

# Validation of whole slide imaging in the primary diagnosis of gynaecological pathology in a University Hospital

Jaume Ordi,<sup>1,2,3</sup> Paola Castillo,<sup>1,3</sup> Adela Saco,<sup>1</sup> Marta del Pino,<sup>4</sup> Oriol Ordi,<sup>2</sup> Leonardo Rodríguez-Carunchio,<sup>1</sup> Jose Ramírez<sup>1,2</sup>

► Additional material is published online only. To view please visit the journal online (<http://dx.doi.org/10.1136/jclinpath-2014-202524>).

<sup>1</sup>Department of Pathology, Hospital Clinic, Barcelona, Spain

<sup>2</sup>University of Barcelona, School of Medicine, Barcelona, Spain

<sup>3</sup>Centre de Recerca en Salut Internacional de Barcelona (CRESIB), Barcelona, Spain

<sup>4</sup>Institute of Gynecology, Obstetrics and Neonatology, Hospital Clinic—Institut d'Investigacions Biomèdiques August Pi I Sunyer (IDIBAPS), Faculty of Medicine-University of Barcelona, Barcelona, Spain

## Correspondence to

Professor Jaume Ordi, Department of Pathology, CRESIB (Centre de Recerca en Salut Internacional de Barcelona) -Hospital Clínic, University of Barcelona Faculty of Medicine, Barcelona, Spain, C/Villarreal 170, 08036-Barcelona, Spain; [jordi@clinic.ub.es](mailto:jordi@clinic.ub.es)

Presented in part at the annual meeting of the USA and Canadian Academy of Pathology; March 2014; San Diego, California, USA.

Received 18 July 2014

Revised 23 September 2014

Accepted 8 October 2014

Published Online First

29 October 2014



CrossMark

**To cite:** Ordi J, Castillo P, Saco A, et al. *J Clin Pathol* 2015;**68**:33–39.

## ABSTRACT

**Aims** Experience in the use of whole slide imaging (WSI) for primary diagnosis in pathology is very limited. We aimed to determine the accuracy of interpretation of WSI compared with conventional light microscopy (CLM) in the diagnosis of routine gynaecological biopsies.

**Methods** All gynaecological specimens (n=452) received over a 2-month period at the Department of Pathology of the Hospital Clinic of Barcelona were analysed blindly by two gynaecological pathologists, one using CLM and the other WSI. All slides were digitised in a Ventana iScan HT (Roche diagnostics) at 200 $\times$ . All discrepant diagnoses were reviewed, and a final consensus diagnosis was established. The results were evaluated by weighted  $\kappa$  statistics for two observers.

**Results** The level of interobserver agreement between WSI and CLM evaluations was almost perfect ( $\kappa$  value: 0.914; 95% CI 0.879 to 0.949) and increased during the study period:  $\kappa$  value 0.890; 95% CI 0.835 to 0.945 in the first period and 0.941; 95% CI 0.899 to 0.983 in the second period. Major discrepancies (differences in clinical management or prognosis) were observed in 9 cases (2.0%). All discrepancies consisted of small lesions (8 high grade squamous intraepithelial lesions of the uterine cervix, one lymph node micrometastasis of an ovarian carcinoma) underdiagnosed or missed in the WSI or the CLM evaluation. Discrepancies with no or minor clinical relevance were identified in 3.8% of the biopsies. No discrepancy was related to the poor quality of the WSI image.

**Conclusions** Diagnosis of gynaecological specimens by WSI is accurate and may be introduced into routine diagnosis.

## INTRODUCTION

Whole slide imaging (WSI), also referred to as virtual microscopy or digital pathology allows digitisation of entire glass slides to achieve the diagnosis of pathological specimens. WSI scanners create a digital slide of the tissue section which, with the use of specific software, can be viewed and magnified in real time across the web very much like the use of a conventional light microscope (CLM).<sup>1–3</sup> WSI has been shown to have many practical applications including education,<sup>4–6</sup> teleconsultation for second opinions,<sup>7–10</sup> intraoperative frozen section consultation<sup>10 11</sup> and quality assurance.<sup>12 13</sup> Potential additional benefits of WSI include improvement of the workflow by eliminating the task of slide distribution and facilitating slide filing and retrieval.<sup>1–3</sup>

The rapid advances in this technology and its many potential benefits will probably result in a progressive shift from conventional to virtual microscopy in the routine diagnosis in pathology. Currently, several commercially available systems are able to digitise glass slides containing tissue sections and produce virtual slides of excellent quality. However, although WSI has been around for more than a decade, its widespread application in primary histological diagnosis still awaits validation as opposed to traditional CLM. Moreover, despite several pilot studies suggesting that WSI is as useful as CLM for diagnostic purposes,<sup>14–22</sup> WSI-based diagnosis has yet to be integrated in routine pathological studies, with a very small number of exceptions. The use of WSI in the routine practice of pathology laboratories is still not common because of difficulties in the integration between the WSI software and the laboratory information systems (LIS), insufficient scanning speed and robustness, and limitations in storage capacity and/or excessive costs of file storage. The lack of systematic validation studies on their use in primary diagnosis is also a major concern that hampers the introduction of this technique.<sup>1 3</sup>

