Interinstitutional Whole Slide Imaging Teleconsultation Service Development

Assessment Using Internal Training and Clinical Consultation Cases

Nicholas C. Jones; Rosalynn M. Nazarian, MD; Lyn M. Duncan, MD; Michal Kamionek, MD; Gregory Y. Lauwers, MD; Rosemary H. Tambouret, MD; Chin-Lee Wu, MD, PhD; G. Petur Nielsen, MD; Elena F. Brachtel, MD; Eugene J. Mark, MD; Peter M. Sadow, MD, PhD; John P. Grabbe, MD; David C. Wilbur, MD

• ***Context.***—**Assessment of accuracy and feasibility of whole slide imaging (WSI) for interinstitutional consultation in surgical pathology.**

Objectives.—**To train technical and pathologist staff in WSI technology, establish and evaluate a WSI workflow using training cases and second-opinion consultations, and assess diagnostic accuracy.**

Design.—**First, WSI training and evaluation using selected subspecialty service cases were performed and compared with the clinical glass slide (GS) diagnosis. Second, WSI and GS diagnoses of consecutive, second-opinion consultation cases were compared. Discrepancies underwent adjudication to determine a reference diagnosis. Participant observations on WSI initiation to practice were gathered.**

Results.—**There were 130 cases evaluated, with 123 correlations (94.6%) and 6 minor (4.6%) and 1 major (0.8%) discrepancies. The 74 consultation cases interpreted had 52 correlations (70.3%) and 18 minor (24.3%), and 4 major (5.4%) discrepancies. The WSI and GS adjusted major discrepancy rates in second-opinion consultations were 2.7% (2 of 74) and 4.1% (3 of 74), respectively. Statistical analysis showed that WSI was not inferior to GS interpretation. Pathologists agreed the software was easy to use and the ~~images~~ were adequate, but more time was spent rendering WSI interpretations.**

Conclusions.—**A significant learning curve was observed in the transition from the training set to clinical consultation cases associated both with WSI interpretation and adjustments to the digital analogs of routine GS workflow. Results from second-opinion consultations indicated that WSI interpretation was as accurate as GS interpretation in properly trained and experienced users. Overall, WSI-based practice appears feasible for second-opinion consultations.**

(*Arch Pathol Lab Med.* doi: 10.5858/arpa.2014-0133-OA)

The use of whole slide imaging (WSI) technology for surgical pathology diagnosis is being investigated in many laboratories around the world. Improvements in the technology during the past several years, coupled with better understanding of pitfalls and caveats of the digital pathology environment and reduced costs for instrumentation, networking, and memory, have allowed for a greater understanding of the promise and the remaining challenges. Many studies have shown potential alternatives or improvements to traditional pathology practice through use of WSI technology in intraoperative frozen section interpretations,[1–3] image analysis applications with hematoxylin-eosin stains[4] and immunohistochemical and special stains,[5] in archiving and quality assurance programs,[6,7] for primary diagnosis,[8,9] for potential laboratory workflow improvements and pathologist ergonomics,[10] with laser-capture microdissection[11] (and thus burgeoning molecular testing methods), and for consultation purposes.[12]

Teleconsultation with WSI could improve consultation services by dramatically decreasing the turnaround time, compared with shipping glass slides; decrease the amount of pathologist labor required in interpretation versus non-robotic or robotic, dynamic (video) telepathology (using either remote personnel or a robotic control of the microscope, respectively); remove the selection bias of static telepathology (transmission of micrographs); improve collaboration between the consult requestor and the consultant (through use of annotation features within WSI viewing software); and provide greater opportunities for quality assurance, archiving, and education. Consultation services

Accepted for publication June 23, 2014.

From the Department of Pathology, Massachusetts General Hospital and Harvard Medical School, Boston (Mr Jones and Drs Nazarian, Duncan, Kamionek, Lauwers, Tambouret, Wu, Nielsen, Brachtel, Mark, Sadow, Grabbe, and Wilbur); and the Department of Pathology, Cambridge Health Alliance, Cambridge, Massachusetts (Dr Grabbe). Dr Kamionek is now with the Department of Pathology, Carolinas Pathology Group, Charlotte, North Carolina.

Dr Wilbur is a member of the scientific advisory board of Corista, LLC and the medical advisory board of Philips Digital Pathology, for no compensation in both cases. The other authors have no relevant financial interest in the products or companies described in this article.

Reprints: Nicholas Jones, Department of Pathology, Massachusetts General Hospital, 55 Fruit St, Gray-Jackson 249C, Boston, MA 02114 (e-mail: ncjones@partners.org).

today vary significantly in clinical context, allowing for a wide range of priorities, difficulty levels, types and complexities of cases, and potential application of WSI.

In some clinical studies or practice, WSI is used to allow for telepathologic diagnosis for the purposes of resource consolidation or to improve pathologist efficiency by limiting travel time.[2,3] In these settings, the difficulty of the cases is not the primary pertinent issue; the rationale for telepathology is based on the logistics and economics of pathologist staffing at multiple sites. In this scenario, the difficulty of cases is likely highly variable and more representative of traditional primary and frozen section microscopic practice. In contrast, the cases seen in a second-opinion consultation service will typically represent a greater difficulty level and constitute a population of cases that should be studied separately. For second opinions, the requesting pathologist is presented with a difficult case and wishes to obtain a subspecialist's opinion before signing out the case. This clinical context requires rapid interpretation of challenging cases. The 2 protocols discussed in this article reflect a specific focus on this latter clinical scenario, namely, the development of a second-opinion WSI teleconsultation service at a large, academic medical center that routinely processes significant numbers of glass-slide (GS) consultations from community pathologists.

The first protocol in this study details data derived from training and evaluation of the use of WSI technology by a digital pathology technologist and several subspecialist pathologists. The second protocol in this study details a clinical trial of a prospective, nonbiased, random sample of second-opinion consultation cases sent from a community hospital to a subspecialty-based, academic pathology service. The former protocol included the development of realistic workflows for everyday clinical practice, whereas the latter evaluates the performance of this established digital-pathology workflow using actual clinical consultation cases. These studies are not to be considered true validations but are investigations into possible issues of WSI diagnosis, case-specific issues, digital pathology workflow, and the overall accuracy of WSI compared with GS interpretation.

