

Original citation:

Snead, David R. J., Tsang, Yee-Wah, Meskiri, Aisha, Kimani, Peter K., Crossman, Richard, Rajpoot, Nasir M., Blessing, Elaine, Chen, Klaus, Gopalakrishnan, Kishore, Matthews, Paul, Momtahan, Navid, Read-Jones, Sarah, Sah, Shatrughan, Simmons, Emma, Sinha, Bidisa, Suortamo, Sari, Yeo, Yen, El Daly, Hesham and Cree, Ian A.. (2015) Validation of digital pathology imaging for primary histopathological diagnosis. *Histopathology* . doi: 10.1111/his.12879

Permanent WRAP url:

<http://wrap.warwick.ac.uk/74831>

Copyright and reuse:

The Warwick Research Archive Portal (WRAP) makes this work of researchers of the University of Warwick available open access under the following conditions.

This article is made available under the Creative Commons Attribution-NonCommercial 4.0 (CC BY-NC 4.0) license and may be reused according to the conditions of the license. For more details see: <http://creativecommons.org/licenses/by-nc/4.0/>

A note on versions:

The version presented in WRAP is the published version, or, version of record, and may be cited as it appears here.

For more information, please contact the WRAP Team at: publications@warwick.ac.uk

warwick**publications**wrap

highlight your research

<http://wrap.warwick.ac.uk>

Validation of digital pathology imaging for primary histopathological diagnosis

David R J Snead,^{1,2} Yee-Wah Tsang,^{1,2} Aisha Meskiri,² Peter K Kimani,³ Richard Crossman,³ Nasir M Rajpoot,^{2,4} Elaine Blessing,¹ Klaus Chen,¹ Kishore Gopalakrishnan,¹ Paul Matthews,¹ Navid Momtahan,^{1,5} Sarah Read-Jones,¹ Shatrughan Sah,¹ Emma Simmons,¹ Bidisa Sinha,¹ Sari Suortamo,¹ Yen Yeo,¹ Hesham El Daly¹ & Ian A Cree^{1,2}

¹Department of Cellular Pathology, University Hospitals of Coventry and Warwickshire NHS Trust, Coventry, UK,

²Centre of Excellence for Digital Pathology, University Hospitals of Coventry and Warwickshire NHS Trust, Coventry, UK, ³Warwick Medical School, University of Warwick, Coventry, UK, ⁴Department of Computer Science, University of Warwick, Coventry, UK, and ⁵Histopathology Department, City Hospital, Birmingham, UK

Date of submission 9 July 2015

Accepted for publication 23 September 2015

Published online Article Accepted 26 September 2015

Snead D R J, Tsang Y-W, Meskiri A, Kimani P K, Crossman R, Rajpoot N M, Blessing E, Chen K, Gopalakrishnan K, Matthews P, Momtahan N, Read-Jones S, Sah S, Simmons E, Sinha B, Suortamo S, Yeo Y, Daly H El & Cree I A

(2015) *Histopathology*. DOI: 10.1111/his.12879

Validation of digital pathology imaging for primary histopathological diagnosis

Aims: Digital pathology (DP) offers advantages over glass slide microscopy (GS), but data demonstrating a statistically valid equivalent (i.e. non-inferior) performance of DP against GS are required to permit its use in diagnosis. The aim of this study is to provide evidence of non-inferiority.

Methods and results: Seventeen pathologists re-reported 3017 cases by DP. Of these, 1009 were re-reported by the same pathologist, and 2008 by a different pathologist. Re-examination of 10 138 scanned slides (2.22 terabytes) produced 72 variances between GS and DP reports, including 21 clinically significant variances. Ground truth lay with GS in 12 cases and with DP in nine cases. These results

are within the 95% confidence interval for existing intraobserver and interobserver variability, proving that DP is non-inferior to GS. In three cases, the digital platform was deemed to be responsible for the variance, including a gastric biopsy, where *Helicobacter pylori* only became visible on slides scanned at the $\times 60$ setting, and a bronchial biopsy and penile biopsy, where dysplasia was reported on DP but was not present on GS.

Conclusions: This is one of the largest studies proving that DP is equivalent to GS for the diagnosis of histopathology specimens. Error rates are similar in both platforms, although some problems e.g. detection of bacteria, are predictable.

Keywords: diagnosis, digital pathology, histopathology, validation, whole slide imaging

Introduction

Digital pathology (DP) is the conversion of the light microscope image of a slide into a set of digitized files that allow the reproduction of the original slide on a

computer workstation; it is also called virtual microscopy or whole slide imaging. Digitization also allows for the manipulation of the image and or its data to derive information of diagnostic, prognostic or therapeutic benefit, which would otherwise not normally be readily available.^{1,2}

Although it is widely used in research and teaching, this technology has only relatively recently become applicable to routine diagnostic work,

Address for correspondence: Dr D Snead, Pathology Department, UHCW NHS Trust, Coventry CV2 2DX, UK. e-mail: david.snead@uhcw.nhs.uk

through the development of large-capacity slide scanners capable of dealing with the workloads generated daily in modern diagnostic laboratories. These benefits include the ability to report cases remotely from the surgical laboratory generating the slides, thereby allowing for increased and more robust subspecialization, and more efficient use of the pathologist's time, the adoption of computerized analysis to aid and improve diagnosis, and the potential to automate some aspects of the pathologist's workload.

