



## Use of whole slide imaging in surgical pathology quality assurance: design and pilot validation studies<sup>☆,☆☆</sup>

Jonhan Ho MD<sup>a</sup>, Anil V. Parwani MD, PhD<sup>b</sup>, Drazen M. Jukic MD, PhD<sup>b,c</sup>,  
Yukako Yagi BS<sup>a</sup>, Leslie Anthony MA<sup>d</sup>, John R. Gilbertson MD<sup>a,e,\*</sup>

<sup>a</sup>Center for Pathology Informatics, Department of Pathology, University of Pittsburgh Medical Center, Pittsburgh, PA 15232, USA

<sup>b</sup>Department of Pathology, University of Pittsburgh School of Medicine, UPMC Presbyterian/Shadyside, Pittsburgh, PA 15232, USA

<sup>c</sup>Department of Dermatology, University of Pittsburgh School of Medicine, UPMC Presbyterian/Shadyside, Pittsburgh, PA 15232, USA

<sup>d</sup>Center for Biomedical Informatics, University of Pittsburgh, Pittsburgh, PA 15261, USA

<sup>e</sup>Center for Oncology Informatics, Department of Pathology, University of Pittsburgh Medical Center, Pittsburgh, PA 15232, USA

Received 3 September 2005; revised 7 November 2005; accepted 9 November 2005

### Keywords:

Whole slide imaging;  
Quality assurance;  
Informatics;  
Imaging

**Summary** By imaging large numbers of slides automatically at high resolution, modern automated whole slide imaging (WSI) systems have the potential to become useful tools in pathology practice. This article describes a pilot validation study for use of automated high-speed WSI systems for surgical pathology quality assurance (QA). This was a retrospective comparative study in which 24 full genitourinary cases (including 47 surgical parts and 391 slides) were independently reviewed with traditional microscopy and whole slide digital images. Approximately half the cases had neoplasia in the diagnostic line. At the end of the study, diagnostic discrepancies were evaluated by a pathology consensus committee. The study pathologists felt that the traditional and WSI methods were comparable for case review. They reported no difference in perceived case complexity or diagnostic confidence between the methods. There were 4 clinically insignificant discrepancies with the signed-out cases: 2 from glass slide and 2 with WSI review. Of the 2 discrepancies reported by the WSI method, the committee agreed with the reviewer once and the original report once. At the end of the study, the participants agreed that automated WSI is a viable potential modality for surgical pathology QA, especially in multifacility health systems that would like to establish interfacility QA. The participants felt that major issues limiting the implementation of WSI-based QA did not involve image acquisition or

<sup>☆</sup> This work was supported by funding from the US Air Force administered by the US Army Medical Research Acquisition Activity, Fort Detrick, Md, Contract No. DAMD17-03-2-0017. The content of the information does not necessarily reflect the position or policy of the US government, and no official endorsement should be inferred.

<sup>☆☆</sup> The University of Pittsburgh Investigational Review Board reviewed this research, No. 0412088, and approved it as exempt from Federal Policy for the Protection of Human Research Subjects.

\* Corresponding author. Centers for Pathology and Oncology Informatics, University of Pittsburgh Medical Center, Pittsburgh, PA 15232, USA.

E-mail address: gilbertsonjr@upmc.edu (J.R. Gilbertson).

quality but rather image management issues such as the pathologist's interface, the hospital's network, and integration with the laboratory information system.  
© 2006 Published by Elsevier Inc.

## 1. Introduction

### 1.1. Quality analysis in surgical pathology

Quality assurance (QA) is an important part of surgical pathology practice and is mandated by the College of American Pathologists (CAP) in their accreditation of laboratories. Although the CAP publishes a series of checklists to guide surgical pathology QA, the specific implementation is intentionally left flexible to accommodate the wide variety of pathology practices that exist across the country. For example, although the CAP checklists require second review of surgical pathology reports in certain circumstances, the method and implementation of this QA are not specified. Typical approaches include (1) intradepartmental review, (2) second opinion requested by patient or clinician, (3) interdepartmental review, (4) extradepartmental review, and (5) cases reviewed (over-read) by a second pathologist as part of departmental QA policies [1].

Although many approaches to surgical pathology case review are possible, case over-reads, in which a second pathologist reviews cases as part of a departmental QA policy, represent the overwhelming majority of surgical pathology QA in general practice. In a typical practice, a percentage of signed-out cases are picked at random and distributed to a second pathologist for review. Discrepancies are noted and resolved between the pathologists, and if necessary, amended reports or addendums are issued. These QA programs can be quite extensive because as many practices review 10% (or more) of cases.

There is good reason for extensive QA programs in pathology. A wide range of anatomic pathology error rates is reported in the literature [1], ranging from 1% to 43% of all received specimens. This broad distribution reflects varied methods used to detect errors as well as differences in error qualification. Raab et al [1] have estimated that the actual error rate likely ranges from 1% to 5%, and in a recent study of self-reported discrepancies among 72 institutions, 6.7% of anatomic pathology diagnoses were found to be discrepant at second review. Significantly, in 1% of these cases (0.067% of all cases), a significant clinical event occurred as a result of these discrepancies.

Statistics based on second review (such as the previous data) likely reflect an underrepresentation of errors because as traditional case over-read is associated with a number of potential biases. For example, the reviewer often knows the original diagnosis and/or the identity of the sign-out pathologist. It has been shown that knowledge of the original diagnosis affects the sensitivity of review in cytopathologic specimens [2], and although not yet studied, it is likely that second review of surgical pathology specimens is similarly

affected. Furthermore, reviews are usually done at the same facility at which the case was signed out. This could hide or even enhance local biases and could be a significant problem, especially for large, multiple facility health systems that would like to establish a uniform level of quality across their enterprise. A major hindrance to establishing a multi-facility QA program is the expense and difficulty of moving and managing slides between facilities, especially if the QA is to be done close to the sign-out date.

