



Published in final edited form as:

Hum Pathol. 2013 November ; 44(11): . doi:10.1016/j.humpath.2013.06.015.

A comparison of cervical histopathology variability using whole slide digitized images versus glass slides: Experience with a statewide registry

Julia C. Gage, PhD, MPH^a, Nancy Joste, MD^b, Brigette M. Ronnett, MD^c, Mark Stoler, MD^d, William C. Hunt, MA^b, Mark Schiffman, MD, MPH^a, Cosette M. Wheeler, PhD^b, and Committee for The New Mexico HPV Pap Registry Steering

^aDivision of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, DHHS, Bethesda, MD, USA

^bDepartment of Pathology, Center for HPV Prevention, University of New Mexico Health Sciences Center, Albuquerque, NM, USA

^cDepartments of Pathology and Gynecology & Obstetrics. The Johns Hopkins University School of Medicine, Baltimore, MD, USA

^dDepartment of Pathology, University of Virginia, Charlottesville, VA, USA

Abstract

Whole slide imaging (WSI) is increasingly used for primary and consultative diagnoses, teaching, telepathology, slide sharing and archiving. We compared pathologist evaluations of glass slides and corresponding digitized images within the context of a statewide surveillance effort. Cervical specimens collected by the New Mexico HPV Pap Registry (NMHPVPR) research program targeted cases diagnosed between 2006–2010. Two samples of 250 slides each were digitized with the ScanScope XT (Aperio, Vista, California) microscope and reviewed with Aperio ImageScope reader. (1) A “random set” had a distribution of community diagnoses: 70% from cases of cervical intraepithelial neoplasia grade 2 (CIN2) or higher, 20% from cases of CIN grade 1 (CIN1) and 10% from negative cases. (2) A “discrepant set” was represented by difficult cases where two study pathologists initially disagreed. Within the regular workflow of the NMHPVPR, 3 pathologists read the slides 2–3 times each without knowledge of clinical history, previous readings or sampling scheme. Pathologists also read each corresponding image twice. For within- and between-reader comparisons we calculated unweighted Kappa statistics and asymmetry Chi-square tests. Across all comparisons, slides and images yielded similar results. For the random set, almost all within-reader and between-reader Kappa values ranged between 0.7–0.8 and 0.6–0.7,

© 2013 Elsevier Inc. All rights reserved.

Corresponding author: Cosette M. Wheeler, University of New Mexico Health Sciences Center, Department of Pathology, House of Prevention Epidemiology (HOPE), 1816 Sigma Chi Rd NE, MSC02-1670, Albuquerque, NM 87131, cwheeler@salud.unm.edu, Tel: (505) 272-5785, Fax: (505) 227-0265.

Disclosure: CMW has received funding through the University of New Mexico from 1) Merck and Co., Inc. and Glaxo SmithKline for HPV vaccine studies and travel reimbursements related to publication activities and 2) equipment and reagents from Roche Molecular Systems for HPV genotyping. Dr. Schiffman and Dr. Gage report working with Qiagen, Inc. on an independent evaluation of non-commercial uses of CareHPV (a low-cost HPV test for low-resource regions) for which they have received research reagents and technical aid from Qiagen at no cost. They have received HPV testing for research at no cost from Roche. The other authors report no conflicts of interest.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

respectively. For the discrepant set, most within- and between-reader Kappa values were 0.4–0.6. As CIN diagnostic terminology changes, pathologists may need to re-read histopathology slides to compare disease trends over time, e.g., before/after introduction of human papillomavirus (HPV) vaccination. Diagnosis of CIN differed little between slides and corresponding digitized images.

Keywords

Cervical intraepithelial neoplasia; whole slide imaging; clinical pathology; reliability; registries

Introduction

Digital pathology is increasingly used in medical practice with the advent of whole slide imaging (WSI), the process of scanning and digitizing glass slides to produce images that can be either reviewed by pathologists or subjected to automated image analysis. WSI is common in a variety of contexts including: primary diagnoses, consultation, telepathology, education and quality assurance (1–4). A collection of research across subspecialty areas suggests WSI can perform similarly to glass slides (1, 5–9) although study designs are heterogeneous.

