Commentary

Commentary: Can pathologists interpret digital images as well as they interpret microscope slides?

Thomas W. Bauer[1]

[1]Department of Anatomic Pathology, The Cleveland Clinic, 9500 Euclid Ave, Cleveland, OH 44195, USA

E-mail: *Dr. Thomas W. Bauer - BauerT@ccf.org
*Corresponding author

Before implementing widespread use of whole slide imaging (WSI) for diagnostic patient care, it is important to determine whether pathologists can interpret digital images just as well as they can interpret glass microscope slides. Often called "validation," this documentation of intraobserver variability for interpreting WSI is best accomplished with an *a priori* understanding of the existing intraobserver variability for reinterpreting microscope slides. Once this variability is known, the sample size needed to test equivalence (or noninferiority) of WSI compared to viewing microscope slides can be calculated. While true validation requires only comparing intraobserver variability between WSI and review of glass slides (GSs), there is also interest in studies that test both intra- and inter-observer variability using an experimental design in which external experts establish "gold standard" diagnoses.

Snead *et al*.[1] recently addressed this topic with the largest validation study to date. Unlike a previous validation study that used published concordance rates to calculate sample size,[2] this pathology group used data collected from prior multidisciplinary team meetings (MDTs) at which microscope slides from a proportion of surgical pathology cases had been reviewed to calculate baseline discrepancy rates. Retrospective data from MDT reviews at the University Hospitals of Coventry and Warwickshire showed a high concordance rate of 98.8% during 2011, from which the authors calculated that, for 95% power, a sample size of 3,014 cases would be needed to demonstrate noninferiority of interpreting digital images compared to GSs.

After a 3-week wash-out period (slightly longer than recommended by a committee of the College of American Pathologists[3] but shorter than a previous study),[2] scanned WSI of microscope slides were interpreted by either the same pathologist (about 1/3 of cases), or a different pathologist than who had interpreted the original microscope slides. This study did not include an arm in which the original microscope slides were rereviewed for comparison to confirm baseline intraobserver variability. Discrepancies between the original and digital image interpretations were evaluated by the same group of pathologists who participated in the study, and for a primary outcome measure, were classified as either: (1) not concordant, or (2) "complete concordance or variance of no clinical significance."

Of the 3,017 cases included in the study, there were 51 cases (1.6%) with minor discrepancies of no clinical significance, and only 21 (0.7%) with major discrepancies (diagnoses that might have resulted in different patient care). When compared to a gold standard diagnosis ("ground truth") determined by consensus review, the best diagnosis was determined to have been based on GSs in slightly more than half of the discrepant cases (57%), while the best diagnosis was actually based on the digital images in the remaining cases (43%). The authors concluded that interpreting these cases using digital pathology was not inferior to interpreting using glass microscope slides.

The authors also made several additional relevant observations. First, several major discrepancies involved distinguishing inflammatory changes from various levels of epithelial dysplasia. Difficulties classifying dysplasia using digital images have been reported in previous studies,[2,4] but is also recognized as a common area of discrepancy in the interpretation of GSs. The authors also noted difficulty recognizing *Helicobacter* on their 40X scanned images and recommended scanning selected types of cases at 60X. Alternatively, one might consider using slides that had been stained for organisms, likely

facilitating recognition of those organisms in scans made at lower magnifications. Scanning microscope slides at 60X take considerably longer and usually results in exponentially larger file sizes than scanning slides at lower magnification. Careful evaluation of the image file sizes at different magnifications as tabulated in the Snead publication also suggests that, perhaps based in part on image compression algorithms, there are differences among scanners and viewing applications such that laboratories may need to determine for themselves an optimum balance between objective lens magnification, scanning time, storage space, and the ability to recognize subtle patterns of inflammation and organisms.

Minor weaknesses of the Snead publication include the use of the same group of pathologists to establish "ground truth" (as opposed to independent external subspecialty experts), the relatively short wash-out interval of 3 weeks, the apparent exclusion of neuropathology cases, and the frequent use of images scanned at 60X (a process that would markedly increase scanning time and file size for most currently available scanners and that has not been necessary in most previous validation studies). Although there was no parallel study arm testing simultaneous discrepancy rates for rereviewing glass microscope slides, the discrepancy rate of 1.2% based on prior multidisciplinary conference reviews serves as an adequate comparison and provided the basis for calculating the sample size. Documentation of all discrepant cases and the recognition that among discrepant cases, the interpretation closest to "ground truth" was based on digital images nearly as frequently as it was based on the original microscope slides are also attributes of the Snead study.

There are many variations in design among studies intended to test our ability to interpret digital images, and the study by Snead *et al.* should add to our confidence that good quality digital images can be safely interpreted and used for patient care.

REFERENCES

1. Snead DR, Tsang YW, Meskiri A, Kimani PK, Crossman R, Rajpoot NM, *et al.* Validation of digital pathology imaging for primary histopathological diagnosis. Histopathology 2015. [doi:10.1111/his.12879].
2. Bauer TW, Schoenfield L, Slaw RJ, Yerian L, Sun Z, Henricks WH. Validation of whole slide imaging for primary diagnosis in surgical pathology. Arch Pathol Lab Med 2013;137:518-24.
3. Pantanowitz L, Sinard JH, Henricks WH, Fatheree LA, Carter AB, Contis L, *et al.* Validating whole slide imaging for diagnostic purposes in pathology: Guideline from the college of american pathologists pathology and laboratory quality center. Arch Pathol Lab Med 2013;137:1710-22.
4. Ordi J, Castillo P, Saco A, Del Pino M, Ordi O, Rodríguez-Carunchio L, *et al.* Validation of whole slide imaging in the primary diagnosis of gynaecological pathology in a university hospital. J Clin Pathol 2015;68:33-9.

Access this article online	
Quick Response Code:	
	Website: www.jpathinformatics.org