Improved diagnostic stratification of digitised Barrett’s oesophagus biopsies by p53 immunohistochemical staining

Myrtle J van der Wel,1,2 Lucas C Duits,2 Roos E Pouw,2 Cornelis A Seldenrijk,3 G J A Offerhaus,4 Mike Visser,5 Fiebo J ten Kate,4 Katharina Biermann,6 Lodewijk A A Brosens,4 Michael Doukas,6 Clement Huysentruyt,7 Arend Karrenbeld,8 Gursah Kats-Ugurlu,8 Jaap S van der Laan,9 G (Ineke) van Lijnschoten,7 Freek C P Moll,10 Ariadne H A G Ooms,11 Hans van der Valk,11 Jan G Tijssen,12 Jacques J Bergman2 & Sybren L Meijer1

1Department of Pathology, Academic Medical Centre, Amsterdam, 2Department of Gastroenterology and Hepatology, Academic Medical Centre, Amsterdam, 3Department of Pathology, Pathology-DNA, St Antonius Hospital, Nieuwegein, 4Department of Pathology, University Medical Centre, Utrecht, 5Department of Pathology, Zymbiant BV, Alkmaar, 6Department of Pathology, Erasmus Medical Centre, Rotterdam, 7Department of Pathology, Stichting PAMM, Eindhoven, 8Department of Pathology, Academic Medical Centre, Groningen, 9Department of Pathology, Haaga Hospital, The Hague, 10Department of Pathology, Isala Clinics, Zwolle, 11Department of Pathology, St Franciscus Vlietland Gasthuis, Pathan BV, Rotterdam, and 12Department of Cardiology, Academic Medical Centre, Amsterdam, The Netherlands

Date of submission 30 October 2017
Accepted for publication 29 December 2017
Published online Article Accepted 3 January 2018


Improved diagnostic stratification of digitised Barrett’s oesophagus biopsies by p53 immunohistochemical staining

Aims: Interobserver agreement for dysplasia in Barrett’s oesophagus (BO) is low, and guidelines advise expert review of dysplastic cases. The aim of this study was to assess the added value of p53 immunohistochemistry (IHC) for the homogeneity within a group of dedicated gastrointestinal (GI) pathologists.

Methods and results: Sixty-single haematoxylin and eosin (HE) slide referral BO cases [20 low-grade dysplasia (LGD); 20 high-grade dysplasia (HGD); and 20 non-dysplastic BO reference cases] were digitalised and independently assessed twice in random order by 10 dedicated GI pathologists. After a ‘wash-out’ period, cases were reassessed with the addition of a corresponding p53 IHC slide. Outcomes were: (i) proportion of ‘indefinite for dysplasia’ (IND) diagnoses; (ii) interobserver agreement; and (iii) diagnostic accuracy as compared with a consensus ‘gold standard’ diagnosis defined at an earlier stage by five core expert BO pathologists after their assessment of this case set. Addition of p53 IHC decreased the mean proportion of IND diagnoses from 10 of 60 to eight of 60 ($P = 0.071$). Mean interobserver agreement increased significantly from 0.45 to 0.57 ($P = 0.0021$). The mean diagnostic accuracy increased significantly from 72% to 82% ($P = 0.0072$) after p53 IHC addition.

Conclusion: Addition of p53 IHC significantly improves the histological assessment of BO biopsies, even within a group of dedicated GI pathologists. It decreases the proportion of IND diagnoses, and increases interobserver agreement and diagnostic accuracy. This justifies the use of accessory p53 IHC within our upcoming national digital review panel for BO biopsy cases.
Introduction