An important initial step in the wider adoption of this technology is the establishment of validation data assessing how effective pathologists are using digital workstations in comparison with conventional light microscopes and glass slide microscopy (GS) when examining cases for primary diagnosis. Previous studies using a range of differing technologies in different clinical settings have reported the comparison of DP with GS,^{3–17} and criteria for such comparisons have been published.¹⁸ However, although most of these studies have shown promise, no single study has been sufficiently powered to demonstrate a statistically significant equivalence (i.e. non-inferiority) of DP against GS.

Histopathology is an interpretive discipline, and pathologists' performance in the analysis of diagnostic surgical specimens naturally varies both within and between observers, as has been measured in previous studies.^{19,20} To achieve a statistically valid outcome, any comparative study must include sufficient specimens to allow for this variation, which should ideally be measured among the group of pathologists taking part in the study. In addition to audit studies,¹⁹ UK National Health Service (NHS) laboratories routinely review a proportion of their cases in multidisciplinary team (MDT) meetings, which are used in the planning of future care for the patient. Recording the number of variances detected when these cases are reviewed at MDT meetings gives a useful measure of a laboratory's baseline variation rate, i.e. GS to GS variance, which can inform the sample size calculation for DP validation studies.

Comparison studies of this type require the establishment of the correct diagnosis or 'ground truth', so that each method can be compared with the ground truth independently, thereby avoiding bias from either method. Finally, whichever method is used first, there must be a period of elapsed time before the same case is re-evaluated with the alternative method, to avoid the pathologist(s) remembering the case from previous viewings, the so-called 'wash-out period'.

In this study, we retrospectively evaluated a series of cases reported on GS and then re-reported on DP

workstations, to establish whether there was a significant difference between the diagnoses achieved beyond what would normally be expected through intraobserver and interobserver variation.

Materials and methods

ETHICS

The study was approved by the National Research Ethics Service London Dulwich 12/LO/0993.

INTRAOBSERVER AND INTEROBSERVER VARIATION

Data from MDT meetings were recorded over a period of 12 months. The total number of specimen requests reviewed and the number of changes made that could, or would, have resulted in altered clinical management were recorded from each MDT meeting. These results were pooled to generate the overall combined interobserver and intraobserver variation across all of the MDT teams.

CASE SELECTION, SLIDE SCANNING, AND VIEWING

Prior to the start of the trial, biomedical scientists and histopathologists were trained in using the DP system for a period of 6 weeks as part of a beta study; during this time, 284 cases were reviewed and re-reported. These cases were not included in the present study, and formed a training set for those taking part in the current study.

Request cards from reported cases area were selected from the filing tray in the surgical laboratory. Each subspecialty area was targeted, and the selection of individual cases within each area was sequential. All cases were scanned on the Omnyx VL4 scanner (LLC 1251; Omnyx, Pittsburgh, PA, USA), with a $\times 40$ (0.274 $\mu\text{m}/\text{pixel}$) setting as a default, changing to $\times 60$ (0.137 $\mu\text{m}/\text{pixel}$) for renal biopsies, and cases that required Giemsa, Ziehl–Neelsen or Gram stains for the detection of microorganisms. The Omnyx Integrated Digital pathology (IDP) system uses a proprietary lossy compression algorithm that varies between slides and typically results in a compression of 20:1. The slides were viewed on the Omnyx IDP workstation (LLC 1251). The workstation is equipped with two Hewlett-Packard (HP ZR2440wL ED Backlit LCD) monitors, with a resolution of 1920 \times 1200 at 60 Hz, and pixel pitch of 0.270 mm (Hewlett-Packard, Palo Alto, CA, USA). Colour calibration was carried out with a Spyder4Pro display calibration unit

(Datacolor, Lawrenceville, NJ, USA). The entire system was installed by use of the existing information technology (IT) network. Scanned whole slide images were uploaded to a server located within the main IT server hall. Connectivity between the scanners, servers and workstations was protected to 1 gigabyte/s. The entire system architecture was maintained within the Trust firewall, and protected by user login and password.

All of the slides for each case were scanned, including any special stains or immunocytochemistry that had been performed. All marks were removed from the slides before scanning, and any cases with scratched or heavily marked coverslips were re-cover-slipped before they were scanned. The scanned images were evaluated by a laboratory technician to assess image quality, focus, and completeness of the case, before the case was released for re-reporting. Once 3 weeks had elapsed from the sign-out of the original report, the case was released to the pathologists' worklist for re-reporting.

The cases were reported by pathologists working within their subspecialty areas. To reflect the departmental practice of the MDT review process, some of the cases were reported by the same pathologist, and the remainder by other pathologists in the same subspecialty team. Pathologists requested repeat scanning or re-scanning at $\times 60$ magnification whenever they felt that this was needed for reporting of the case.

