

# Accuracy of Digital Pathologic Analysis vs Traditional Microscopy in the Interpretation of Melanocytic Lesions

Tracy Onega, PhD, MA, MS; Raymond L. Barnhill, MD; Michael W. Piepkorn, MD, PhD; Gary M. Longton, MS; David E. Elder, MBChB, FRCPA; Martin A. Weinstock, MD, PhD; Stevan R. Knezevich, MD, PhD; Lisa M. Reisch, PhD; Patricia A. Carney, PhD; Heidi D. Nelson, MD, MPH, MACP; Andrea C. Radick, MS; Joann G. Elmore, MD, MPH

[+ Supplemental content](#)

**IMPORTANCE** Use of digital whole-slide imaging (WSI) for dermatopathology in general has been noted to be similar to traditional microscopy (TM); however, concern has been noted that WSI is inferior for interpretation of melanocytic lesions. Since approximately 1 of every 4 skin biopsies is of a melanocytic lesion, the use of WSI requires verification before use in clinical practice.

**OBJECTIVE** To compare pathologists' accuracy and reproducibility in diagnosing melanocytic lesions using Melanocytic Pathology Assessment Tool and Hierarchy for Diagnosis (MPATH-Dx) categories when analyzing by TM vs WSI.

**DESIGN, SETTING, AND PARTICIPANTS** A total of 87 pathologists in community-based and academic settings from 10 US states were randomized with stratification based on clinical experience to interpret in TM format 180 skin biopsy cases of melanocytic lesions, including 90 invasive melanoma, divided into 5 sets of 36 cases (phase 1). The pathologists were then randomized via stratified permuted block randomization with block size 2 to interpret cases in either TM (n = 46) or WSI format (n = 41), with each pathologist interpreting the same 36 cases on 2 separate occasions (phase 2). Diagnoses were categorized as MPATH-Dx categories I through V, with I indicating the least severe and V the most severe.

**MAIN OUTCOMES AND MEASURES** Accuracy with respect to a consensus reference diagnosis and the reproducibility of repeated interpretations of the same cases.

**RESULTS** Of the 87 pathologists in the study, 46% (40) were women and the mean (SD) age was 50.7 (10.2) years. Except for class III melanocytic lesions, the diagnostic categories showed no significant differences in diagnostic accuracy between TM and WSI interpretation. Discordance was lower among class III lesions for the TM interpretation arm (51%; 95% CI, 46%-57%) than for the WSI arm (61%; 95% CI, 53%-69%) ( $P = .05$ ). This difference is likely to have clinical significance, because 6% of TM vs 11% of WSI class III lesions were interpreted as invasive melanoma. Reproducibility was similar between the traditional and digital formats overall (66.4%; 95% CI, 63.3%-69.3%; and 62.7%; 95% CI, 59.5%-65.7%, respectively), and for all classes, although class III showed a nonsignificant lower intraobserver agreement for digital. Significantly more mitotic figures were detected with TM compared with WSI: mean (SD) TM, 6.72 (2.89); WSI, 5.84 (2.56);  $P = .002$ .

**CONCLUSIONS AND RELEVANCE** Interpretive accuracy for melanocytic lesions was similar for WSI and TM slides except for class III lesions. We found no clinically meaningful differences in reproducibility for any of the diagnostic classes.

JAMA Dermatol. 2018;154(10):1159-1166. doi:10.1001/jamadermatol.2018.2388  
Published online August 15, 2018.

**Author Affiliations:** Author affiliations are listed at the end of this article.

**Corresponding Author:** Tracy Onega, PhD, MA, MS, One Medical Center Drive, Lebanon, NH 03756 (tracy.l.onega@dartmouth.edu).

Digital whole-slide imaging (WSI) has only very recently been approved by the US Food and Drug Administration (FDA) for primary diagnosis in 1 system (ie, Philips).<sup>1</sup> Prior to FDA approval, use of WSI in the United States was growing rapidly for other clinical and educational purposes.<sup>2,3</sup> Internationally, guidelines that provide technical specifications on WSI and other elements of digital pathology have been published by professional groups from the United States, Canada, Europe, and Australia.<sup>4-6</sup> Adoption of WSI for primary interpretation is contingent on technical needs, regulatory policies, and digital pathology workflow processes, as well as evidence demonstrating comparable interpretive performance between traditional microscopy (TM) and digital images, particularly for specific tissue types.

Evidence that WSI is as accurate as TM is beginning to accumulate but is still limited. While several published studies strongly suggest that WSI interpretation is comparable to TM,<sup>7-9</sup> this effect has not been consistently observed.<sup>10</sup> The majority of these studies reported comparable accuracy, but the need for subspecialty studies is crucial to fill substantial gaps in evidence. One large study reported equivalent diagnostic accuracy for TM and WSI; however, dermatopathology was only 1 of the 10 subspecialties studied, and clinically significant discordance between TM and WSI formats was found for 2 cases of melanoma.<sup>7</sup>

Dermatopathology can present unique challenges to pathological assessment, such as ascertainment of cytonuclear characteristics and mitotic figures. Although a recent study reported similar diagnostic accuracy of WSI compared with TM in the interpretation of skin biopsies, of concern was the finding that interpretation of melanocytic lesions were substantially less accurate when performed in the digital WSI format.<sup>11</sup> Of note, this important study was not designed to specifically address WSI accuracy for interpretation of melanocytic lesions. Discordance in TM of this challenging group is broadly recognized,<sup>12</sup> making such investigation difficult. Given that melanocytic lesions represent about 1 of every 4 skin biopsies,<sup>13</sup> this topic is especially relevant to clinical practice and the field of dermatology.