Barrett’s oesophagus (BO) is a known precursor lesion for the development of oesophageal adenocarcinoma (OAC), and is characterised by the replacement of stratified squamous epithelium with metaplastic intestinal epithelium at the distal oesophagus. OAC can eventually develop through a sequence of events, the metaplasia–dysplasia–carcinoma sequence. Current guidelines recommend endoscopic surveillance of patients with BO with biopsies for proper risk stratification. Histopathological diagnosis of low-grade dysplasia (LGD) is the only accepted predictor for progression. However, this is complicated by interobserver variability between pathologists, and conflicting results in reporting rates of progression to OAC for LGD have been reported. Studies suggest that, when LGD is confirmed by an expert pathologist, the risk of progression increases sharply. Various international BO guidelines also advocate the use of expert review of dysplastic cases by a pathologist with special interest and extensive experience in interpretation of BO-associated neoplasia. To facilitate these revisions in The Netherlands, our goal is to set up a national review panel consisting of expert BO pathologists. For practical purposes, we want to use only digitalised slides. The use of digital microscopy has been validated previously. To optimise the panel (consensus) diagnosis, we wanted to investigate the use of an adjunct diagnostic marker. Earlier studies have shown that aberrant nuclear immunohistochemical staining of the protein encoded by the tumour suppressor gene TP53 (p53) may improve the histological assessment of BO. TP53 is a tumour suppressor gene acting as the guardian of genetic stability. The half-life of a normal p53 protein is short, resulting in weak immunohistochemical staining (wild-type expression). However, the half-life of mutated p53 is prolonged, which, together with cellular feedback mechanisms attempting DNA damage repair, leads to nuclear accumulation of (mutated) p53, visible by the use of immunohistochemistry (IHC). In the case of a nonsense mutation or homozygous deletion of the TP53 locus, failed translation of the protein, or translation into a truncated protein lacking the p53 antibody epitope, a complete absence of staining (null mutation) occurs. Currently, guidelines are contradictory concerning the use of p53 IHC in BO diagnostic work-up, and their advice to use p53 IHC as a diagnostic aid is tentative at most.

At present, 10 dedicated gastrointestinal (GI) pathologists from The Netherlands are participating in a standardised self-assessment training programme. We want to know whether p53 IHC can improve the homogeneity of this group of pathologists, defined according to the following outcome parameters: (i) the proportion of diagnoses ‘indefinite for dysplasia’ (IND) per pathologist; (ii) the interobserver agreement per pathologist; and (iii) the diagnostic accuracy per pathologist as compared with the consensus gold standard diagnosis. The main aim of this study was therefore to test the added value of p53 IHC in the diagnostic work-up of a case set consisting of mainly dysplastic BO cases.

Materials and methods

The Medical Ethical Committee of the Academic Medical Centre (AMC) waived the need for approval for this study.

Case selection, assessors, and study design

For this diagnostic study, we selected 60 single haematoxylin and eosin (HE)-stained slides from 60 individual BO biopsy cases referred to the BO surveillance programme at the Amsterdam AMC for pathology review by the local Barrett expert panel between 2007 and 2013. The referring diagnosis was LGD in 20 cases and high-grade dysplasia (HGD) or OAC in 20 cases. These cases were supplemented with 20 non-dysplastic BO (NDBO) reference cases. From each individual case, a single representative HE slide and a concomitant p53 IHC slide were selected and digitalised by the study coordinator, on the basis of detailed review of the pathology report. For each case, a consensus gold standard diagnosis had been generated previously by five ‘core’ expert BO pathologists through multiple group discussions (NDBO, n = 20; LGD, n = 29; HGD, n = 11). These core experts currently constitute the national digital review panel for dysplastic BO. They are all working at one of the eight BO expert centres in The Netherlands.
Netherlands (range of experience: 10–30 years). In The Netherlands, care for patients with dysplastic BO is centralised in these eight centres. The expert BO pathologists handle a case load of 5–10 BO cases per week, of which 25% are dysplastic. They are considered to be experts among their peers, and each has co-authored >10 peer-reviewed publications in this field. The assessors were 10 other dedicated GI pathologists, also working at the eight BO expert centres in The Netherlands. All 10 pathologists independently assessed the case set twice, with a wash-out time of at least 1 month between the assessment rounds. During the assessments, they were blinded to the consensus gold standard diagnoses. The first assessment round consisted of only the HE slide, and in the second assessment round the HE slides were examined in tandem with the p53 IHC slides. All individual diagnoses from each round were recorded on a case record form.