ESTABLISHMENT OF GROUND TRUTH AND VARIANCE, AND ERROR GRADING

The ground truth was established for each case following double reporting. Once the DP report had been issued, this was compared with the original GS report by the research assistant (A.M.). Cases in which any of the GS and DP reports, including, where appropriate, frozen sections, differed were reviewed at a fortnightly steering group meeting, consisting of the research assistant and the participating pathologists. The steering group agreed which variances were real, i.e. not simply terminological differences, and graded these into clinically significant, i.e. would or could result in a change to clinical management, or clinically insignificant. The concordant diagnoses reached by the pathologist(s) formed the ground truth in cases without variance. When the diagnoses did not concur, the reporting pathologist(s) reviewed both the GS and DP sections together to reach a consensus diagnosis. The ground truth then lay with the platform on which that diagnosis was

made. When two pathologists failed to agree, cases were referred to a third pathologist for arbitration.

SAMPLE SIZE CALCULATION

The sample size was based on a non-inferiority hypothesis test. The audit of MDT reviews for 2011 at the University Hospitals of Coventry and Warwickshire NHS Trust indicated that initial and review diagnoses are concordant for 98.78% of the time over all subspecialty areas, and so, for the sample size calculation, we assumed that the percentage of agreement between DP and GS is 98.8%. DP was considered to be non-inferior if the lower 95% confidence interval (CI) for the percentage agreement was $>98\%$. We concluded that, for 95% power, we required 3014 cases.

STATISTICAL ANALYSIS

In line with the departmental MDT review practice, the DP and GS results for some samples were reported by the same pathologist, and those for some samples were reported by different pathologists. In the analysis, we considered two forms of concordance, i.e. two endpoints: concordance defined as complete concordance or variance of no clinical significance (primary endpoint); or complete concordance (secondary endpoint). For the primary analysis, we performed three sets of analyses, with the first consisting of all samples, the second consisting of samples for which the DP and GS diagnoses were reported by the same pathologist, and the third consisting of samples for which the DP and GS diagnoses were reported by different pathologists. For each set of analyses, we calculated the percentage of samples for which DP and GS diagnoses were concordant and the 95% CI. The 95% CI was calculated by using the normal approximation to the binomial distribution. For the secondary analysis, we calculated the percentage of samples for which DP and GS diagnoses were concordant or the ground truth was provided by DP and the corresponding 95% CI.

Results

A total of 3103 cases were selected and scanned. The study closed when 3017 cases had been double-reported, leaving 86 cases that were incomplete at the conclusion of the study, comprising 24 that had been examined on the digital system, but required additional work (re-scanning of one or more slide,

and cases incomplete at the time of scanning), and 62 that were awaiting re-reporting on the digital system.

The proportion of cases in each subspecialty group is shown in Table 1.

Ninety-seven (3.2%) cases required re-scanning before a DP report could be offered. No cases required additional stains or deeper sections before the DP report was issued.

The 3017 cases generated 10 138 slides, which, when scanned, resulted in a digital archive of 2.22 terabytes. Details of the mean scanning time and the size of the files generated are shown in Table 2.

The cases were re-reported by the same pathologist in 1009 (33.4%) instances and by a different pathologist in 2008 (66.6%) instances. A total of 17 pathologists took part in the study, comprising all substantive consultant histopathologists within the department, two visiting consultant pathologists, and one senior trainee. One pathologist experienced eye strain because of glare when using the digital workstation, owing to a longstanding eye condition, which was successfully alleviated by placing a filter (3M Privacy Filter; 24.0-inch Widescreen 16:10; PF24.0W) over the workstation screens. The other contributing pathologists reported no difficulties in using the system.

A comparison of diagnoses made with GS and DP is shown in Table 3. A total of 72 cases (2.3%) showed variance between GS and DP reports. Of

these, 21 (0.7%) were deemed to be of clinical significance, and the details of these are shown in Table 4. Of these discrepancies, the ground truth lay with GS in 12 (57%) cases and with DP in nine (43%) cases. In the majority of cases, the variances were considered to probably represent normal intraobserver and interobserver variation. There were three occurrences for which the reporting pathologists felt that the digital platform was directly responsible for the variance. These were a gastric biopsy in which *Helicobacter pylori* was not visible on the DP images, and only became so when the slides were re-scanned at the $\times 60$ setting, a bronchial biopsy showing squamous metaplasia in which moderate dysplasia was reported on DP, and a penile biopsy showing human papilloma virus changes in which penile intraepithelial neoplasia was reported on DP.

