“Validation of Whole Slide Imaging for Histopathological Diagnosis: 
Current State”

Adela Saco MD ¹, Josep Ramírez MD, PhD ¹, Natalia Rakislova MD¹, Aurea Mira MD, PhD ¹, Jaume Ordi MD, PhD ¹.²

¹Department of Pathology, Hospital Clínic, University of Barcelona, School of Medicine, Barcelona, Spain
²ISGlobal, Barcelona Ctr. Int. Health Res. (CRESIB), Barcelona, Spain.

Running title: Validation of WSI in Primary Diagnosis

Corresponding author: Jaume Ordi. Department of Pathology, Hospital Clínic, University of Barcelona, Spain, C/Villarroel 170, 08036-Barcelona, Spain. Tel: +34 93 2275450 Fax: +34 93 2275717. E-mail address: jordi@clinic.ub.es

Key words: Virtual Microscopy, Whole-slide images, validation, primary diagnosis, routine diagnosis

Abstract word count: 209

Total word count: 3660

Number of figures: 4

Number of references: 90
Abstract

Rapid advances in informatics and technological improvements have led to the development of high throughput whole slide imaging (WSI) scanners able to produce high quality digital images, which allow achieving a correct diagnosis of the biopsies using virtual viewers. This technology is currently prepared to be introduced in the departments of pathology for routine diagnosis. The aim of this review is to analyze the current evidence regarding the use of WSI in primary or routine diagnosis in the different subspecialties of pathology. An increasing number of studies have shown almost perfect inter- and intra-observer agreement between the diagnoses obtained with WSI and the classical diagnosis based on conventional light microscopy. The only exception seems to be cytology, which still requires some technological development. Although validation studies are needed in some areas of pathology, growing evidence indicate that WSI is a reliable tool for routine diagnosis. Pathologists have a positive perception of the ergonomics of the workstations, the low magnification of WSI, and the possibility of making annotations, and measurements. WSI can be used from any device and anywhere, thereby providing great opportunities for teleconsultation. New technologies such as the recognition of histopathology patterns using image analysis may facilitate diagnosis and improve the reproducibility among pathologists in the future.
Introduction and historical perspective

For more than a century conventional light microscopy (CLM) has been the basic tool for tissue evaluation and has played a pivotal role in pathological diagnosis. Until the incorporation of non-morphological molecular technologies into routine practice in recent years, the standard of diagnosis for pathologists was morphology, and especially morphological criteria evaluated under CLM. Indeed, the evaluation of most specimens submitted to Pathology laboratories today still relies on the interpretation of images by CLM, complemented by gross examination and a number of ancillary molecular techniques, most of which (histochemistry and immunohistochemistry) are also evaluated with CLM. Asking experts or other colleagues for diagnostic opinions required sending glass slides or paraffin blocks for examination by CLM. Teaching pathology to undergraduates and residents and continuing medical education for certified pathologists also depended on the use of CLM.

This scenario slowly started to change a few decades ago [1-3]. Static digital images allowed teaching and, to a certain degree, teleconsultation, but limitations in image quality and, particularly, the inability to navigate and use different optical objectives made the substitution of CLM unfeasible[4]. Dynamic real-time telepathology systems with videocameras integrated to the CLM were used for intraoperative frozen biopsies because they allowed an image to be sent to an expert located remotely. This capacity was extraordinarily useful for small hospitals as it provided a quick diagnostic approach for difficult cases[5-12]. However, the relatively poor image quality and the impossibility to remotely conduct the navigation through the slide made the system inadequate for routine diagnosis.
Rapid advances in informatics as well as technological improvements led to the development of scanners able to create digital reproductions from whole glass slides, which appeared one decade ago [1;2]. These scanners are the basis of virtual microscopy or whole-slide imaging (WSI), which allows navigation across the virtual slide and visualization at different magnifications, allowing the computer to be used as a CLM. However, the image quality of the initial scanners was limited, and the costs of implementation of the technology, including the scanner, monitors and suitable computers were very high, thereby restricting the use of WSI to certain areas such as teaching and teleconsultation and excluding routine diagnosis [10;13-17].

Currently, a number of high throughput scanners able to produce high quality images are available on the market. These scanners allow correct diagnosis of the biopsies using virtual viewers. The cost of implementation of WSI has significantly decreased, and the speed of visualization has notably increased[17-22]. Constant Improvements in this technology have led to an important expansion in the use of WSI in routine diagnosis in recent years. The aim of this review is to evaluate the current evidence on the validation of WSI in routine diagnosis.