The College of American Pathologists Pathology and Laboratory Quality Center has recently published the guidelines for validating the use of WSI for diagnostic purposes.<sup>3</sup> According to these guidelines all laboratories should conduct their own validation studies, and the validation should include at least 60 samples. However, very few large validation studies have focused on pathology subspecialties. Indeed, only one study has been published on gynaecological disease, evaluating the correlation between WSI and CLM in the assessment of the diagnoses of frozen sections of 52 ovarian lesions.<sup>11</sup> No previous studies have evaluated the accuracy of WSI diagnosis in the routine practice of gynaecological pathology. In the present study we aimed to determine the accuracy of interpretation using WSI as compared with CLM in a series of consecutive gynaecological specimens, representative of the spectrum of specimen types and diagnoses encountered in the routine practice of a large academic department.

## MATERIALS AND METHODS

### Characteristics of the institution

This study was performed at the Department of Pathology, Hospital Clinic (Barcelona, Spain), a large academic department composed of 15 staff pathologists, 8 residents and additional fellows.

There are 14 subspecialties, and most pathologists limit their practice to one or two subspecialty areas. In 2013 the Department handled 41 928 specimens with 83 619 blocks. The number of gynaecological specimens analysed during 2013 was 4909, with 16 805 blocks of gynaecological specimens and with a median number of slides per case of 1.

### Sample size calculation

Based on previous reports,<sup>15 23</sup> the major rate of discrepancy between the original diagnosis by CLM and that by WSI was calculated to be 3%, with a non-inferiority margin for WSI review of 4%. A one-sided binomial test was used for comparison at a level of significance of 0.05. The power to be achieved was 80%, and the level of significance was 0.05. Based on these assumptions, it was calculated that 450 cases would need to be reviewed to establish non-inferiority.

### Specimens included in the study

All gynaecological specimens consecutively received over a 2-month period (July–August 2013) were included in the study at the Department of Pathology of the Hospital Clinic of Barcelona (n=452). This represented 9.21% (452/4909) of the total number of gynaecological specimens evaluated in 2013: 353/452 (78.1%) specimens were evaluated in the 1st month (July) and 99/452 (21.9%) in the 2nd month (August). Table 1 shows a summary of the location and type (either biopsy or resection) of the specimens, and the number of cases included for each type. The number of blocks per case ranged from 1 to 30 (median 1, IQR 1–3). The overall number of slides scanned was 1253.

To evaluate possible changes associated with increasing experience in the use of WSI over the study period and the agreement between WSI and CLM observers, the 456 specimens were divided into two periods: one including the first 226 specimens and the second including the last 226 specimens.

The Department of Gynecology of our hospital has a very active referral Colposcopy Clinic.<sup>24</sup> Consequently, the series included 157 biopsies or excisions of the uterine cervix from patients referred to colposcopy because of an abnormal Pap smear. In addition to evaluation in the general analysis, these specimens were evaluated independently.

The study was approved by the Hospital Clinic Institutional Ethical Review Board.

**Table 1** Location and type (either small biopsies or surgical resections) of the specimens evaluated in the study

Location	Number	(%)	Biopsies	Surgical specimens
Vulva	19	(4)	13	6
Vagina	12	(3)	12	1
Uterine cervix	125	(28)	93	32
Endocervix	46	(10)	46	0
Endometrium	60	(13)	52	8
Uterus	69	(15)	0	69
Fallopian tube	12	(3)	0	12
Ovary	36	(8)	0	36
Lymph nodes	44	(10)	0	44
Peritoneum	14	(3)	0	14
Other *	15	(3)	0	15

\*Includes biopsies of the abdominal wall (1), large bowel (9), urinary bladder (4) and ureter (1).

### Scanning process and characteristics of the WSI display

In July and August 2013 all the slides of gynaecological pathology were scanned daily after diagnosis by light microscopy. Scanning was performed on a Ventana iScan HT (Roche Diagnostics, Sant Cugat, Spain) at a magnification of 200×. The system creates high-resolution digital images of the tissue sections. The whole scanning process runs automatically (including selection of the area of the slide that contains tissue, placing focus points, calibration, etc). In cases with step sections of a sample on a single slide the system scans all the sections. No specific quality control of the slides scanned was made by the technicians prior to evaluation by the pathologist. The WSI produced are stored in a dedicated mass storage environment and linked to the pathology report, based on the recognition of a quick response (QR) code on the slide label. Although WSI can be accessed through the pathology LIS software (Novopath, Vitrosoft, Sevilla, Spain), for the purposes of this study the accession to the WSI was made through the viewer.