Automated whole slide imaging (WSI), in which all the slides in a case are imaged in their entirety at high resolution and made available to pathologists on a network, is a modality that may prove useful in surgical pathology case review. A digitized case could allow a QA system to hide the original sign-out pathologist and, if desired, original diagnosis. More importantly, however, digital slides available on a network can mitigate the problems of glass slide logistics and, by so doing, enable routine multifacility surgical pathology QA.

In this article, we report the results of a pilot study in which surgical pathology case review based on digital slides created by an automated high-speed WSI robot is compared with traditional surgical pathology review based on direct microscopic examination of the glass slides.

### 1.2. Automated high-speed WSI

Early attempts to incorporate digital imaging in surgical pathology practice centered on microscope-mounted cameras. These systems became popular in the early 1990s with the availability of moderately priced video/digital cameras, reasonably sized hard drives (tens to hundreds of megabytes), local area networks, and the Internet. For example, in 1994, Schubert et al [3] presented a "pathologist designed imaging system for anatomic pathology sign out, teaching and research." The system, designed for permanent storage of static images and their incorporation in sign out and reporting, involved multiple cameras (640 × 480 pixels with 24-bit color), network connectivity to an image server and the pathologists' workstations, a RAID 5 (Dell, Inc, Round Rock, Tex) storage device and database, image capture, and image display software. The system was later integrated into the CoPath Laboratory Information System (LIS) (Cerner DHT, Inc, Waltham, Mass) [4]. Similar systems [5] are still widely used today for gross image management and for microscopic single-field documentation. However, their use is limited; at our institution, less than 1% of cases get a microscopic image. There are numerous reasons for this [6], but perhaps, the most important is that single-field, camera-on-microscope systems do not document the entire slide, forcing the

pathologist to find and capture limited fields of interest. This causes the pathologist to become a photographer, takes significant time, and results in images that subsample the case and do not necessarily stand on their own.

Another approach to imaging in pathology was robotic microscopy; these systems have been reasonably successful in telepathology with tens of thousands of successful cases documented since the mid-1990s [7-9]. However, these are largely real time, remote control systems that require hands on involvement of the pathologist and do not provide (as part of the core operations) permanent image storage.

In the late 1990s, pathologists began to experiment with systems that imaged and permanently stored the entire slide (or parts of the slide) at reasonably high resolutions. In 1997, Joel Saltz and his group (Ferreira et al [10]), presented a system for “Enhanced Field Microscopy” in which a robotic microscope captured a large area of a slide, field by field, and a computer then “knitted” the individual fields together into a montage. The development was a significant advancement and was recognized as such by awards at the 1997 American Medical Informatics Association Fall Symposium. The system did have significant limitations, however, most significant was the long time (often many hours) required to capture a single extended field. That said, “slow” WSI based on traditional robotic microscopes are still used today for education and proficiency testing [11,12].

Automated high-speed WSI began in 1999 when Art Wetzel and John Gilbertson, then of Interscope Technologies (Pittsburgh, Pa), developed a fully automated, high-speed device that can image entire slides at high resolution and at a reasonable cost. In late 1999, a prototype with a fully functional robot was developed at the University of Pittsburgh Medical Center (UPMC), and in March 2000, it was presented at the International Academy of Pathology Meeting in Nagoya. It was based on traditional microscope optics, a strobe light linked to a precision stage, and a digital video camera. With a primary magnification of  $\times 20$ , a numerical aperture of 0.7, and square  $6.6\text{-}\mu\text{m}$  pixels, it had a spatial sampling period (pixel size/optical magnification, a measure of resolution) of  $0.33\ \mu\text{m}/\text{pixel}$  and can image a slide in 5 to 10 minutes depending on the size of the tissue section and the amount of image compression desired [6].

Since 2000, there has been a small explosion of companies producing increasingly capable automated high-speed whole slide imagers. A typical imaging robot today can run in batch mode (reading barcodes on slides) and can capture and compress an image of a slide with a tissue section measuring  $1.5 \times 1.5\ \text{cm}$  in approximately 6 minutes with spatial sampling periods of between 0.3 to  $0.5\ \mu\text{m}/\text{pixel}$ . The high-speed WSI robot industry is becoming highly diverse with a wide range of optics, detectors, slide handling devices, and software, resulting in an increasing range of capabilities and costs. Newer devices are implementing nontraditional optics, illumination, and sensors designed specifically for very high speed image capture [13] and should result in significant improvements

in speed, throughput, and resolution in the months and years ahead, with different manufacturers eventually focusing on different aspects of the market.

The rise of automated, high-speed, high-resolution WSI should have a significant impact on pathology practice because, for the first time, pathologists have a device that can digitize large numbers of slides automatically. A digital slide could allow pathologists to better apply computational power and network connectivity to pathology practice, resulting in new capabilities and greater productivity similar or greater than those seen with the rise of the LIS and digitalization of the pathology report. However, although there are several published studies describing the use of WSI in education and research [14-16], there are only limited papers describing automated high-speed WSI in clinical practice and only a handful describing a direct comparison of automated WSI with traditional microscope examination in clinical activities [13]. This is particularly disappointing in that there are some clinical activities in which the advantages of a digital slide (eg, the ability to access the image across a wide network) may make WSI useful even if the image quality is not quite as good as direct examination of a slide under the microscope.