HISTOLOGY, IHC, AND DIGITALISING OF SLIDES

The process of staining and digitalising of slides is described in Data S1. The p53 IHC slides were scored according to international reported criteria as one of three staining patterns: normal background staining (wild-type expression), overexpression (nuclear accumulation), or complete absence of staining (null-mutation). Normal background staining served as an internal control in cases with complete absence of staining. Examples of different staining patterns are shown in Figure 1.

OUTCOME MEASUREMENTS

The outcomes of this study were: the proportion of IND diagnoses without and with p53 IHC; the interobserver agreement of the pathologists without and with p53 IHC; and the diagnostic accuracy as compared with the consensus gold standard diagnosis, without and with p53 IHC.

GROUP DISCUSSION

After finishing the two assessment rounds, the pathologists met in a group discussion to discuss discrepant cases in relation to the gold standard diagnosis. They especially discussed the interpretation of different p53 staining patterns, and formulated a consensus decision rule, which is shown in Figure 2.

STATISTICAL ANALYSIS

The proportion of IND diagnoses per pathologist was counted per assessment round, and the difference between the two rounds was calculated. The statistical significance of this difference was calculated with the paired t-test, with the outcome being dichotomised to compare the proportions of IND diagnoses per pathologist per round. The statistical significance of the mean difference was also calculated with the paired t-test. The interobserver agreement was measured as weighted Cohen’s kappa (κ), with two diagnostic categories (NDBO + IND and LGD + HGD). The 10 pathologists assessed the case set twice, yielding a total of 20 assessment rounds with nine pairwise assessments per pathologist (pathologist 1 against pathologist 2, pathologist 1 against pathologist 3, pathologist 1 against pathologist 4, pathologist 1 against pathologist 5, pathologist 1 against pathologist 6, pathologist 1 against pathologist 7, pathologist 1 against pathologist 8, pathologist 1 against pathologist 9, and pathologist 1 against pathologist 10). The interobserver agreement per pathologist was defined as the mean κ of these nine pairwise assessments, per round. The mean interobserver agreement was defined as the mean of all 45 pairwise assessments. Because of the possibility

Figure 1. Examples of p53 immunohistochemistry wild-type expression (A), overexpression (B) and null-mutation (C) from the case set. Note the normal background staining in (C) (arrowhead) serving as an internal control.
of skewed marginal totals, the maximum possible $\kappa$ per cross-table does not always equal 1. Therefore, the agreement calculated as a fraction of the maximum possible $\kappa$ is also depicted. Strength of agreement was traditionally categorised as follows: a value of zero or less indicates agreement no better than chance alone (‘poor’); 0.00–0.20 indicates ‘slight’ agreement; 0.21–0.40 indicates ‘fair’ agreement; 0.41–0.60 indicates ‘moderate’ agreement; 0.61–0.80 indicates ‘substantial’ agreement; and 0.81–1.00 indicates ‘almost perfect’ agreement. The difference between before and after p53 IHC addition was calculated. The statistical significance of this difference was calculated with the paired $t$-test for paired samples, with comparison of either two times nine pairs per pathologist, or two times 45 pairs for the mean interobserver agreement. The diagnostic accuracy was calculated for the discrimination between non-dysplastic and dysplastic BO of the individual pathologists with respect to the consensus gold standard diagnosis, and compared between the two assessment rounds. The outcomes were first dichotomised into ‘dysplasia’ (LGD + HGD) and ‘no dysplasia’ (NDBO + IND). The ‘true-positive cases’ and ‘true-negative cases’ combined were compared with the outcomes of the pathologists.