MATERIALS AND METHODS

In both study protocols, WSIs were created using a Mirax MIDI scanner (Zeiss, Oberkochen, Germany, currently sold as Pannoramic MIDI, PerkinElmer, Waltham, Massachusetts, produced by 3DHISTECH, Budapest, Hungary), which uses a ×20 Zeiss 0.8 numerical aperture plan apochromat objective, yielding images with 0.32 μm/pixel. The WSIs were initially reviewed for quality control on the scanning computer by the WSI technologist using the Mirax Viewer software. The quality control process is estimated at a minute per slide and includes the technician checking the entire WSI at low power and checking some areas at high power. The process screens for any missed tissue, out-of-focus regions, or miscellaneous technical issues. When images were found to be inadequate, they were rescanned until either an adequate image was made or the case was excluded because of technical issues.

Once adequate images were created and quality controlled, the images were uploaded to the Corista Digital Pathology Platform (varying versions, Corista LLC, Concord, Massachusetts). The Corista platform (see Figures 1a, 1b, and 2) functions as a digital-pathology case-management platform, encompassing case accessioning, assignment, pathologist review, annotation, diagnosis, case reporting, archiving, and querying; in essence, the platform provides some functions of an anatomic pathology laboratory information system, in addition to its basic function as a networked WSI viewer (also known as a *cloud viewer* or a *browser-based viewer*). A networked viewer has images saved on a server and dynamically serves images to a pathologist as they navigate the case, in contrast to a traditional (*local* or *stand-alone*) WSI viewer, in which WSIs are saved and viewed on an individual's personal computer.

During image upload, the WSI technologist accessioned cases into the Corista platform, including patient and clinical information, case and specimen type, gross description, and any case-specific, administrative or technical notes. Additional clinical information can be scanned and attached to a case in any document format. After data entry, the case was assigned to a pathologist and appeared on their work list. Pathologist workstations in the first protocol included personal computers with dual-core 1.86 GHz or faster processors, 3.5 GB RAM, and consumer-grade, 17-inch monitors capable of 1280 × 1024 resolution. The pathologist logged into the system through a Web browser and performed the case evaluation, having access to all entered patient and specimen information. The pathologist could also request digital intradepartmental consultations through the Corista system. The pathologist completed the case by entering a final diagnosis into the system.

The training protocol consisted of scanning and interpreting cases from each participating pathologist's clinical rotations. Cases were selected on an ad hoc basis from daily workload. The pathologist reviewed the WSI first, rendering a diagnosis (the WSI interpretation), then viewed the glass slide or slides per routine practice (the *GS interpretation*). No specific "wash-out" period (time between the WSI and GS review) was used in this protocol. The imaging technician providing initial training sessions to the pathologists and was available at all times while the WSI reviews were taking place to ensure adequate training and system performance and to gather ongoing information for system and operation improvement. Emphasis was placed on evaluating differences in subjective appraisals of the tissue on WSI versus GS and on image quality of the WSIs, and those data were gathered through dictation to the technician or by written comments. Concordance rates were determined between the WSI and GS interpretations.

The second protocol was conducted under appropriate institutional review board approvals for the participating facilities. In this protocol, a nonbiased, random sample of actual second-opinion consultation cases being sent from a community hospital to a subspecialized, academic hospital was scanned prospectively upon receipt. All consultations relevant to the subspecialty areas of the participating pathologists were selected during the study period. By the time of the second protocol, pathologist workstations included personal computers with dual-core 2.66 GHz or faster processors, 3.5 GB RAM, and consumer-grade 24-inch monitors capable of 1920 × 1080 resolution. The consultation slides proceeded through the usual clinical workflow for GS interpretation and were reported in the routine fashion. A small subset of cases was scanned after GS interpretation when scanning upon receipt was not possible because of a patient care-related timing issue. The WSI cases were assigned to study pathologists based on an assessment of the subspecialty area and the availability of the appropriate subspecialty pathologist. If, after review of the initial scanned slides, the pathologist requested a particular stain (eg, immunohistochemistry or a special stain), the initial interpretation was saved. If the stain became available through the routine clinical workflow, it was then scanned and made available for the WSI pathologist to use to add to or modify his or her initial interpretation. If no additional stains were available, the initial interpretation was recorded as the WSI interpretation. Stains obtained in the routine workflow were not scanned and presented to the WSI pathologist if they had not been specifically requested. From the WSI pathologists' perspective, the workflow was effectively equivalent to a WSI consultation service in which cases are scanned at the community hospital and sent digitally to the consulting institution (see Figure 3 for a diagram of the intended workflow).

Cases in both protocols were adjudicated and categorized into 3 categories: no discrepancy (complete correlation between 2 diagnoses), minor discrepancy (a noteworthy difference between the 2 cases that would not affect patient care), and major