As mentioned previously, anatomic pathology QA offers such a potential niche. Anatomic pathology QA is an important clinical activity, and most institutions have policies that mandate some percentage of cases be reviewed (over-read) by a second pathologist. If discrepancies are noted, the case is discussed, and if necessary, the report is supplemented with an amendment or an addendum. Quality assurance is normally operated as an intrafacility activity even in institutions with multiple facilities such as UPMC or the US Air Force largely because of difficulties in glass slide logistics. A strong argument could be made that some amount of interfacility over-reads would be desirable and may help standardize pathology practice and reporting across large enterprises. Furthermore, it is reasonable to postulate that digital representations of the pathology cases, created by automated systems and readily available over networks, may be the tool necessary to make interfacility over-reads a practical reality. Finally, anatomic pathology QA is a mandated real-world clinical activity that has a set of protocols and outcomes that can act as a realistic background for the evaluation of automated WSI systems and the clinical space.

In this article, we report the results of a pilot study in which surgical pathology case review based on digital slides created by an automated high-speed WSI robot is compared with traditional surgical pathology review based on direct microscopic examination of the glass slides.

## 2. Materials and methods

### 2.1. Study design and participants

This was a retrospective comparative study evaluating the use of automated high-speed WSI in a routine clinical

activity, surgical pathology QA. The study took place in the Division of Anatomic Pathology at UPMC, Shadyside Hospital, a 486-bed tertiary care hospital. The division has 9 board certified, faculty (sign-out) pathologists with between 2 and 30 years of practice experience (mean, 12.8 years). Faculty offices and sign-out rooms are equipped with Olympus BX45 microscopes (Olympus, Melville, NY). The objective lens were Plan Achromat (magnifications range from  $\times 2$  to  $\times 100$ ), and the scopes are professionally calibrated twice a year.

Three anatomic pathologists with some prior experience reading digital slides and a special interest in this research volunteered participation as study pathologists. These included a board certified pathologist/dermatopathologist with 5 years of practice experience, a board certified pathologist with 2 years of practice experience (both had additional training in GU pathology), and a pathology fellow in his fifth year of postgraduate training.

In addition to the study pathologists, a team of individuals with diverse and distinct backgrounds worked on this study, including

- a project manager and a principal faculty investigator, responsible for the management, integration, and overall execution of the project;
- evaluators with the University of Pittsburgh Center for Biomedical Informatics, responsible for institutional review board approvals, data management, and focus group interviews;
- the Quality Assurance Division of the Department of Pathology, responsible for case selection and the formal resolution of any diagnostic discrepancies discovered during the study;
- an honest broker, responsible for deidentifying case material; and
- an imaging team, responsible for whole slide image capture, storage, and presentation.

## 2.2. Procedures and apparatus

The study was designed to run as an extension of the standard QA process existing for the GU bench at a tertiary care hospital at UPMC. Case selection was performed by the Department of Pathology Quality Assurance Division using standard operating procedures and was not influenced by study personnel. Cases were selected randomly using a computer program in the LIS. Once QA personnel assured that study pathologists were not previously associated with selected cases, they printed pathology reports, retrieved case slides from archives, and delivered case packets (reports and slides) to the honest broker.

The honest broker deidentified the cases by eliminating all patient, pathologist, accession, and clinician identifiers (as well any other Health Insurance Portability and Accountability Act safe harbor data) from the reports. Study numbers were assigned to each case, and new slide labels

with bar codes (for use in imaging), study numbers, block and slide numbers, and stain information were placed over the original labels. Slides were sent for imaging (*vide infra*), and then the deidentified reports and slides were sent to the project manager.

The project manager randomly assigned each case to 2 of the 3 study pathologists; one pathologist received the glass slides, and the other received the whole slide images. Case assignments were not shared with study pathologists. For each case, study pathologists received the same deidentified clinical report, along with a standard QA form and an evaluation form. Data collection included typical QA data points, levels of diagnostic concordance, diagnostic confidence, case complexity, time to complete, and system performance metrics. This generated 3 data sets for each case: (1) the original signed-out pathology report diagnosis, (2) the QA results on the basis of glass slide review, and (3) the QA results on the basis of whole slide image review.

Once a pathologist completed a case, QA and evaluation forms were delivered to the evaluation team and slides were returned to the project manager. Pathologists did not discuss their cases during the study, and the evaluation team held the QA and evaluation data in escrow until the consensus meeting at the end of the project (*vide infra*).

The study was performed on 24 pathology cases representing 47 diagnostic parts and 391 slides in a 4-week period (see [Table 1](#)). Of these cases, 12 had a signed-out diagnosis of cancer or in situ neoplasia. Before the study began, there was a prestudy dry run on 6 cases to ensure the functionality of the workflow and to make sure that the technology was working properly and that data would be handled appropriately.

After the pathologists completed all their cases, the entire team (project management, evaluation, and imaging teams and the study pathologists) met as a group and discussed each case, focusing on those in which there had been disagreements. Disagreements were possible between (1) the original report and the glass slide over-read, (2) the original report and the imaging over-read, or (3) the glass slide over-read, the imaging over-read, and the original report (although

**Table 1** Type and complexity of cases

Case type	No. of cases	No. of parts	No. of slides
Prostate biopsy	9	29	214
Bladder biopsy	5	6	21
TURP	3	3	23
Radical prostatectomy	2	4	98
Nephrectomy	2	2	24
Vasectomy	2	2	9
Excision of a scrotal lesion	1	1	2
Total	24	47	391

NOTE. Cases were selected randomly through a QA program in the LIS. Half of the cases (12) had a signed-out diagnosis of neoplasia. Abbreviation: TURP, transurethral resection of the prostate.

the latter did not occur). Discussion included detailed examination of the original pathology reports, the digital slides, and the glass slides under a multiheaded microscope. Study pathologists reached consensus on all discrepant cases. As a group, the pathologists rated all disagreements as mild, moderate, or severe, and clinically irrelevant or relevant using what they felt represented normal surgical pathology QA practice at our institution. All results were reported to the Quality Assurance Division, and cases in which there were differences between the consensus of the study pathologists and the original report were resolved through standard QA procedures.