WSI can be of particular benefit for pathology surveillance projects where the process of storing and retrieving a large volume of glass slides can be time-consuming and expensive. Slide retrieval is of particular importance in the statewide New Mexico HPV Pap Registry (NMHPVPR) which was established in 2006 to compare the statewide prevalence of cervical intraepithelial neoplasia (CIN) and human papillomavirus (HPV) infections before and after introduction of the HPV vaccine. A valid before-after comparison will require utilization of similar histopathology diagnostic criteria across both time periods. Since histopathology diagnosis of CIN is variable (10–15) and CIN nomenclature has recently changed (16), we may see secular trends in diagnostic criteria, which could confuse evaluations of the impact of vaccination. Therefore, many years from now, early slides will need to be retrieved and re-read concurrently with after-vaccination slides to ensure similar diagnostic criteria across time periods. WSI would be an attractive alternative to standard slide archiving.

In the NMHPVPR, we sought to measure within- and between-reader agreement among experienced pathologists reading glass slides and their digitized images. In particular, we evaluated whether the most difficult cases (poor between-reader agreement) were either harder or easier to interpret using digitized images vs. glass slides.

Materials and Methods

The New Mexico HPV Pap Registry (NMHPVPR) is a public health surveillance activity established to evaluate the continuum of cervical cancer prevention throughout the state (17). All Pap and HPV tests and all cervical, vulvar and vaginal pathology results are reportable under the New Mexico Notifiable Diseases and Conditions (<http://nmhealth.org/ERD/healthdata/documents/NotifiableDiseasesConditions022912final.pdf>). As part of an ongoing HPV genotyping research study of cervical biopsy specimens, a histology review scheme was established. Cervical biopsy cases are randomly selected by the NMHPVPR from the time period 2006–2009, within strata defined by the community pathologist diagnosis (Negative, CIN grade 1 [CIN1], and CIN grade 2 [CIN2] or higher). Blocks from participating laboratories are de-identified and new tissue sections are obtained for Hematoxylin and Eosin (H&E) staining and histologic review. The review panel consists of three experienced gynecologic pathologists with academic affiliations (NJ, BMR, and MS).

One slide from each case is initially reviewed by two of the three study pathologists. If they agree on the histologic diagnosis, no further review is undertaken. When they disagree, the third study pathologist reviews the slide and serves as the final adjudicator. All reviewers are blinded to the diagnoses of the other study pathologists, the initial community diagnosis, and all patient information.

Two sets of cases from a single laboratory were established to evaluate within- and between-reader variability. All slides in each set were read three times by each of the three reviewing pathologists over a period of approximately 12 months. Slides were mixed into the routine registry caseload with altered identifiers so that the reviewing pathologists had no knowledge of which slides were repeated. One set of cases (the “random” set) consisted of a stratified random sample of 250 cases selected with strata based on the community diagnosis: 70% from cases of CIN2 or higher, 20% from cases of CIN1 and 10% from negative cases. To examine more closely cases that were especially difficult, a second sample of 250 cases (the “discrepant” set) was selected. These were cases where the initial community diagnosis was CIN2 or higher, and the first two study reviewers disagreed with each other as to exact diagnosis.

All slides in both sets were digitized using the ScanScope XT (Aperio, Vista, California) microscope and reviewed using the Aperio ImageScope reader, a free software image viewer. The resulting digitized images were then read twice by each reviewer over a period of approximately six months. Because the only slides that were digitized were part of the reproducibility study, reviewers were aware of this when reading the digitized images, unlike the study slides which were mixed into the routine registry caseload.

For this analysis, we first tested for within- and between-reader differences in the marginal distributions of the readings using Bhapkar’s test of marginal homogeneity (18). We then quantified within-reader and between-reader agreement using the unweighted Kappa statistic and a 95% confidence interval. For reference, a Kappa value of 0.4–0.6 indicates moderate agreement while a value of 0.6–0.8 indicates substantial agreement and 0.8–1.0 indicates almost perfect agreement. Comparisons of Kappa values were made using a bootstrap estimate (1,000 replicates) of the standard error of the difference in Kappa values (19). All analyses were performed using SAS 9.3 (SAS Institute Inc., Cary, North Carolina). SAS Procedure CATMOD was used to compute tests of marginal homogeneity (20) and procedure FREQ was used to compute kappa statistics. This research study was approved by the University of New Mexico Human Research Review Committee.