A $P$-value of $\leq 0.05$ was considered to be statistically significant for all tests, and a reason to reject $H_0$. Statistical analyses were performed with spss 24.0 (IBM Corp., Armonk, New York, NY, USA). The weighted $\kappa$ was developed with the self-automated program AGREESTAT (version 24; Advanced Analytics, LCC, Gaithersburg, MD, USA).

**Results**

**Proportion of IND diagnoses**

Table 1 shows the proportion of IND diagnoses per pathologist. It can be seen that the proportion of IND diagnoses decreased for all pathologists except one (pathologist 6) after p53 IHC addition. The proportional differences between assessment rounds 1 and 2 pathologists 6 and 7 were significant ($P = 0.034$). The median proportion of IND diagnoses was 10 of 60 cases (17%) in assessment round 1, before p53 IHC addition. After p53 IHC addition, this proportion decreased to a median of seven of 60 (12%) IND diagnoses in assessment round 2. There was a trend towards a lower proportion of IND diagnoses after p53 IHC addition ($P = 0.071$).

**Interobserver agreement**

Table 2 shows the interobserver agreement for dysplasia versus no dysplasia before and after p53 IHC addition. It can be seen that all individual $\kappa$ scores improved after p53 IHC addition. Six of 10 $\kappa$ scores improved significantly (see Table 2 for values). Before p53 IHC addition, the mean interobserver agreement was 0.45, increasing to 0.57 after p53 IHC addition (both ‘moderate’, $P = 0.0021$), which...
is a significant improvement of 0.12. The maximum possible values for these weighted \( \kappa \) scores were all <1. After correction, the mean fractions of maximum possible \( \kappa \) scores were 0.59 (‘moderate’) for assessment round 1 and 0.73 (‘substantial’) for assessment round 2.

Table 1. Proportion of cases diagnosed as ‘indefinite for dysplasia’ (IND) before and after addition of p53 immunohistochemistry (IHC)

<table>
<thead>
<tr>
<th>Pathologist</th>
<th>IND before p53 IHC</th>
<th>%</th>
<th>IND after p53 IHC</th>
<th>%</th>
<th>Difference</th>
<th>Difference (%)</th>
<th>( P )-value*</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>8/60</td>
<td>13</td>
<td>5/60</td>
<td>8</td>
<td>-3</td>
<td>-5</td>
<td>0.26</td>
</tr>
<tr>
<td>2</td>
<td>8/60</td>
<td>13</td>
<td>3/60</td>
<td>5</td>
<td>-5</td>
<td>-8</td>
<td>0.13</td>
</tr>
<tr>
<td>3</td>
<td>10/60</td>
<td>17</td>
<td>5/60</td>
<td>8</td>
<td>-3</td>
<td>-8</td>
<td>0.13</td>
</tr>
<tr>
<td>4</td>
<td>11/60</td>
<td>18</td>
<td>9/60</td>
<td>15</td>
<td>-2</td>
<td>-3</td>
<td>0.62</td>
</tr>
<tr>
<td>5</td>
<td>2/60</td>
<td>3</td>
<td>1/60</td>
<td>2</td>
<td>-1</td>
<td>-2</td>
<td>0.57</td>
</tr>
<tr>
<td>6</td>
<td>8/60</td>
<td>13</td>
<td>15/60</td>
<td>25</td>
<td>7</td>
<td>12</td>
<td>0.034</td>
</tr>
<tr>
<td>7</td>
<td>9/60</td>
<td>15</td>
<td>2/60</td>
<td>3</td>
<td>-7</td>
<td>-12</td>
<td>0.034</td>
</tr>
<tr>
<td>8</td>
<td>22/60</td>
<td>37</td>
<td>17/60</td>
<td>28</td>
<td>-5</td>
<td>-8</td>
<td>0.096</td>
</tr>
<tr>
<td>9</td>
<td>13/60</td>
<td>22</td>
<td>10/60</td>
<td>17</td>
<td>-3</td>
<td>-5</td>
<td>0.41</td>
</tr>
<tr>
<td>10</td>
<td>10/60</td>
<td>17</td>
<td>9/60</td>
<td>15</td>
<td>-1</td>
<td>-2</td>
<td>0.74</td>
</tr>
<tr>
<td>Mean</td>
<td>10/60</td>
<td>17%</td>
<td>8/60</td>
<td>13%</td>
<td>-2</td>
<td>-3%</td>
<td>0.071</td>
</tr>
</tbody>
</table>

*Paired \( t \)-test (two-tailed), significant when \( P \leq 0.05 \).