### 2.3. Poststudy focus group

At the conclusion of the study, the evaluation team presented preliminary findings and conducted a focus group interview with study pathologists and the principal investigator (also a pathologist). Following a round-robin format, participants were asked to contribute their thoughts on the key take-away lessons from this project. Responses were audio-recorded and transcribed. The evaluation team assessed the transcripts summarizing individual statements and categorized statements into emerging themes.

### 2.4. Whole slide imaging

Automated whole slide image capture was performed on an Aperio T2 slide scanner (Aperio Technologies, Vista, Calif) outfitted with a Nikon Plan Fluor  $\times 20$ , 0.7 numerical aperture objective lens (Nikon Instruments, Melville, NY), and Basler L301 “trilinear-array” line scan camera (Basler Vision Components, Ahrensburg, Germany). The system’s spatial sampling period (the area of tissue section subtended by a pixel) was approximately  $0.46 \mu\text{m}/\text{pixel}$ . The system ran in automated batch mode with automated focus and tissue finding. Images were compressed during the capture process in a multilayered JPG2000 format using a Matrox Morphis compression board (Matrox Incorporated, Montreal, Quebec, Canada) with a quality setting of 30 (high), resulting in file sizes ranging from 10 to 300 MB. The image server (Dell) was equipped with dual Xeon processors (Intel Corporation, Santa Clara, Calif), 2 GB of RAM, and 1 TB of hard drive storage running Microsoft Windows Server 2000 (Microsoft Corporation, Redmond, Wash), Internet Information Services, and Aperio Image Server software version 5.6.

After automated image capture, the imaging technician visually examined whole slide images for general image quality. On the basis of this evaluation, 2% of slides (9/391) were reimaged in manual mode.

Pathologists viewed the whole slide images through network connections on remote workstations (located at their offices or homes). Intrahospital connections were physically limited to a maximum of 10 Mbps, and off-site connections were physically limited to Internet service provider connection speeds (approximately 1.5 Mbps). In

virtually all cases, workstations and the server were on different subnetworks. Workstations (and displays) used to view the whole slide images varied greatly, ranging from previous-generation laptops to typical current-generation hospital workstations. There were 2 home-grown, Web-based whole slide image viewers available to the pathologists depending on local personal computer management and network security rules. One was based on Active-X technology, and the other was based on Zoomify (Flash) Web technology.

## 3. Results

### 3.1. Case detail

The 24 study cases are summarized in Table 1. The cases represented 47 parts and 391 whole slide images, 235 of which were from prostate needle biopsies, transurethral prostate resections, or radical prostatectomies. The remaining 66 slides were from urinary bladder biopsies, sections of vas deferens, 2 partial nephrectomies, and 1 scrotal lesion. Hematoxylin-eosin–stained sections (336), hematoxylin-eosin recuts/levels (12), immunohistochemical studies (29), frozen sections (2), and touch preparations (2) were represented. The immunohistochemical stains included primarily racemase (9), p63 (8), and high-molecular weight keratin (4). The number of slides associated with each case ranged from 2 to 55, with a mean of 15.9 and a median value of 10 slides per case. Half of the cases (12) had a signed-out diagnosis of cancer or in situ neoplasia in at least 1 part. Cases were examined in 4 contiguous 1-week periods, 6 glass slide reviews, and 6 WSI reviews per period, with each pathologist performing 4 reviews per period. One hundred forty slides/images were reviewed in the first week (23.3 per case), 106 slides/images in the second week (17.7 per case), 110 slides/images in the third week (18.3 per case), and 25 slides/images in the fourth week (4.2 per case).

### 3.2. Comparative assessments

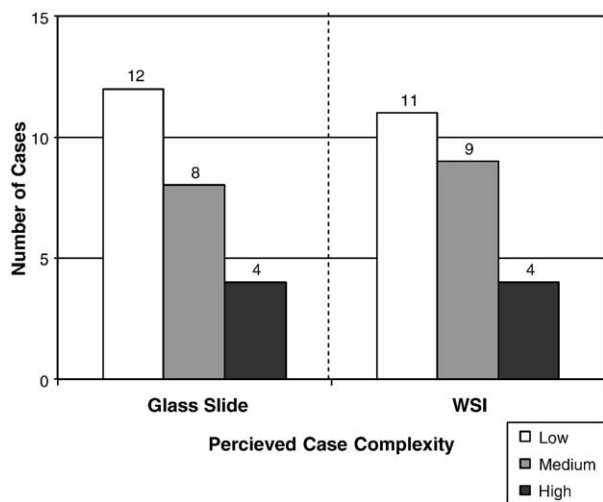
#### 3.2.1. Diagnostic concordance

There was good concordance between the original report and both the glass and WSI reviews. Over the 24 cases in the study, the glass QA assessment (review) resulted in 22 cases of agreement (with the signed-out report), 1 case of moderate, clinically insignificant disagreement (case 16), and 1 case of mild, clinically insignificant disagreement (case 21). With WSI reviews, there were 21 cases of complete agreement (with the original report) and 3 cases of mild, clinically irrelevant disagreements (cases 10, 19, and 29).

All 5 cases that generated a disagreement are discussed in detail hereinafter.