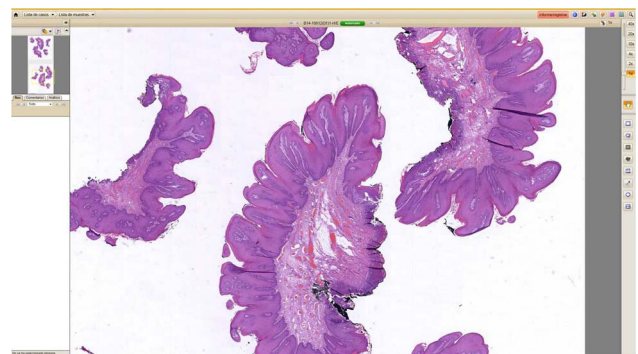
The images are viewed in the Virtuoso viewer (Roche), which works as a web browser and simulates a conventional microscope. The images are shown using the same structure provided by the LIS. No specific software installation is required to visualise the WSI. The images scanned can be viewed up to a real magnification of 200× and up to 400× with digital zoom, are always in focus, with optimised contrast and adjusted illumination. The viewer shows a thumbnail of the whole slide, which allows confirmation that all the material present on the glass slide has been included in the digital image and helps in the navigation through the slide. Figure 1 shows the appearance of the virtual microscope display.

The WSI were displayed on a 30" Coronis fusion MDC4130 monitor which has a resolution of four megapixels (Barco Electronic Systems, Barcelona, Spain).

### WSI and CLM diagnosis

All cases were analysed blindly by two gynaecological pathologists, one using CLM and the other WSI. The pathologist doing the WSI evaluations had previously had a 1-week training course on the use of WSI. WSI were presented to the pathologist per case, together with the original clinical information in order to emulate the real clinical environment, and blinded to the original report based on CLM. For the purposes of this study only the H&E slides were evaluated.

The original CLM and the WSI-based diagnoses were compared by an independent gynaecologist, who judged the concordance of the two diagnoses as: (A) complete agreement between the original diagnosis and that determined with WSI; (B) minor discrepancy (mild differences which would not have



**Figure 1** Screenshot of the virtual microscope display.

any clinical or prognostic implications); and (C) major discrepancy (differences with clinical and/or prognostic implications for the patient).

### Final gold standard diagnosis

The gold standard was considered as the concordant diagnosis in all cases with complete agreement in both evaluations. Each case with a discrepant result was reviewed by the two pathologists involved in the study. The revision was made using CLM, and a final consensus diagnosis was established. In this final adjudication process, the H&E slides, as well as the immunohistochemical stains (when necessary) were used to achieve the final diagnosis.

### Statistical analysis

The SPSS statistical programme (SPSS TM140, V.18, Chicago, Illinois, USA) was used for statistical analysis. The results for categorical variables are expressed as absolute numbers and percentages and 95% CIs. The  $\chi^2$  or the Fisher's exact tests were used to compare the variables. The results were evaluated by weighted  $\kappa$  statistics for two raters. This measure calculates the degree of agreement in classification over that which would be expected by chance and is scored as a number between 0 and 1. Following the Landis-Koch benchmarks the strength of agreement of the  $\kappa$  values is: 0 poor; 0–0.20 slight; 0.21–0.40 fair; 0.41–0.60 moderate; 0.61–0.80 substantial; 0.81–1.00 almost perfect.<sup>25</sup> For the purposes of weighted  $\kappa$  calculation the diagnoses were categorised from normal to cancer as 1: normal tissue or reactive lesions; 2: benign tumours; 3: low-grade premalignant lesions; 4: high-grade premalignant lesions; and 5: malignant tumours.

## RESULTS

### Final diagnoses

Overall, 218/452 specimens (48.2%) were evaluated as being composed of normal tissue or showing reactive lesions; 130 (28.8%) were benign tumours, 18 (4.0%) were low-grade premalignant lesions, 48 (10.6%) high grade premalignant lesions and 38 (8.4%) showed malignant tumours. Table 2 shows the distribution of these diagnoses for each specific site.

### Agreement between WSI and CLM evaluations

Interpretations by WSI and CLM completely agreed in 94.2% of the biopsies (95% CI 91.7 to 96.0) while major discrepancies were observed in 9/452 (2.0%) and minor discrepancies were identified in 17/452 (3.8%) of the biopsies.

The final consensus diagnosis achieved after the adjudication meeting was in agreement with the CLM evaluation in 7/9 (77.8%) major discrepancies and in 11/17 (64.7%) minor discrepancies. Discrepancy in interpretations between WSI and CLM evaluations occurred in only two settings. Eight out of the nine major discrepancies (88.9%) observed in the study were related to the diagnosis of a high-grade squamous intraepithelial lesion (H-SIL) of the uterine cervix, as a low-grade squamous intraepithelial lesion (L-SIL) or as negative (four cases each, figure 2). The consensus diagnosis was in keeping with the CLM evaluation in six of eight cases and with the WSI evaluation in two of eight cases. The last case was a small lymph node metastasis of an ovarian carcinoma missed in the WSI evaluation (figure 3). In this latter case a small lipogranuloma was identified close to the small metastatic nest missed in the WSI evaluation. Thirteen out of the 17 minor discrepancies were related to overdiagnosis or underdiagnosis of L-SIL. In 10 cases the L-SIL lesions were missed in the evaluation (8 in the

WSI and 2 in the CLM evaluation). Three cases were reactive changes in the uterine cervix overdiagnosed as L-SIL (three biopsies, two overdiagnosed in the WSI and one in the CLM evaluation). Two cases of endometrial polyps were missed (one case missed in the WSI evaluation) or overdiagnosed (one case, overdiagnosed in the CLM evaluation). The other two minor discrepancies were two small foci of endometriosis (one in the ovary, one in the Fallopian tube) missed in the CLM evaluation. None of the discrepancies was related to the poor quality of the WSI image or to insufficient magnification.