Table 2. Interobserver agreement in two diagnostic categories*, before and after the addition of p53 immunohistochemistry (IHC)

<table>
<thead>
<tr>
<th>Pathologist</th>
<th>Before p53 IHC</th>
<th>After p53 IHC</th>
<th>Difference (mean weighted ( \kappa ))</th>
<th>( P )-value†</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean weighted</td>
<td>Mean maximum</td>
<td>Weighted ( \kappa )</td>
<td>Mean maximum</td>
</tr>
<tr>
<td>1</td>
<td>0.52</td>
<td>0.81</td>
<td>0.62</td>
<td>0.64</td>
</tr>
<tr>
<td>2</td>
<td>0.58</td>
<td>0.77</td>
<td>0.75</td>
<td>0.64</td>
</tr>
<tr>
<td>3</td>
<td>0.52</td>
<td>0.83</td>
<td>0.63</td>
<td>0.58</td>
</tr>
<tr>
<td>4</td>
<td>0.35</td>
<td>0.83</td>
<td>0.43</td>
<td>0.54</td>
</tr>
<tr>
<td>5</td>
<td>0.54</td>
<td>0.81</td>
<td>0.66</td>
<td>0.63</td>
</tr>
<tr>
<td>6</td>
<td>0.42</td>
<td>0.64</td>
<td>0.65</td>
<td>0.49</td>
</tr>
<tr>
<td>7</td>
<td>0.43</td>
<td>0.83</td>
<td>0.54</td>
<td>0.62</td>
</tr>
<tr>
<td>8</td>
<td>0.26</td>
<td>0.71</td>
<td>0.39</td>
<td>0.44</td>
</tr>
<tr>
<td>9</td>
<td>0.46</td>
<td>0.77</td>
<td>0.62</td>
<td>0.52</td>
</tr>
<tr>
<td>10</td>
<td>0.42</td>
<td>0.81</td>
<td>0.53</td>
<td>0.59</td>
</tr>
<tr>
<td>Mean‡</td>
<td>0.45</td>
<td>0.78</td>
<td>0.59</td>
<td>0.57</td>
</tr>
</tbody>
</table>

*Non-dysplastic Barrett’s oesophagus + indefinite for dysplasia; low-grade dysplasia + high-grade dysplasia.
†Paired \( t \)-test (two-tailed), significant when \( P \leq 0.05 \).
‡Total of 45 pairwise \( \kappa \) scores.
Table 3. Diagnostic accuracy of non-dysplastic Barrett's oesophagus versus dysplasia, as compared with the gold standard diagnosis before and after the addition of p53 immunohistochemistry (IHC)