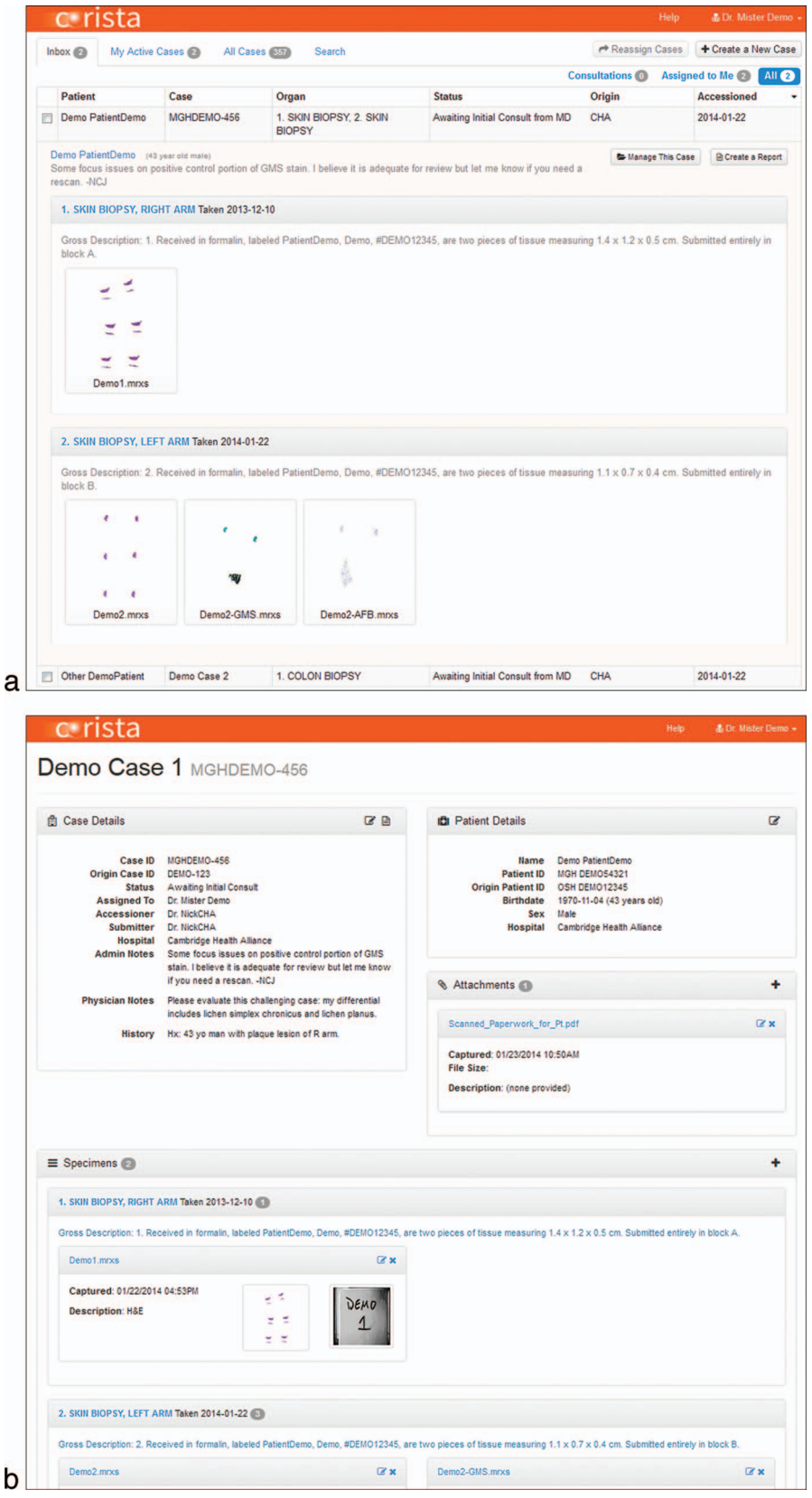


Figure 1. *a, Example of a pathologist's case queue in the Corista Digital Pathology Platform (Corista LLC, Concord, Massachusetts) (demonstration image only). b, Example of the case management page (demonstration case only).*

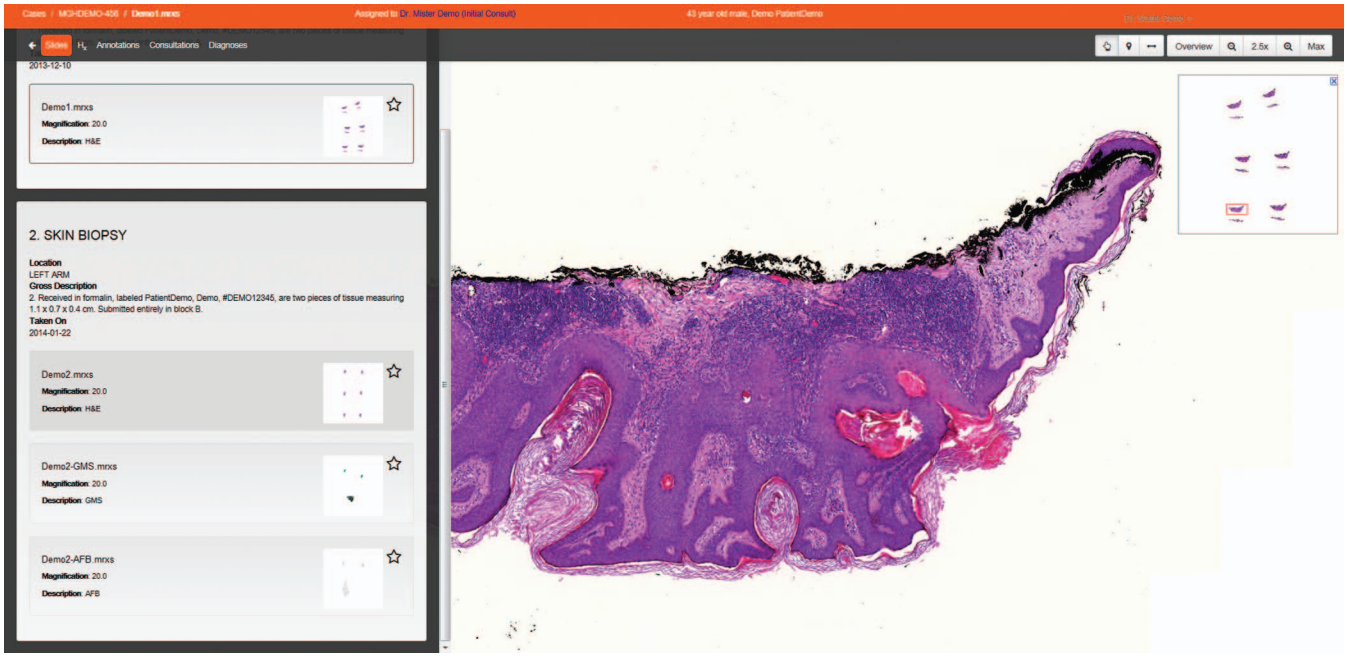


Figure 2. *Screen shot of the whole slide image viewing screen (demonstration image only) (hematoxylin-eosin, original magnification ×20 at 0.32 μm/pixel; downsampled image presented at ×2.5).*