Accurate pathologic diagnosis is critical for staging, prognosis, and management. Thus, as WSI use continues to expand, particularly for primary diagnostic interpretation, data on validation studies of WSI compared with TM specific to melanocytic lesions are needed.

We compare TM vs WSI diagnoses focused directly on the clinical issue of melanocytic lesions, and present results of a study that compared the interpretive accuracy and reproducibility of diagnoses by pathologists viewing melanocytic skin lesions in both WSI and TM. Both accuracy (defined as agreement with a reference consensus diagnosis) and reproducibility (intraobserver concordance) were analyzed to determine if differences were evident. We investigated whether major clinical discrepancies were noted between the 2 interpretive techniques based on accuracy.

## Methods

All procedures were compliant with the Health Insurance and Accountability Act and approved by the institutional review

### Key Points

**Question** Are pathologists' diagnoses of melanocytic lesions as accurate and reproducible when using digital whole-slide imaging (WSI) vs traditional microscopy (TM)?

**Findings** This study surveyed 87 pathologists randomized with stratification based on clinical experience to compare digital WSI with TM to diagnose melanocytic lesions. Interpretive accuracy for melanocytic lesions was similar for digital WSI and TM.

**Meaning** These results add to the growing evidence of accuracy and reproducibility of WSI interpretation compared with TM and support the US Food and Drug Administration's recent approval of WSI for primary diagnosis.

boards of the University of Washington, Fred Hutchinson Cancer Research Center, Oregon Health & Science University, Rhode Island Hospital, and Dartmouth College. Participating pathologists provided written informed consent.

### Biopsy Case Development

Skin biopsy case development and melanoma pathology study design have been previously described.<sup>14,15</sup> Briefly, 240 skin biopsy specimens were randomly selected from Dermatopathology Northwest, the largest physician-owned dermatopathology practice in the Northwest with 6 board-certified dermatopathologists, in Bellevue, Washington. Each case was of a melanocytic lesion and included standardized data on the patient's age at biopsy, biopsy location, and biopsy type. Cases were selected with stratification based on patient age (20-49 years, 50-64 years, or  $\geq 65$  years) and medical record documentation of the original diagnosis. The 240 cases were used to assemble 5 equivalent sets of 48 cases for the larger melanoma pathology study.<sup>14,15</sup> For the purposes of this study, a smaller subset of 36 cases from each of the 5 sets was used, for a total of 180 cases.

Each TM slide was scanned using a Hamamatsu NanoZoomer 2.0-RS digital slide scanner in magnification  $\times 40$  high-resolution mode. A technician and an experienced dermatopathologist reviewed each digital image, rescanning as needed to obtain the highest quality. A custom online digital slide viewer was built using HD View SL, Microsoft's open source Silverlight gigapixel image viewer. A full description of HD View SL is available at <http://hdviews1.codeplex.com/>. The viewer, like popular online mapping applications and industry-sponsored WSI viewers, allowed pathologists to pan the image and zoom (up to 40 times the actual scanned magnification with additional digital magnification for a final maximum magnification of  $\times 60$ ). Additional tools were available for measuring lesion size and counting mitotic figures.

### Determination of Reference Standard

Three experienced dermatopathologists developed a reference interpretation, defined as complete agreement among the 3 dermatopathologists, through a consensus process for each case in TM format using standardized Melanocytic Pathology Assessment Tool and Hierarchy for Diagnosis (MPATH-Dx)

categories.<sup>14</sup> The 180 cases had intentionally higher proportions in MPATH-Dx classes II through V than typically encountered in practice: 8.3%, class I; 16.7%, class II; 25.0%, class III; 25.0%, class IV; and 25.0%, class V. We present all data in comparison with the TM slide reference diagnoses. The MPATH-Dx diagnostic classes have been described in detail in prior work.<sup>15</sup> Briefly, MPATH-Dx class I indicates nevus/mild atypia (no further treatment required); class II indicates moderate atypia/dysplasia (consider narrow but complete excision margin <5 mm); class III indicates severe dysplasia/melanoma in situ (excision with 5-mm margins); class IV indicates stage pT1a invasive melanoma (wide excision ≥1-cm margin); and class V indicates stage pT1b or higher invasive melanoma (wide excision ≥1 cm with possible additional treatment). These examples are not inclusive of the many terms that can be used in diagnosis of melanocytic lesions and are subject to further development and revision by consensus groups.

### Pathologist Recruitment, Selection, and Baseline Data Collection

Study pathologists were recruited from 10 US states (California, Connecticut, Hawaii, Iowa, Kentucky, Louisiana, New Jersey, New Mexico, Utah, and Washington), had completed residency and/or fellowship training, had interpreted skin specimens (including melanocytic skin lesions) for at least 1 year in their clinical practices, and intended to continue interpreting skin specimens (including melanocytic skin lesions) for at least 1 year. Pathologists were invited to participate via email(s), subsequent mailed invitations, and telephone calls. After enrolling, pathologists completed a demographic and practice characteristic survey.