**Advantages and challenges of WSI for routine diagnosis**

Routine histopathological diagnosis can benefit from the multiple advantages of WSI. WSI workstations are more ergonomic (Figure 1). WSI has a much larger field of vision than CLM and allows a wider range of magnifications, thus providing easier navigation. In particular, WSI allows very low magnifications (<100 x) to be used, which is very useful in the evaluation of surgical specimens. The computer tools allow making annotations and
measurements. WSI viewers can simultaneously show and synchronously move several slides of a case, which is particularly helpful in the evaluation of immunohistochemically (IHC) stained slides (Figure 2). Indeed, studies evaluating the opinion of the pathologists have revealed a positive perception of image quality, and have pointed out the utility of the measurement and annotation tools, as well as the ergonomics and usability of the viewer[22]. WSI can be used from any device and anywhere, thereby providing great opportunities for teleconsultation and remote work. Indeed, portability is one of the major advantages of WSI and this will probably be further improved in the near future when the current viewers are fully adapted to portable devices such as tablets and smartphones [23-25]. Moreover, the need for standardization in the diagnosis and in the evaluation of immunohistochemical biomarkers predicting the outcome of specific therapies will probably boost the implementation of WSI.

Finally, WSI allow automatic quantification of IHC slides. These diagnostic algorithms facilitate quantification of IHC positivity resulting in a more objective evaluation, which is extremely useful in the evaluation of some biological markers. Algorithms of evaluation of IHC stains are variably used depending on the subspecialties and are particularly useful in cases of breast cancer [26-29].

In contrast with these positive opinions many pathologists still prefer using CLM. The most criticized feature of WSI is the speed in uploading the image. Indeed, most pathologists feel that more time is required to make a diagnosis with WSI. However, some studies have shown that although diagnosis with WSI is initially more time consuming, this time quickly decreases as pathologists become familiar with the use of the WSI viewer [30-35]. Thus, there is a
learning curve in the use of WSI and the time required for making a diagnosis, and a recent study conducted at our institution confirmed that the diagnostic performance improved with practice [36]. Another limitation of WSI is the relatively high cost of the equipment. The basic needs for a WSI system which is adequate for routine diagnosis include not only high throughput scanners but also high resolution monitors [37;38]. This is a common concern since, despite the reduction in the price of the equipment in the last few years, it still represents a considerably high investment, which has a relatively low added value for many pathologists as the basic functions of WSI are already being confidently achieved with the old CLM. Finally, WSI requires a significant investment in high capacity servers; the files generated by WSI scanners are huge, with sizes frequently over 2GB per slide. Thus, strategies to reduce the size of the files, such as scanning at relatively low magnification (200x instead of 400x or 600x) are frequently used[37].

The need for validation studies

The number of studies aimed at validating WSI in primary or routine diagnosis is rapidly increasing. However, whereas relatively abundant information is available in some areas, validation studies are very scant in several subspecialties and completely absent in others. Some validation studies include biopsies from several subspecialties, instead of analyzing biopsies with similar characteristics [33;39-43]. This relative absence of validation studies has led to reluctance in the implementation of WSI in routine clinical practice. Nevertheless, the number of centers implementing this technology is increasing due to the positive experiences reported in many departments [41;42;44;45].
Below we review the current evidence on the validation of WSI versus CLM in the different subspecialties of pathology.

Breast Pathology

WSI has been validated in the diagnosis of breast pathology in a number of studies conducted by different groups. Most of these studies analyzed a relatively small number of routine biopsies (between 100 and 150), including either only needle biopsies or both needle and surgical specimens [32;46;47]. Although scanning at 400x was recommended in one of the studies [32], in two of the studies a scanning magnification of 200x was considered as sufficient[46;47].

The intra- and inter-observer agreement between CLM and WSI is excellent in all the studies, with values ranging between 90% and 99%. Most of the discrepancies detected did not have clinical repercussion. Interestingly, in two of the reports the WSI diagnosis was more frequently considered as correct compared to the diagnosis performed with CLM [32;46]. A study specifically dealing with the distinction between hyperplasia and cancer reported interobserver concordance in the diagnosis of 90.2%. Major discrepancies appeared in 2.3% of the cases, which, in most cases, were solved with IHC stains[48].

A major advantage of digitization in breast pathology is the possibility to use image analysis to improve the accuracy and reproducibility of HER-2, estrogen and progesterone receptors and Ki-67 scoring, which have a crucial role in the planning of treatment strategies [27-29;49]. Moreover, the evaluation may be improved with the use of automatic quantification algorithms (Figure 3).
Cytopathology

The use of WSI in cytopathology has shown some advantages in second opinions, quality assurance, slide archiving, proficiency testing and education. However, a number of significant weaknesses of the current WSI scanners such as the difficulties in focusing at different z-axes are a major limitation for the introduction of this technology in routine diagnosis [50;51]. Improvements in informatics may allow multiplane focusing using the z axis but they still need to be validated [21;52;53].