#### 3.2.2. Diagnostic confidence

The pathologists expressed high confidence in both the glass and WSI over-reads. Of the 24 glass over-reads, 23 were



**Fig. 1** Distribution of perceived case complexity. Reported case complexity was virtually identical for microscopic glass slide and whole slide image interpretation, indicating that WSI did not cause cases to be perceived as more complex.

reported as high confidence, whereas 1 (case 27) was reported as medium confidence. Of the 24 WSI over-reads, 21 were reported as high confidence and 3 (cases 10, 27, and 28) were reported as medium confidence. Case 10 was a prostate needle biopsy, and case 28 was a vasectomy. Case 27, a complex bladder biopsy, was reported as medium confidence in both modalities.

### 3.2.3. Case complexity

The overall distribution of perceived case complexity (as reported by the pathologists) was virtually identical between glass and WSI examination. Average reported case complexity was 1.79 for glass examination and 1.71 for WSI examination (1, low; 2, medium; 3, high), indicating that the WSI modality did not cause the cases to be perceived as more complex (Fig. 1). However, there was significant variation in perceived complexity between pathologists at the level of the individual case. More than half of the cases (13/24) had discrepant perceived complexity between glass and WSI examination, and 2 of them differed by 2 degrees (high-low). This was possibly due to variations in the training and experience of the study pathologists.

### 3.2.4. Time to case completion

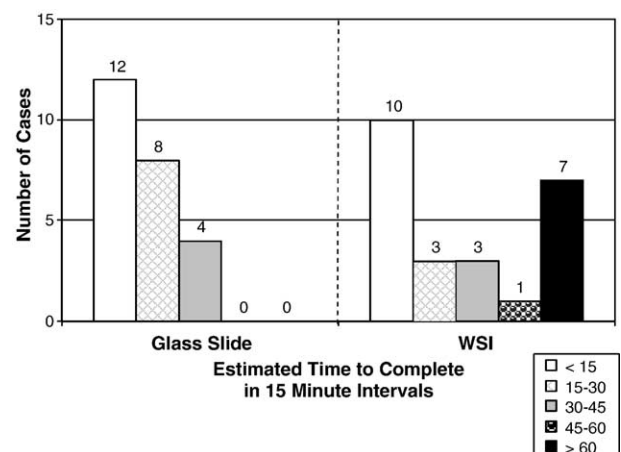
Whole slide imaging over-reads, on average, took longer to complete than glass over-reads (Fig. 2). The main discrepancy was a set of 8 cases that took more than 45 minutes to evaluate with the imaging system. Significantly, however, 7 of the long WSI over-reads were among the first 10 cases in the study, possibly indicating a learning curve. Furthermore, there were also a series of network problems in the first week of the study that the pathologists reported impacted both time to complete and perceived image quality (vide infra). This is discussed in detail in the Discussion section.

### 3.2.5. Discussion of discrepant cases

All microscope glass slide and whole slide image reviews reported as discrepant (cases 10, 16, 19, 21, and 29) were reviewed again by the 3 participating pathologists as a group at the conclusion of the study. The review was conducted in a conference room equipped with a multiheaded microscope as well as an image projector, and a consensus was achieved with each discrepant case. A summary of discrepant findings is presented in Table 2. There were no severe disagreements, and after review, 3 cases were adjusted to full agreement with the original diagnosis, 1 case remained at moderate, clinically insignificant disagreement, and 1 case remained at mild, clinically irrelevant disagreement. Importantly, the cause of initial disagreement in 1 case (case 29) was a spelling error in the report and had nothing to do with histologic interpretation.

Case 10 was a 2-part (right and left) prostate needle biopsy. One of the parts was signed out as prostatic adenocarcinoma with a Gleason score of  $3 + 3 = 6$ . The study pathologist doing the glass slide review agreed with that diagnosis, whereas the pathologist doing the WSI review reported a mild disagreement, favoring a Gleason score of  $3 + 4 = 7$ . Consensus review focused on 2 to 3 small glands, and all pathologists concurred that the difference was a matter of reasonable morphological interpretation and was not significant enough to alter the signed-out report. This reported discrepancy was adjusted to full agreement with the original diagnosis (Fig. 3).

Case 16 was a 2-part (right and left) prostate needle biopsy signed out as having a focus of high-grade prostatic intraepithelial neoplasia (PIN) in one of the right core biopsies and a microscopic focus of moderately differentiated prostatic adenocarcinoma (Gleason score  $3 + 3 = 6$ )



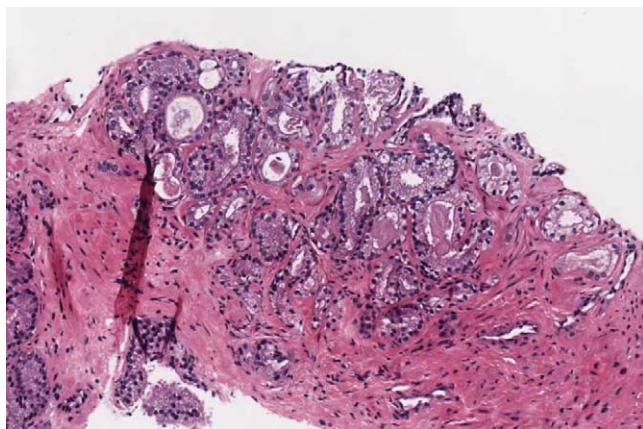
**Fig. 2** Time to case completion. Whole slide image review took significantly longer than glass slide review, with 8 WSI reviews taking more than 45 minutes. However, these longer reviews occurred early in the study. Some of this effect was likely due to (1) network problems that occurred in the first week of the study and possibly (2) a learning curve. By the end of the study, the difference in speed had decreased significantly.