### Biopsy Case Interpretations

The biopsy case interpretations were analyzed in 2 phases. In phase 1, pathologists were randomly assigned to 1 of 5 sets of 36 cases, with randomization stratified by clinical expertise and based on possession of 1 or more of the following characteristics from the self-reported survey: fellowship trained or board certified in dermatopathology, considered by colleagues to be an expert in melanocytic lesions, or 10% or more of usual case-load included cutaneous melanocytic lesions.<sup>12</sup> Participants were told that the cases in phase 1 were not representative of clinical practice in terms of prevalence of particular diagnoses. All phase 1 interpretations were obtained in TM-slide format, with the same TM slide interpreted for each case.

After a washout period of at least 8 months, in which pathologists continued in clinical practice but had no study-related activities, the pathologists were invited to continue interpreting cases in phase 2 of the study. Pathologists were invited at this time to volunteer for the study of digital vs traditional interpretations. Pathologists volunteering for the digital vs traditional study were randomly assigned to interpretive method in phase 2 (36 traditional slides vs 36 digital slides). This randomization to interpretive method was performed with stratification based on clinical expertise (eFigure 1 in the Supplement). The phase 2 participant allocation to TM or WSI used a stratified permuted block randomization with block size 2. There were 10 strata defined by the 5 slide sets, as well as

by expertise. Regardless of interpretive method, pathologists were scheduled for interpretations at a mutually convenient time and were given 1 week to complete interpretations.

The pathologists interpreted the same cases in both phases; however, the cases were randomly reordered for each participant and also for each phase. The participants were not informed that the cases in phase 2 were a subset of the same, but fewer, reordered cases that they had already interpreted in phase 1. The pathologists used a web-based form to document interpretations and used their own microscopes, computers, and monitors for this activity. They received up to 20 hours of category 1 continuing medical education credits after participating in the interpretation and subsequent educational program.

### Statistical Analyses

We calculated case discordance rates with the consensus reference diagnoses as a measure of accuracy for interpretations. Tests for discordance rates and confidence intervals (CIs) accounted for both within- and between-participant variability by using variance estimates of the form  $(\text{VarRateP} + [\text{AvgRateP} \times (1 - \text{AvgRateP})]/nc)/np$ , where VarRateP is the sample variance among pathologists, AvgRateP is the average rate among pathologists, nc is the number of cases interpreted by each pathologist, and np is the number of pathologists. Pathologist characteristics (eg, expertise, digital experience) as possible confounders were examined. We used logistic regression to examine if the accuracy of TM vs digital WSI format remained after adjusting for pathologist characteristics.

To assess reproducibility, pathologists' interpretations in phase 2 were compared with their interpretations of the same cases in phase 1. All participants interpreted TM format slides in phase 1. Thus, the relevant reproducibility comparison was between participants who were randomized to interpret TM slides in both phases and those who differentially interpreted TM slides in phase 1 and WSI in phase 2. This study design element is important given the extent of intraobserver variability noted when both interpretations were in the TM slide method. We considered 3 comparisons to represent clinically significant differences, based on the likely change in case management and/or outcomes: (1) notable difference between class III compared with class I, IV, or V; (2) notable difference between class IV and any other class; and (3) notable difference between class V and any other class. Agreement rates and CIs were based on logit models using a robust estimator of the variance to account for correlation of case interpretations from the same pathologist. Hypothesis tests were based on Wald tests of logit model coefficients distinguishing between interpretations made on different combinations of diagnostic formats.

Participants recorded several measures of diagnostic certainty for each assessment, including ratings of diagnostic difficulty and assessment confidence, and also noted if a second opinion was desired and if an assessment was considered borderline between 2 diagnoses. Differences in these measures between TM and digital WSI interpretations were examined with logistic regression models.

**Table 1. Characteristics of Study Pathologists Randomized to TM or WSI for Interpretation of Melanocytic Lesions**

Characteristic	Pathologists, No. (%) (N = 87)	
	Digital WSI	Glass
Overall	41	46
Age, y		
<40	7 (17)	8 (17)
40-49	10 (24)	19 (41)
50-59	15 (37)	15 (33)
≥60	9 (22)	4 (9)
Sex		
Female	22 (54)	18 (39)
Male	19 (46)	28 (61)
Affiliation with academic medical center		
No	28 (68)	35 (76)
Yes	13 (32)	11 (24)
Board certified or fellowship trained in dermatopathology <sup>a</sup>		
No	27 (66)	22 (48)
Yes	14 (34)	24 (52)
Years interpreting melanocytic skin lesions		
<10	18 (44)	23 (50)
≥10	23 (56)	23 (50)
Caseload interpreting melanocytic skin lesions, No./mo		
<60	22 (54)	23 (50)
≥60	19 (46)	23 (50)
Considered self to be an expert in melanocytic skin lesions in the eyes of colleagues		
No	27 (66)	18 (39)
Yes	14 (34)	28 (61)

Abbreviations: TM, traditional microscopy; WSI, whole-slide imaging.

<sup>a</sup> This category consists of physicians with single or multiple fellowships that include dermatopathology. Also includes physicians with single or multiple board certifications that include dermatopathology.