Results

Table 1 shows the distribution of community diagnoses as well as diagnoses by reader for the random set of cases. For each reviewer the three readings of each slide and two readings of each digitized image are presented. Nine of the 250 cases were excluded because one or more of the reviewers indicated that the slide or image was technically unsatisfactory for diagnosis. The marginal distributions varied significantly across the three slide readings for readers 2 and 3 and across the two image readings for readers 1 and 3, although the differences were modest. Only reader 2 showed any significant difference between slides and images, with a greater proportion of negative and a lower proportion of CIN3 calls for the images ($p < 0.001$). Significant variation was observed between all three readers ($p < 0.001$). Reader 2 tended to call CIN2 more frequently (22.0%–25.7%) while reader 1 called CIN2 less frequently (9.1%–12.9%), and reader 3 was most variable with CIN2 calls (13.7%–23.7%).

Figure 1 presents the within-reader and between-reader agreement for the random case set. For reader 1 and reader 3, within-reader Kappa values ranged between 0.71 and 0.84 with overlapping confidence intervals for all comparisons (Figure 1a). Reader 2 had similarly high agreement for comparisons between slides and between images but lower Kappa values (0.64 and 0.61) for comparisons between the slide and image readings, which is consistent with this reader's tendency to call more negatives and fewer CIN3 for images compared to slides. Between-reader agreement (Figure 1b) tended to be lower than within-reader agreement with all but one Kappa value between 0.62 and 0.74.

The discrepant set was selected from cases with a community diagnosis of CIN2 or higher and all three readers showed considerable downgrading from the community diagnosis (Table 2). Additionally, the distribution of diagnoses for the discrepant set showed more variability, both within-reader and between-reader, and all but one test of marginal homogeneity were statistically significant. All three readers tended to call fewer CIN2 or higher for images compared to slides ($p < .001$ for all comparisons).

Compared to the random slide set, readers had lower within-reader agreement for slides with Kappa values ranging between 0.37 and 0.61 (Figure 2a). Comparisons of slide readings vs. image readings showed similarly low Kappa values of 0.33 to 0.53. Figure 2b shows between-reader agreement for the discrepant set. The first slide reading was not included in the between-reader analysis because this reading defined the discrepant set and, by definition, there was complete disagreement between two of the readers. Comparisons of subsequent between-reader agreement showed that readers 1 and 2 had better agreement with each other when reviewing slides and images with Kappa values between 0.38 and 0.55 while reader 3 tended to disagree more with both readers 1 and 2 with Kappa values between 0.22 and 0.35. This difference was due to reader 3's greater tendency to diagnose more cases as CIN2 and CIN3 (Table 2). The between-reader agreement was similar for slides and images.

Discussion

In a sample of cervical histopathology slides from a statewide registry and its associated research program, we found very good reproducibility both within and between-readers when evaluating glass slides and their digitized images. Similarly, reproducibility between glass slide and digitized images was also very good. Among difficult cases, the agreement for digitized images was similar to agreement for glass slides with a slight tendency towards downgrading for images and slides. The between-reader Kappa values for the random set ranged from 0.6 to 0.7, a range similar to some unweighted Kappa values previously reported for cervical histopathology(15), but higher than others (11, 13, 14, 21).

Our study of almost 500 cases is one of the larger studies examining the comparative performance of WSI vs. traditional glass slides and the first of which we are aware that is dedicated to cervical histopathology. Our findings generally coincide with results from other WSI studies, although there are differences in design (1, 5) and they targeted other organ sites (9, 22, 23). Of note, this study was performed in the context of a large pathology surveillance project and not clinical practice, therefore no additional patient information was provided and there was no further review of challenging cases. In addition, the sampling design of this study was not selected to represent the spectrum of cases observed in a routine clinical practice. Yet, we note that reviewers were unaware that they were reviewing slides and images multiple times or even in a random vs. discrepant set. Slides were introduced into the routine registry caseload.

Given that imaging systems have a variety of organizational archiving and image viewing options as well as image capture settings, WSI scanning can require a significant investment of time for laboratories to establish a system. Once imaging system management and processes are established validation approaches must be considered as previously described (<http://www.cap.org/apps/docs/membership/transformation/new/validating.pdf>). We found that imaging itself is a very labor intensive activity even in an experienced laboratory using automation capable of capturing up to 120 slides or greater per imaging session. Imaging time varies depending on the magnification, image focus points and image format selected by the user and can take many hours per 100 slides. Difficulties in the imaging process can result in additional investments of time on a regular basis (e.g. reimaging can be required to obtain satisfactory images for a variety of reasons including variable thickness across the tissue specimen or when system barcode reading errors are encountered). Approximately 3–5% of images required reimaging by the NMHPVPR. A manual technical review process was established for individual images and the need for customized solutions to address imaging failures had a significant impact on workflow.