Indeed, the current evidence of validation in cytology is almost limited to real-time dynamic digital microscopy using a video-camera connected to the optic microscopy and not to WSI. The intra-observer agreement of this approach with the final diagnosis is high (92%)[54], and in some studies is better than with CLM [53-55]. One study evaluating 192 liquid based-cervical cytology slides showed good intra-observer concordance (89- 97%), but the inter-observer concordance was better for CLM than for WSI (94% vs. 82%)[52].

Dermatopathology

Only two studies have focused on the validation of skin biopsies evaluating routine specimens. Although both studies included a small number of cases (100 and 79, respectively) the intra-observer agreement was high (94% for WSI and 96% for CLM, respectively)[30;56]. A study limited to tumor and tumor-like skin lesions showed agreement in the diagnosis by WSI and CLM, with a Kappa value of 0.93 for both methods[57]. Another study evaluated inflammatory and melanocytic lesions, with good agreement between CLM and
WSI (only one discordant diagnosis in the inflammatory biopsies and 100% concordance in the melanocytic specimens), but the number of patients included was very limited (24 cases). In this study it was concluded that in most cases scanning at 200x is sufficient to achieve a correct diagnosis [56].

Interestingly WSI has shown to be suitable for teleconsultation in skin biopsies and may reduce the time of response in expert diagnosis from 5-10 days to a few hours or even minutes [57].

**Gastrointestinal Pathology**

A few studies have shown that the diagnosis of gastrointestinal biopsies using WSI or CLM provides comparable results [58;59]. Two reports analyzed consecutive routine biopsies, but one was limited to gastric and colonic biopsies [59]. The intra-observer concordance between WSI and CLM was 95% in both studies and scanning at 200x was considered as adequate. One study compared WSI and CLM in the evaluation of polyps in surgical specimens. Although the intra-and inter-observer agreement was excellent for both methods in terms of diagnosis, WSI facilitated the quantification of the polyps due to the very low magnification that allows a panoramic view of the complete sample [60]. A study focused on Barret`s dysplasia and neoplasia showed good diagnostic agreement between WSI and CLM, but the consensus neoplasia score was lower using WSI and the time spent in making the diagnosis was longer. These results were probably due to the lack of confidence and experience in the manipulation of the WSI viewer and seemed to improve with familiarity and practice [34].

**Genitourinary Pathology**
Prostatic biopsies, particularly needle biopsies, are good candidates for digitization for a number of reasons: the tissue size is small and the images generated are lighter; multiple measurements are frequently required and informatic tools can facilitate these; and WSI allows a global view to more easily establish the Gleason score (figure 4)[61]. An additional advantage of WSI is the possibility to synchronize hematoxilin-eosin stains and p63 IHC in the same screen, thereby allowing the comparison of the two images and facilitating the diagnostic and teaching process [62].

Thus, the current evidence on the validation of WSI in the diagnosis of prostatic biopsies is more extensive than in other areas. A number of studies including between 50 and over 800 cases have been focused on the evaluation of the Gleason score in needle biopsies. Scanning at 200x was considered sufficient to make the diagnosis. The Kappa values for diagnosis ranged between 0.586 and 0.813 [63-65], but one of the reports included only difficult biopsies with a borderline Gleason score. Concordance between WSI and CLM seems to be higher for primary (Kappa values 0.65 - 0.96) than for secondary Gleason scores (Kappa values 0.53 - 0.75), and most discordances have no impact on patient management [66]. Tumor size is better evaluated with WSI, and other parameters such as perineural invasion show similar values with WSI and CLM [66].

Two additional studies focused on genitourinary biopsies included a mixture of prostatic and urinary tract biopsies and showed good intra-observer concordance (90% and 87.5%, respectively) [67;68].

Gynecological Pathology
Studies on the validation of WSI in gynecological biopsies are scant. Only one study conducted at our institution analyzed inter-observer agreement in 452 routine gynecological specimens showing a Kappa index of 0.914 (almost perfect concordance). Interestingly, the agreement between WSI and CLM increased in this study in parallel with time, suggesting that there is a learning curve in the use of WSI and that experience in the use of WSI viewers improves the results obtained. Major discrepancies were found in only 2% of the cases, and none was related to poor image quality. Most discrepancies in this study were observed in biopsies of premalignant lesions of the uterine cervix, an area which has shown high inter- and intra-observer variability rates using CLM[36]. The magnification used in the study was 200x and higher magnification did not seem to be required.