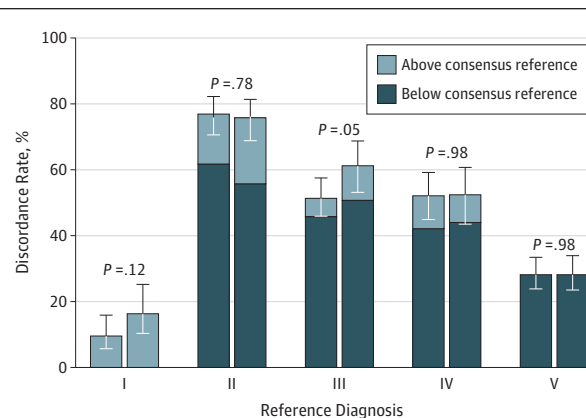
We also examined mitotic rate and Breslow depth for cases by traditional and digital format interpretation. These comparisons were performed in cases with a consensus reference diagnosis of invasive melanoma and when participants diagnosed the case as invasive melanoma, because in all of these cases the pathologists provided data on mitotic rate and Breslow depth. The nonparametric Wilcoxon signed-rank test was used to test for differences in mitotic rate, and the paired *t* test was used for Breslow depth.

The study design and analysis meets or exceeds the guidelines set forth by the College of American Pathologists to validate the diagnostic use of digital WSI.<sup>16-18</sup>

## Results

Eighty-seven pathologists (46 assigned to TM and 41 to WSI) agreed to participate in the randomized study and completed their interpretations. Although randomization was stratified on a composite measure of participant expertise, and the resulting proportion of experts defined in this way was reason-

**Figure 1. Discordance of Melanocytic Lesion Interpretations With a Reference Diagnosis**



This graph shows similar results for traditional and digital formats across the Melanocytic Pathology Assessment Tool and Hierarchy for Diagnosis categories,<sup>14</sup> except for a significantly lower discordance for class III diagnoses. Error bars indicate 95% confidence intervals.

ably similar across arms (70% in the TM arm [n = 32], 63% in the WSI arm [n = 26]), there was some imbalance in individual measures of expertise. In particular, pathologists randomized to TM included a higher proportion of those with dermatopathology board certification or fellowship training (24 [52%] vs 14 [34%]) and a higher proportion of those who considered themselves to be experts in the eyes of their peers (28 [61%] vs 14 [34%]) (Table 1).

We found no significant differences between the accuracy of interpretations in traditional vs digital format across the 5 MPATH-Dx diagnostic categories, with the exception of class III lesions. It was reassuring that the discordance rates were the same for all high-stage invasive melanoma cases (MPATH-Dx class V, defined as T1b or higher stage) for both interpretive formats (Figure 1 and Table 2). Diagnostic errors (ie, discordance with the reference) were lower among class III lesions for the TM interpretation arm (51%; 95% CI, 46%-57%) than for the WSI arm (61%; 95% CI, 53%-69%) (P = .05). Specifically, for cases with a class III reference diagnosis (eg, melanoma in situ), we found that in the TM format, 5.8% (24/414) of the interpretations given by participants were upgraded to class IV or V (ie, invasive melanoma); when these same cases were interpreted in the digital WSI format, 10.6% (39/369) of the interpretations were upgraded to invasive melanoma (Table 2). Given that class III lesions do not represent invasive melanoma, while classes IV and V do, we defined this upgrading as a major clinical discrepancy because patient care and possibly outcomes would be different. A model for the effect of TM format relative to WSI on accuracy across all 5 MPATH-Dx diagnostic classes yields an odds ratio (OR) of 1.12 (95% CI, 0.92-1.37; P = .27). Adjusting for dermatopathology board certification and/or fellowship training reduced the estimated OR to 1.04 (95% CI, 0.87-1.25; P = .64).

Overall, intraobserver agreement was slightly higher but not significantly different when the cases were interpreted in TM format in both phases (66.9%; 95% CI, 65.1%-68.6%) vs when

**Table 2. Agreement in Skin Pathology Interpretation and Assessment for Clinical Discrepancy Between TM and Digital WSI in 5 Sets of Melanocytic Lesion Cases**

Reference Diagnosis <sup>a</sup>	I	II	III	IV	V	Total No. of Interpretations	Agreement of Pathologists With Consensus Reference, % (95% CI)
Participating pathologists' interpretation with TM							
I	125 <sup>b</sup>	7	5	0	1	138	91 (84-95)
II	170	64 <sup>b</sup>	32	6	4	276	23 (18-30)
III	109	80	201 <sup>b</sup>	21 <sup>c</sup>	3 <sup>c</sup>	414	49 (43-54)
IV	18 <sup>d</sup>	37 <sup>d</sup>	119 <sup>d</sup>	198 <sup>b</sup>	42	414	48 (41-55)
V	9 <sup>e</sup>	6 <sup>e</sup>	41 <sup>e</sup>	61 <sup>e</sup>	297 <sup>b</sup>	414	72 (67-76)
Participating pathologists' interpretation with WSI							
I	103 <sup>b</sup>	11	7	2	0	123	84 (75-90)
II	137	60 <sup>b</sup>	32	11	6	246	24 (19-31)
III	110	77	143 <sup>b</sup>	38 <sup>c</sup>	1 <sup>c</sup>	369	39 (31-47)
IV	34 <sup>d</sup>	30 <sup>d</sup>	98 <sup>d</sup>	176 <sup>b</sup>	31	369	48 (39-56)
V	11 <sup>e</sup>	8 <sup>e</sup>	32 <sup>e</sup>	53 <sup>e</sup>	265 <sup>b</sup>	369	72 (66-77)

Abbreviations: TM, traditional microscopy; WSI, whole-slide imaging.

