Histopathology

Comparing virtual with conventional microscopy for the consensus diagnosis of Barrett's neoplasia in the AspECT Barrett's chemoprevention trial

Journal:	*Histopathology*
Manuscript ID:	HISTOP-01-12-0011.R1
Wiley - Manuscript type:	Original Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Sanders, Scott; Warwick Hospital, Pathology Grabsch, Heike; Leeds Institute of Molecular Medicine, University of Leeds, Pathology & tumour Biology Harrison, Rebecca; Dept of Pathology Bateman, Adrian; Southampton General Hospital, Pathology Going, James; Glasgow Royal Infirmary, Pathology Goldin, Robert; Imperial College, Centre for Pathology Mapstone, Nicholas; Royal Lancaster Infirmary, Department of Histopathology Novelli, Marco; University College London, Pathology Walker, Marjorie; Imperial College, London, Histopathology Jankowski, Janusz; Leicester Royal Infirmary, Medicine
Keywords:	Barrett's, oesophagus, dysplasia, virtual microscopy, telepathology

Comparing virtual with conventional microscopy for the consensus diagnosis of Barrett's neoplasia in the AspECT Barrett's chemoprevention trial pathology audit

DSA Sanders [1], H Grabsch [2], R Harrision [3], A Bateman [4], J Going [5], R Goldin [6], N Mapstone [7], M Novelli [8], MM Walker [6], J Jankowski [9]

(On behalf of the AspECT trial management group and trial principle investigators)

Dept. of Cellular Pathology, Warwick Hospital [1], Leeds Institute for Molecular Medicine, St James' University Hospital, Leeds [2], Dept. of Pathology, University Hospital Leicester [3], Pathology Dept., Southampton General Hospital [4], Cellular Pathology, Glasgow Royal Infirmary [5], Pathology Dept. St Mary's Hospital, London [6], Dept. of Histopathology, Royal Lancaster Hospital [7], Dept. of Histopathology, UHCL, London [8], Digestive Diseases Centre, Leicester Royal Infirmary [9]

Address for correspondence:

Dr DSA Sanders, Dept of Cellular Pathology, Warwick Hospital, Lakin Rd, Warwick CV34 5BJ
email; scott.sanders@nhs.net

Running Title; Comparing conventional with virtual microscopy for the diagnosis of Barrett's neoplasia

Aims: To compare the diagnostic accuracy of conventional versus virtual microscopy for the diagnosis of Barrett's neoplasia

Methods and results: 61 biopsies from 35 AspECT trial patients were given a Barrett's neoplasia score (1-5) by a panel of 5 pathologists using

conventional microscopy. 33 biopsies positive for neoplasia were digitised and blindly re-scored by virtual microscopy. Diagnostic reliability was compared between conventional and virtual microscopy using Fleiss' kappa. There was substantial reliability of diagnostic agreement (kappa = 0.712) scoring the 61 biopsies and moderate agreement scoring the subgroup of 33 'positive' biopsies with both conventional microscopy (kappa = 0.598) and virtual microscopy (kappa = 0.436). Interobserver diagnostic agreement between 2 pathologists by virtual microscopy was substantial (Kappa = 0.76). Comparison of panel consensus neoplasia scores between conventional and virtual microscopy was almost perfect (kappa = 0.8769). However with virtual microscopy there was lowering of the consensus neoplasia score in 9 biopsies

Conclusions: Diagnostic agreement with virtual microscopy compares favourably with conventional microscopy in what is recognised to be a challenging area of diagnostic practice. However this study highlights possible limitations for this method in the primary diagnostic setting.

Key words; Barrett's, oesophagus, dysplasia, virtual microscopy, telepathology

Introduction

Barrett's oesophagus is a pre-malignant condition affecting 2% of the population characterised by columnar metaplasia of the lining of the distal oesophagus and with a small, but recognised, risk of development of adenocarcinoma. AspECT (ASPirin Esomeprazole ChemopreventionTrial) is a multicentre trial which has recruited 2,500 patients to investigate the chemopreventative role of aspirin in combination with acid suppression in all causes of mortality, including malignancy, in patients with Barrett's oesophagus. Trial entry, identification of neoplastic progression (dysplasia), and trial endpoint of identification of conversion to high grade dysplasia or malignancy, together with decisions around patient intervention and surveillance protocols are based on a histological diagnosis made on endoscopic biopsies by a broad group of gastrointestinal (GI) and non-gastrointestinal pathologists at the multiple trial centres. Interobserver and intraobserver histological agreement has been shown to be poor in the diagnosis of GI tract dysplasia, [1, 2] Barrett's dysplasia, and early upper GI malignancy [3-8] and therefore robust central audit for the trial duration of all pathological diagnoses of Barrett's neoplasia from all trial centres is an essential part of the trial management to ensure validity and high quality of the pathological data and trial outcomes.