The primary analysis results are summarized in Figure 1. For concordance defined as complete concordance or variance of no clinical significance, the conclusions for analyses based on all samples, samples diagnosed by same pathologists and samples diagnosed by different pathologists were similar. The lower CI bounds were above the non-inferiority boundary (98%), and so DP was non-inferior to GS. The performances of the same pathologist and different pathologists were also similar with regard to complete concordance against all variances. The secondary analysis results are given in Figure 2. As expected from the primary analysis results, for the percentage of samples for which there was complete

Table 1. Distribution of cases across different subspecialty teams

Specialty	Cases, <i>n</i> (%)
Breast	253 (8.4)
Dermatopathology	539 (18)
ENT	257 (8.5)
GIT	405 (13.4)
General pathology	487 (16.1)
Gynaecological	377 (12.5)
Lymphoreticular	166 (5.5)
Renal	94 (3.1)
Respiratory	197 (6.5)
Urology	242 (8)
Total	3017

ENT, Ear, nose and throat; GIT, gastrointestinal tract.

Table 2. Summary of slides scanned and the data generated in the validation study

Cases	3017
Biopsies	2666
Resections	340
Frozen sections	11
Total number of slides	11 522
Mean slides per case	3.8
Most slides per case	37
Slides scanned at $\times 40$	10 138
Slides scanned at $\times 60$	1384
Range of data per slide	3.51–3183 MB
Mean data per slide	189 MB
Total data generated	2.22 TB

MB, megabytes; TB, terabytes.

Table 3. Summary of the results

Did DP and LM give the same diagnosis?	Were DP and GS reported by the same pathologist?		Total, <i>n</i> (%)
	Yes, <i>n</i> (%)	No, <i>n</i> (%)	
Yes	981 (32.5)	1964 (65)	2945 (97.6)
No			
No clinical difference	19 (0.6) [11, 8]*	32 (1) [14, 18]*	51 (1.7) [25, 26]*
Clinical difference	9 (0.3) [1, 8]*	12 (0.4) [8, 4]*	21 (0.7) [9, 12]*
Total	1009 (33.4)	2008 (66.6)	3017

DP, Digital pathology; GS, glass slide microscopy; LM, light microscopy.

*The numbers in brackets correspond to the numbers of samples for which the ground truth was provided by DP and GS, respectively.

concordance, there was variance of no clinical significance, or DP provided the ground truth, DP was non-inferior to GS. For the percentage of samples for which there was complete concordance or DP provided the ground truth, for analysis based on all samples and analysis based on samples diagnosed by different pathologists, DP was non-inferior to GS. However, the results were indeterminate for analysis based on samples diagnosed by the same pathologists.

Discussion

This study is the largest to date of a pathologist's ability to assess histopathology samples on DP as opposed to GS. In line with previous guidelines, we instigated a small pilot study prior to this validation, with the aim of giving each of the pathologists experience in using the DP system.¹⁸ Measurement of the departmental observer variability prior to the commencement of the study ensured that it was adequately powered to confirm that there is no difference in reports made on glass slides and those made on computer workstations. All of the pathologists were able to use the workstations for reporting, and, with the notable exception of oversized slides, the cases included in the study covered the full spectrum of histopathology specimens in the department, including frozen sections, immunocytochemistry, and special stains.

From 3017 cases, differences in diagnosis were identified in 72 (2.3%) cases, of which 21 (0.7%) would have, or were likely to have, resulted in a difference in patient management. These cases are all listed in Table 4, along with the number of slides that they contained; approximately half were single-slide cases, and the others were multi-slide cases. The majority of these variances mapped to a single slide

and were very similar to our observed variability prior to the study commencing, as indicated by the ground truth being approximately equally distributed between GS and DP. Several variances were attributable to instances in which important morphological clues, such as *Candida* hyphae, had been missed on the GS, and it is reassuring that these were picked up on review with DP. The observer variation in our department was 1.22%. This is smaller than the figure used in some comparable studies,¹³ and hence necessitated a larger number of cases to achieve the 95% power targeted. Nevertheless, the observed variation is in line with other studies.²¹ Interestingly, the variation observed in the study was less than that seen in the MDT review audit. There are several reasons why this might be so. First, the MDT review audit looked at both macroscopic and microscopic sections of the report, whereas the validation study reported here only examined differences in the microscopic report. Second, many of the cases included in the validation study had already been through MDT meetings prior to recruitment, and had therefore been subjected to one review prior to the study review. Finally, the case population for this study was consecutively selected, whereas MDT review cases are selected clinically from patients suspected of having malignancy, where the incidence of clinical significant variances might be expected to be higher.

However, two situations were identified that gave cause for concern regarding the DP system. First, two cases were reported on DP as showing dysplasia that was not present. These two cases were reported by different pathologists, but with the same pathologist reporting both the GS and DP sections. In both instances, no satisfactory reason for the difference was apparent, apart from the fact that the nuclei appeared much darker on DP than on GS, giving an

Table 4. A list of the clinically significant discordant cases, summarizing the glass slide microscopy (GS) and digital pathology (DP) diagnoses and where the ground truth lay