**Table 2** Summary of discrepant findings

Case	Case type	Disagreement	Consensus
10	Prostate needle biopsy	WSI review: Gleason 3 + 4 = 7 Signed-out report: Gleason 3 + 3 = 6	Agreement with the signed-out report
16	Prostate needle biopsy	Glass review: foci of carcinoma Signed-out report: no foci of cancer in the core in question	Moderate, clinically insignificant disagreement with signed-out report
19	Radical prostatectomy	WSI review: high-grade PIN Signed-out report: no mention of high-grade PIN	Mild, clinically insignificant disagreement with the signed-out report
21	Prostate needle biopsy	Glass review: no evidence of atypia Signed-out report: foci of atypia	Agreement with the signed-out report
29	TURP	WSI review: spelling error in report	Spelling error present but not clinically significant

NOTE. Disagreements between WSI review, glass slide review, and signed-out diagnosis. There were 5 disagreements in the 48 reviews (24 glass slide and 24 WSI reviews) cases. Of the 5, 1 was a spelling error in the report. Of the 4 disagreements involving histologic interpretation, all were relatively minor and all were considered clinically insignificant. Two disagreements were “called” by the glass review process, and 2 were called by the WSI review process. Consensus agreed with the WSI review once and the glass slide review once.

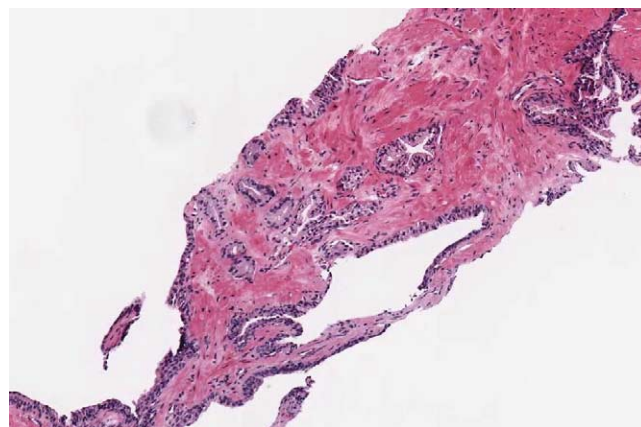
with concurrent high-grade PIN in the left core biopsies. The study pathologist doing glass slide review noticed an atypical group of glands suspicious for invasive carcinoma in one of the right sided cores that was not noticed by the pathologist reviewing the whole slide images. Upon consensus review of the glass slides, it was agreed that this focus was suspicious enough to warrant further workup and that it was most likely carcinoma. However, consensus review of the whole slide images revealed that several features of carcinoma present on the glass slides were less apparent on the whole slide image. Specifically, the “bluish cast” and the “blue-tinged” mucus in the cytoplasm often seen in malignant glands were much less pronounced on whole slide image (Fig. 4). All 3 of the study pathologists agreed that although discernible on glass slide, they would likely have missed the lesion while scanning the whole slide image. Consensus review determined this case to be of



**Fig. 3** Case 10, focus of disputed Gleason grading, area of the whole slide image, moderate magnification. The image shows a small focus of prostatic adenocarcinoma. The area in question is in the upper left where several glands appear to fuse, creating a small area that the pathologist reviewing the case with WSI considered a focus of higher Gleason grade (grade 4), with an overall grade of Gleason score 3 + 4 = 7.

moderate, clinically insignificant disagreement with the original diagnosis and recommended that this case be returned to the original sign-out pathologist for further evaluation and workup. Because prostatic adenocarcinoma had already been reported in the left core biopsies, the pathologists felt that patient management would not be changed even if the added diagnosis of concurrent right-sided disease were confirmed.

Case 19 was a radical prostatectomy specimen signed out with a diagnosis of prostatic adenocarcinoma (Gleason score 3 + 4 = 7). The diagnostic line did not mention the presence of high-grade PIN. The pathologist doing the whole slide image review reported a mild disagreement, noting that high-grade PIN was present in the specimen, whereas the glass slide review did not report on the presence or absence of high-grade PIN. Consensus review found abundant evidence of



**Fig. 4** Case 16, area of whole slide image, moderate magnification. The area in question is in the center of the field. These 5 to 6 small glands with an infiltrative growth pattern are suspicious for invasive adenocarcinoma under the microscope but appear less worrisome on the whole slide image. The original signed-out report mentioned high-grade PIN, but not invasive cancer, in this part of the case. This case had established areas of adenocarcinoma on additional cores.

high-grade PIN, and the case was determined to be a mild, clinically insignificant disagreement with the original diagnosis. It was felt that the presence or absence of PIN in the final report in this case would not have changed patient management. However, the consensus recommendation was to return the case to the original sign-out pathologist for further review.

Case 21 was a set of prostatic needle core biopsies (right and left) with one of the cores originally signed out as “atypical small acinar proliferation, suspicious for carcinoma.” The study pathologist doing glass slide review on the case did not identify atypical areas and reported a mild disagreement, whereas the whole slide image review pathologist saw a suspicious area and reported an agreement. Consensus review examined a single focus of prostatic glands with cytologic and architectural atypia suspicious for adenocarcinoma. Corresponding areas at different levels were also examined closely, and after this examination, the study pathologists were in agreement with the original diagnosis.

Case 29 was a urinary bladder biopsy originally labeled as “prostate, turansarethral [sic] resection.” On glass slide review, the study pathologist reported agreement, whereas the whole slide image pathologist reported a mild disagreement citing the spelling error of the word *transurethral* in the original report (turansarethral). The consensus was not to report a disagreement because the meaning of the misspelled word was clear in the context of the report.

### 3.2.6. Poststudy focus group

At the end of the study, the evaluation team convened a round table discussion with the study pathologists (vide supra, Methods). Several themes emerged from the focus group, including image quality, presentation and navigation speed, implications of archived digital images, and study workflow.