Participating pathologists noted that a learning period will be important to implement image review processes and may vary among individual pathologists. Given that most experience and current standards of clinical practice use glass slide review, it is likely that, at least initially, more time will be required to review images than slides. Potential difficulties associated with visualizing small lesions in the image format should be considered when undertaking future evaluations.

In conclusion, the discrepancies within and between pathologists were similar whether reading slides or images. Reading digitized images as opposed to glass slides will likely produce similar results and difficult cases will still be difficult to interpret. These findings suggest that WSI could be useful for archiving within the context of a pathology surveillance project but significant efforts are required. This study's methodology may prove useful as pathology laboratories undertake in-house validation of WSI prior to use as a tool for patient care activities such as primary diagnosis or consultation.

Acknowledgments

Supported by R01CA134779 (CMW) and in part by the Intramural Research Program of the National Cancer Institute, National Institutes of Health, Department of Health and Human Services.

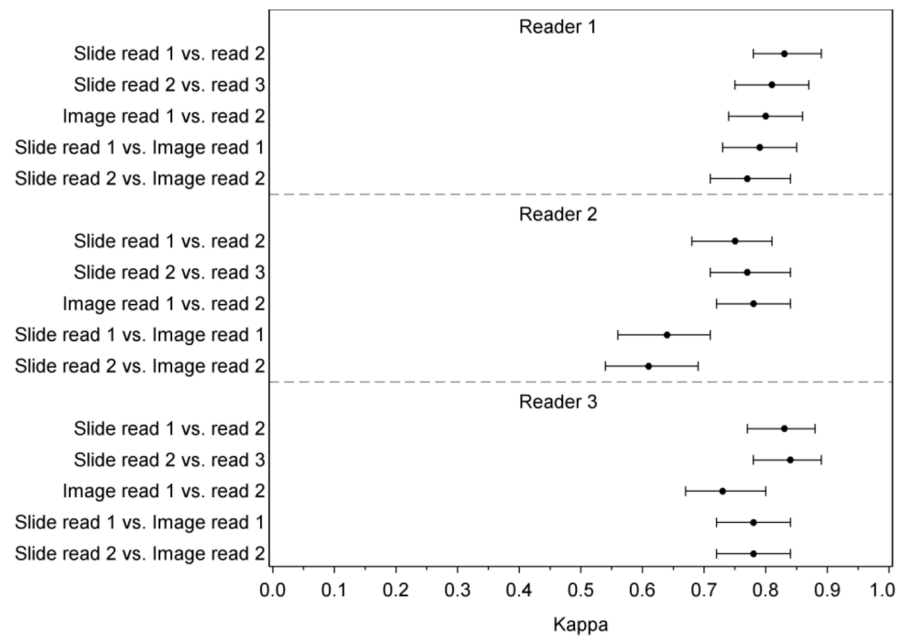
References

1. 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; 43:1739–1744. [PubMed: 22591912]
2. Pantanowitz L. Digital images and the future of digital pathology. *Journal of pathology informatics.* 2010;1. [PubMed: 20805964]
3. Pantanowitz L, Valenstein PN, Evans AJ, Kaplan KJ, Pfeifer JD, Wilbur DC, Collins LC, Colgan TJ. Review of the current state of whole slide imaging in pathology. *Journal of pathology informatics.* 2011; 2:36. [PubMed: 21886892]
4. Weinstein RS, Graham AR, Richter LC, Barker GP, Krupinski EA, Lopez AM, Erps KA, Bhattacharyya AK, Yagi Y, Gilbertson JR. Overview of telepathology, virtual microscopy, and whole slide imaging: prospects for the future. *Hum Pathol.* 2009; 40:1057–1069. [PubMed: 19552937]
5. Gilbertson JR, Ho J, Anthony L, Jukic DM, Yagi Y, Parwani AV. Primary histologic diagnosis using automated whole slide imaging: a validation study. *BMC clinical pathology.* 2006; 6:4. [PubMed: 16643664]