Consensus diagnosis by a panel of gastrointestinal pathologists has been reported to have high specificity and high predictive value for the identification of neoplastic progression in Barrett's oesophagus. In this respect one study showed that a consensus diagnosis of dysplasia from at least 4 or 5 pathologists was most likely to predict progression [9]. Initially a 'gold standard'

approach was favoured in the AspECT pathology audit by obtaining a consensus neoplasia score from a panel of 5 expert gastrointestinal pathologists on review of trial centre endoscopic biopsies by conventional microscopy. The logistics of this approach were however challenging and found to be untenable. Timely rotation of conventional glass slides between pathologists proved to be problematic and expensive, with the potential for slide damage and loss in transit.

Histological diagnosis on screen based images of digitally scanned slides, 'virtual microscopy' or 'telepathology', is gaining acceptance in some settings, and has a substantial logistical advantage over conventional microscopy. Early experience of using telepathology for the diagnosis of dysplasia in ulcerative colitis, [2, 10] grading dysplasia in the quality assurance of screen detected colorectal polyps [11], and as a method to evaluate diagnostic consistency in the external quality assurance scheme for prostate biopsies [12] has been reported but there are scant data on its utility in the setting of upper GI tract neoplasia and, in particular, Barrett's oesophagus. The aim of this study was to compare the feasibility of using virtual microscopy in the consensus approach to the diagnosis of Barrett's neoplasia.

Methods

Glass slides of 61 biopsies from different levels of the oesophagus from 35 AspECT patients were divided between 2 panels of 5 expert GI pathologists, including biopsies taken at 3 separate endoscopies from one patient and including section levels when available. Members of the panel allocated each

Deleted: .

Deleted:

slide a diagnostic neoplasia score from 1 to 5 (as defined by the modified Vienna classification of upper GI tract neoplasia) [13] (Table 1) by conventional microscopy. An average consensus neoplasia score was calculated for each slide. The majority of biopsies (n=28) with a consensus neoplasia score of 1 ('easier cases' with no dysplasia) were excluded from the virtual microscopy study with 33 remaining biopsies scanned at x40 magnification using the Aperio T3 scanning platform. Images were made available on-line at www.virtualpathology.leeds.ac.uk, and viewed on-line using freely available Aperio Imagescope software. Re-scoring of the virtual slides which included available section levels was carried out by a panel of 5 expert pathologists at least 6 months after scoring the same slides using conventional microscopy. All pathologists were blinded to any clinical and endoscopic features, and identifying slide features for the virtual part of the study, and used either work or home- based computers and monitors for on-screen diagnosis.

Statistical analysis utilised the methodology of Fleiss'(generalised) kappa [14] to assess the reliability of diagnostic agreement between multiple raters (5 panel pathologists) for conventional and virtual microscopy using a freely downloadable modified excel type spreadsheet (http://www.ccit.bcm.tmc.edu/jking/homepage/). Fleiss' (generalised) kappa can be used as a statistical measure for assessing the reliability of agreement between the 5 'raters' to a fixed number of items. Fleiss' kappa specifically assumes that although there a fixed number of raters, i.e. 5 panel pathologists, different items can be rated by different individuals. Cohen's kappa [15] was used to assess diagnostic agreement between the two methodologies for 2 trial audit pathologists (SAS, RH) and to compare the

diagnostic consensus between conventional and virtual microscopy. Cohen's kappa statistic works for two raters rating the same items, or items rated by two different methodologies.

