individual workload, reduce lead time and obtain nationwide coverage while maintaining the assessment quality and homogeneity.<sup>12</sup> For this, 10 other gastrointestinal (GI) pathologists from the eight BO expert centres assessed digitalised slides of 60 endoscopy procedures, enriched for dysplasia. This case set had also been assessed by the five core pathologists. Three of the 10 pathologists met the stringent benchmark quality criteria for the case set, as established based on the results of the five core pathologists. Therefore, these three pathologists



© Author(s) (or their employer(s)) 2020. Re-use permitted under CC BY. Published by BMJ.

**To cite:** Klaver E, van der Wel M, Duits L, et al. *J Clin Pathol* Epub ahead of print: [please include Day Month Year]. doi:10.1136/jclinpath-2020-206511

**Table 1** Patient and endoscopy characteristics

	Cases phase I (n=41)	Cases phase II (n=39)
Age at endoscopy in years (median, IQR)	67 (61–72)	66 (59–70)
Sex (male) (n, %)	26 (68)	31 (80%)
Levels biopsied (median, IQR)	2 (1–4)	2 (1–3)
Total biopsies (median, IQR)	8 (4–12)	7 (3–11)
Initial community diagnosis, n (%)		
IND	16 (39)	16 (41)
LGD	25 (61)	23 (59)

IND, indefinite for dysplasia; IQR, interquartile range; LGD, low-grade dysplasia.

were considered to be homogeneous in their histological assessment with the five core pathologists and joined the core group of the expert panel. The majority of the other assessors showed only limited deviation from the preset benchmark scores. After this structured assessment, all pathologists participated in face-to-face group meetings where discrepant cases of the assessment rounds were discussed. Although we speculate that these assessments and group discussions will likely have improved their assessments and homogeneity, no formal evaluation has been performed on their performance.

Meanwhile, the national review panel has become operational. This platform is driven by reviews of the eight core pathologists and concentrates on revisions for alleged LGD and indefinite for dysplasia (IND) from centres throughout the Netherlands.

The aim of this study was to evaluate the performance of all pathologists of the eight BO expert centres in the assessment of the first 80 prospective LGD reviews submitted to the Dutch Barrett's Pathology Review panel and to assess if the 'non-core' pathologists were homogeneous in their assessments with the eight core pathologists.

## METHODS

### Case submission and scanning

Eighty BO surveillance cases with a diagnosis of IND or LGD during surveillance endoscopy were submitted to the Dutch Barrett's Pathology Review panel (table 1).

Requests for consultation were submitted by gastroenterologists or pathologists via a dedicated website ([www.barrett.nl](http://www.barrett.nl)), on which pathology slides were requested by the back office of the review panel. All slides were digitalised, using a scanner with a  $\times 20$  microscope objective (Slide, Olympus, Tokyo, Japan). Digitalised slides were checked for focus and acuity by the study coordinator, rescanned if necessary and stored on a secure server. Subsequently pathologists were invited by email to review the slides.

### Histological assessment

All 15 pathologists independently assessed the digital cases through a virtual slide system (Digital Slidebox 4.5, Leica Microsystems, Dublin, Ireland) allowing them to assess the slides similarly to their conventional microscopic diagnostic practice. Cases were scored according to the modified Vienna criteria for GI neoplasms.<sup>13</sup> Diagnostic categories were: NDBO, LGD, high-grade dysplasia (HGD) or more, or IND.

The five initial core expert BO pathologists individually assessed all cases at an earlier stage. Group meetings were organised to discuss all cases in which their individual scores did not show a 4/5 or 5/5 agreement. As such a consensus diagnosis was generated as a reference for all pathologists' individual assessments.

All pathologists assessed the first 41 cases in phase I, after which a group discussion was held to discuss discrepant cases. Cases were considered discrepant if three or more pathologists disagreed with the consensus diagnosis. The group discussion consisted of two sessions with one teleconference and one face-to-face meeting. The cases were shown on screen and discussed plenary. A total of nine cases was discussed. Pathologists had access to their own score and the consensus diagnosis after the group discussion but were unaware about their relative scores. After the group discussion, pathologists individually assessed another 39 cases in the phase II.