A second study described the usefulness of WSI in the evaluation of 52 frozen ovarian sections showing 96% inter-observer agreement. Interestingly, in this study no clinical information was provided to the pathologists and the time spent per case was 3-5 minutes [7].

**Head and Neck Pathology**

To date no validation study including the complete spectrum of samples of this subspecialty is available. Only one study focused on premalignant laryngeal lesions has been published. This study concluded that WSI is a valid alternative to CLM. Although the correlation with the final diagnosis was slightly lower with WSI than with CLM, the differences were not statistically significant[69].

**Neuropathology**
Validation studies of neuropathology are limited to intraoperative biopsies and smears [5;8;70]. Agreement between the diagnosis with WSI and the final diagnosis using CLM is very good, even with low scan magnification (100x). The studies conclude that 200x magnification is sufficient to obtain a diagnosis. In one study the diagnosis achieved with WSI was concordant with CLM in 29 of the 30 cases evaluated, and the discordant diagnosis did not lead to changes in the management of the patient [8]. A second study included 126 frozen sections that were evaluated by 4 different pathologists. The diagnosis was discordant with the final report in only 8 cases. In this study the diagnosis of the frozen section scanned and diagnosed using WSI, was compared with the final diagnosis obtained in formalin-fixed, paraffin embedded tissue [70].

Algorithms are currently being developed to identify the hot spots in Ki67 stained sections to automatically quantify the proliferative activity in tumors of the central nervous system [71;72].

**Pediatric Pathology**

Two studies have validated the use of WSI in pediatric pathology. One included 80 routine biopsies of patients under 18 years of age and 20 placentas. The intra-observer concordance between the diagnoses with WSI and CLM was 90% in pediatric biopsies and 93% in placental specimens. Major discrepancies were observed in only 2% of the cases. A scanning magnification of 200x generated an image quality allowing correct diagnosis, except for the identification of nucleated red blood cells; which is very difficult even when the slides are scanned at a magnification of 400x [73].
The second study evaluated WSI in 60 cases selected to include the whole spectrum of the diagnostic complexity of pediatric biopsies. The surgical specimens were digitized at 200x magnification, whereas small biopsies and cytological samples were digitized at 400x. The intra-observer agreement was almost perfect with only one discordant case. The scanning process of two cytological smears was unsatisfactory because the material was very scanty [74].

**Pulmonary Pathology**

One study validating WSI in the diagnosis of intraoperative pulmonary specimens included a variety of samples, with 114 frozen sections from tumors, lymph nodes and bronchial margins, 174 fine needle aspiration slides, 3 exfoliative smears and 13 small biopsies. This study evaluated both a dynamic real-time telepathology system and WSI and found very good agreement, which was better for WSI than for the real-time telepathology system (100% in consultation and frozen biopsies) [75]. A second study analyzed the use of WSI in 20 tumor biopsies and surgical specimens sent for consultation. Complete inter-observer agreement was achieved in 85% of the cases, even at a scanning magnification of 100x [75].

**Renal Pathology**

Validation studies of WSI in the diagnosis of renal pathology biopsies are scarce and include few cases. A report including 50 routine renal biopsies showed complete intra-observer agreement in 84% of the cases. Five major discrepancies (with clinical repercussion for the patient) were found and in two cases the correct diagnosis was that made with WSI. In this study renal
transplant biopsies showed significantly more discrepancies at a magnification of 200x[67]. Another study using a magnification of 400x reported good agreement in renal transplant biopsies, but the time spent for obtaining the diagnosis was longer with WSI than with CLM [76]. Finally, one study evaluated the concordance between 96 pathologists in the diagnosis of 12 renal biopsies using WSI and CLM and found no significant differences between the two methods [77].

**Frozen intraoperative diagnosis**

A number of studies have evaluated dynamic real-time telepathology in intraoperative sections, showing a good correlation with CLM diagnosis. They emphasize the learning curve in the use of the WSI technology, which typically involves longer diagnostic time at the beginning but rapid improvement with practice [78]. A validation study using WSI in frozen intraoperative sections from different anatomical sites has shown almost perfect agreement with a Kappa index of 0.85. The mean time spent on diagnosis was 2 min 50 sec per case. The quality of the image was considered excellent in 98% of cases [9]. Studies using WSI in frozen intraoperative sections from specific specialties have been discussed above. Another study evaluated the diagnosis of 67 consecutive frozen intraoperative sections viewing the virtual slides in a portable device (iPad tablet). The slides were scanned at 200x, and all cases were shown together with the clinical information. The concordance between the diagnoses achieved with WSI and CLM was good with a Kappa value of 0.85. The mean time to achieving a diagnosis using WSI was 2 minutes and 46 seconds [79].