Subspecialty	GS	DP	Ground truth	Same path, yes/no	No. of slides	B/R
Dermatology	Melanoma with microsatellite	Melanoma without microsatellite	GS	Yes	3	B
Dermatology	Melanoma <i>in situ</i>	Melanoma, radial growth phase	GS	Yes	6	B
Dermatology	Frictional keratosis	Actinic keratosis	DP	No	1	B
Dermatology	Hyperkeratosis	Actinic keratosis	DP	Yes	1	B
H&N	Colloid nodule	Papillary microcarcinoma	DP	No	5	R
H&N	Inflammation	Oral candidiasis	DP	Yes	2	B
H&N	Inflammation	Oral candidiasis	DP	Yes	2	B
H&N	Non-erupted tooth	Inflammation NOS	DP	No	2	B
GIT	<i>Helicobacter pylori</i> gastritis	Gastritis	GS	Yes	1	B
GIT	Small-bowel adhesions	Ischaemic bowel	DP	No	7	R
Gynaecological	CIN1 + HPV*	HPV	GS	No	1	B
General pool	Rheumatoid nodule	Organizing haemorrhage	DP	No	16	R
Respiratory	Metaplasia	Moderate dysplasia	GS	Yes	1	B
Respiratory	NSCLC NOS	NSCLC, favours squamous	GS	Yes	10	B
Respiratory	Suspicious for SCC	Inflammation	GS	Yes	1	B
Urological	HPV with atypia	Penile intraepithelial neoplasia	GS	Yes	1	B
Urological	Prostate core suspicious	Prostate core benign	GS	Yes	10	B
Urological	Prostatic carcinoma Gleason grade 3 + 4 = 7	Prostatic carcinoma Gleason grade 3 + 3 = 6	GS	Yes	3	B
Urological	TCC, WHO grade 1 Low grade	TCC, WHO grade 2 High grade	GS	No	1	B
Urological	TCC with no CIS	TCC with CIS	DP	No	1	B
Urological	TCC non-invasive	TCC early invasion	GS	Yes	1	B

B, Biopsy; CIN, cervical intraepithelial neoplasia; CIS, carcinoma *in situ*; GIT, gastrointestinal tract; H&N, head and neck; HPV, human papilloma virus; NOS, not otherwise specified; NSCLC, non-small-cell lung cancer; R, resection; SCC, squamous cell carcinoma; TCC, transitional cell carcinoma; WHO, World Health Organization.

*Note that some healthcare systems do not distinguish between HPV and CIN1.

erroneous impression of dysplasia. This potential problem may benefit from further investigation, as we are aware of other groups who have encountered similar problems in Barrett's oesophagus (Dr D Treanor, personal communication) and cervical squamous dysplasia.¹⁴ The other situation concerned the detection of microorganisms. A case of *H. pylori*-positive gastritis was missed because the organisms were only visible on the $\times 60$ scans. This led to the adoption of scanning rules whereby all such cases are scanned at $\times 60$ by default. Other studies examining

this problem have shown that z-plane focusing is also an effective means of improving detection of these organisms.²²

The intensity of the red colour in some special stains, namely diastase periodic acid schiff (DPAS) positivity for fungi and positive staining for Ziehl-Neelsen in mycobacteria, was noticed to be not as strong on DP as on GS. This problem is related to the colour balance of the DP image, and other studies have also identified this as a problem for identification of eosinophils.¹³ Once this problem had been identi-