Results

Of 61 slides examined with conventional microscopy 33 slides (54%) had a consensus neoplasia score of 1, 4 slides (6.6%) scored 2, 7 slides (11%) scored 3, 15 slides (25%) scored 4, and 2 slides (3.3%) scored 5
Of the subgroup of 33 slides examined with conventional microscopy 4 slides (12%) had a consensus neoplasia score of 1, 4 slides (12%) scored 2, 8 slides (24%) scored 3, 15 slides (45%) scored 4, and 2 slides (6%) scored 5
Scoring of 61 biopsies using conventional microscopy showed substantial reliability of diagnostic agreement between 5 panel pathologists (kappa = 0.712) Conventional microscopy assessment and neoplasia scoring of glass slides for the subgroup of 33 cases (with proportionately fewer 'easier' negative cases) showed moderate reliability of diagnostic agreement between 5 panel pathologists (kappa = 0.598) 18 slides had discrepant scores. (Table 2)

Table 2 here

Virtual microscopy assessment of the 33 digitised slides also showed moderate reliability of diagnostic agreement between panel pathologists, but with a lower Kappa score (kappa = 0.436) (Table 3) 25 slides had discrepant scores. However of 6 cases with a consensus neoplasia score of 1, 4 showed complete panel agreement

Table 3 here

Interobserver diagnostic agreement between the 2 principal trial pathologists using virtual microscopy was substantial (n=43) (Kappa = 0.76)

Diagnostic agreement comparing consensus neoplasia scores between conventional and virtual microscopy showed almost perfect reliability (kappa = 0.8769) (Table 4)

Table 4 here

Intraobserver agreement for the 2 principle pathologists comparing conventional and virtual microscopy was almost perfect (SAS - n=33 kappa = 0.926, RH - n=22 Kappa = 0.963) Comparison of consensus neoplasia scores between conventional and subsequent virtual microscopy (n= 33) showed full agreement in 70% of cases (n=23), a lower neoplasia score in 27% of cases (n=9) (8 cases by 1 point, and 1 case by 2 points) and a higher neoplasia score in 3% of cases (n=1) by 1 point.

Discussion

Diagnosis of Barrett's neoplasia using virtual slides has been used as a tool to study causation of diagnostic error in histopathology with poor diagnostic agreement (kappa 0.38) reported between 2 expert GI pathologists [16]. Poor interobserver diagnostic agreement between 7 pathologists has been reported for the diagnosis of dysplasia in ulcerative colitis irrespective of using virtual or conventional microscopy. Interestingly, using virtual slides, the diagnosis was changed in 51% usually downgrading dysplasia, and hence the validity of the virtual approach in the diagnostic setting was questioned [10]. Virtual microscopy has however been shown to be an effective tool for

diagnostic quality assurance of polyps detected in colorectal cancer screening [11] , assessment of prostatic biopsies in the external quality assurance scheme [12], and forms the basis of the pathology bowel cancer screening polyp External Quality Assurance (EQA) scheme in the UK (www.gieqa.org.uk. Scheme organiser Dr Nick Mapstone).

The practical difficulty of shipping glass slides between groups of pathologists for assessment and neoplasia scoring has limited the total number of cases reviewed in this study. However 'substantial' diagnostic agreement was achieved between 5 pathologists using conventional microscopy when the study cases included a larger proportion of slides negative for dysplasia (easier cases). Excluding the majority of the 'easier' negative slides (Vienna score 1) left a group of 33 proportionately more difficult slides for which Interobserver diagnostic agreement was reduced to moderate with conventional microscopy for the panel. This is consistent with previous reports that pathologists have, in particular, difficulties in reproducibly differentiating between indeterminate or borderline dysplasia (score 2) and true low grade dysplasia (score 3) [4, 5]. The panel diagnostic agreement remained moderate for the more difficult 33 slides using virtual microscopy, but with a lower kappa score and with only 8 slides achieving complete diagnostic agreement amongst the pathologist panel compared with 15 slides with complete agreement for conventional microscopy. A comparison of consensus neoplasia scores between conventional and virtual microscopy was almost perfect, however the consensus neoplasia score was lower in 27% of cases using virtual microscopy indicative of lower diagnostic confidence after scrutiny of virtual slides. Comments from participating

pathologists indicated that the required reporting time was increased and they felt more uncertain about making a diagnosis using virtual microscopy, which seemed to improve with familiarity and practice. For some pathologists this was compounded by hardware and software issues, internet access speeds, poor computer specification and small monitor size for on-screen assessment.