<sup>a</sup> Consensus Melanocytic Pathology Assessment Tool and Hierarchy for Diagnosis (MPATH-Dx) reference diagnosis was obtained using the TM format; roman numerals indicate MPATH-Dx classes, with I indicating the least severe and V the most severe.

<sup>b</sup> Reference diagnosis is concordant with participating pathologists' interpretation.

<sup>c</sup> Reference diagnosis is class III; participating pathologists' interpretation as class IV or V: traditional 6%, digital 11%.

<sup>d</sup> Reference diagnosis is class IV; participating pathologists' interpretation as class I, II, or III: traditional 42%, digital 44%.

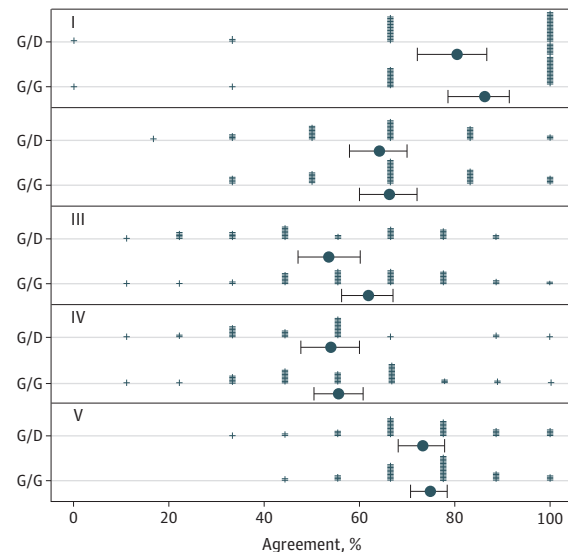
<sup>e</sup> Reference diagnosis is class V; participating pathologists' interpretation as class I, II, III, or IV: traditional 28%, digital 28%.

the format changed from TM to digital WSI format between phases (62.7%; 95% CI, 59.5%-65.7%) ( $P = .10$ ). Adjustment for dermatopathology board certification and/or fellowship training reduced the estimated effect of WSI on reproducibility and increased the associated  $P$  value to .24 (OR, 1.11; 95% CI, 0.92-1.33) (eFigure 2 in the Supplement). Intraobserver agreement was highest for class I melanocytic lesions for both the TM/TM arm (86.2%; 95% CI, 78.7%-91.4%) and TM/WSI arm (80.5%; 95% CI, 72.3%-86.7%) ( $P = .24$ ), and lowest for classes III and IV (Figure 2). None of the reproducibility analyses stratified by MPATH-Dx lesion class showed significant differences between the TM and digital WSI arms.

The small observed decrease in assessment confidence and increase in difficulty ratings among digital interpretations relative to traditional interpretations were not significant, and the estimated effect of digital WSI diminished when dermatopathology board certification was accounted for (eTable in the Supplement). The proportion of assessments considered to be borderline was slightly higher for digital than traditional interpretations (28% vs 22%). The estimated effect (borderline diagnosis without adjustment for dermatopathology board certification [OR, 1.37; 95% CI, 0.97-1.94] and with adjustment [OR, 1.37; 95% CI, 0.96-1.96];  $P = .08$ ) were unchanged by adjustment for dermatopathology board certification. The observed proportion of assessments for which a second opinion was desired was equivalent among traditional and digital interpretations (44%) (eTable in the Supplement).

Given the underlying variability in mitotic figures and Breslow depth between cases, we performed a case-level analysis of whether the average number of mitotic figures for TM