### Quality criteria

As described in our previous study,<sup>12</sup> we used the following outcome parameters to quantify the quality of the individual histological assessment of each pathologist:

1. Percentage of diagnosis of IND per pathologist.
2. Percentage agreement with the consensus diagnosis per pathologist, using three diagnostic categories (NDBO, IND and LGD+HGD).
3. Percentage of cases with a consensus diagnosis of LGD or HGD underdiagnosed as NDBO per pathologist.

For these three quality criteria, we created benchmark scores based on the individual scores of the initial five core pathologists supplemented by the individual scores of the three pathologists who qualified as a core member based on their scores in the aforementioned study in which digitalised slides of 60 surveillance endoscopies were assessed.<sup>12</sup>

Benchmark ranges for each of the three criteria were based on a 95% prediction interval (PI) of the individual scores of these eight core pathologists. The 95% PI was calculated as the mean score from the eight pathologists  $\pm 2.365$  times the SD (based on a t-distribution with 7 df, since  $n=8$  pathologists). The upper or lower limit of these ranges, depending on the criterion, were used as benchmark values. Pathologists who met the benchmark score for all three quality criteria during phase I were added to the core, and benchmark scores for phase II were calculated based on the new extended core panel.

Anonymised materials were used in this study, hereby waiving the need for ethical approval by the medical ethical committee of the Amsterdam UMC or obtaining informed consent.

The statistical analyses were performed using the Statistical Package for the Social Sciences (SPSS V.24.0).

## RESULTS

### Baseline characteristics of samples in case sets

Baseline characteristics of patients, endoscopic findings, number of biopsies obtained and original histological diagnoses of phase I and phase II were similar and are shown in table 1.

### Performance of pathologists in phase I

For the percentage of IND cases, six of seven aspirant panel members met the benchmark value of 28% with percentages ranging from 7.3% to 14.6%. Only one pathologist did not meet the required performance score and diagnosed 31.7% of cases as IND (table 2).

For the percentage agreement with the consensus diagnosis, one pathologist met the benchmark lower limit of 69%, while three others just fell outside the range with a score of 68.3%.

The benchmark value for the percentage of consensus LGD and HGD cases underdiagnosed as NDBO was 11%, and six pathologists fell within this value. One pathologist clearly did

**Table 2** Results of all pathologists for phase I (n=41)

Pathologist	Percentage of cases 'indefinite for dysplasia'	Percentage agreement (three categories)*	Consensus LGD and HGD cases underdiagnosed as NDBO (n (%)) n=28
Core pathologists			
1	24.4	82.9	0
2	9.8	80.5	2 (7.1)
3	22.0	80.5	1 (3.6)
4	7.3	87.8	1 (3.6)
5	9.8	78.0	3 (10.7)
New core pathologists			
B	12.2	78.0	1 (3.6)
E	12.2	75.6	1 (3.6)
J	17.1	73.2	1 (3.6)
Aspirant panel members			
A	12.2	39.0	16 (57.0)
C	14.6	58.5	1 (3.6)
D	31.7	65.9	2 (7.1)
F	12.2	70.7	0
G	12.2	68.3	2 (7.1)
H	7.3	68.3	1 (3.6)
I	9.8	68.3	1 (3.6)
Benchmark value†	≤28	≥69	≤3 (11)

red, score does not fall within benchmark values

green, score falls within benchmark values

\*NDBO/IND/LGD+HGD.

†Based on eight core pathologists.

HGD, high-grade dysplasia; IND, indefinite for dysplasia; LGD, low-grade dysplasia;

NDBO, non-dysplastic Barrett's oesophagus.

not qualify for this criterion, underdiagnosing 16 of the 28 (57%) dysplasia cases as NDBO.

If all three quality criteria were taken into account, one pathologist met all benchmark scores and thus qualified as a core panel member. This pathologist was added to the core and incorporated in the benchmark ranges calculated for phase II. Four pathologists qualified for two criteria but showed a small deviation from the benchmark score for the percentage agreement with the consensus diagnosis. Two pathologists did not meet the required benchmark scores for two of the three quality criteria.

