



HHS Public Access

Author manuscript

J Cutan Pathol. Author manuscript; available in PMC 2017 September 11.

Published in final edited form as:

J Cutan Pathol. 2016 October ; 43(10): 830–837. doi:10.1111/cup.12751.

Achieving consensus for the histopathologic diagnosis of melanocytic lesions: use of the modified Delphi method

Patricia A. Carney¹, Lisa M. Reisch², Michael W. Piepkorn^{3,4}, Raymond L. Barnhill^{5,6}, David E. Elder⁷, Stevan Knezevich^{8,9,10}, Berta M. Geller¹¹, Gary Longton¹², and Joann G. Elmore¹³

¹Professor of Family Medicine and of Public Health & Preventive Medicine, Oregon Health & Science University, Portland, OR, USA

²Department of Internal Medicine, University of Washington School of Medicine, Seattle, WA, USA

³Division of Dermatology, Department of Medicine, University of Washington School of Medicine, Seattle, WA, USA

⁴Dermatopathology Northwest, Bellevue, WA, USA

⁵Department of Pathology, University of California, Los Angeles, CA, USA

⁶Institut Curie and University of Paris René Descartes, Paris, France

⁷Department of Pathology, University of Pennsylvania, Philadelphia, PA, USA

⁸Department of Pathology, The Veterans Affairs Medical Center, Seattle, WA, USA

⁹Assistant Professor of Pathology, University of Washington, Seattle, WA, USA

¹⁰Pathology Associates, Clovis, CA, USA

¹¹Department of Family Medicine, University of Vermont, Burlington, VT, USA

¹²Fred Hutchison Cancer Research Center, Seattle, WA, USA

¹³Professor of Internal Medicine, University of Washington, Seattle, WA, USA

Abstract

Objective—To understand the sophisticated nature of coming to consensus when diagnosing complex melanocytic lesions among a panel of experienced dermatopathologists.

Methods—A total of 240 melanocytic lesions were assessed independently by three experienced dermatopathologists with their diagnoses mapped into one of five Melanocytic Pathology Assessment Tool and Hierarchy for Diagnosis (MPATH-DX) categories: (I) nevus/mild atypia, (II) moderate atypia, (III) severe atypia/melanoma *in situ*, (IV) T1a invasive melanoma and (V) T1b invasive melanoma. The dermatopathologists then discussed the cases, using a modified Delphi method to facilitated consensus building for cases with discordant diagnoses.

Results—For most cases, a majority of interpretations (two or three of three) agreed with the consensus diagnosis in 95% of Category I, 64% of Category II, 84% of Category III, 88% for Category IV and 100% of Category V cases. Disagreements were typically due to diagnostic threshold differences (64.5%), differing contents on slides even though the slides were sequential cuts (18.5%), and missed findings (15.3%). Disagreements were resolved via discussion of histopathologic features and their significance while reviewing the slides using a multi-headed microscope, considering treatment recommendations, citing existing literature, reviewing additional slides for a case, and choosing a provisional/borderline diagnosis to capture diverse opinions. All experienced pathologists participating in this study reported that the process of coming to consensus was challenging for borderline cases and may have represented compromise rather than consensus. They also reported the process changed their approaches to diagnosing complex melanocytic lesions.

Conclusions—The most frequent reason for disagreement of experienced dermatopathologists was differences in diagnostic thresholds related to observer viewpoints. A range of approaches was needed to come to consensus, and this may guide pathology groups who do not currently hold consensus conferences.

Keywords

dermatopathology; interpretive accuracy; melanocytic lesions

In the histopathologic diagnosis of melanocytic lesions, especially challenging cases exist such as distinguishing between what is interpreted by some as high-grade dysplasia vs. early melanoma and in the evaluation of atypical spitzoid lesions.¹⁻⁴ However, key questions about the diagnostic process remain, including why experienced dermatopathologists disagree on diagnoses and whether examining the consensus process in detail^{5,6} can reveal how diagnostic interpretations could be improved, especially for pathology groups not yet using consensus conferences. We undertook an observational study of experienced dermatopathologists' agreement when interpreting 240 melanocytic lesions to assess the types of disagreement that occurred, why they occurred, and the processes used to achieve consensus.