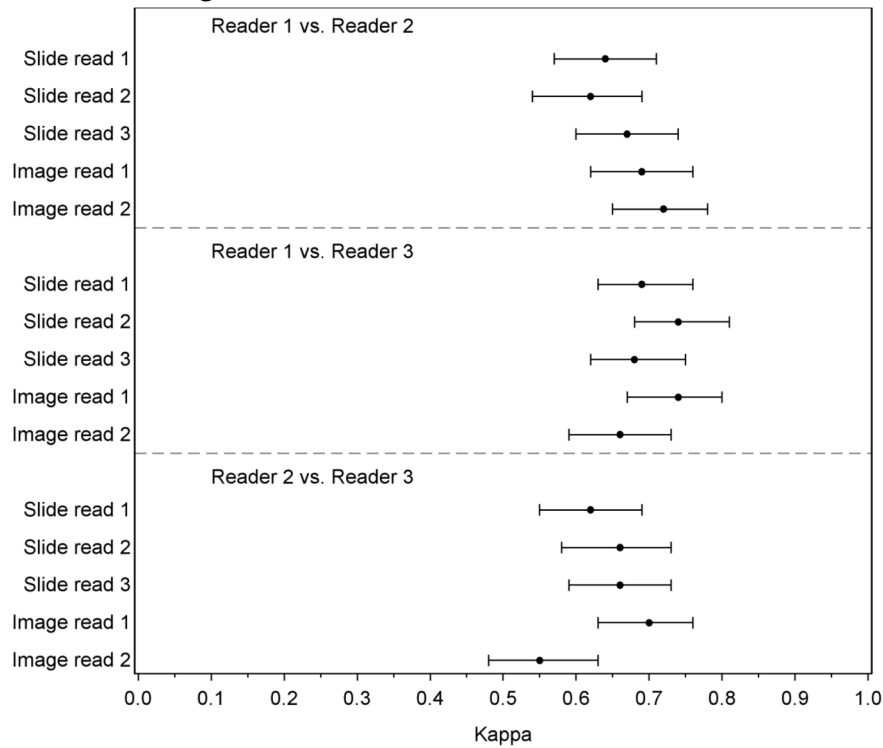
6. Jara-Lazaro AR, Thamboo TP, Teh M, Tan PH. Digital pathology: exploring its applications in diagnostic surgical pathology practice. *Pathology*. 2010; 42:512–518. [PubMed: 20854068]
7. Jukic DM, Drogowski LM, Martina J, Parwani AV. Clinical examination and validation of primary diagnosis in anatomic pathology using whole slide digital images. *Archives of pathology & laboratory medicine*. 2011; 135:372–378. [PubMed: 21366463]
8. Massone C, Soyer HP, Lozzi GP, Di Stefani A, Leinweber B, Gabler G, Asgari M, Boldrini R, Bugatti L, Canzonieri V, Ferrara G, Kodama K, Mehregan D, Rongioletti F, Janjua SA, Mashayekhi V, Vassilaki I, Zelger B, Zgavec B, Cerroni L, Kerl H. Feasibility and diagnostic agreement in teledermatopathology using a virtual slide system. *Hum Pathol*. 2007; 38:546–554. [PubMed: 17270240]
9. Nielsen PS, Lindebjerg J, Rasmussen J, Starklint H, Waldstrom M, Nielsen B. Virtual microscopy: an evaluation of its validity and diagnostic performance in routine histologic diagnosis of skin tumors. *Hum Pathol*. 2010; 41:1770–1776. [PubMed: 20869750]
10. Ismail SM, Colclough AB, Dinnen JS, Eakins D, Evans DM, Gradwell E, O'Sullivan JP, Summerell JM, Newcombe RG. Observer variation in histopathological diagnosis and grading of cervical intraepithelial neoplasia. *BMJ*. 1989; 298:707–710. [PubMed: 2496816]
11. de Vet HC, Knipschild PG, Schouten HJ, Koudstaal J, Kwee WS, Willebrand D, Sturmans F, Arends JW. Interobserver variation in histopathological grading of cervical dysplasia. *J Clin Epidemiol*. 1990; 43:1395–1398. [PubMed: 2254777]
12. Creagh T, Bridger JE, Kupek E, Fish DE, Martin-Bates E, Wilkins MJ. Pathologist variation in reporting cervical borderline epithelial abnormalities and cervical intraepithelial neoplasia. *J Clin Pathol*. 1995; 48:59–60. [PubMed: 7706521]
13. McCluggage WG, Bharucha H, Caughley LM, Date A, Hamilton PW, Thornton CM, Walsh MY. Interobserver variation in the reporting of cervical colposcopic biopsy specimens: comparison of grading systems. *J Clin Pathol*. 1996; 49:833–835. [PubMed: 8943751]
14. Stoler MH, Schiffman M. Interobserver reproducibility of cervical cytologic and histologic interpretations: realistic estimates from the ASCUS-LSIL Triage Study. *JAMA*. 2001; 285:1500–1505. [PubMed: 11255427]
15. Malpica A, Matisic JP, Niekirk DV, Crum CP, Staerckel GA, Yamal JM, Guillaud MH, Cox DD, Atkinson EN, Adler-Storthz K, Poulin NM, Macaulay CA, Follen M. Kappa statistics to measure interrater and intrarater agreement for 1790 cervical biopsy specimens among twelve pathologists: qualitative histopathologic analysis and methodologic issues. *Gynecol Oncol*. 2005; 99:S38–52. [PubMed: 16183106]
16. Darragh TM, Colgan TJ, Cox JT, Heller DS, Henry MR, Luff RD, McCalmont T, Nayar R, Palefsky JM, Stoler MH, Wilkinson EJ, Zaino RJ, Wilbur DC. The Lower Anogenital Squamous Terminology Standardization Project for HPV-Associated Lesions: Background and Consensus Recommendations From the College of American Pathologists and the American Society for Colposcopy and Cervical Pathology. *Journal of lower genital tract disease*. 2012; 16:205–242. [PubMed: 22820980]
17. Wheeler CM, Hunt WC, Cuzick J, Langsfeld E, Pearse A, Montoya GD, Robertson M, Shearman CA, Castle PE. A population-based study of human papillomavirus genotype prevalence in the United States: Baseline measures prior to mass human papillomavirus vaccination. *Int J Cancer*. 2013; 132:198–207. [PubMed: 22532127]
18. Bhapkar VP. A Note on Equivalence of 2 Test Criteria for Hypotheses in Categorical Data. *J Am Stat Assoc*. 1966; 61:228–235.
19. Vanbelle S, Albert A. A bootstrap method for comparing correlated kappa coefficients. *J Stat Comput Sim*. 2008; 78:1009–1015.
20. Sun, X.; Yang, Z. Generalized McNemar's test for homogeneity of the marginal distributions. *SAS Global Forum*; San Antonio, Texas. 2008.
21. de Vet HC, Knipschild PG, Schouten HJ, Koudstaal J, Kwee WS, Willebrand D, Sturmans F, Arends JW. Sources of interobserver variation in histopathological grading of cervical dysplasia. *J Clin Epidemiol*. 1992; 45:785–790. [PubMed: 1619458]
22. Fine JL, Grzybicki DM, Silowash R, Ho J, Gilbertson JR, Anthony L, Wilson R, Parwani AV, Bastacky SI, Epstein JI, Jukic DM. Evaluation of whole slide image immunohistochemistry