### Performance of pathologists in phase II

After the group discussion to discuss discrepant cases assessed in phase I and, subsequently, completing reassessment of the next 39 cases, all pathologists fell within the benchmark score of ≤38% for the percentage of IND cases in phase II; scores ranged from 2.6% to 15.4% (table 3).

All six remaining aspirant members fell within the benchmark range for percentage agreement with the consensus diagnosis.

Five of the six pathologists met the benchmark value for percentage of cases with a consensus diagnosis of LGD or HGD underdiagnosed as NDBO (table 3). They had zero (n=2), one (n=2) or two (n=1) underdiagnosed cases (median percentage 4.3% vs 4.3% for core pathologists).

One pathologist persisted in underdiagnosing a significant number of dysplastic cases as NDBO (7/23: 30.4%).

At the end of phase II, five of the six remaining aspirant panel members met all quality criteria.

**Table 3** Results of all pathologist for phase II (n=39)

Pathologist	Percentage of cases 'indefinite for dysplasia'	Percentage agreement (three categories)*	Consensus LGD and HGD cases underdiagnosed as NDBO (n (%)) n=23
Core pathologists			
1	17.9	89.7	0
2	10.3	87.2	2 (8.7)
3	28.2	71.8	2 (8.7)
4	20.5	87.2	2 (8.7)
5	17.9	82.1	0
New core pathologists			
B	7.7	74.4	2 (8.7)
E	25.6	69.2	1 (4.3)
J	12.8	71.8	1 (4.3)
New core pathologist after phase 1			
F	33.3	61.5	1 (4.3)
Aspirant panel members			
A	2.6	56.4	7 (30.4)
C	7.7	66.7	0
D	12.8	64.1	0
G	12.8	66.7	2 (8.7)
H	12.8	69.2	1 (4.3)
I	15.4	76.9	1 (4.3)
Benchmark value†	≤38	≥56	≤3 (13)

red, score does not fall within benchmark values

green, score falls within benchmark values

\*NDBO/IND/LGD+HGD.

†Based on nine core pathologists.

HGD, high-grade dysplasia; IND, indefinite for dysplasia; LGD, low-grade dysplasia;

NDBO, non-dysplastic Barrett's oesophagus.

### DISCUSSION

The aim of this study was to evaluate the performance of the GI pathologists of the eight BO expert centres in the assessment of the first 80 consecutive review cases submitted to the Dutch Barrett's Pathology Review panel with an initial diagnosis of IND or LGD. In our ambition to expand the current panel of eight core pathologists, while maintaining assessment quality and homogeneity, benchmark quality criteria based on the results of the core pathologists were used. These criteria enabled us to assess the ongoing process of homogenising future panel members. Assessments were done in two phases, with a group discussion to discuss discrepancies after phase I in order to further homogenise assessments.

When comparing the performance scores of the seven non-core pathologists to the benchmark scores, one out of seven aspirant panel members met all three benchmark quality criteria in phase I and was added to the core. Five out of the remaining six pathologists did not meet the benchmark value for only one of the three performance scores, percentage agreement, with three pathologists scoring only marginally below the required benchmark value (68.3% vs 69%). After the group discussion, five of the six remaining non-core pathologists met the benchmark values for all quality criteria during phase II.

This study is the third in a series of studies with this group of pathologists and includes the third independent set of histology cases. As with the previous studies, individual assessments were complemented with face-to-face group discussions, discussing discrepant cases from the slide set.<sup>9-12</sup> The current study demonstrates that 14 pathologists can be considered homogenous in



their assessments, which implies that these 14 pathologists are interchangeable as panel members of the Dutch Barrett's Pathology Review panel.

In contrast to our earlier studies, we did not assess intraobserver agreement since the assessors only assessed all cases once. We decided not to use this variable as a benchmark quality criterion. Since kappa resembles the agreement percentage corrected for chance, kappa is influenced by the prevalence of the diagnosis and thus the variance over the different categories. Since the cases submitted for review by Dutch Barrett's Pathology Review panel mainly consisted of dysplastic cases the variance is low, leading to a high agreement by chance. This results in low kappas and is therefore not a reliable representation of the quality of the assessments.