The practicality and efficacy of the technique of virtual microscopy as a means of evaluating slides by large groups of pathologists has been essentially previously validated in the setting of External quality Assurance (EQA) [12] and is upheld in our study, within a central pathology audit. On the one hand it is encouraging that moderate diagnostic agreement can be achieved with virtual microscopy in what is recognised to be a challenging area of diagnostic practice. However on the other hand the negative impact on diagnostic reliability/agreement and tendency for downgrading of neoplasia in the setting of Barrett's oesophagus in this study highlights possible limitations of this technology and it may not be suitable for neoplasia diagnosis as patient surveillance intervals and treatment would be based on this primary diagnosis. Lack of familiarity with on-screen diagnosis using virtual slides and technical issues would need to be further addressed. Given the majority of neoplasia score 1 ('normal') cases achieved full panel consensus (table 3) there may be a role for virtual microscopy in screening out cases that would then not require further scrutiny by conventional microscopy. To what extent the moderate diagnostic reproducibility achieved in this study could be replicated by general pathologists in the diagnostic setting is also open to question. We were unable to assess the impact on diagnostic reliability of the utility of immunohistochemistry, in particular p53

Deleted: for its

Deleted: use in primary

immunostaining, in this current study as trial centre pathologists had rarely carried out p53 immunostains in their diagnostic practice.

In the setting of the AspECT trial pathology audit panel, consensus neoplasia diagnosis using virtual slides provides a practical and valid approach but, on the basis of the data presented, a review of cases by conventional microscopy may be prudent where there is significant diagnostic discrepancy at a future point in the trial between central review and trial centre diagnoses.

Conflict of interests

The authors state that there are no competing interests to declare

Acknowledgements

The authors would like to thank the Virtual Scanning Team, Leeds Institute of Molecular Medicine, St. James's University Hospital, Leeds for digitisation of the glass slides and and staff in the Oncology Clinical Trials Office (OCTO), Department of Oncology, Old Road Campus Research Building, University of Oxford for on-going support of the AspECT trial and pathology audit.

Collaborators;

AspECT Trial Management Group (TMG);

J de Caestecker

S Love

P Watson

S Attwood

H Barr

C Brooks

P Moayyedi

P Bhandari

D Johnson

Advisory non executive TMG members;

M Anderson

C Edwards

D Morris

C Rogers

M Gibbons

K Ragunath

AspECT Trial Principal Investigators;

I Penman,

Y Ang,

N Church

References

1) Eaden J, Abrams K, MCKay H, Denley H, Mayberry J. Inter-observer variation between general and specialist gastrointestinal pathologists when grading dysplasia in ulcerative colitis. J Pathol 2001;1994:152-157

2) Odze RD, Goldblum J, Noffsinger A, Alsaigh N, Rybicki LA, Fogt F. Interobserver variability in the diagnosis of ulcerative colitis-associated dysplasia by telepathology. Mod Pathol 2002;15:379-386

3) Reid BJ et al. Observer variation in the diagnosis of dysplasia in Barrett's esophagus Human Pathol 1988;19:166-178

4) Montgomery E, Bronner MP, Goldblum JR et al. Reproducibility of the diagnosis of dysplasia in Barrett's esophagus. Hum Pathol 2001;32:368-378

5) Montgomery E, Goldblum JR, Greenson JK et al. Dysplasia as a predictive marker for invasive carcinoma in Barrett esophagus: a follow up study based on 138 cases from a diagnostic variability study. Human Pathol 2001;32:379-388

6) 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;50:920-927

7) Odze RD. Diagnosis and grading of dysplasia in Barrett's oesophagus. J Clin Path 2006;59:1029-1038

8) Wani S, Falk G, Post J, et al. Risk factors for the progression of low grade dysplasia in patients with Barrett's esophagus. Gastroenterology 2011; 141:1179-1186

9) Kaye PV, Haider SA, Ilyas M, et al. Barrett's dysplasia and the Vienna classification: reproducibility, prediction of progression and impact of consensus reporting and p53 immunohistochemistry. Histopathology 2009;54:699-712

10) Odze RD, Tomaszewski JE, Furth EE, et al. Variability in the diagnosis of dysplasia in ulcerative colitis by dynamic telepathology. Oncol Rep 2006:16:1123-1129

11) Risio M, Bussolati G, Senore C, et al. Virtual microscopy for histology quality assurance of screen detected polyps. J Clin Pathol 2010;63:916-920

12) Harnden P, Coleman D, Moss S, et al. Evaluation of the use of digital images for a national prostate core external quality assurance scheme. Histopath 2011;59:703-709

13) Schlemper RJ, Kato Y, Stolte M. Review of histological classifications of gastrointestinal epithelial neoplasia: differences in diagnosis of early carcinomas between Japenese and Western pathologists. J Gastroenterol 2002;36:445-456

14) Fleiss, JL. Measuring nominal scale agreement among many raters. Psychological Bulletin 1971;76:378-382

15) Cohen J. A coefficient of agreement for nominal scales. Educ. Psych. Meas.1960; 20:37-46.