Overall the level of interobserver agreement between the WSI and CLM evaluations was almost perfect ( $\kappa$  value: 0.914; 95% CI 0.879 to 0.949).

### Concordance in biopsies of the uterine cervix and in other samples

In the subset of 157 biopsies or excisions of the uterine cervix from patients referred to colposcopy because of abnormal Pap smear, complete agreement was observed between the WSI and CLM interpretations in 86.6% (95% CI 80.3 to 91.5) of the biopsies. Major discrepancies were observed in 8/157 (5.1%) and minor discrepancies in 13/157 (8.3%) of the samples. The  $\kappa$  value for this subset of samples was 0.832 (95% CI 0.757 to 0.906).

In the 295 gynaecological specimens representing tissues other than cervical biopsies and excisions, complete agreement between WSI and CLM was observed in 98.3% (95% CI 96.1 to 99.4) of the biopsies. Major discrepancies were observed in 1/295 (0.3%) and minor discrepancies in 4/295 (1.4%) of the samples. The  $\kappa$  value for this subset of samples was 0.976 (95% CI 0.950 to 1).

### $\kappa$ Analysis and discrepant diagnoses in the two study periods

Interobserver agreement increased during the study period,  $\kappa$  value: 0.890 (95% CI 0.835 to 0.945) in the first period, and  $\kappa$  value: 0.941 (95% CI 0.899 to 0.983) in the second period. In the first period of the study 5/226 (2.2%) major discrepancies and 12/226 (5.3%) minor discrepancies were detected. The number of major and minor discrepancies in the second period was 4/226 (1.80%) and 5/226 (2.2%), respectively. Interestingly, whereas the consensus gold standard diagnosis was in keeping with the CLM diagnosis in 14/17 (82.4%) discrepancies observed in the initial period, the consensus was in keeping with the WSI evaluation in 5/9 discrepancies (55.5%) observed in the second period ( $p=0.078$ ).

## DISCUSSION

The results of our study show a high concordance between WSI and CLM evaluations (over 94%) in the diagnosis of a large series of routine gynaecological specimens. The  $\kappa$  value, considered as a measure of the level of agreement among observers corrected by chance, was at the almost perfect level (0.914). Thus, our results confirm that WSI may safely be used for performing primary histological diagnoses of gynaecological specimens.

The results of our study are comparable with other validation studies conducted on skin,<sup>5 19</sup> breast,<sup>14 26</sup> prostate,<sup>17 27</sup> urinary bladder,<sup>18</sup> gastrointestinal<sup>8 20</sup> and paediatric pathology,<sup>28</sup> which show a high rate of concordance between WSI and CLM-based diagnoses. However, no previous studies have evaluated the accuracy of WSI diagnosis in the routine practice of gynaecological pathology, and neither are there any data about intraobserver or interobserver agreement in the evaluation of routine gynaecological specimens using CLM. The rate



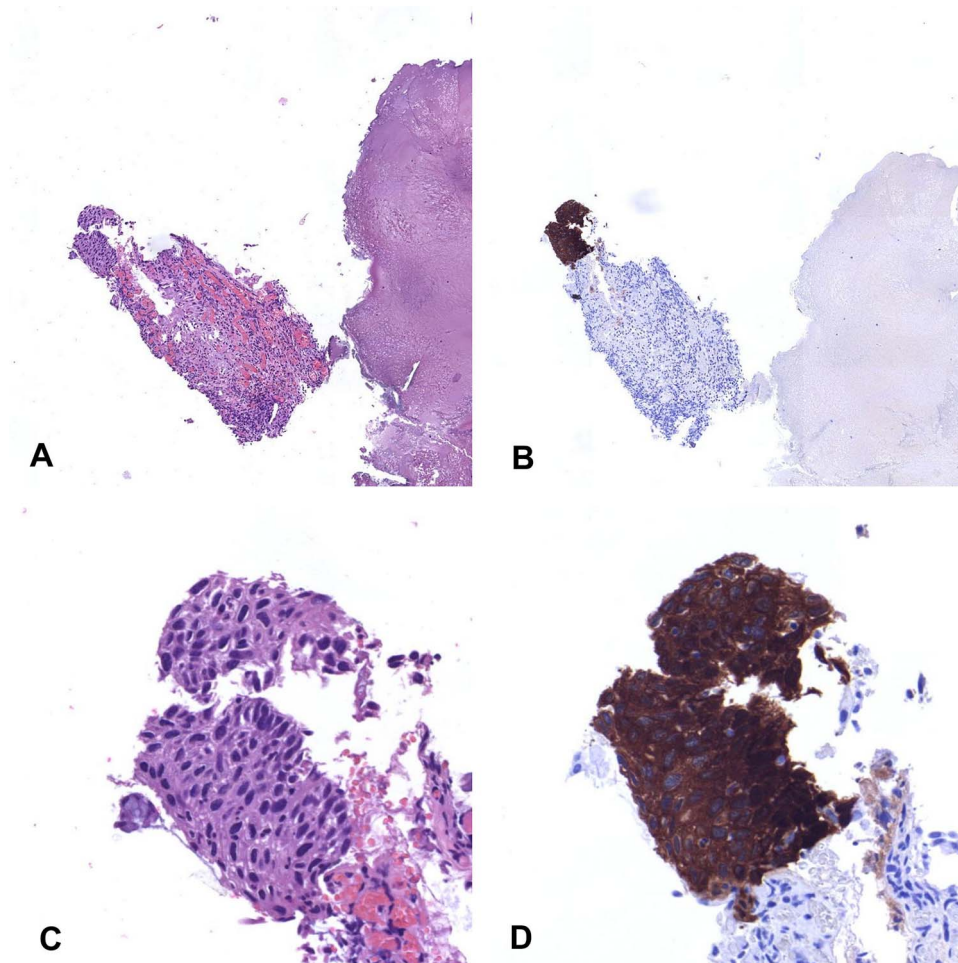
**Table 2** Absolute numbers and percentages (in parenthesis) of specimens showing normal/reactive lesions, benign tumours, low-grade premalignant, high-grade premalignant and malignant tumours in each location