Methods

The University of Washington's institutional review board approved all study activities (IRB # 41700), which were Health Insurance Portability and Accountability Act (HIPAA) compliant. The study involved creating a reference diagnosis for each case in a test set designed to be used to assess concordance in a larger study. Three experienced dermatopathologists (hereafter called the reference panel) generated independent and final consensus diagnoses for each case, and the consensus process was captured for analysis. Members of the reference panel were senior faculty members at university-based medical centers, all with specialty training in pathology and/or dermatology. Two completed additional specialty training in melanocytic lesions, and all three reported interpreting these lesions for 20 years.

Test set development

Melanocytic skin lesions biopsied between January 1, 2010 and December 31, 2011 from patients aged 20 years and older were obtained from a dermatopathology practice in Washington State. Shave, punch and excisional biopsies were included, while consultative cases and re-excisions were excluded (Fig. 1). Based on the original pathologist's interpretation, eligible skin lesions (n = 27,481) were divided into five interpretive categories, according to corresponding clinical management groupings using the Melanocytic Pathology Assessment Tool and Hierarchy for Diagnosis (MPATH-Dx).⁷ The order of eligible patient cases was randomly generated within the five categories. Cases considered for inclusion in the study were selected from the top of each list until stratified diagnostic categories by patient age groups (20–49; 50–64; 65) were filled (Table 1). Five new sequentially cut slides were made for each case from tissue blocks. Other exclusions are outlined in Fig. 1. Members of the reference panel independently reviewed the glass slides on their own microscopes while blinded to others' interpretations and entered their findings into an online MPATH-Dx pathology assessment form, specifically designed for this purpose.⁷

Modified Delphi method

The Delphi method is a systematic, interactive approach of structuring group communication using a facilitator.^{8–11} We modified the traditional Delphi approach for this study because it was not possible to maintain anonymity following the initial round of independent interpretations. Following the independent case reviews, six full-day consensus meetings occurred between February 17, 2012 and August 24, 2012. The panel reviewed the histopathologic features of all 240 final test cases using a multi-headed microscope. Each panel member provided the rationale for their initial interpretation and critiqued the interpretations of their colleagues. In some cases, the discussions involved consideration of treatment recommendations, use of existing literature, review of additional slides for the case or choice of a provisional diagnosis to capture diverse opinions. A Consensus Data Collection form recorded standardized data on reasons for disagreement and how consensus was achieved. Due to decision fatigue, some complex cases required several discussions across meetings to come to final consensus. The trained facilitator with extensive expertise (author PC) ensured that during the discussion of the cases no single pathologist dominated the discussion and that each was allowed to voice their views on the case and discuss them in a balanced way. She also confirmed the accuracy of decisions made as a validation step prior to moving on to the next case or deciding to revisit the case during another consensus meeting.

Data analysis

Contents on 37 cases reviewed by the reference panel members during their independent review differed enough that calculating agreement in the analytic file would artificially skew results. Two additional cases were removed from evaluation because independent assessments from at least one reference panel member were not available due to technical problems with the web-based data collection system. These 39 cases were removed from the statistical analysis on agreement. For the remaining 201 cases, kappa coefficients were

computed among the three panel members and across broad diagnostic categories for clinically significant interpretive disagreements that would result in treatment differences per MPATH-Dx.⁷

Results

The final composition of cases according to patient age is outlined in Table 1. The test set was weighted for severe atypia and melanoma compared with common dermatopathology practice. Table 2 depicts the full extent of agreement and where disagreements occurred among panel members; this ranges from no agreement to complete agreement according to the diagnostic categories. Pairwise kappa coefficients ranged from 0.54 to 0.61 with a three-way kappa of 0.58 [95% confidence interval (CI): 0.52–0.64] indicating moderate concordance.¹²

Reasons for agreement or disagreement were categorized as: (i) *No disagreement*: all three panel members agreed on the case diagnosis within major MPATH-Dx diagnostic categories; (ii) *differences in terms/nomenclature*: differing use of terms resulted in apparent disagreement; (iii) *threshold differences*: features of a given case resulted in a subjective determination as to whether the degree of abnormalities was sufficient for a particular diagnostic category or grade of atypia; (iv) *missed finding*: one or more panel members overlooked a feature on the slide that contributed to a different diagnosis by other panel members; (v) *contents on slides differ*: the slides were not comparable, even though each panel member had sequential sections of the block and (vi) *philosophical/conceptual differences*: a difference in style of interpretive thinking derived from training and/or experience in clinical practice. The extent of disagreement represented either minor non-clinically significant disagreements within the five diagnostic categories or clinically significant differences crossing one or more of them.