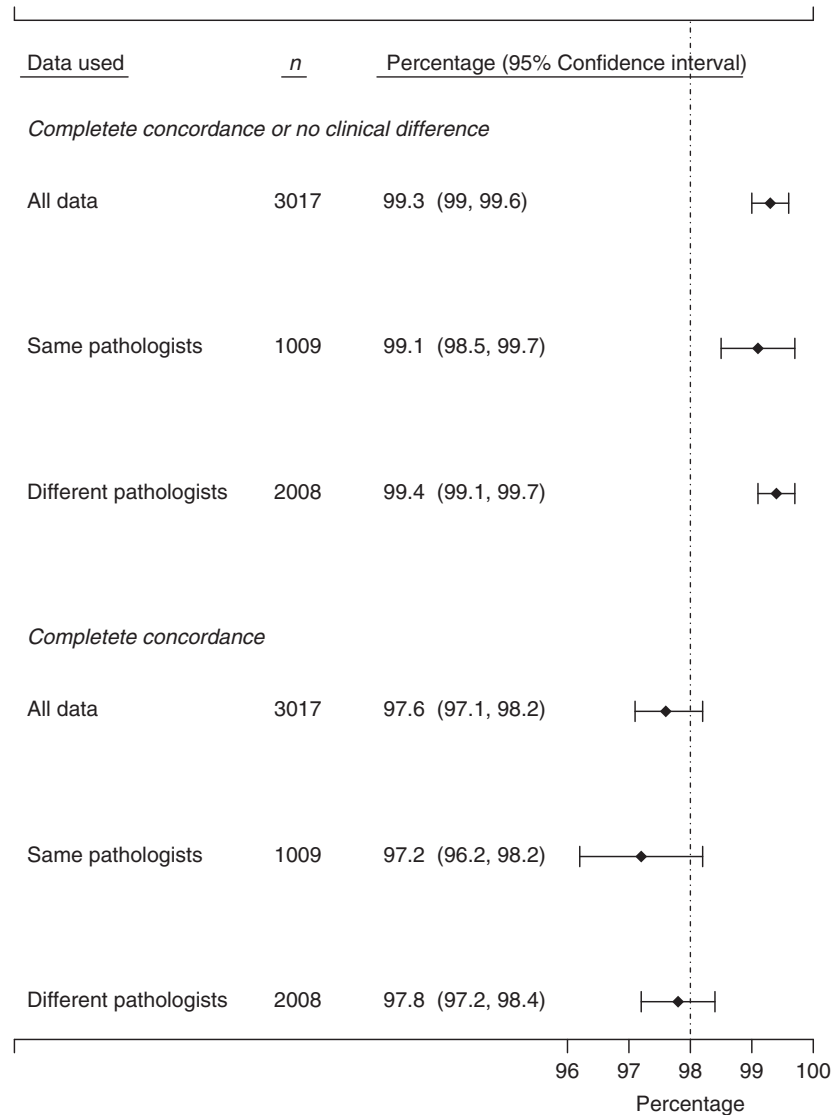


Figure 1. Primary analysis results. The vertical dashed line at 98% indicates the lower boundary of the confidence interval generated from intraobserver and interobserver variation audit data. The upper half of the figure shows variances of no clinical significance merged with completely concordant cases, equivalent to these audit data, clearly indicating that DP is not inferior to GS. The lower half shows the percentage of cases that were completely concordant.

fied, all of the monitors in use were checked, and the colour balance was formally standardized. No further problems were encountered. As with all special stains, it is imperative to include appropriate positive controls to educate the observer about the intensity of the positive stain achieved in any particular run. These events support the findings of other groups that have highlighted specific areas in which DP may differ from conventional GS, and highlight the need for local validation before its implementation.²³

In our series, we did not encounter any problems with z-plane (vertical) focusing, a weak point of several DP systems that do not allow focusing through the plane of the section. This has been cited as a cause for error in some of the older studies,²⁴ and, although this factor may still be relevant for cytology

samples, it does not seem to be a common problem in histopathology samples, where sections are now routinely cut by most departments at 4 μ m.

The DP system gives the operator some advantages, but also imposes some limitations. Although not formally assessed in this study, some aspects of routine reporting were improved by DP; for example, measurements of tumour depth, diameter or margin of clearance appeared to be much quicker and more accurate with the DP system than with GS. Likewise, DP keeps the entire case together, so there is no time lost in searching for individual slides of a case. There is clearly potential for rapid review of previous cases held on an individual patient as the digital archive grows, which will reduce the time lost in waiting for slides to be retrieved from the file. When the laboratory is

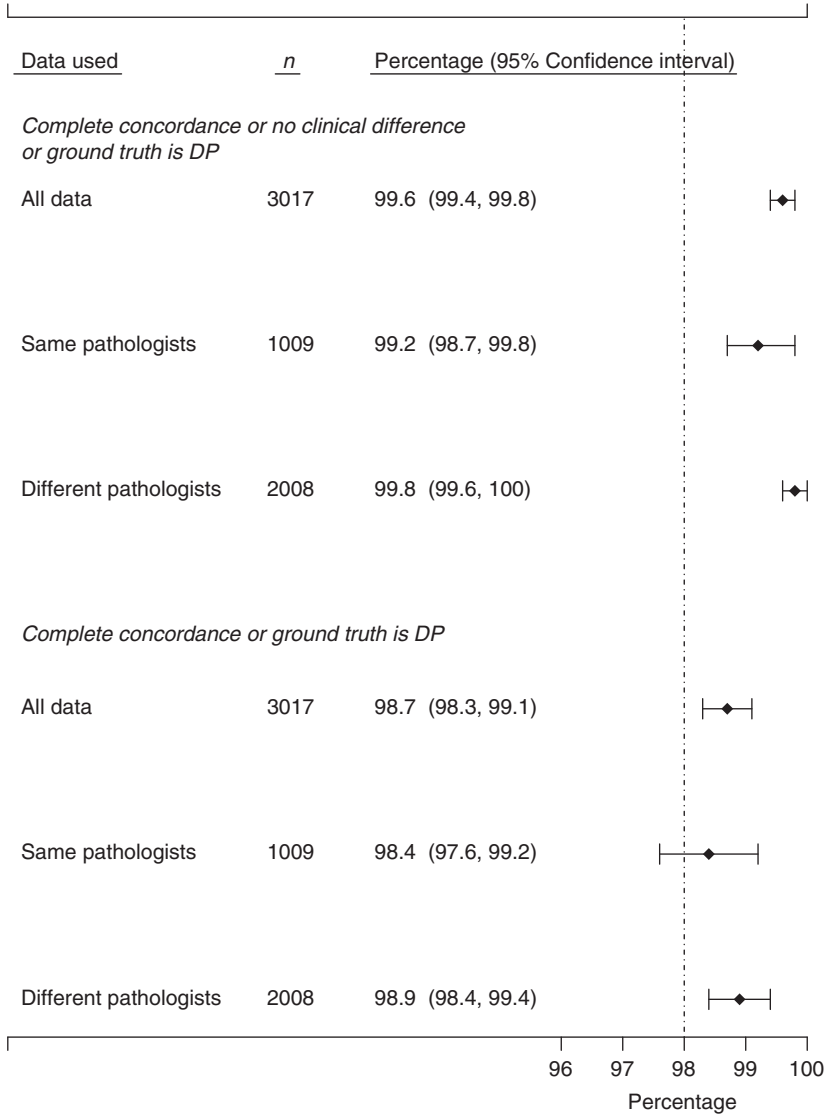


Figure 2. Secondary analysis results. The variances for which the ground truth lay with DP are merged with complete concordance, with (top) and without (bottom) variances that had no clinical relevance.