- interpretation in challenging prostate needle biopsies. *Hum Pathol.* 2008; 39:564–572. [PubMed: 18234276]
23. Helin H, Lundin M, Lundin J, Martikainen P, Tammela T, van der Kwast T, Isola J. Web-based virtual microscopy in teaching and standardizing Gleason grading. *Hum Pathol.* 2005; 36:381–386. [PubMed: 15891999]

1a. Within-reader agreement



1b. Between-reader agreement



Excludes 9 cases where slide or image was judged unsatisfactory for at least one reader.

Figure 1.

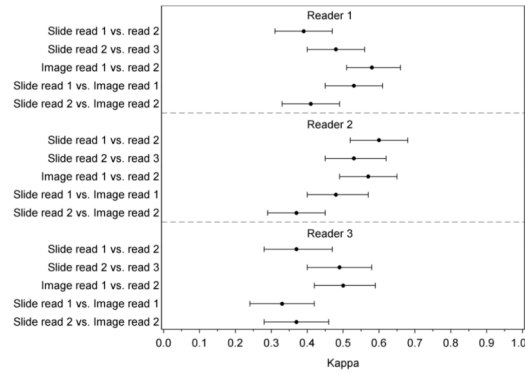
Within-and between-reader agreement for 3 readings of glass slides and 2 readings of digitized images among a stratified random sample of 241 cases

1a. Within-reader agreement

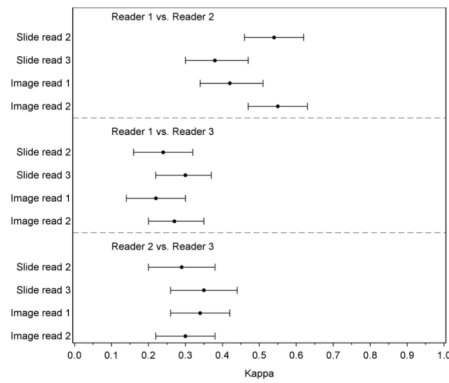
1b. Between-reader agreement

Excludes 9 cases where slide or image was judged unsatisfactory for at least one reader.