Location	Normal/reactive	Benign tumours	Low-grade premalignant lesions	High-grade premalignant lesions	Malignant tumours
Vulva	7 (36.8)	6 (31.6)	0 (0)	5 (26.3)	1 (5.3)
Vagina	8 (66.7)	0 (0)	0 (0)	0 (0)	4 (33.3)
Uterine cervix	64 (51.2)	3 (2.4)	16 (12.8)	37 (29.6)	5 (4.0)
Endocervix	30 (65.2)	9 (19.6)	2 (4.3)	3 (6.6)	2 (4.3)
Endometrium	29 (48.3)	28 (46.7)	0 (0)	1 (1.7)	2 (3.3)
Uterus	13 (18.8)	46 (66.7)	0 (0)	2 (2.9)	8 (11.6)
Fallopian tube	7 (58.3)	5 (41.7)	0 (0)	0 (0)	0 (0)
Ovary	8 (22.2)	24 (66.7)	0 (0)	0 (0)	4 (11.1)
Lymph nodes	41 (93.2)	0 (0)	0 (0)	0 (0)	3 (6.8)
Peritoneum	6 (42.9)	0 (0)	0 (0)	0 (0)	8 (57.1)
Other *	5 (33.3)	9 (60.0)	0 (0)	0 (0)	1 (6.7)

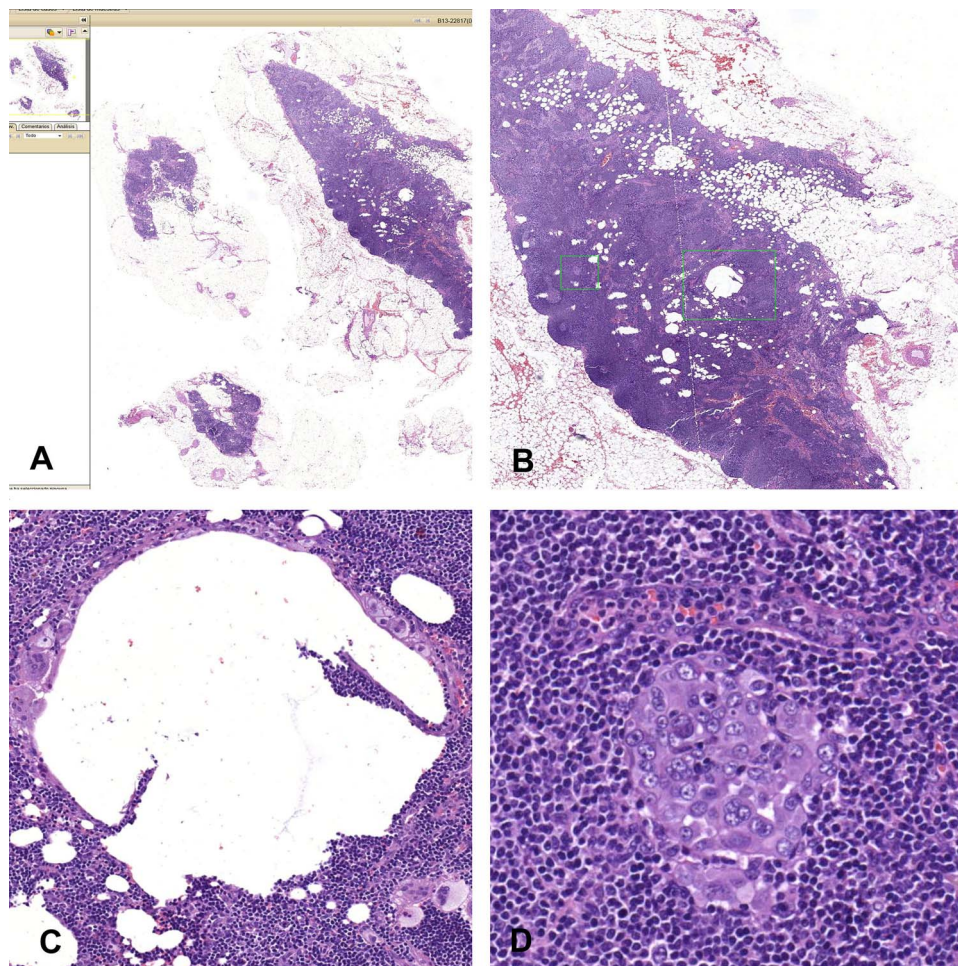
\*Includes biopsies of the abdominal wall (1), large bowel (9), urinary bladder (4) and ureter (1).