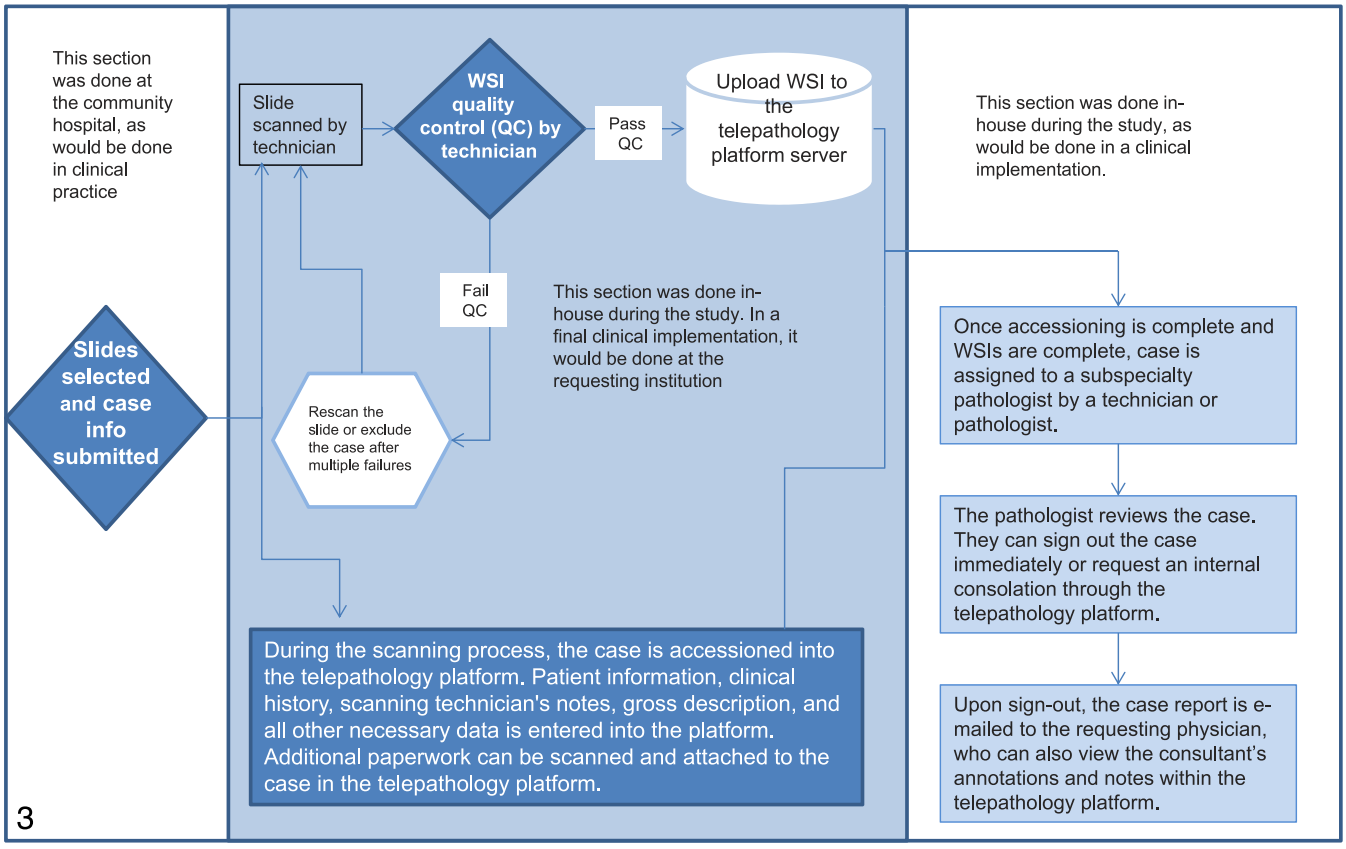
discrepancy (a difference between the 2 diagnoses that would affect patient care). Adjudication of the training protocol was performed by the senior author and included data from the pathologists' reviews of each case and consultation with the appropriate subspecialist, if indicated. In the second protocol, the WSI and GS interpretations were reviewed by the senior author and classified as no discrepancy, minor discrepancy, or major discrepancy or were sent for further subspecialist review when an initial assessment was not possible. All major and selected minor discrepancies were sent for further adjudication to determine a reference diagnosis for each case. Adjudication cases were reviewed by the clinical (GS) and study (WSI) pathologists or by a third subspecialist pathologist. The adjudication pathologists were blinded to the original diagnoses. Attempts were made to determine the nature and cause for each discrepancy. Because of inherent limitations of pathologist availability a few cases in the second protocol were reviewed digitally and routinely by the same pathologist. Cases diagnosed by the same pathologist within 2 weeks of each other were excluded from statistical analysis. All included cases of this type, therefore, had an at least 2-week washout period between WSI and GS reviews. In discrepant cases, the reference diagnosis was compared with the WSI and GS interpretations to determine which diagnosis was to be considered correct for determining the error rate in each arm of the study.

Statistical analysis was not used on the training set because of the selection bias inherent in the protocol. For the second protocol, which was a nonbiased, random sample of clinical consultation cases, the McNemar test was used to determine the equivalence of WSI and GS interpretations by evaluating comparative, adjusted major discrepancy rates. The adjusted major discrepancy rates were calculated by subtracting the major discrepancies that were superior by that method from the total major discrepancies. This test was used because the WSI and GS interpretations are not truly independent because of the study design of matched pairs of diagnoses; the same cases were diagnosed by both methods. The test's confidence interval was used to evaluate the potential inferiority of WSI at an α of .05. The noninferiority margin was chosen at 4%, similar to prior evaluations used for a primary diagnosis validation.[8] Minor discrepancy rates were not evaluated statistically.

RESULTS

In the training set, 130 cases composed of 170 specimens with 357 slides were reviewed by 9 pathologists. Two cases had been excluded by the technician for technical reasons. The technician's rescan rate was 13.4% (48 of 357). The mean number of cases interpreted by each pathologist was 19. Of the 130 cases, there were 6 minor discrepancies (4.6%) and one major discrepancy (0.8%); the remaining 123 cases (94.6%) were classified as no discrepancy. The minor discrepancies showed several occasions of a higher degree of confidence on GS and slight differences in terminology. The one major discrepancy occurred when a pathologist viewing their second digital case signed out the case containing 2 WSIs after only viewing a single slide, thereby, missing important pathology present on the second slide. Most discrepancies occurred within the first 7 cases a pathologist diagnosed digitally. In the clinical consultation trial, 78 cases were initially accessioned. One case was excluded from the study by the technician because of inadequate scans caused by scanner-maintenance issues. No cases were deferred by the pathologists because of image-quality problems, although they were given the option to do so. Three single-specimen cases (all no discrepancy) were excluded because they were interpreted by the same individual within a 2 week period, and, therefore, did not meet the wash-out criteria, leaving 74 cases consisting of 86 specimens and 347 slides for evaluation. (See Table 1 for a breakdown of cases by subspecialty.)