**Figure 2. Diagnosis Reproducibility for Melanocytic Lesions in Study Phase 2**

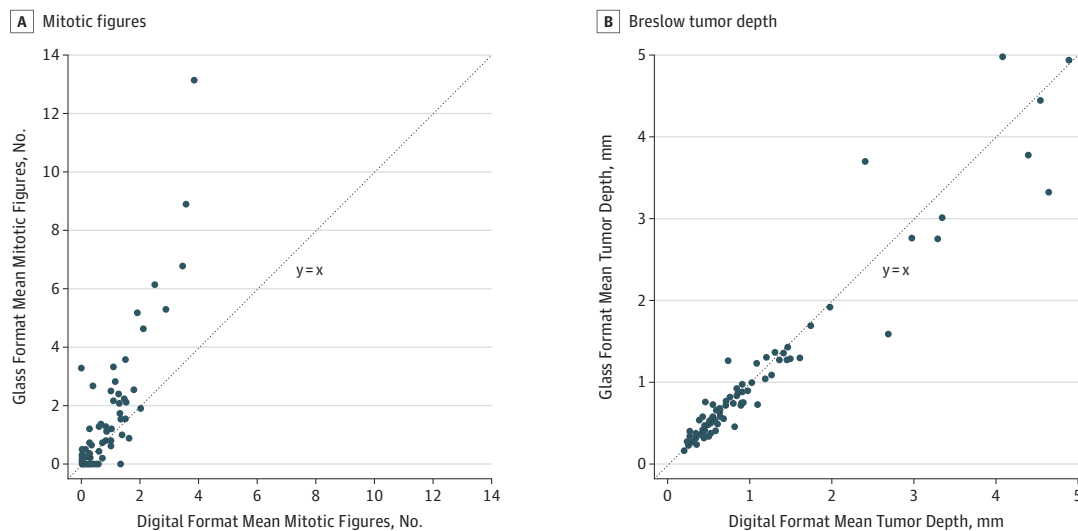


These results show that for all 5 Melanocytic Pathology Assessment Tool and Hierarchy for Diagnosis (MPATH-Dx) categories,<sup>14</sup> diagnosis reproducibility did not differ significantly by traditional microscopy (M) vs digital whole-slide imaging (D). Roman numerals indicate MPATH-Dx diagnostic categories; plus signs, participating pathologists' observed agreement; black dots, mean percent agreement; and error bars, SD for percent agreement.

was the same as for digital WSI, and whether Breslow depth differed by format. Mitotic rates tended to be higher in traditional interpretations (mean [SD]: TM, 6.72 [2.89]; WSI, 5.84



Figure 3. Case-Level Comparisons of Interpretive Features Using Traditional vs Digital Format



These figures show that there is a higher number of mitotic figures seen using traditional microscopy but no difference in Breslow tumor depth. A, The 63 points represent the 87 cases included; 21 cases had 0 mitotic rate for all

traditional and digital interpretations, and 4 other points represent 2 cases each. B, The 87 points represent the 87 cases included. All traditional-digital combinations of mean case-level Breslow depth are distinct.

[2.56];  $P = .002$ ). Assessment of mitotic rates were significantly different between the 2 formats, with traditional slides having almost twice as many mitotic figures detected as with digital (mean: TM, 1.20; WSI, 0.69;  $P = .02$ ). Case-level mean mitotic rates were increased in traditional relative to digital interpretations for 42 of 87 evaluable invasive melanoma cases, decreased in 23 cases, and equivalent in 22 cases (Figure 3A). No significant differences were found in Breslow depth between traditional and digital interpretation in the case-level analysis (Figure 3B).

## Discussion

To our knowledge, this study of melanocytic lesions is the largest to date in dermatopathology and provides the first evidence that interpretive accuracy and reproducibility are similar when comparing TM and WSI formats within the challenging and clinically important area of melanocytic skin lesions. The low levels of agreement for some of the intermediate classes of melanocytic skin lesions underscores the diagnostic challenges with these types of lesions.<sup>12</sup> The relatively small and insignificant differences in using TM and digital WSI are reassuring as implementation of WSI expands. However, for class III lesions such as a severe dysplastic nevus or melanoma in situ, the accuracy decreased when the second interpretation was in the digital format, with more repeat interpretations upgraded to a diagnosis of invasive melanoma.

A common view is that digital WSI pathology will overtake the field, much as it has in radiology, a notion that may be more likely given the FDA approval of WSI for primary interpretations in at least 1 system to date.<sup>19,20</sup> Thus, generating evidence of its comparative effectiveness is critically needed and is best served

with evidence specific to individual subspecialties, particularly those with high population impacts such as the dermatopathology of melanocytic lesions.

This study's results showing similar performance for both TM and WSI formats in the arena of melanocytic skin lesions support existing evidence. The findings are an important addition to those of a strong, prior study that reported viability of WSI for primary diagnosis in the field of dermatopathology.<sup>11</sup> However, this previous study noted concerning major clinically significant discrepancies for interpretation of melanocytic lesions that fell outside of the 4% noninferiority margin.<sup>11</sup> This previous study was limited to only 3 pathologists and 15 cases of invasive melanoma and thus was not adequately powered for a substudy of melanocytic lesions. In comparison, the present study of 87 pathologists and 180 melanocytic lesions (with 90 cases of invasive melanoma) had more power to examine these lesions. In addition, the WSI scanning magnification was  $\times 20$  in the previous study, while the present study used  $\times 40$ , which is thought to produce a superior image.<sup>21</sup>

Given the extent of variability noted among pathologists interpreting these challenging melanocytic lesions,<sup>12</sup> we also added a randomized study arm to compare intraobserver agreement between phase 1 and phase 2 interpretations. Specifically, we compared diagnostic interpretation of pathologists who first read TM and who then were randomized to either TM or WSI for the second phase of interpretation; thus we could compare agreement among the TM/TM arm vs the TM/WSI arm. Other unique strengths of this study compared with prior reports are its ability to adjust for clinical experience of the pathologists and its inclusion of data on assessment of mitoses and Breslow depth.

The findings demonstrate potential difficulty in detecting mitotic figures in digital format: the average mitotic rate in traditional images was twice that for mitoses identified in the digital format. The American Joint Committee on Cancer's newly revised and

evidence-based *Cancer Staging Manual, 8th Edition* removes mitogenicity as a microstage modifier,<sup>22</sup> but mitogenicity remains an important prognostic factor by multivariate analysis.