of discrepancies observed in our study is within the range of generally observed intraobserver variability in pathology.<sup>27 29</sup> Interestingly, the final consensus diagnosis was in agreement with the WSI evaluation in 22.2% of the major discrepancies and in 35.3% of the minor discrepancies. None of the

discrepancies was related to the poor quality of the WSI image or to insufficient magnification, but rather were mostly associated with different interpretations of difficult or borderline cases or with the presence of small lesions overlooked in the evaluation.



**Figure 2** One of the major discrepancies identified in the study: a small area of high-grade squamous intraepithelial lesion (H-SIL) involving the squamous epithelium of the uterine cervix diagnosed as reactive changes in the whole slide imaging (WSI) evaluation. In the consensus meeting a p16 staining was requested, which confirmed the diagnosis of H-SIL rendered by conventional light microscopy (A and C, H&E; B and D p16 immunostaining; screenshots of the WSI image).



**Figure 3** Screenshots of the only major discrepancy observed in the study not related to cervical pathology; a small lymph node metastasis of an ovarian carcinoma, which was missed in the whole slide imaging (WSI) evaluation. (A) Low power magnification showing several lymph nodes included in the slide; (B) higher magnification showing the two areas of interest located in the centre of the largest node (green squares); (C) higher magnification of the area identified with the large green square in B, showing a lipogranuloma with a large fatty vacuole and giant macrophages in the periphery; (D) higher magnification of the area identified in the small green square in B, showing a small nest of metastatic carcinoma cells, which were missed in the WSI evaluation.

Eight out of the nine discrepancies (88.9%) observed in the study were related to the diagnosis of H-SIL as L-SIL or as negative or reactive changes in the uterine cervix (four cases each). Similarly, 13 out of the 17 minor discrepancies (76.5%) were related to discrepancies in the diagnosis of L-SIL lesions versus normal/reactive cervical epithelium. Consequently, the  $\kappa$  value for the subset of cervical biopsies or excisions from patients referred because of abnormal Pap smear was 0.832, clearly lower than the general value. A number of studies using CLM have shown that there is a substantial variation between and within-observers in the interpretation of squamous intraepithelial lesions on H&E-stained tissue sections. Indeed,  $\kappa$  coefficients are typically found within the range of 0.45–0.50,<sup>29–36</sup> indicating moderate agreement. Estimates for the reproducibility of histological cervical specimen interpretations performed in the course of the ASCUS/L-SIL Triage Study comparing the diagnostic results of the original clinical centre pathologists with the results from a quality control review<sup>35</sup> showed that the reproducibility of histological interpretations of biopsy specimens was moderate ( $\kappa < 0.5$ ). In the latter study, the lack of reproducibility was substantially higher for punch biopsy specimens than loop excision procedure specimens, and variability was more prominent in the low-grade abnormalities, similar to what was observed in our

study. The p16 immunohistochemical stain, strongly expressed in almost all H-SIL and not in reactive lesions,<sup>36</sup> has recently been recommended by the College of American Pathologists to reduce interobserver variability, particularly in cases of professional disagreement.<sup>37</sup> This technique was used in our series to achieve the final consensus diagnosis in all disagreements between CLM and WSI detected in biopsies of the uterine cervix.

The pathologist who performed the WSI evaluation had previously had very little experience in the use of WSI, although this did not severely affect the reproducibility, even in the initial period of the study. Nevertheless, a clear increase in the rate of reproducibility was observed during the study period. This suggests that minor difficulties may arise in the initial periods of using this new tool and that increasing experience with WSI increases the accuracy of the diagnosis. Moreover, whereas the consensus diagnosis reached in the discrepant cases was in keeping with the CLM diagnosis in most discrepancies observed in the initial period (82.4%), in the second period there was a tendency towards a more balanced situation and in 55.5% of the cases the gold standard diagnosis was in keeping with the WSI evaluation.

The pathologist working with WSI did not report difficulties in rendering the diagnosis at the magnification of 200 $\times$  applied,



indicating that a higher magnification does not seem to be relevant for most cases. This scanning strategy which is used in most validation studies,<sup>15 17 19–21 23 28</sup> significantly saves scanning time and storage requirements. However, when WSI is routinely used for primary diagnosis it is very likely that the pathologist would require an increase in the scanning magnification in a small percentage of cases to safely establish the diagnosis. Although no formal timing was performed, the pathologist doing the WSI evaluation perceived the diagnostic process to be a little slower.

The main strength of our study is that it is the largest validation study focused on gynaecological pathology that includes a sufficient number of cases to allow robust statistical power. Indeed, only one published study has analysed the correlation between WSI and CLM in the evaluation of frozen section diagnosis of 52 ovarian lesions, showing that, as observed in our study, the correlation between CLM and WSI diagnoses is very good.<sup>11</sup> The second strength of our study is that the participating pathologists were aware of the clinical information which could affect the diagnostic outcome.