In the clinical trial set, the technician's rescan rate was 5.4% (19 of 347). Of the 74 cases, 4 (5.4%) were found to have major discrepancies. The first, a gastric antrum biopsy, had a GS interpretation of high-grade dysplasia and a WSI interpretation of severe reactive changes. The adjudication process indicated the reference diagnosis was severe reactive changes. The second, an abdominal skin-shave biopsy, had a difference in the determination of the extent of cytologic



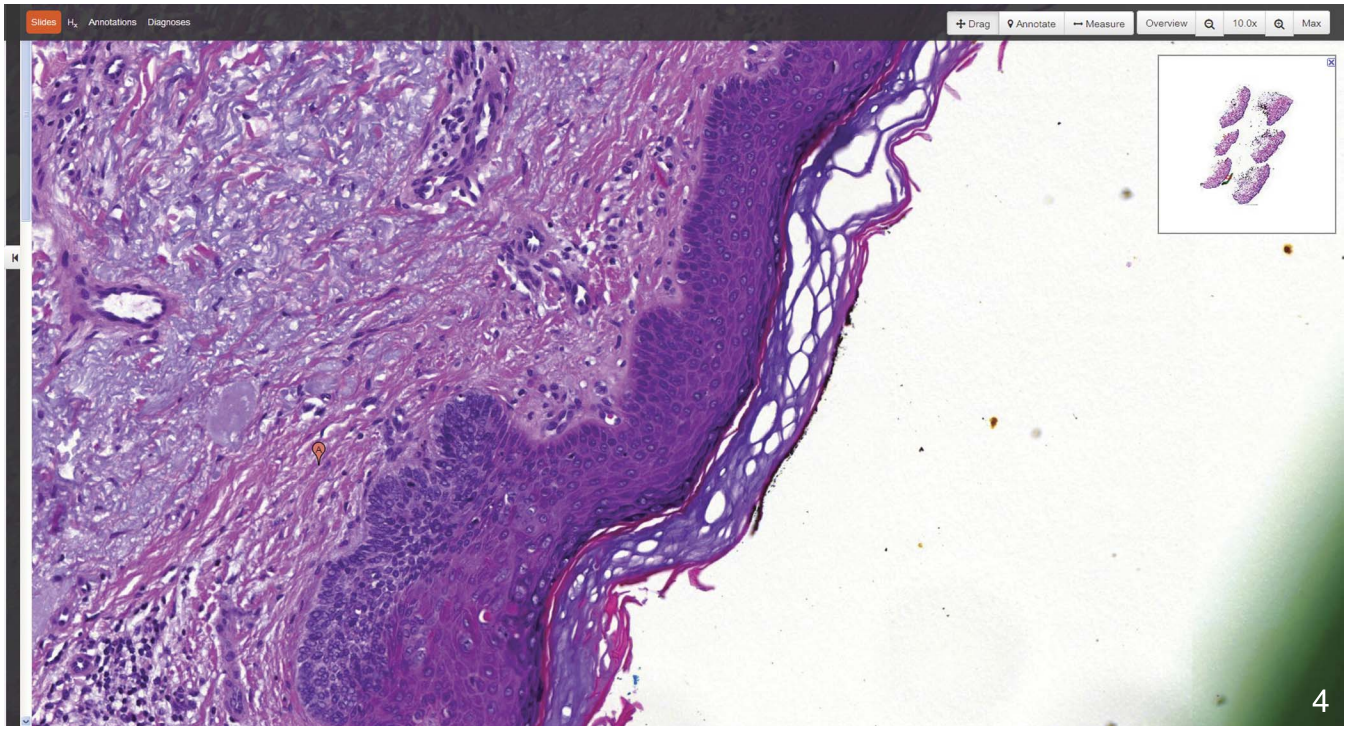


Figure 3. *The proposed service workflow is shown. In the clinical trial, all scanning was done within the academic medical center, but in final clinical practice, the slides would be scanned at the requesting institution.*

Figure 4. *Basal cell carcinoma missed on whole slide image review. Adjudication determined this slide was likely never reviewed (or reviewed carefully) in the 18-slide case (hematoxylin-eosin, original magnification ×20 at 0.32 µ/pixel; downsampled image presented at ×10).*

Table 1. Cases by Subspecialty in the Clinical Trial Protocol and Cases Interpreted Correctly in Each Arm

Subspecialty	Cases, No. (%)	WSI Correct, No. (%)	GS Correct, No. (%)
Dermatopathology	52 (70)	51 (98)	51 (98)
Gastrointestinal	10 (14)	10 (100)	9 (90)
Gynecologic	6 (8)	6 (100)	6 (100)
Bone and soft tissue	2 (3)	2 (100)	2 (100)
Head and neck	2 (3)	2 (100)	2 (100)
Breast	1 (1)	1 (100)	1 (100)
Pulmonary	1 (1)	0 (0)	0 (0)
Total	**74** (100)	**72** (97)	**71** (96)

Abbreviations: GS, glass slide; WSI, whole slide imaging.

atypia in a lentiginous compound dysplastic nevus; the GS interpretation classified it as severe cytologic atypia, whereas the WSI interpretation classified it as mild cytologic atypia. Upon adjudication, the reference diagnosis was determined to be mild cytologic atypia. The third case was a skin excision from the left leg of a patient with a history of basal cell carcinoma and squamous cell carcinoma in situ in that area. The GS interpretation included a residual superficial basal cell carcinoma and scar tissue, whereas the WSI interpretation included only the scar and did not note any residual carcinoma. Upon adjudication, it was noted that, on both the GS and the WSI, the basal cell carcinoma was clearly and easily noted, especially because it had been dotted by the requesting pathologist. The digital slide was apparently missed by the reviewer among the 18 slide case because the third-party adjudicator noted that the lesion, which was only present on one slide, was clearly and adequately identifiable (see Figure 4). The fourth case was a wedge biopsy of the right upper lobe of the lung in a patient with a history of smoking, asthma, and chronic obstructive pulmonary disease. The GS interpretation was respiratory bronchiolitis, whereas the WSI interpretation was pulmonary hemorrhage with interstitial infiltrate, possibly representing capillaritis. Adjudication by 3 pathologists led to 3 different diagnoses, with the final reference diagnosis being "the pathology is not definitive" because an accurate