<table>
<thead>
<tr>
<th>Pathologist</th>
<th>Before p53 IHC (%)</th>
<th>After p53 IHC (%)</th>
<th>Difference (%)</th>
<th>P-value*</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>82</td>
<td>83</td>
<td>1</td>
<td>0.57</td>
</tr>
<tr>
<td>2</td>
<td>85</td>
<td>93</td>
<td>8</td>
<td>0.096</td>
</tr>
<tr>
<td>3</td>
<td>77</td>
<td>85</td>
<td>8</td>
<td>0.096</td>
</tr>
<tr>
<td>4</td>
<td>63</td>
<td>73</td>
<td>10</td>
<td>0.14</td>
</tr>
<tr>
<td>5</td>
<td>82</td>
<td>95</td>
<td>13</td>
<td>0.01</td>
</tr>
<tr>
<td>6</td>
<td>80</td>
<td>70</td>
<td>-10</td>
<td>0.16</td>
</tr>
<tr>
<td>7</td>
<td>65</td>
<td>85</td>
<td>20</td>
<td>0.002</td>
</tr>
<tr>
<td>8</td>
<td>63</td>
<td>78</td>
<td>15</td>
<td>0.002</td>
</tr>
<tr>
<td>9</td>
<td>63</td>
<td>77</td>
<td>14</td>
<td>0.004</td>
</tr>
<tr>
<td>10</td>
<td>63</td>
<td>83</td>
<td>20</td>
<td>0.001</td>
</tr>
<tr>
<td>Mean</td>
<td>72</td>
<td>82</td>
<td>10</td>
<td>0.0072</td>
</tr>
</tbody>
</table>

*Paired t-test (two-tailed), significant when $P \leq 0.05$. 

**Discussion**

In this diagnostic study, 10 dedicated GI pathologists assessed a histological single-slide case set of 60, mainly dysplastic, BO biopsy cases. The addition of p53 IHC in assessment round 2 decreases the median proportion of IND diagnoses ($P = 0.071$), significantly increases the mean interobserver agreement ($P = 0.0021$) and significantly increases the mean diagnostic accuracy ($P = 0.0072$) of this large group of GI pathologists. This signiﬁes that p53 IHC appears to aid pathologists in the accurate stratification of BO patients as compared with a consensus gold standard diagnosis generated by ﬁve core expert pathologists who are current members of the national digital review panel for BO.14 The positive effect on interobserver agreement observed in our study is comparable to what is shown in the existing literature. Kaye et al. performed two histopathological studies to investigate the added value of p53 IHC on observer agreement among two groups of pathologists. The ﬁrst study consisted of a group of ﬁve pathologists assessing 186 single-slide cases of BO,17 and the second study consisted of a group of 10 pathologists assessing 72 BO cases.18 The ﬁrst study showed improvement in interobserver weighted $\kappa$ scores after addition of p53 IHC, from 0.42 to 0.48 (both 'moderate').17 The second study investigated the generalisability of p53 IHC, and showed that p53 IHC interpretation was more reliable than HE interpretation for diagnosing BO + dysplasia cases, and improved the mean interobserver weighted $\kappa$ score from 0.47 to 0.55 (both ‘moderate’).18 Both studies also showed aberrant p53 expression to be an independent predictor of disease progression. However, unlike in our study, the pathologists did not meet in group discussions to discuss discrepant cases, and did not have a consensus diagnosis for comparison. Despite the positive effect of p53 IHC on BO diagnostics in these studies, guidelines still only half-heartedly advise the use of p53 IHC.

This study has a number of unique features. First, a face-to-face group discussion with all dedicated GI pathologists was held after the two assessment rounds were ﬁnished. This gave the group the opportunity to discuss all discrepant cases in relation to p53 IHC expression, and establish a consensus decision rule for the interpretation of p53 IHC, if it is performed, as shown in Figure 2. Second, the set-up of this study was thorough. The study set was enriched for difﬁcult dysplastic cases, it was carried out with the help of digitalised slides for maximum efﬁciency, and all pathologists assessed the study set twice and independently in a randomised fashion, with a ‘washout’ phase. Third, to calculate the accuracy of the pathologists, we used a consensus gold standard diagnosis, generated earlier by the ﬁve ‘core’ expert pathologists of the national digital review panel. They are considered to be true experts by their peers, and have proven this in many earlier studies.14