2b. Within-reader agreement



2b. Between-reader agreement



Excludes 2 cases where slide or image was judged unsatisfactory for at least one reader.

Figure 2. Within-and between-reader agreement for 3 readings of glass slides and 2 readings of digitized images among 248 cases where readers initially disagreed
 2b. Within-reader agreement
 2b. Between-reader agreement
 Excludes 2 cases where slide or image was judged unsatisfactory for at least one reader.

Distribution of diagnoses by reader for 3 readings of glass slides and 2 readings of digitized images among a stratified random sample of 241 cases

Table 1

	Negative %	CIN1 %	CIN2 %	CIN3/AIS %	Cancer %	Marginal homogeneity p-value
Community diagnosis	9.5	20.7	23.7	31.1	14.9	
Reader 1 Slide	24.9	21.2	9.5	32.8	11.6	
read 1	24.9	22.0	11.6	30.3	11.2	
read 2	24.1	25.3	9.1	29.9	11.6	0.1
read 3	22.8	23.2	12.9	29.5	11.6	
Image	28.6	23.2	10.0	26.6	11.6	0.002 ^a
read 1	22.0	17.4	25.7	23.7	11.2	
read 2	21.6	15.8	23.7	27.8	11.2	
read 3	21.2	21.2	22.0	24.1	11.6	0.02 ^a
Image	27.0	19.1	23.2	19.5	11.2	
read 1	29.9	18.3	22.8	17.4	11.6	0.22
read 2	19.5	15.4	23.7	30.3	11.2	
read 3	19.9	17.4	17.8	32.8	12.0	
Image	20.3	16.6	15.8	35.7	11.6	0.01 ^a
read 1	22.8	17.0	19.9	28.6	11.6	
read 2	21.6	16.6	13.7	36.9	11.2	0.003 ^a

Excludes 9 cases where slide or image was judged unsatisfactory for at least one reader.

The distribution of diagnoses was significantly different between all readers for both slides and images, $p < 0.001$. Reader 2 showed a significant difference in the distribution of diagnoses between slides and images, $p < 0.001$.

^a Statistically significant at $p < .05$ threshold.

Distribution of diagnoses by reader for 3 readings of glass slides and 2 readings of digitized images among 248 cases where readers initially disagreed

Table 2

		Negative %	CIN1 %	CIN2 %	CIN3/AIS %	Cancer %	Marginal homogeneity p-value
Community diagnosis		0.0	0.0	69.0	29.8	1.2	
Reader 1	Slide	18.1	51.6	14.9	13.7	1.6	
	read 2	20.6	34.3	28.6	15.3	1.2	
	read 3	21.8	43.1	23.0	10.5	1.6	<0.001 ^a
Image	read 1	20.2	45.2	20.2	13.7	0.8	
	read 2	28.2	46.0	19.0	5.6	1.2	<0.001 ^a
Reader 2	Slide	21.0	30.2	36.3	11.7	0.8	
	read 2	17.7	22.6	44.0	15.7	0.0	
	read 3	17.7	23.0	48.4	10.1	0.8	<0.001 ^a
Image	read 1	27.4	30.2	33.9	8.1	0.4	
	read 2	31.0	34.3	29.4	4.8	0.4	0.002 ^a
Reader 3	Slide	13.7	1.6	63.7	21.0	0.0	
	read 2	9.7	9.3	56.5	24.6	0.0	
	read 3	10.1	18.5	45.6	25.0	0.8	<0.001 ^a
Image	read 1	10.1	25.0	46.0	19.0	0.0	
	read 2	15.3	22.2	41.9	20.2	0.4	0.06

Excludes 2 cases where slide or image was judged unsatisfactory for at least one reader.

The distribution of diagnoses was significantly different between all readers for both slides and images, p <0.001. All readers showed significant differences in the distribution of diagnoses between slides and images, p <0.001.

^a Statistically significant at p<.05 threshold.