Table 2. Minor Discrepancies in the Consultation Study

Specialty	Specimen Type	WSI Diagnosis	GS Diagnosis	Reason for Difference	Adjudication Preference
Breast	Breast biopsy	Fibroepithelial lesion	Hamartoma	Interobserver variability	Neither
Dermatology	Skin biopsy	Hyperkeratosis, parakeratosis, dermal fibrosis	Excoriated seborrheic keratosis	Interobserver variability	Neither
Dermatology	Skin biopsy	Hyperkeratosis, hypergranulosis, acanthosis, perivascular chronic inflammation	Fibrous papulae with granulation tissue and fibrosis	Interobserver variability	WSI
Dermatology	Skin biopsy	Acral lentiginous nevus with slight atypia	Acral junctional nevus	Thresholding	Neither
Dermatology	Skin biopsy	Lentiginous compound nevus	Lentiginous compound nevus with moderate atypia	Thresholding	WSI
Dermatology	Skin biopsy	Lichenoid dermatitis	Hypertrophic actinic keratosis	Interobserver variability	WSI
Dermatology	Skin biopsy	Spongiotic and interface dermatitis	Suggestive of morphea	Interobserver variability	GS
Dermatology	Skin biopsy	C/w epidermal nevus	C/w fibrokeratoma	Interobserver variability	GS
Dermatology	Skin biopsy	Lentiginous compound nevus with slight atypia	Lentiginous compound nevus	Thresholding	WSI
Dermatology	Labial biopsy	Nonspecific inflammatory changes, favor lichen simplex chronicus, cannot exclude lichen sclerosus	Inflammatory changes, differential diagnosis includes lichen sclerosus	Interobserver variability	WSI
Dermatology	Skin biopsy	C/w verruca vulgaris	C/w elephantiasis nostra verrucosa	Interobserver variability	GS
Dermatology	Skin biopsy	C/w actinic keratosis	Lichenoid hypersensitivity reaction	Interobserver variability in the interpretation of a focal finding	GS
Dermatology	Skin biopsy	Compound nevus	Dermal nevus	Interobserver variability in the interpretation of a focal finding	Neither
Dermatology	Skin biopsy	Lichenoid actinic keratosis	Prurigo nodule	Interobserver variability	WSI
GI	Gastric biopsy	Polypoid inflamed mucosa with intestinal metaplasia	Polypoid intestinal metaplasia with reactive atypia	Thresholding	GS
GI	Liver biopsy	Marked steatosis, steatohepatitis with marked fibrosis	Moderate steatosis, question evolving cirrhosis	Thresholding	Neither
GI	Gastric biopsy	Inflamed polypoid mucosa. Request *H pylori* stain	*H pylori*	WSI requested stain—not available	GS
Gynecology	Endometrial biopsy	Endometrial polyp	Endometrial polyp with nonatypical hyperplasia	Thresholding	GS

Abbreviations: C/w, consistent with; GI, gastrointestinal; GS, glass slide; *H pylori*, *Helicobacter pylori*; WSI, whole slide image.

Table 3. Case Analysis From the Second Protocol for McNemar Test[a]

	GS Correct, No. (%)	GS Incorrect, No. (%)	Total, No. (%)
WSI correct	70 (94.6)	2 (2.7)	72 (97.3)
WSI incorrect	1 (1.4)	1 (1.4)	2 (2.7)
Total	**71** (95.9)	**3** (4.1)	**74** (100)

Abbreviations: GS, glass slide; WSI, whole slide imaging.

[a] For this analysis, *correct* was defined as no discrepancies plus minor discrepancies, and *incorrect* was defined as major discrepancies.

Table 4. Statement of Hypothesis Tests and Confidence Intervals

McNemar Test (2-Sided)
H0, *GS-adjusted major discrepancy rate − WSI-adjusted major discrepancy rate* = 0
HA, *GS-adjusted major discrepancy rate − WSI-adjusted major discrepancy rate* ≠ 0
P value = .46; we do not reject the null hypothesis
McNemar's 95% CI
GS-adjusted major discrepancy rate − WSI-adjusted major discrepancy rate = −3.24 to 5.94
This suggests the noninferiority of WSI by the 4% margin with 95% confidence.

Abbreviations: 95% CI, 95% confidence interval; GS, glass slide; WSI, whole slide imaging.

reference diagnosis could not be determined. The closest thing to a reference diagnosis would include both interpretations as differential diagnoses; for the purposes of the study, both the WSI and GS interpretations were considered incorrect. All 3 pulmonary pathologists agreed that the case was not ideal for this type of study, but because the sampling method for case inclusion was not contrived, it was considered a "nondefinitive" case, a diagnosis that would be occasionally expected in a study of actual second-opinion consultation cases.

Eighteen of the 74 cases (24%) had minor discrepancies, which are more difficult to categorize (see Table 2). Of the 18 cases, adjudicators preferred 6 (33%) of the WSI interpretations, 6 (33%) of the GS interpretations, and showed no preference on the remaining 6 (33%). The most common reason for minor discrepancies were interobserver difference and thresholding or slight differences in emphasis on different features in the specimens. One gastric biopsy was notable in that the WSI interpretation noted chronic inflammation, and a stain for *Helicobacter pylori* was requested, whereas the GS pathologist had noted the *H pylori* on the hematoxylin-eosin stain. Although the presence of *H pylori* on a hematoxylin-eosin stain was found to be extremely subtle, the third-party pathologist on the case noted that an *H pylori* stain would usually be ordered in that type of case. This case was notable because the study protocol led to the minor discrepancy; the study pathologist was unable to order the stain because it had not been previously made for routine clinical care. Differences in thresholding were perhaps most notable in the dermatopathology cases, which accounted for 11 of the 18 minor discrepancies (61%). Among those 11 cases, 5 WSI interpretations (45%) were preferred and 4 GS interpretations (36%) were preferred, with no preference in the other 2 (18%). It was the opinion of the adjudicators that the differences were not related to the modality or the image quality but were due to interobserver variability caused by inherent differences in interpretation of histopathologic findings, nonequivalent thresholding, and terminology in dermatopathology.

Statistical analysis of the data is based on the McNemar test (see Tables 3 and 4). The test uses the null hypothesis that the major discrepancy rates of WSI and GS methods are equal. We did not reject the null hypothesis ($P = .46$), although we did not seek to; our intent was to show the noninferiority of WSI through use of the test's confidence interval. Our 95% confidence interval (95% CI) was –3.24 to 5.94. The 95% CI suggests that the WSI was, at worst, 3.24% more likely to have a major discrepancy than GS was and, at best, was 5.94% less likely to have a major discrepancy. This meets our intended objective of testing statistically with a 95% CI that WSI was noninferior because the noninferiority margin was chosen as 4%. These results, however, only suggest that this is true for the population of cases sent for consultation, and the authors make no specific claims about the results in any of the subspecialties not represented in the study.