The first limitation of this study concerns the interpretability of p53 IHC. International guidelines are not uniform regarding the use of p53 IHC, and no objective parameters for the morphological
assessment, or for the application frequency of p53 IHC in BO biopsies, exist. Both a quantitative interpretation and a qualitative interpretation are combined with the assessment of morphological features on the HE slide. There is no compelling evidence to support the use of p53 IHC on every single BO biopsy. However, as it is useful in diagnostic work-up and improves diagnostic agreement, we do advocate a low threshold for p53 IHC use, while still recognising its limits. The morphological diagnosis on HE should take precedence, and aberrant staining can be used to support a diagnosis of dysplasia. In this setting, aberrant p53 IHC staining can also be used as a diagnostic biomarker for comparison of future biopsies, or as a marker for progression in clinical studies. A non-aberrant p53 IHC staining pattern in a morphologically dysplastic specimen, however, does not exclude the presence of dysplasia. When p53 IHC shows a focus of unexpectedly aberrant staining, this may cause diagnostic confusion. The exact clinical meaning of such a focus is unknown, but when it is unequivocally aberrant, we consider it best to regard it as IND, which has been found to be associated with progression.

Using the results from our group discussion, we summarised the above findings in a p53 IHC decision rule (Figure 2), in order to increase the uniformity of p53 IHC interpretation in the future. For optimal p53 IHC staining, adhering to external quality programmes and using standardised staining protocols with adequate controls is a prerequisite. Furthermore, training in interpretation of the different staining patterns as compared with background staining is necessary. Second, we are aware that our study shows a large range of outcome measures for both the proportion of IND diagnoses and the interobserver agreement. We attribute this to the fact that the group of dedicated GI pathologists have just started a standardised training programme for dysplastic BO, but have never before received training together as a group. Nonetheless, for all pathologists except one, we see a trend towards fewer IND diagnoses when p53 IHC is added. However, when the pathologist with more IND diagnoses after p53 IHC is excluded from the analysis, the rest of the group show significantly fewer IND diagnoses after p53 IHC addition (P = 0.0009; Table S1). These results confirm the need for uniform interpretation of p53 IHC. The mean interobserver agreement and mean diagnostic accuracy of pathologist 6 do show a statistically significant improvement when p53 IHC is added. The last limitation concerns the artificial set-up of the study, because each case consisted of only a single slide. Most endoscopic procedures generate biopsies on more than one level, leading to more than one slide per case. Therefore, the 10 GI pathologists will now assess a second study set, consisting of all slides of all tissue blocks generated during one endoscopic procedure.

In conclusion, our results show that p53 IHC addition significantly improved the proportion of IND diagnoses, the interobserver agreement and the diagnostic accuracy within a large group of 10 dedicated GI pathologists. This justifies the use of p53 IHC within our upcoming national digital review panel for BO biopsy cases. It can thereby improve the stratification of patients in order to optimise diagnostic work-up and (endoscopic) treatment.

Acknowledgements

We thank Eelco Roos for his help with the layout of the figures.

Funding: no external sources of funding were used for this study.

Author contributions


Conflicts of interest

J J Bergman: receives research support from Olympus Endoscopy, Cook Medical, Boston Scientific, GI Solutions Covidien, Erbe, Fractyl and Ninepoint Medical, Fuji Film, Cernostics, Interpace; receives financial support for training programmes from GI Solutions Covidien; and receives honorarium/consultancy/speakers’ fees from Cook Medical, Boston Scientific, and GI Solutions. The other authors state that they have no conflicts of interest.
References

32. van der Wel MJ, Duits LC, Klaver E et al. Development of benchmark quality criteria by a panel of expert pathologists assessing whole-endoscopy Barrett’s esophagus biopsy cases. United European Gastroenterol J. 2018 (Accepted for publication).
36. Kestens C, Leenders M, Offerhaus GJ, van Baal JW, Siersma PD. Risk of neoplastic progression in Barrett’s esophagus

**Supporting Information**

Additional Supporting Information may be found in the online version of this article:

**Data S1.** Supplementary methods.
**Table S1.** Proportion of cases diagnosed as ‘indefinite for dysplasia’ before and after addition of p53 immunohistochemistry after exclusion of pathologist 6.