Overall, panel members achieved agreement, with all three independent diagnoses agreeing, for 115 of the 239 cases (48.1%) and disagreeing on 124 (51.9%) (Table 3). Agreement was highest for nevus/mild atypia and invasive melanoma T1b (56 and 75.4%, respectively) and was lowest for moderate atypia (14.3%). Agreement for severe atypia/melanoma *in situ* and T1a melanoma were similar (45.0 and 39.7%, respectively). In 17 cases (13.7%), the disagreement was due to more than one reason. Cases with disagreements had differences in threshold noted (80/124 cases or 64.5%), differing contents on slides (23/124 or 18.5%), missed finding by one or more panel members (19/124 or 15.3%) and philosophical/conceptual differences (17/124 or 13.7%).

Resolution of disagreements always involved a re-review of the case with discussion among panel members (Table 4). For mild atypia cases, nearly all disagreements were resolved by re-reviewing slides and discussing them. For both moderate and severe atypia, panel members additionally chose to indicate that the case was of a borderline phenotype. Much of the basis for this interpretive variability stemmed from threshold differences. For invasive melanoma cases, disagreements were often addressed with review of additional slides to confirm or rule out mitotic figures.

Discussion

This study is novel in its detailed examination of types and causes of diagnostic disagreements and how they can be resolved for a set of melanocytic lesions in a test set. This approach provides an important model for how consensus could be achieved, as we found that members of the expert panel undertook several approaches to come to agreement that could be applied when obtaining second opinions, such as considering treatment recommendations or consulting recent literature.

Our findings for overall interpretive agreement among experienced pathologists were similar to existing literature,¹⁻⁴ which is reassuring for determining the presence or absence of AJCC 7th Edition Stage 1b or greater primary melanoma. Lower levels of agreement occurred across diagnostic categories of lesser severity but have treatment consequences, especially for re-excisions. Threshold differences (subjectively set levels at which histopathologic observations are considered sufficiently abnormal to constitute definitive criteria for classification) were the most common cause for disagreement, especially for distinguishing between moderate atypia vs. severe atypia/melanoma *in situ*; determining presence or absence of microinvasive melanoma; and confirming or ruling out mitotic forms for discriminating AJCC Stage 1a vs. 1b melanoma.

The act of coming to consensus reflects a collective processing of differing observers' opinions regarding asserted histopathologic findings (e.g. whether disorder among junctional nests of melanocytes and nuclear enlargement and pleomorphism in junctional melanocytes are sufficient for a diagnostic category of moderate vs. mild dysplasia; whether a small superficial aggregate of atypical cells is attached to or detached from the epidermis, such as *in situ* vs. invasive melanoma; or whether a basophilic structure is an apoptotic cell vs. a mitotic form, such as AJCC Stage 1a vs. 1b melanoma). The consensus process started with identification of disagreements from independent review, followed by successive discussions and histopathologic reviews to assess the verities or absence of them from one observer's point of view relative to the collective perception. In instances where the strength of asserted verities waned, other points of view correspondingly gained credibility. Following iterative cycles, group consensus often emerged.

The panel did report the consensus process likely changed their interpretive practices going forward. For example, one member was surprised by the inclusion of junctional spitzoid proliferations into the differential diagnosis for some lesions where that member had favored a diagnosis of severe dysplasia. Although some level of disagreement had been expected, members were surprised by the extent of variation in diagnostic terminology revealed by the study. When the process was complete, participants were comfortable with the consensus diagnoses achieved in most cases.