**Surgical pathology**
A number of studies have evaluated a variety of different specimens from the routine practice of a Department of Pathology, including between 25 and 607 samples [22;39-41;80;81]. Inter- and intra-observer agreement between WSI and CLM varied from 75% to 97.7% depending on the study. Most studies conclude that a magnification of 200x provides images with adequate quality for diagnosis [22;39-41;80;81]. The inter-observer agreement between WSI and CLM was 95%, and all discrepancies were minor. However, although the general opinion of the pathologists was positive, some felt that the WSI system was slower than CLM, and most of the pathologists interviewed were reluctant to completely move from CLM to WSI in routine diagnosis [33]. One study suggested that the inter-observer agreement was better for neoplastic than for non-neoplastic diseases [16]. It has been suggested that a scanning magnification of 200x may not be sufficient to allow correct diagnosis in inflammatory lesions [82].

Finally two studies included only consultation biopsies of different organs. The inter-observer agreement between WSI and CLM diagnosis in these studies was greater than 91%, and most of the discrepancies were due to the intrinsic difficulty in the diagnosis of some cases [15;16;82].

Current recommendations for validation of WSI

Validate of WSI at each institution has been recommended before its implementation in routine diagnosis. Several professional associations have developed guidelines with recommendations for the introduction of WSI in routine diagnosis in a Department of Pathology. The first guidelines were developed by the College of American Pathologists and the American Telemedicine Association and includes some recommendations and
suggestions to be followed before using WSI for diagnosis [2;10;17]. It is recommended to include a variety of different biopsies representative of the complexity of the surgical specimens usually analyzed in the center. The guidelines state that it is not necessary to validate each subspecialty because the results from one specialty can be extrapolated to others with similar features. Each specific type of specimen with significant differences requires an internal validation. The guidelines recommend measuring intra-observer agreement between WSI and CLM, using a “washout period” of 2 weeks. Finally, it is recommended that a pathologist with experience in WSI should be involved in the process of validation.

**Conclusions**

In conclusion, independently of the subspecialty, all the validation studies published show a very good correlation between diagnoses achieved with WSI and CLM. Thus, WSI seems to be an adequate tool for histological diagnosis in routine practice and has several advantages over CLM. However, although good evidence demonstrating that WSI can be reliably used for routine diagnosis has been provided for several specialties, there are a number of areas of pathology, such as liver, endocrine, hematopathology, or bone and soft-tissue pathology for which no study has yet been published. Although some of these areas may be considered similar to others already validated, specific validation studies are needed in other areas with many differences such as liver biopsies or hematopathology. These validations are necessary before the use of WSI can be extended to these subspecialties with the aim of going fully digital in Pathological Services in the future.
Notwithstanding, as with many other new tools, the use of WSI has a learning curve, and the time spent on the diagnosis and, to a lesser extent inter- and intra-observer agreement, may be suboptimal in the initial phases of its use. Cytology seems to be an exception; the application of WSI in this area is more controversial due to the impossibility of focusing on different planes.

However, the introduction of WSI in routine diagnosis faces some difficulties, mainly related to the reluctance of pathologists to abandon CLM and to the costs associated with the acquisition of the equipment and the storage of the images generated. New technologies that allow creating 3D reconstruction from two dimensional biopsies may help to improve the understanding of the growth patterns and the spatial arrangement of diseased cells [21;83]. Another area that will markedly expand in the next few years is that of histopathology pattern recognition using image analysis, which can facilitate the diagnostic tasks and improve the reproducibility among pathologists in many subspecialties [65;84-90].
References


LEGENDS OF FIGURES

Figure 1. WSI work stations for primary diagnosis typically include two screens, one displaying the WSI viewer and the other the laboratory information system and the clinical records or other clinical or imaging information. This physical structure has shown to be highly ergonomic. Additional advantages of WSI viewers are a much larger field of vision than CLM and the possibility of using a very low magnification.

Figure 2. WSI viewers may simultaneously show and synchronously move several slides of a case, which is particularly helpful in the evaluation of immunohistochemically stained slides.

Figure 3. A major advantage of digitization in breast pathology is the possibility to use image analysis in improving the accuracy and reliability of HER-2, estrogen and progesterone receptors and Ki-67 scoring, which have a crucial role in the planning of treatment strategies.

Figure 4. Prostatic biopsies often require multiple measurements. The tools of WSI viewers allow these measurements to be easily performing. WSI allows a global view to more adequately establish the Gleason score.