16) Treanor D, Chee Hooi Lim, Derek Magee, Andy Bulpitt and Phil Quirke. Tracking with virtual slides: a tool to study diagnostic error in histopathology Histopathology 2009; 55:37-45

Category	Diagnosis
1	Negative for dysplasia
2	Indefinite for dysplasia
3	Low grade dysplasia
4	High grade dysplasia, intramucosal carcinoma, Suspicious of invasive carcinoma, carcinoma-in-situ
5	Submucosal invasion by carcinoma

Tabel 1 – Modified Vienna classification of GI tract neoplasia

biopsy	1	2	3	4	5	Consensus neoplasia score
1	5	0	0	0	0	**1**
2	5	0	0	0	0	**1**
3	0	0	1	4	0	**4**
4	0	0	0	5	0	**4**
5	5	0	0	0	0	**1**
6	0	0	0	5	0	**4**
7	0	0	1	4	0	**4**
8	5	0	0	0	0	**1**
9	0	0	4	1	0	**3**
10	0	0	0	5	0	**4**
11	3	1	1	0	0	**2**
12	2	2	1	0	0	**2**
13	0	0	1	4	0	**4**
14	0	0	0	5	0	**4**
15	4	1	0	0	0	**1**
16	0	0	5	0	0	**3**
17	0	0	5	0	0	**3**
18	0	0	5	0	0	**3**
19	0	0	0	3	2	**4**
20	0	1	0	4	0	**4**
21	0	0	0	1	4	**5**
22	0	0	0	5	0	**4**
23	0	3	2	0	0	**2**
24	0	1	4	0	0	**3**
25	0	0	0	3	2	**4**
26	0	0	0	0	5	**5**
27	0	0	0	3	2	**4**
28	0	0	5	0	0	**3**
29	0	0	2	3	0	**4**
30	0	0	0	5	0	**4**
31	0	0	0	3	2	**4**
32	2	1	2	0	0	**2**
33	0	0	3	2	0	**3**

Table 2. Diagnostic consensus for conventional microscopy. Top row in bold = neoplasia score (1-5), left column in bold = biopsy number (n=33). Rows represent number of pathologists (between 0 and 5) for each neoplasia score. Cases with discrepant scores highlighted in yellow (n=18). Consensus neoplasia scores in bold in right hand column. Kappa = 0.598

biopsy	1	2	3	4	5	Consensus neoplasia score
1	5	0	0	0	0	**1**
2	4	1	0	0	0	**1**
3	0	1	3	1	0	**3**
4	0	0	1	4	0	**4**
5	5	0	0	0	0	**1**
6	0	0	0	3	2	**4**
7	0	0	3	2	0	**3**
8	5	0	0	0	0	**1**
9	2	0	3	0	0	**2**
10	0	0	0	5	0	**4**
11	2	3	0	0	0	**2**
12	2	0	3	0	0	**2**
13	0	1	0	2	2	**4**
14	0	0	0	5	0	**4**
15	5	0	0	0	0	**1**
16	0	1	4	0	0	**3**
17	0	3	2	0	0	**2**
18	1	0	4	0	0	**3**
19	0	0	0	2	3	**5**
20	0	1	0	4	0	**4**
21	0	0	0	1	4	**5**
22	0	0	0	3	2	**4**
23	3	1	1	0	0	**2**
24	2	2	1	0	0	**2**
25	0	0	0	4	1	**4**
26	0	0	0	0	5	**5**
27	0	0	0	2	3	**5**
28	2	1	2	0	0	**2**
29	0	3	2	0	0	**2**
30	0	0	0	4	1	**4**
31	0	0	1	4	0	**4**
32	4	0	1	0	0	**1**
33	0	0	5	0	0	**3**

Table 3. Diagnostic consensus for virtual microscopy. Top row in bold = neoplasia score (1-5), Left column in bold = biopsy number (n=33). Rows represent number of pathologists (between 0 and 5) for each neoplasia score. Cases with discrepant scores highlighted in yellow (n=25). Consensus neoplasia scores in bold in right hand column. Kappa = 0.436

	S1	S2	S3	S4	S5	Total
S1	5	0	0	0	0	5
S2	1	3	0	0	0	4
S3	0	4	3	0	0	7
S4	0	1	2	10	2	15
S5	0	0	0	0	2	2
Total	6	8	5	10	4	33

Cohen's Kappa = 0.877

Table 4. Diagnostic agreement comparing consensus neoplasia scores between conventional and virtual microscopy (n=33) (kappa = 0.877)