Each of the pathologists stated that image quality was very high, and there was uniform agreement that the images were of diagnostic quality and that subtle features can generally be picked up. Case 16 was singled out as an exception that underscores the need for further evaluation, calibration, and standardization. Significantly, several pathologists felt that case 16 would be difficult with anything but high-end microscopes with top of the line optics.

All pathologists agreed that slide presentation speed and navigation were considered the greatest limitation of the technology. Navigation through a digital slide was slower than navigation on a microscope. Furthermore, there were significant selection and presentation delays between slides. Pathologists wanted a workflow that more resembled the microscope. They desired faster image retrieval and the ability to quickly switch between both objectives and whole slide images. Overall, they felt that current display and navigation speed would be a significant limitation to routine case reviews based on WSI. The pathologists also had

numerous suggestions of how to improve the layout and functionality of the user interface.

The pathologists expressed enthusiasm about the advantages of digitally archiving slides, potentially easing the process currently required to retrieve glass slides from an offsite repository. An image repository can also serve several different purposes, including consultation, management of prior cases, distribution of special stains and recuts across the enterprise, and resident education. Importantly, it was felt that the digital images could and should be managed in conjunction with LIS data, simplifying and improving departmental workflows. In addition, whole slide images would not fade with time, providing, in some ways, a more permanent record of a slide.

## 4. Discussion

This study was envisioned as a pilot study to evaluate the potential of automated high-speed WSI as a modality for surgical pathology QA. As such, the study was designed to fit entirely within the existing protocols and operations of the Department of Pathology Quality Assurance Division. As discussed in the Methods section, cases were selected by QA division protocol (not by the study managers), and the output of the study (concordant and discrepant findings) was handled within normal QA procedures. In fact, with the exception of questions on case review time, the glass slide arm of the study was virtually equivalent to normal QA operations at our institution. The WSI arm was very similar, save for evaluation questions on presentation speed, system performance, and image quality.

### 4.1. Automated high-speed WSI as a modality for surgical pathology QA

The results suggest that high-speed WSI is a promising candidate for further consideration as a modality for surgical pathology QA. Although the number of cases (24) was small, the case difficulty was significant, representing 47 individual diagnostic parts and 391 individual slides. Most of the work involved prostate and bladder biopsies, and half the cases had the diagnosis of cancer or in situ neoplasia in the signed-out diagnostic line. Glass slide review revealed 2 relatively mild, clinically insignificant discrepancies with the signed-out report, and consensus favored the glass slide-based review once and the signed-out report once. Similarly, WSI review reported 2 histology interpretation-based discrepancies, both mild, and consensus agreed once with the WSI review once and once with the signed-out report.

In 1 case, however (case 16, Fig. 4), the study pathologist doing the glass over-read found a subtle area of atypia that was missed by the pathologist doing the WSI over-read as well as by the original signed-out report. The study consensus found that the finding was real and probably represented cancer, and the case was referred back to the sign-out pathologist through standard institutional QA protocols. The

committee also concluded that, in this case, subtle but important morphological clues were hard if not impossible to pick up in the whole slide image. This is a significant finding that cannot be ignored. The fundamental problem appears to be one of color fidelity in the image (*vide supra*, Results). We are examining this situation further and are testing a mechanism to calibrate color between batches in the WSI robot. We are also examining whether it is possible and/or desirable for color calibration to be consistently applied across the entire system, including the monitors in the pathologist's office. Although case 16 indicates that image quality in current automated high-speed WSI has limitations compared with the glass slide (a finding that should not surprise anyone), it is important to stress that WSI is an evolving technology, that the morphological finding in question was subtle, and that it was also missed by the original (glass slide) examination of the sign-out pathologist. Because cancer was diagnosed on a separate core on the same case, the additional finding did not affect the treatment for the patient.

In addition to the primary finding discussed previously, there were secondary findings that also argue for automated high-speed WSI as a potential modality in surgical pathology QA. In particular, there was little or no difference in perceived case complexity or diagnostic confidence between WSI and glass examination (*vide supra*, Results). We will be examining complexity, confidence, and system performance in more detail during more extensive evaluations next year.

Importantly, at the end of the study, the pathologists expressed the belief that the main limitations of the WSI system for QA were not image capture or image quality but rather slide navigation, user interface, clinical data integration, and presentation speed. This was somewhat unexpected, and we had collected little data in this area, limiting ourselves to the time to case completion on whole image and glass slide reviews. Whole slide imaging over-reads, on average, took longer to complete than glass over-reads. All glass over-reads were completed in less than 45 minutes, whereas 8 WSI over-reads took more than 45 minutes and 7 of those took longer than an hour (Fig. 2). However, the longer WSI case completion times were seen in the early part of the study (the first 10 cases). Toward the end of the study, differences in completion times (WSI versus glass) decreased considerably and cannot be differentiated with our rough scale of 15, 30, 45, and 60 minutes (although we believe that WSI still took longer). These findings indicate the possibility of a learning curve; however, interpretation was limited by a series of network and server problems in the first weeks of the study that definitely increased case completion times during that period.

The importance that the study pathologists put on data integration implies the need for WSI-capable pathology picture archiving and communications systems similar to those currently in use in radiology. Significantly, Digital Imaging and Communications in Medicine, the main imaging standards organization for clinical medicine, has

recently approved a pathology working group (WG 26) to specifically investigate and develop standards for secure WSI-capable picture archiving and communications systems. The development of these standards is an important initial step in the development of useful image and data integration in pathology. We expect that system performance, data integration, and user interfaces will be complex and will require significant research, development, and testing in years ahead.

## 5. Conclusion

The primary goal of this study was to evaluate WSI as a potential platform for surgical pathology QA. It was hypothesized that digital slides could be implemented within an existing QA workflow and might be useful in establishing timely interfacility QA across a multifacility health system.