Class III melanocytic lesions have clinical significance when diagnosed as class I, IV, or V owing to the potential ramifications of treatment and management pathways. For example, a mildly dysplastic nevus (class I) diagnosed as melanoma in situ (class III) would likely be overtreated by excision, while invasive melanoma (class IV or V) diagnosed as class III would be undertreated without oncologic management and have the potential for a poor outcome. The study's findings demonstrate only a marginal decrease in accuracy with digital WSI compared with TM, so the clinical implications are uncertain. As use of digital WSI for primary diagnosis continues to expand, these findings may help shape development of machine learning, deep learning, or artificial intelligence algorithms that are increasingly being applied to image classification in pathology and radiology. Training these algorithms on the most challenging and/or discrepant cases, such as class III-like melanocytic lesions, may allow such cases to be triaged for more focused review among dermatopathologists, thus enhancing diagnostic yield, which in turn improves patient care.

### Limitations

Although, to our knowledge, this study is the first to report comparative performance of TM and digital WSI by such a large

and diverse group of participants (N = 87 pathologists) in the interpretation of melanocytic skin lesions, 2 potential limitations in the study design are notable. First, the test set environment does not fully replicate actual clinical practice. For example, the biopsy cases were enriched for more advanced diagnostic categories and were not representative of typical case prevalence distributions. Second, only a single slide was available for each case. However, if these factors had any effect on the results, we would expect that it would have the same effect on both TM and WSI interpretation.

### Conclusions

Overall, we found no clinically meaningful differences in diagnostic accuracy or reproducibility between pathologists randomized to TM vs digital WSI interpretation of melanocytic skin lesions, although a marginally lower accuracy with digital was seen for class III MPATH-Dx lesions (eg, melanoma in situ, severely dysplastic nevi). Results from this study add to the growing evidence of similarity of WSI interpretation compared with TM and provide large-scale evidence specifically for the diagnosis of melanocytic skin lesions.

#### ARTICLE INFORMATION

**Accepted for Publication:** May 29, 2018.

**Published Online:** August 15, 2018.

doi:10.1001/jamadermatol.2018.2388

**Author Affiliations:** Departments of Medicine and Community and Family Medicine, The Dartmouth Institute for Health Policy and Clinical Practice, Hanover, New Hampshire (Onega); Geisel School of Medicine at Dartmouth, Hanover, New Hampshire (Onega); Norris Cotton Cancer Center, Lebanon, New Hampshire (Onega); Department of Pathology, Institut Curie, Paris Sciences and Lettres Research University, and Faculty of Medicine University of Paris Descartes, Paris, France (Barnhill); Division of Dermatology, Department of Medicine, University of Washington School of Medicine, Seattle (Piepkorn); Dermatopathology Northwest, Bellevue, Washington (Piepkorn); Program in Biostatistics and Biomathematics, Fred Hutchinson Cancer Research Center, Seattle, Washington (Longton); Department of Pathology and Laboratory Medicine, Hospital of the University of Pennsylvania, Philadelphia (Elder); Center for Dermatoepidemiology, Providence VA Medical Center, Providence, Rhode Island (Weinstock); Departments of Dermatology and Epidemiology, Brown University, Providence, Rhode Island (Weinstock); Pathology Associates, Clovis, California (Knezevich); Department of Medicine, University of Washington School of Medicine, Seattle (Reisch, Radick); Department of Family Medicine, School of Medicine, Oregon Health & Science University, Portland (Carney); Department of Medical Informatics and Clinical Epidemiology, School of Medicine, Oregon Health & Science University, Portland (Nelson); Department of Medicine, School of Medicine, Oregon Health & Science University, Portland (Nelson); Providence Cancer Center, Providence Health and Services,

Portland, Oregon (Nelson); Department of Medicine, David Geffen School of Medicine at UCLA, Los Angeles, California (Elmore).

**Author Contributions:** Drs Onega and Elmore had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

**Study concept and design:** Onega, Barnhill, Elder, Carney, Nelson, Elmore.

**Acquisition, analysis, or interpretation of data:** Onega, Barnhill, Piepkorn, Longton, Elder, Weinstock, Knezevich, Reisch, Carney, Nelson, Radick.

**Drafting of the manuscript:** Onega, Reisch, Radick.

**Review and revision of manuscript:** Weinstock.

**Critical revision of the manuscript for important intellectual content:** Onega, Barnhill, Piepkorn, Longton, Elder, Weinstock, Knezevich, Reisch, Carney, Nelson, Elmore.

**Statistical analysis:** Onega, Longton.

**Obtained funding:** Carney, Elmore.

**Administrative, technical, or material support:** Onega, Piepkorn, Reisch, Carney, Radick.

**Study supervision:** Onega, Elmore.

**Conflict of Interest Disclosures:** Dr Barnhill reports a financial relationship with Myriad Genetics outside of the submitted work, and Dr Longton reports grants from Fred Hutchinson Cancer Research Center during the conduct of the study. No other disclosures are reported.

**Funding/Support:** Research reported in this publication was supported by the National Cancer Institute of the National Institutes of Health under award numbers R01 CA151306, R01 CA200690, and R01 CA201376.

**Role of the Funder/Sponsor:** The funders had no role in the design and conduct of the study; collection, management, analysis, and

interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

**Disclaimer:** The content is solely the responsibility of the authors and does not necessarily represent the views of the National Cancer Institute or the National Institutes of Health.