COMMENT

Although the training protocol was not intended for statistical analysis, the lessons we learned were that WSIs were generally adequate for confident interpretation, that cases that were difficult to diagnose using GS microscopy were also difficult on WSI, that pathologists were most likely to make an error in their first several cases using WSI, and that the administrative and user interface aspects of WSI diagnosis are as likely to lead to error as are any issues associated with image adequacy. For both the training set and the clinical trial set, only 2 of the 74 cases (2.7%) were found to have major discrepancies in the WSI arms, and in both cases, it was determined that the pathologist did not actually review the digital slide containing the diagnostic focus. Based on that observation, the software vendor implemented a feature in the viewer to help prevent that simple, but important, error, by changing the appearance of each digital slide once it has been viewed. It should be noted that a pathologist can miss examining an individual glass slide in a multislide case and that is avoided by careful repetition and development of consistently good habits (such as flipping the slide in the tray). Our experiences suggest that similar repetition, along with a well-conceived user interface design, should be able to prevent this type of error.

The results from the second-phase clinical trial suggest that the WSI is adequate for interinstitutional, second-opinion consultations in surgical pathology with a 95% degree of confidence. Study scope limitations were unable to separate interobserver variability from intermodality variability, but subjective opinions from the pathologists involved in the adjudication process suggested that the variability observed appeared to be predominantly interobserver. None of the cases involved in the clinical trial had an image inadequate for diagnosis that led to a major discrepancy. The investigators were specifically trying to determine whether image quality led to misdiagnoses; the closest result to that was the gastric biopsy in which the WSI pathologist requested the *H pylori* stain, but it was unavailable, leading to a minor discrepancy. Again, third-party adjudication specified that, in most instances, a pathologist would have ordered that stain on traditional microscopic examination, so that minor discrepancy was felt to be artificial in nature. The clinical trial sample's adjusted

major discrepancy rate for WSI was 2.7% (2 of 74), whereas the adjusted major discrepancy rate of GS was 4.1% (3 of 74). The authors do not believe that difference is indicative of the superiority of WSI, but rather that the true population's mean adjusted discrepancy rates were so close that random chance provided the seemingly superior results for WSI. Other possible explanations for that result include changes related to practice behavior, such as spending more time on the WSI, novelty-induced interest, or, potentially, viewing cases at higher magnifications.

Studies of nonconsult-grade cases among multiple subspecialties,[7–9,13,14] and consult-grade cases among a single subspecialty[3,15,16] have been published previously. The only other study[12] that addresses consult-grade cases of multiple subspecialties, to our knowledge, showed a 91% agreement rate (48 of 53) between WSI and GS interpretation for a population of cases selected for difficulty from an outside institution. That study had not emphasized training, case information presentation, issues of workflow, or the role of the imaging technician. In the present study, we noted a significant improvement in accuracy, which we believed to be the result of improved training and experience when the efforts changed from a feasibility study toward the development of an actual clinical service. The adjusted WSI major discrepancy rate in our study of 2.7% (2 of 74) by case is similar to the Bauer et al[8] results of 1.65% (5 of 303) and to the Campbell et al[13] results of 1.5% (3 of 212) for primary diagnosis. The results of these studies and ours are notably better than the average agreement rates found in the Pantanowitz et al[17] statistical meta-analysis of the publications to date, given in the College of American Pathologists guidelines for WSI validation, which showed WSI to be 3% (89% versus 92%) less accurate than GS examination, which is noteworthy because the clinical trial presented here was on intrinsically more difficult, consultation cases. There are a number of potential explanations for these differences. First, the number of minor discrepancies was fairly high in the current study (24.3%; 18 of 74), representing a larger pool of minor interobserver disagreements among pathologists on challenging cases. Second, our experiences suggest that pathologists need a certain amount of practice training with WSI diagnosis. Because this trial was preceded by an elaborate training protocol, the WSI arm may have had fewer errors. Although a pathologist can be shown how to use a WSI viewer in a matter of minutes, it may require significant practice to adjust one's case-viewing routines and to develop confidence in using this method of slide review. Pantanowitz et al[17] described similar results in the meta-analysis, noting that studies that mentioned any sort of training procedure averaged 95% accuracy with WSI as opposed to a 79% average for studies that did not mention training. Third, the high accuracy rate may be due to technical issues, such as high-quality scans with routine image-technician quality control. The scan resolutions used in our study were notably higher than industry expectations for ×20 objective magnifications (0.32 μm/pixel versus industry expectations of 0.46–0.50 μm/pixel[18]), hence, yielding image quality closer to ×40 (0.23–0.25 μm/pixel[18]) original magnification. Routine quality control for each image at the time of scanning produced excellent results with no images rejected by pathologists as inadequate for interpretation. We must also note the technician's rescan rate changed from 13.5% (48 of 357) in the training set to 5.4% (4 of 74) in the clinical trial, indicating a significant learning curve for the technician as well. The authors concluded that having scans that are well controlled for individual quality led to greater confidence in diagnosis. The interactive nature of the training methods used in this series of studies may have assisted as well. The imaging technician sat with the pathologists during training, allowing the technician to teach the features of the system to the pathologists during their initial use of the system. This interaction also enabled the technician to achieve a better understanding of pathology practice and the individual methodologic variations among pathologists as they approach their reviews, which allowed for improvements in procedure development and workflow assessment. A more highly trained technician is, therefore, able to assist with service improvements that can potentially lead to improved case accuracy. Similar quality-control processes have been described in WSI operations at the University of Pittsburgh Medical Center[19] and in 2 facilities in Sweden.[10] Furthermore, the College of American Pathologists' guidelines recommend this assessment as part of the validation process[17]; our recommendation is to include this assessment as part of the clinical scanning workflow as well. Finally, the quality of WSIs is known to be linked to the quality of the GS preparation. The source institution was known to produce optimally prepared GS preparations. Less-optimal slides may have contributed to the lower apparent accuracy rates for WSI reported by others.

We may also note that a large portion of the cases in the second protocol were dermatopathology cases. Our samples' adjusted major discrepancy rates for WSI and GS in the second protocol were both 1.9% (1 of 52 cases); there was no difference in sampling accuracy between the 2 methods. Massone et al[20] reported, in 2007, significant problems with a WSI system for inflammatory cases; they noted that further system and technology development and additional training might improve accuracy. In 2012, Al Habeeb et al[15] reported concordance rates of 96% (76 of 79) and 100% (12 of 12) in a multiple-armed study, and they concluded modern WSI systems were adequate for diagnosis of challenging dermatopathology cases; our results agree with those findings. The feedback from the participating dermatopathologists in our study was entirely positive regarding image quality; their only concern was with speed of diagnosis in a subspecialty known for large workloads.