Although the study was small (24 cases representing 47 parts and 391 slides), the results are encouraging. In 2 cases (cases 10 and 19), the pathologists using WSI review reported 2 minor discrepancies with the signed-out (microscope based) report. In 1 case, consensus favored the WSI review, and in the other, consensus favored the original report. These results are virtually identical to the glass slide-based review of the same cases in which glass slide-based review reported 2 discrepant cases and consensus agreed with the review once and the original report once. The study pathologists expressed a high degree of diagnostic confidence in digital slides; the use of digital slides did not result in an increase in perceived case complexity, and at the end of the study, all pathologists felt that QA can be done effectively with WSI.

There is, however, reason to be cautious. In 1 case (on 1 slide), both the signed-out report and the digital over-read missed a focus in a prostate biopsy highly suspicious for cancer. The consensus of the study pathologists was that the finding would be very hard if not impossible to pick up on the digital slide, citing poor color fidelity in the image. In summary, current WSI technology is worthy of more extensive evaluation in surgical pathology QA. It provides a potential mechanism to share QA between facilities, with the caveat that image quality (and image calibration) is very good, but not yet perfect. Exactly when WSI will be accepted as a tool for QA will depend on both advances in technology and clear definition of the value of the QA task at hand. Whole slide imaging may be accepted long before it is perfect if it can provide valued services (potentially such as timely inter-health system QA or timely interhospital confirmation of malignancy by a second pathologist) better and cheaper than glass-based alternatives.

There is a least 1 other important conclusion. Operating WSI as a clinical service (QA) within a working medical center will require more than an imaging robot. In the experience of this study, factors such as the stability of the health system's network, the speed of servers, the perfor-

mance of the pathologist's workstation and monitor, and even the functionality of the image presentation and navigation software proved more important (caused more difficulties) for the pathologists than the performance of the whole slide imager itself. Although this study was not designed to measure those parameters explicitly, it is clear that future studies will have to. Furthermore, it appears that, for WSI to reach its potential, informatics groups will need to have better control (or at least visibility) across the entire IT environment, from imager to desktop as well as image integration with the LIS.

In summary, current WSI systems seem capable of providing a useful level of surgical pathology review across a distributed health system and they will only get better. Much work needs to be done, however, especially in slide navigation, presentation speed, and data integration for WSI to research full potential in the clinical space.

## Acknowledgments

The authors would like to thank Michael Kistler, Jon Duboy, Sandy Krills, Ellen Kokal, and Maria Bond for their hard work and insight on this project.

## References

- [1] Raab SS, Nakhleh RE, Ruby SG. Patient safety in anatomic pathology: measuring discrepancy frequencies and causes. *Arch Pathol Lab Med* 2005;129:459-66.
- [2] Renshaw AA, Lezon KM, Wilbur DC. The human false-negative rate of rescreening Pap tests. Measured in a two-arm prospective clinical trial. *Cancer* 2001;93:106-10.
- [3] Schubert E, Gross W, Sideris RH, et al. A pathologist-designed imaging system for anatomic pathology signout, teaching, and research. *Semin Diagn Pathol* 1994;11:263-74.
- [4] Gilbertson J, Becich M. Imaging in pathology. *Adv Lab* 1999;8:55-8.
- [5] Belanger AJ, Lopes AE, Sinard JH. Implementation of a practical digital imaging system for routine gross photography in an autopsy environment. *Arch Pathol Lab Med* 2000;124:160-5.
- [6] Gilbertson JR, Patel A, Yagi Y. Clinical slide digitization—whole slide imaging in clinical practice. In: Gu J, Ogilvie RW, editors. *Virtual microscopy and virtual slides in teaching, diagnosis and research*. Boca Raton: Taylor & Francis; 2005.
- [7] Dunn BE, Choi H, Almagro UA, et al. Routine surgical telepathology in the Department of Veterans Affairs: experience-related improvements in pathologist performance in 2200 cases. *Telemed J* 1999;5:323-37.
- [8] Kaplan KJ, Burgess JR, Sandberg GD, et al. Use of robotic telepathology for frozen-section diagnosis: a retrospective trial of a telepathology system for intraoperative consultation. *Mod Pathol* 2002;15:1197-204.
- [9] Kayser K, Szymas J, Weinstein R. *Telepathology: telecommunication, electronic education and publication in pathology*. Berlin: Springer; 1999.
- [10] Ferreira R, Moon B, Humphries J, et al. The virtual microscope. *Proc AMIA Annu Fall Symp* 1997;449-53.
- [11] Demichelis F, Della Mea V, Forti S, et al. Digital storage of glass slides for quality assurance in histopathology and cytopathology. *J Telmed* 2002;8:138-42.
- [12] Leong FJ, McGee J. Automated complete slide digitization: a medium for simultaneous viewing by multiple pathologists. *J Pathology* 2001;195:508-14.
- [13] Weinstein RS, Descour MR, Liang C, et al. An array microscope for ultrarapid virtual slide processing and telepathology. Design, fabrication, and validation study. *HUM PATHOL* 2004;35:1303-14.
- [14] Blake CA, Lavoie HA, Millette CF. Teaching medical histology at the University of South Carolina School of Medicine: transition to virtual slides and virtual microscopes. *Anat Rec B New Anat* 2003;275:196-206.
- [15] Fujita K, Crowley RS. The Virtual Slide Set—a curriculum development system for digital microscopy. *AMIA Annu Symp Proc* 2003;846.
- [16] Kumar RK, Velon GM, Korell SO, et al. Virtual microscopy for learning and assessment in pathology. *J Pathol* 2004;204:613-8.