**Additional Contributions:** The authors thank Hamamatsu Photonics for use of the NanoZoomer 2.0-RS digital slide scanner C10730 WSI system. No one from Hamamatsu Photonics was compensated for their contributions.

**Additional Information:** All research described in this article is solely the product of the authors with no input from Hamamatsu Photonics. Melanocytic Pathology Assessment Tool and Hierarchy for Diagnosis (MPATH-Dx). 2014, 2018 University of Washington. All rights reserved.

#### REFERENCES

1. FDA allows marketing of first whole slide imaging system for digital pathology [news release]. Silver Spring, MD: US Food and Drug Administration; April 12, 2017. <https://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm552742.htm>. Accessed July 10, 2017.
2. Griffin J, Treanor D. Digital pathology in clinical use: where are we now and what is holding us back? *Histopathology*. 2017;70(1):134-145. doi:10.1111/his.12993
3. Ghaznavi F, Evans A, Madabhushi A, Feldman M. Digital imaging in pathology: whole-slide imaging and beyond. *Annu Rev Pathol*. 2013;8:331-359. doi:10.1146/annurev-pathol-011811-120902
4. Gamechangers. *Ulster Med J*. 2016;85(1):58-59.
5. García-Rojo M. International clinical guidelines for the adoption of digital pathology: a review of

- technical aspects. *Pathobiology*. 2016;83(2-3):99-109. doi:10.1159/000441192
6. 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(1):14. doi:10.4103/2153-3539.129452
7. Snead DR, Tsang YW, Meskiri A, et al. Validation of digital pathology imaging for primary histopathological diagnosis. *Histopathology*. 2016;68(7):1063-1072. doi:10.1111/his.12879
8. Vyas NS, Markow M, Prieto-Granada C, et al. Comparing whole slide digital images versus traditional glass slides in the detection of common microscopic features seen in dermatitis. *J Pathol Inform*. 2016;7:30. doi:10.4103/2153-3539.186909
9. Wilbur DC, Brachtel EF, Gilbertson JR, Jones NC, Vallone JG, Krishnamurthy S. Whole slide imaging for human epidermal growth factor receptor 2 immunohistochemistry interpretation: accuracy, precision, and reproducibility studies for digital manual and paired glass slide manual interpretation. *J Pathol Inform*. 2015;6:22. doi:10.4103/2153-3539.157788
10. Elmore JG, Longton GM, Pepe MS, et al. A randomized study comparing digital imaging to traditional glass slide microscopy for breast biopsy and cancer diagnosis. *J Pathol Inform*. 2017;8:12. doi:10.4103/2153-3539.201920
11. Kent MN, Olsen TG, Feeser TA, et al. Diagnostic accuracy of virtual pathology vs traditional microscopy in a large dermatopathology study. *JAMA Dermatol*. 2017;153(12):1285-1291. doi:10.1001/jamadermatol.2017.3284
12. Elmore JG, Barnhill RL, Elder DE, et al. Pathologists' diagnosis of invasive melanoma and melanocytic proliferations: observer accuracy and reproducibility study. *BMJ*. 2017;357:j2813. doi:10.1136/bmj.j2813
13. Lott JP, Boudreau DM, Barnhill RL, et al. Population-based analysis of histologically confirmed melanocytic proliferations using natural language processing. *JAMA Dermatol*. 2018;154(1):24-29. doi:10.1001/jamadermatol.2017.4060
14. Piepkorn MW, Barnhill RL, Elder DE, et al. The MPATH-Dx reporting schema for melanocytic proliferations and melanoma. *J Am Acad Dermatol*. 2014;70(1):131-141. doi:10.1016/j.jaad.2013.07.027
15. Carney PA, Reisch LM, Piepkorn MW, et al. Achieving consensus for the histopathologic diagnosis of melanocytic lesions: use of the modified Delphi method. *J Cutan Pathol*. 2016;43(10):830-837. doi:10.1111/cup.12751
16. 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. doi:10.5858/arpa.2011-0678-OA
17. Bauer TW, Slaw RJ, McKenney JK, Patil DT. Validation of whole slide imaging for frozen section diagnosis in surgical pathology. *J Pathol Inform*. 2015;6:49. doi:10.4103/2153-3539.163988
18. Pantanowitz L, Sinarid 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. doi:10.5858/arpa.2013-0093-CP
19. Bueno G, Fernández-Carrolles MM, Deniz O, García-Rojo M. New trends of emerging technologies in digital pathology. *Pathobiology*. 2016;83(2-3):61-69. doi:10.1159/000443482
20. Houghton JP, Smoller BR, Leonard N, Stevenson MR, Dornan T. Diagnostic performance on briefly presented digital pathology images. *J Pathol Inform*. 2015;6:56. doi:10.4103/2153-3539.168517
21. 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. doi:10.4103/2153-3539.93399
22. Gershenwald JE, Soong SJ, Balch CM; American Joint Committee on Cancer (AJCC) Melanoma Staging Committee. 2010 TNM staging system for cutaneous melanoma...and beyond. *Ann Surg Oncol*. 2010;17(6):1475-1477. doi:10.1245/s10434-010-0986-3