Strengths of this study are the use of a digitalised case set of consecutive BO cases with IND/LGD submitted for review from all over the Netherlands. Therefore, this case selection reflects the daily workload of the Dutch Barrett's Pathology Review panel. This study builds on three earlier studies in which pathologists were trained through structured assessments and group discussions. All pathologists participating in this study come from the eight Barrett Expert Centers in the Netherlands and thus represent the full potential for histological reviews of the Dutch Barrett's Pathology Review panel.

A limitation is that slides come from different laboratories around the Netherlands. This may have affected slide interpretability, especially for pathologists that are relatively new to reviewing cases from outside their own centre. However, this reflects the reality of our review panel, which per definition will receive consultations from different pathology laboratories.

Another point worth mentioning is that we used the consensus diagnosis made by our five core pathologists as training reference for the other pathologists. This means that our approach aimed to homogenise the assessments of the panel's members. Our study set lacked follow-up information of the cases used and therefore we cannot prove that this homogenisation indeed improves outcomes. The consensus gold standard diagnosis in this study is based on five core pathologists who have extensive experience and an international reputation in the field of BO and have collaborated intensively in the Dutch Barrett advisory committee over the course of many years.<sup>9 14 15</sup> Moreover, their assessments have been validated by comparing their diagnosis to the histological outcome during follow-up in multiple studies, demonstrating their diagnoses to be predictive of progression risk.<sup>1-3</sup> Based on aforementioned data, we assume that the consensus diagnosis of our five core pathologists is a justified reference for the remaining panellists and that homogenising our panel based on this reference will likely improve outcomes.

In the future, several important steps will be taken while the Dutch Barrett's Pathology Review panel proceeds.

A prediction model will be set up to establish how many pathologist's assessments are needed in order to obtain a reliable diagnosis. The algorithm will incorporate the performance score of the pathologist in previous assessments with the outcome of his or her review to decide how many additional pathologists will have to review that specific case in order to retrieve a consensus diagnosis with the same reliability as used in the current study. This will enable an efficient and equal distribution of the workload.

Online group discussions will be continued to discuss cases without a majority diagnosis. In addition, annual trainings will be held, and assessment of homogeneity of all panel members will be renewed periodically.

This slide set of the first 80 consecutive review cases will be used in an online training programme for other pathologists and pathology residents to improve the histopathological assessment of BO. This training module will be accredited and freely available. Information from all study sets and group discussions will be incorporated. Pathologists and residents with an interest in BO will be able to improve their skills and compare their performance to our panel.

A lot of time and effort has been put in the training sets, assessment of digital pathology, development of the quality criteria and the infrastructure of the platform. Other future panels can benefit from these steps we have taken. In systems where endoscopic surveillance for Barrett's patients is performed, there should be enough financial ability to run an expert panel since the panel will recoup its costs. Expert pathology will significantly improve the efficacy of Barrett surveillance by preventing unnecessary expensive endoscopies in patients with a low risk of progression and will also identify the subgroup with a truly increased risk of progression.

### Take home messages

- ▶ Guidelines dictate expert review for a diagnosis of low-grade dysplasia in Barrett's oesophagus (BO).
- ▶ A web-based digital expert review panel has been set up in the Netherlands.
- ▶ Pathologists from Dutch BO expert centres have participated in structured assessments and group discussions to homogenise assessments.
- ▶ This study shows that the expert review panel could be expanded to a total of 14 panel members while maintaining assessment quality and homogeneity by using predefined benchmark quality criteria for histological assessment of biopsies with low-grade dysplasia.

### Author affiliations

<sup>1</sup>Department of Gastroenterology and Hepatology, Amsterdam UMC – Locatie AMC, Amsterdam, The Netherlands

<sup>2</sup>Department of Pathology, Amsterdam UMC – Locatie AMC, Amsterdam, The Netherlands

<sup>3</sup>Department of Pathology, Pathologie-DNA BV, Sint Antonius Hospital, Nieuwegein, The Netherlands

<sup>4</sup>Department of Pathology, UMC Utrecht, Utrecht, The Netherlands

<sup>5</sup>Department of Pathology, Symbiant BV, Alkmaar, The Netherlands

<sup>6</sup>Department of Pathology, Erasmus Medical Center, Rotterdam, Zuid-Holland, The Netherlands