The main limitation of this study is that intraobserver variability, which is considered the best design to confirm the reproducibility of the results obtained with two different techniques, was not evaluated.<sup>38</sup> However, the very good interobserver reproducibility observed in our series suggests that very similar data would be obtained when cases are evaluated by the same observer.

A potential advantage of WSI is that it allows performing image analysis. This may assist in the objective evaluation of the size of the tumours and the depth of the invasion, which are relevant information for tumours of the vulva, cervix and endometrium. Eventually, this may even permit computer-assisted diagnosis that could help improve the diagnosis and decrease interobserver variability. Several legal issues have arisen from the use of WSI for primary diagnosis related to image quality, image presentation (monitor quality), storage space, adequate backup, document transfer, patient confidentiality and the confidence of the pathologist to sign out a pathology report depending on WSI. Most of these issues will probably be settled in the near future. Several digital pathology vendors are currently seeking approval from the US Food and Drug Administration to use WSI in primary diagnosis, which will definitely encourage the general use of this technique.

### Take home messages

- ▶ Interobserver agreement between whole slide imaging (WSI) and conventional light microscopy (CLM) evaluations is very good, with  $\kappa$  values of over 0.91.
- ▶ Although the accuracy of WSI diagnosis is good even in users with limited experience, the reproducibility between WSI and CLM improves over time, indicating that increasing experience with WSI increases the accuracy of the diagnosis.
- ▶ WSI may safely be used for performing primary histological diagnosis of gynaecological specimens using current scanning technology and viewing interfaces.
- ▶ Scanning slides at the magnification of 20 $\times$  is sufficient to achieve a correct diagnosis in most gynaecological biopsies. This scanning strategy significantly saves scanning time and storage requirements.

In conclusion, the diagnosis of gynaecological specimens using WSI shows a high concordance with the results of CLM evaluation. Our results confirm that WSI may safely be used for performing primary histological diagnosis of gynaecological specimens.

**Acknowledgements** The authors are grateful to Rosana Millan, Margarita Mainar, Berta Coloma, Ingrid Rubio, Olga Ten, Gemma Laguna, Silvia Moya, and Arantxa Sánchez, members of the technical and administrative staff of the department of Pathology, for their support in the scanning of the slides. The authors thank Antonio Teruel, Anna Rubi, Jaume Barberi, Jose Antonio Collados and Enric Vidal, members of the staff of Roche for their technical support. The authors thank Donna Pringle for English revision of the manuscript.

**Contributors** JO: conception, design and conduct of the study, analysis and writing of the manuscript, approval of the final version of the manuscript. PC, AS, LR-C: conduct of the study, writing of the manuscript, approval of the final version of the manuscript. OO: conception, design and writing of the manuscript, approval of the final version of the manuscript. MdP: conception of the study, the statistical analysis and writing of the manuscript, approval of the final version of the manuscript. JR: design of the study, writing of the manuscript, approval of the final version of the manuscript.

**Competing interests** None.

**Ethics approval** Ethical committee of clinical research of Hospital Clinic of Barcelona.

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

### REFERENCES

- 1 Al-Janabi S, Huisman A, Van Diest PJ. Digital pathology: current status and future perspectives. *Histopathology* 2012;61:1–9.
- 2 Brachtel E, Yagi Y. Digital imaging in pathology—current applications and challenges. *J Biophotonics* 2012;5:327–35.
- 3 Pantanowitz L, Valenstein PN, Evans AJ, et al. Review of the current state of whole slide imaging in pathology. *J Pathol Inform* 2011;2:36.
- 4 Pantanowitz L, Szymas J, Yagi Y, et al. Whole slide imaging for educational purposes. *J Pathol Inform* 2012;3:46.
- 5 Brick KE, Comfere NI, Broeren MD, et al. The application of virtual microscopy in a dermatopathology educational setting: assessment of attitudes among dermatopathologists. *Int J Dermatol* 2014;53:224–7.
- 6 Carlson AM, McPhail ED, Rodriguez V, et al. A prospective, randomized crossover study comparing direct inspection by light microscopy versus projected images for teaching of hematopathology to medical students. *Anat Sci Educ* 2014;7:130–4.
- 7 Leong FJ, McGee JO. Automated complete slide digitization: a medium for simultaneous viewing by multiple pathologists. *J Pathol* 2001;195:508–14.
- 8 Singson RP, Natarajan S, Greenson JK, et al. Virtual microscopy and the Internet as telepathology consultation tools. A study of gastrointestinal biopsy specimens. *Am J Clin Pathol* 1999;111:792–5.
- 9 Wilbur DC, Madi K, Colvin RB, et al. Whole-slide imaging digital pathology as a platform for teleconsultation: a pilot study using paired subspecialist correlations. *Arch Pathol Lab Med* 2009;133:1949–53.
- 10 Winokur TS, McClellan S, Siegal GP, et al. A prospective trial of telepathology for intraoperative consultation (frozen sections). *Hum Pathol* 2000;31:781–5.
- 11 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;134:1020–3.
- 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. *Histopathology* 2011;59:703–9.
- 13 Ho J, Parwani AV, Jukic DM, et al. Use of whole slide imaging in surgical pathology quality assurance: design and pilot validation studies. *Hum Pathol* 2006;37:322–31.
- 14 Reyes C, Ipkatt OF, Nadjji M, et al. Intra-observer reproducibility of whole slide imaging for the primary diagnosis of breast needle biopsies. *J Pathol Inform* 2014;5:5.
- 15 Bauer TW, Schoenfield L, Slaw RJ, et al. Validation of whole slide imaging for primary diagnosis in surgical pathology. *Arch Pathol Lab Med* 2013;137:518–24.
- 16 Onega T, Weaver D, Geller B, et al. Digitized whole slides for breast pathology interpretation: current practices and perceptions. *J Digit Imaging* 2014;27:642–8.
- 17 Camparo P, Egevad L, Algaba F, et al. Utility of whole slide imaging and virtual microscopy in prostate pathology. *APMIS* 2012;120:298–304.
- 18 Comperat E, Egevad L, Lopez-Beltran A, et al. An interobserver reproducibility study on invasiveness of bladder cancer using virtual microscopy and heatmaps. *Histopathology* 2013;63:756–66.
- 19 Al-Janabi S, Huisman A, Vink A, et al. Whole slide images for primary diagnostics in dermatopathology: a feasibility study. *J Clin Pathol* 2012;65:152–8.