working fully digitally, there is great flexibility in distributing the reporting workload across the consultant workforce and in sharing of difficult cases, which is likely to be particularly relevant in highly complex small-volume areas of pathology, such as transplant and renal pathology.²⁵ The impacts that these advantages may have on the diagnostic service are quite complex and outside the scope of this study. Naturally, the switch from GS to DP takes time to adapt to, but all of the pathologists in this study became sufficiently confident in using the system to be happy in ‘signing out’ their diagnoses. The examination of multi-block resection cases did not cause any problems, and nor did these cases appear to be more cumbersome or time-consuming to report on DP than GS.

However, some pathologists found the systematic screening of slides at medium power ($\times 10$ magnifica-

tion) to be more difficult with the DP track ball used in this study than with a conventional microscope stage. The measurement of mitotic rates requires the operator to record mitoses per unit area, which is, in any case, more precise than high-power field areas, which should not, in general, be used. DP permits this to be calculated relatively easily by showing the size of the viewing area on the screen, and offers the potential for automating this function through algorithm development. It is not possible to examine slides with polarized filters with the current generation of DP systems, and immunofluorescent slides could not be scanned on the system used in this study. Of all of the samples studied, these limitations affected the assessments of renal biopsies most. The renal biopsies in this study were scanned at $\times 60$, but, even with this resolution, the interpretation of subtle features of membranous

change was difficult. However, immunohistochemistry, which is routinely used to increase the sensitivity for membranous nephropathy in any case, was clearly interpretable on the DP system. Congo red stains could not be examined under polarized light on the DP system, but areas of positive staining could nevertheless be detected, if not confirmed, on DP. It is possible that alternative strategies for identifying amyloid, such as immunocytochemistry for protein P, may assume dominance in the digital era.

Histopathology is a scientific discipline based on the visual observation of the variance of tissues from normal, aligned with knowledge of the pathological processes likely to cause the changes observed, and the astute use of additional tests on the tissue samples to provide additional information of use in the making of a diagnosis. DP enables pathologists to make the same judgements on a computer screen as they would with a microscope. DP image quality at high power is not as good as that obtained with a microscope, to the extent that identifying bacteria, small-cell carcinoma and granulocytic inflammatory cells was best performed by scanning the slides at $\times 60$. Similar observations were made in the study by Bauer *et al.*,¹³ suggesting that, in future, many specimens will have to be scanned at $\times 60$ for diagnosis. However, DP images are clearly adequate for reporting the majority of cases. Retention of the facility to use GS when needed is clearly important at the moment. It remains to be seen whether this will continue to be the case as the ability of slide scanners progresses and alternative strategies to circumvent the existing problems are developed.

In conclusion, this study shows that DP is equivalent to GS for the vast majority of tissue specimens taken for diagnosis. Error rates are similar, and, although there may be some differences in the types of error encountered, some of these can be mitigated by acquiring experience with the system and the appropriate use of higher-resolution scans.

Acknowledgements

This study was supported by an educational grant from Omnyx LLC, 1251 Waterfront Place, Pittsburgh, PA 15222, USA.

Author contributions

D. Snead designed the study, took part in the GS and DP reporting, took part in the steering group meetings, assisted with the database management, and wrote the manuscript. Y. W. Tsang contributed to

the study design and slide scanning, took part in the GS and DP reporting, assisted with the database, and contributed to writing the manuscript. A. Meskiri scanned the slides, reviewed GS and DP reports, managed the database and trial records, and chaired the steering group. P. Kimani and R. Crossman carried out the statistical analysis, and contributed to writing the manuscript. N. Rajpoot advised on the study design, and contributed to writing the manuscript. K. Chen, P. Matthews, N. Momtahan, S. Read-Jones, S. Sah, E. Simmons, B. Sinha, S. Suortamo, Y. Yeo and H. El Daly reported GS and DP, and took part in the steering group meetings. I. Cree advised on the study design, took part in the GS and DP reporting, took part in the steering group meetings, and contributed to writing the manuscript.