<sup>7</sup>Department of Pathology, Stichting PAMM, Eindhoven, The Netherlands

<sup>8</sup>Department of Pathology, UMCG, Groningen, The Netherlands

<sup>9</sup>Department of Pathology, Haga Hospital, Den Haag, The Netherlands

<sup>10</sup>Department of Pathology, Isala Klinieken, Zwolle, The Netherlands

<sup>11</sup>Department of Pathology, Pathan BV, Sint Franciscus Vlietland Groep, Rotterdam, The Netherlands

<sup>12</sup>Department of Cardiology, Amsterdam UMC – Locatie AMC, Amsterdam, The Netherlands

**Handling editor** Dharendra Govender.

**Contributors** The following author contributions were made: study concept and design: EK, MW, SM and JB. Acquisition of data: EK, MW, CS, GO, MV, FK, KB, LB, MD, CH, AK, GK, JL, GvL, FM, AO and SM. Analysis and interpretation of data: EK, MW, JT, JB and SM. Drafting of the manuscript: EK, MW, LD, RP, JT, JB and SM. Critical revision for important intellectual content and final approval of the manuscript: all authors. Study supervision: SM and JB.

**Funding** The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

**Competing interests** None declared.

**Patient consent for publication** Not required.

**Ethics approval** The need for approval by the Medical Ethical Committee was waived.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data availability statement** Data are available on reasonable request.

**Open access** This is an open access article distributed in accordance with the Creative Commons Attribution 4.0 Unported (CC BY 4.0) license, which permits others to copy, redistribute, remix, transform and build upon this work for any purpose, provided the original work is properly cited, a link to the licence is given, and indication of whether changes were made. See: <https://creativecommons.org/licenses/by/4.0/>.

#### ORCID iDs

Esther Klaver <http://orcid.org/0000-0003-4981-7540>

Myrtle van der Wel <http://orcid.org/0000-0001-5825-3236>

Lodewijk Brosens <http://orcid.org/0000-0003-1341-8994>

Arend Karrenbeld <http://orcid.org/0000-0002-3518-2305>

#### REFERENCES

- Curvers WL, ten Kate FJ, Krishnadath KK, *et al*. Low-grade dysplasia in Barrett's esophagus: overdiagnosed and underestimated. *Am J Gastroenterol* 2010;105:1523–30.
- 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* 2015;64:700–6.
- 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;152:993–1001.
- Weusten B, Bisschops R, Coron E, *et al*. Endoscopic management of Barrett's esophagus: European Society of gastrointestinal endoscopy (ESGE) position statement. *Endoscopy* 2017;49:191–8.
- Wani S, Rubenstein JH, Vieth M, *et al*. Diagnosis and management of low-grade dysplasia in Barrett's esophagus: expert review from the clinical practice updates committee of the American Gastroenterological Association. *Gastroenterology* 2016;151:822–35.
- 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.
- di Pietro M, Fitzgerald RC, BSG Barrett's guidelines working group. Revised British Society of gastroenterology recommendation on the diagnosis and management of Barrett's oesophagus with low-grade dysplasia. *Gut* 2018;67:392–3.
- van der Wel MJ, Jansen M, Vieth M, *et al*. What makes an expert Barrett's Histopathologist? *Adv Exp Med Biol* 2016;908:137–59.
- 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. *Dis Esophagus* 2017;30:1–7.
- 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 Gastroenterol J* 2018;6:830–7.
- van der Wel MJ, Duits LC, Pouw RE, *et al*. Improved diagnostic stratification of digitised Barrett's oesophagus biopsies by p53 immunohistochemical staining. *Histopathology* 2018;72:1015–23.
- van der Wel MJ, Klaver E, Duits LC, *et al*. 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. *United European Gastroenterol J* 2019;7:889–96.
- 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. *J Gastroenterol Hepatol* 2000;15(Suppl):G49–57.
- Hulscher JB, Haringsma J, Benraad J, *et al*. Comprehensive cancer centre Amsterdam Barrett Advisory Committee: first results. *Neth J Med* 2001;58:3–8.
- Offerhaus GJA, Correa P, van Eeden S, *et al*. Report of an Amsterdam Working group on Barrett esophagus. *Virchows Arch* 2003;443:602–8.