- 20 Al-Janabi S, Huisman A, Vink A, *et al.* Whole slide images for primary diagnostics of gastrointestinal tract pathology: a feasibility study. *Hum Pathol* 2012;43:702–7.
- 21 Gilbertson JR, Ho J, Anthony L, *et al.* Primary histologic diagnosis using automated whole slide imaging: a validation study. *BMC Clin Pathol* 2006;6:4.
- 22 Cornish TC, Swapp RE, Kaplan KJ. Whole-slide imaging: routine pathologic diagnosis. *Adv Anat Pathol* 2012;19:152–9.
- 23 Bauer TW, Slaw RJ. Validating Whole-Slide Imaging for Consultation Diagnoses in Surgical Pathology. *Arch Pathol Lab Med* Published Online First: 9 May 2014.
- 24 Ordi J, Sagasta A, Munmany M, *et al.* Usefulness of p16/Ki67 immunostaining in the triage of women referred to colposcopy. *Cancer Cytopathol* 2014;122:227–35.
- 25 Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159–74.
- 26 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;137:1733–9.
- 27 Rodriguez-Urrego PA, Cronin AM, Al-Ahmadie HA, *et al.* Interobserver and intraobserver reproducibility in digital and routine microscopic assessment of prostate needle biopsies. *Hum Pathol* 2011;42:68–74.
- 28 Al-Janabi S, Huisman A, Nikkels PG, *et al.* Whole slide images for primary diagnostics of paediatric pathology specimens: a feasibility study. *J Clin Pathol* 2013;66:218–23.
- 29 Nelson D, Ziv A, Bandali KS. Going glass to digital: virtual microscopy as a simulation-based revolution in pathology and laboratory science. *J Clin Pathol* 2012;65:877–81.
- 30 Creagh T, Bridger JE, Kupek E, *et al.* Pathologist variation in reporting cervical borderline epithelial abnormalities and cervical intraepithelial neoplasia. *J Clin Pathol* 1995;48:59–60.
- 31 de Vet HC, Koudstaal J, Kwee WS, *et al.* Efforts to improve interobserver agreement in histopathological grading. *J Clin Epidemiol* 1995;48:869–73.
- 32 McCluggage WG, Walsh MY, Thornton CM, *et al.* Inter- and intra-observer variation in the histopathological reporting of cervical squamous intraepithelial lesions using a modified Bethesda grading system. *Br J Obstet Gynaecol* 1998;105:206–10.
- 33 McCluggage WG, Bharucha H, Caughley LM, *et al.* Interobserver variation in the reporting of cervical colposcopic biopsy specimens: comparison of grading systems. *J Clin Pathol* 1996;49:833–5.
- 34 Robertson AJ, Anderson JM, Beck JS, *et al.* Observer variability in histopathological reporting of cervical biopsy specimens. *J Clin Pathol* 1989;42:231–8.
- 35 Stoler MH, Schiffman M. Interobserver reproducibility of cervical cytologic and histologic interpretations: realistic estimates from the ASCUS-L SIL Triage Study. *JAMA* 2001;285:1500–5.
- 36 Bergeron C, Ordi J, Schmidt D, *et al.* Conjunctive p16INK4a testing significantly increases accuracy in diagnosing high-grade cervical intraepithelial neoplasia. *Am J Clin Pathol* 2010;133:395–406.
- 37 Darragh TM, Colgan TJ, Thomas CJ, *et al.* The Lower Anogenital Squamous Terminology Standardization project for HPV-associated lesions: background and consensus recommendations from the College of American Pathologists and the American Society for Colposcopy and Cervical Pathology. *Int J Gynecol Pathol* 2013;32:76–115.
- 38 Pantanowitz L, Sinarid 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;137:1710–22.