Although there was no attempt at measuring time of diagnosis for the 2 methods, subjective feedback suggests that the time to interpret WSI cases was greater than it was for GS. With improved computer hardware, network speeds, improved user interfaces, and pathologist experience, the authors believe that there will be substantial improvements in overall interpretation speed for WSI in the future. For instance, regarding the user interface, Yagi et al[21] showed that a game controller could provide a smoother and more ergonomic interface, which could substantially improve the user experience. Velez et al[22] showed that differences in the time of pathologist interpretation between optical and digital analysis decreased at greater complexities.

Although the results of these studies suggest that the technical and clinical aspects of WSI interpretation can be sufficient for clinical use with second-opinion consultation cases, one significant technical limitation remains. In this study, the consultation slides were scanned at the receiving institution, not at the referring institution as would be done in a real-world situation. Therefore, there are still potential issues related to interinstitutional image-file transmissions that have yet to be studied, such as network speeds among

institutions, and other issues associated with interinstitutional information-technology collaboration. In addition, in the intended clinical workflow, laboratory staff at the community hospital would scan the slides; success of that endeavor would also be dependent on the adequate training of, and implementation by, those technicians. The next major step to show true clinical viability is for the outside institution to scan the cases and send them digitally to the receiving institution. Such a study is currently underway using the same second-opinion consultation practice protocol as was used in the present study.

The results of this study showed that excellent accuracy can be achieved for WSI second-opinion consultations if issues of image quality, training, and user interfaces are satisfactorily addressed. Second-opinion consultation is considered an ideal candidate for the adoption of digital methods because rapid turnaround times can be achieved, and synchronous viewing between consultant and consulter has the potential for maximizing the educational interaction, which should ultimately lead to improved pathology interpretation and, hence, better patient outcomes.

Nicholas C. Jones and Rosalynn M. Nazarian, MD contributed equally to this work, and both must be considered first authors. Statistical analysis was done per standard formulas (McNemar Test) and was reviewed by Jason Baron, MD, Department of Pathology, Massachusetts General Hospital.

References

1. 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*. 134(7): 1020–1023.

2. Evans AJ, Chetty R, Clarke BA, et al. Primary frozen section diagnosis by robotic microscopy and virtual slide telepathology: the University Health Network experience. *Hum Pathol*. 2009;40(8):1070–1081.

3. Slodkowska J, Pankowski J, Siemiatkowska K, Chyczewski L. Use of the virtual slide and the dynamic real-time telepathology systems for a consultation and the frozen section intra-operative diagnosis in the thoracic/pulmonary pathology. *Folia Histochem Cytobiol*. 2009;47(4):679–684.

4. Beck AH, Sangoi AR, Leung S, et al. Systematic analysis of breast cancer morphology uncovers stromal features associated with survival. *Sci Transl Med*. 2011;3:108–113.

5. Gavrielides MA, Gallas BD, Lenz P, Badano A, Hewitt SM. Observer variability in the interpretation of HER2/neu immunohistochemical expression with unaided and computer-aided digital microscopy. *Arch Pathol Lab Med*. 2011;135(2):233–242.

6. Ho J, Parwani AV, Jukic DM, Yagi Y, Anthony L, Gilbertson JR. Use of whole slide imaging in surgical pathology quality assurance: design and pilot validation studies. *Hum Pathol*. 2006;37(3):322–331.

7. Graham AR, Bhattacharyya AK, Scott KM, et al. Virtual slide telepathology for an academic teaching hospital surgical pathology quality assurance program. *Hum Pathol*. 2009;40(8):1129–1136.

8. 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(4):518–524.

9. Al-Janabi S, Huisman A, Nap M, Clarjis R, van Diest PJ. Whole slide images as a platform for initial diagnostics in histopathology in a medium-sized routine laboratory. *J Clin Pathol*. 2012;65(12):1107–1111.

10. Thorstenson S, Molin J, Lundström C. Implementation of large-scale routine diagnostics using whole slide imaging in Sweden: digital pathology experiences 2006–2013. *J Pathol Inform*. 2014;5:14. doi:10.4103/2153–3539.129452.

11. Hipp J, Cheng J, Hanscon JC, et al. SIVQ-aided laser capture microdissection: a tool for high-throughput expression profiling. *J Pathol Inform*. 2011;2: 19. doi:10.4103/2153–3539.78500.

12. 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(12):1949–1953.

13. 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(10):1739–1744.

14. Gilbertson JR, Ho J, Anthony L, et al. Primary histologic diagnosis using automated whole slide imaging: a validation study. *BMC Clin Pathol*. 2006; 27;6: 4. doi:10.1186/1472-6890-6-4.

15. Al Habeeb A, Evans A, Ghazarian D. Virtual microscopy using whole-slide imaging as an enabler for teledermatopathology: a paired consultant validation study. *J Pathol Inform*. 2012;3:2.

16. 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(1):68–74. doi:10.4103/2153–3539.93399.

17. Pantanowitz L, Sinard JH, Henricks WH, et al; College of American Pathologists Pathology and Laboratory Quality Center. 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(12):1710–1722.

18. Weinstein RS, Graham AR, Richer LC, et al. Overview of telepathology, virtual microscopy, and whole slide imaging: prospects for the future. *Hum Pathol*. 2009;40(8):1057–1069.

19. Pantanowitz L, Wiley CA, Demetris A, et al. Experience with multimodality telepathology at the University of Pittsburgh Medical Center. *J Pathol Inform*. 2012;3:45. doi:10.4103/2153–3539.104907.

20. Massone C, Soyer HP, Lozzi GP, et al. Feasibility and diagnostic agreement in teledermatopathology using a virtual slide system. *Hum Pathol*. 2007;38(4): 546–554.

21. Yagi Y, Yoshioka S, Kyusojn H, et al. An ultra-high speed whole slide image viewing system. *Stud Health Technol Inform*. 2012;179:239–249.

22. Velez BS, Jukic D, Ho J. Evaluation of 2 whole-slide imaging applications in dermatopathology. *Hum Pathol*. 2008;39(9):1341–1349.