References

1. Furness PN. The use of digital images in pathology. *J. Pathol.* 1997; **183**: 253–263.
2. Gurcan MN, Boucheron LE, Can A, Madabhushi A, Rajpoot NM, Yener B. Histopathological image analysis: a review. *IEEE Rev. Biomed. Eng.* 2009; **2**: 147–171.
3. Weinberg DS, Allaert FA, Dusserre P *et al.* Telepathology diagnosis by means of digital still images: an international validation study. *Hum. Pathol.* 1996; **27**: 111–118.
4. Kronz JD, Westra WH, Epstein JL. Mandatory second opinion surgical pathology at a large referral hospital. *Cancer* 1999; **86**: 2426–2435.
5. Leong FJ, Nicholson AG, McGee JO. Robotic telepathology: efficacy and usability in pulmonary pathology. *J. Pathol.* 2002; **197**: 211–217.
6. Singh N, Akbar N, Sowter C, Lea KG, Wells CA. Telepathology in a routine clinical environment: implementation and accuracy of diagnosis by robotic microscopy in a one-stop breast clinic. *J. Pathol.* 2002; **196**: 351–355.
7. Li X, Liu J, Xu H *et al.* A feasibility study of virtual slides in surgical pathology in China. *Hum. Pathol.* 2007; **38**: 1842–1848.
8. Fine JL, Grzybicki DM, Silowash R *et al.* Evaluation of whole slide image immunohistochemistry interpretation in challenging prostate needle biopsies. *Hum. Pathol.* 2008; **39**: 564–572.
9. Fallon MA, Wilbur DC, Prasad M. Ovarian frozen section diagnosis: use of whole-slide imaging shows excellent correlation between virtual slide and original interpretations in a large series of cases. *Arch. Pathol. Lab. Med.* 2010; **134**: 1020–1023.
10. Jukic DM, Drogowski LM, Martina J, Parwani AV. Clinical examination and validation of primary diagnosis in anatomic pathology using whole slide digital images. *Arch. Pathol. Lab. Med.* 2011; **135**: 372–378.
11. Fonyad L, Krenacs T, Nagy P *et al.* Validation of diagnostic accuracy using digital slides in routine histopathology. *Diagn. Pathol.* 2012; **7**: 35.
12. Shaw EC, Hanby AM, Wheeler K *et al.* Observer agreement comparing the use of virtual slides with glass slides in the pathology review component of the POSH breast cancer cohort study. *J. Clin. Pathol.* 2012; **65**: 403–408.
13. Bauer TW, Schoenfield L, Slaw RJ, Yerian L, Sun Z, Henricks WH. Validation of whole slide imaging for primary diagnosis in

- surgical pathology. *Arch. Pathol. Lab. Med.* 2013; **137**: 518–524.
14. Ordi J, Castillo P, Saco A *et al.* Validation of whole slide imaging in the primary diagnosis of gynaecological pathology in a university hospital. *J. Clin. Pathol.* 2015; **68**: 33–39.
 15. Campbell WS, Lele SM, West WW, Lazenby AJ, Smith LM, Hinrichs SH. Concordance between whole-slide imaging and light microscopy for routine surgical pathology. *Hum. Pathol.* 2012; **43**: 1739–1744.
 16. Krishnamurthy S, Mathews K, McClure S *et al.* Multi-institutional comparison of whole slide digital imaging and optical microscopy for interpretation of hematoxylin–eosin-stained breast tissue sections. *Arch. Pathol. Lab. Med.* 2013; **137**: 1733–1739.
 17. Campbell WS, Hinrichs SH, Lele SM *et al.* Whole slide imaging diagnostic concordance with light microscopy for breast needle biopsies. *Hum. Pathol.* 2014; **45**: 1713–1721.
 18. Pantanowitz L, Sinard JH, Henricks WH *et al.* Validating whole slide imaging for diagnostic purposes in pathology: guideline from the College of American Pathologists Pathology and Laboratory Quality Center. *Arch. Pathol. Lab. Med.* 2013; **137**: 1710–1722.
 19. Cree IA, Guthrie W, Anderson JM *et al.* Departmental audit in histopathology. *Pathol. Res. Pract.* 1993; **189**: 453–457.
 20. Furness PN, Lauder I. A questionnaire-based survey of errors in diagnostic histopathology throughout the United Kingdom. *J. Clin. Pathol.* 1997; **50**: 457–460.
 21. Hocking GR, Niteckis VN, Cairns BJ, Hayman JA. Departmental audit in surgical anatomical pathology. *Pathology* 1997; **29**: 418–421.
 22. Kalinski T, Zwonitzer R, Sel S *et al.* Virtual 3D microscopy using multiplane whole slide images in diagnostic pathology. *Am. J. Clin. Pathol.* 2008; **130**: 259–264.
 23. Kondo Y, Iijima T, Noguchi M. Evaluation of immunohistochemical staining using whole-slide imaging for HER2 scoring of breast cancer in comparison with real glass slides. *Pathol. Int.* 2012; **62**: 592–599.
 24. Molnar B, Berczi L, Diczhazy C *et al.* Digital slide and virtual microscopy based routine and telepathology evaluation of routine gastrointestinal biopsy specimens. *J. Clin. Pathol.* 2003; **56**: 433–438.
 25. Ozluk Y, Blanco PL, Mengel M, Solecz K, Halloran PF, Sis B. Superiority of virtual microscopy versus light microscopy in transplantation pathology. *Clin. Transplant.* 2012; **26**: 336–344.