An inherent limitation in histopathologic interpretation of many melanocytic lesions is lack of information about their biological nature. As evidenced here, the histopathologic findings in many lesions are quite subtle and can be prone to a non-reproducible range of interpretations, which are difficult to resolve into discrete categories. Further, because of the lack of objective tools for assessment, observers may use different or variably weighted

histopathologic and cytological attributes ('criteria') for interpretation and grading of lesions, which results in inevitable discordance for many lesions. Resolving all or most disagreements sometimes required more than one review and discussion period on different days to mitigate decision fatigue, which is another important phenomenon that has received little attention in pathology interpretation. Differences in the content between slides despite being sequential cuts off the same case accounted for some disagreements; the frequency of this occurrence was unexpected, which culminated in a sub-study to examine significant differences in mitotic counts on sequential tissue cuts.¹³

In cases intractable to complete consensus, the utility of assigning provisional assessment terms proved useful to convey a level of diagnostic uncertainty. These assessment terms (e.g. MELTUMP and SAMPUS)⁷ reflect limitations in the current criteria-based diagnostic paradigms that await future advancements in molecular testing or other technologies for definitive resolution as to the actual biological nature of the melanocytic proliferations. Clearly, challenges remain in elucidating complex lesions.

The strengths of this study are the detailed data we captured on diagnostic agreement for over 200 melanocytic lesions, particularly regarding types of disagreement and how consensus was achieved. Limitations include weighting of the test set with higher grades of abnormality than is common in dermatopathology practice and that the true biological nature of grey-zone proliferations is unknown. Achieving consensus does not translate into an understanding of the true biological nature of grey-zone proliferations. Other limitations that should lead to caution in generalizing these results include that only a single slide was examined for the initial reading, and additional special studies used to support diagnostic interpretations (including the use of deeper levels) were not available.

In conclusion, diagnostic disagreement exists across the spectrum of melanocytic lesions, even among experienced dermatopathologists. The majority of differences in diagnoses were threshold differences, often resolved on group re-review of slides, using existing literature, consideration of treatment options, and use of provisional assessment terms (e.g. Melanocytic Tumors of Uncertain Malignant Potential (MELTUMP) and Superficial Atypical Melanocytic Proliferation of Uncertain Significance (SAMPUS), melanocytic neoplasm – indeterminate). If feasible, discussing difficult cases systematically, as we describe here, using consideration of treatment options, current literature and discussion of terminology about borderline lesions could enhance discussion related to second opinions, may provide an important mechanism for resolving diagnostic discordance, especially for pathology groups that do not routinely use consensus conferences. When consensus proves to be impossible to achieve for a single diagnosis, the use of terminology that reflects the uncertainty of the diagnosis may be the most appropriate resolution of the process, with agreement for treatment options often being easier to achieve than for any specific diagnosis, and with patient safety and appropriate management being the most important considerations.

Acknowledgments

This work was supported by the National Cancer Institute (R01 CA151306). 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.

References

1. Braun RP, Gutkowitz-Krusin D, Rabinovitz H, et al. Agreement of dermatopathologists in the evaluation of clinically difficult melanocytic lesions: how golden is the 'gold standard'? *Dermatology*. 2012; 224:51. [PubMed: 22433231]
2. Ferrara G, Argenziano G, Soyer HP, et al. Dermoscopic and histopathologic diagnosis of equivocal melanocytic skin lesions - An interdisciplinary study on 107 cases. *Cancer*. 2002; 95:1094. [PubMed: 12209696]
3. Edwards SL, Blessing K. Problematic pigmented lesions: approach to diagnosis. *J Clin Pathol*. 2000; 53:409. [PubMed: 10911797]
4. Morales-Callaghan AM, Castrodeza-Sanz J, Martínez-García G, Peral-Martínez I, Miranda-Romero A. Correlation between clinical, dermoscopic, and histopathologic variables in atypical melanocytic nevi. *Actas Dermosifiliogr*. 2008; 99:380. [PubMed: 18501170]
5. Jones J, Hunter D. Qualitative Research: consensus methods for medical and health services research. *BMJ*. 1995; 311:376. [PubMed: 7640549]
6. Wolf, FM. Meta-analysis: quantitative methods for research synthesis. Sage Publishing; Thousands Oaks, CA: 1986. p. 59
7. Piepkorn MW, Barnhill RL, Elder DE, et al. The MPATH-Dx reporting schema for melanocytic proliferations and melanoma. *J Am Acad Dermatol*. 2014; 70:131. [PubMed: 24176521]
8. Linstone, HA., Turoff, M. [accessed on 6 April 2013] The Delphi method: techniques and applications. 2002. <http://is.njit.edu/pubs/delphibook/delphibook.pdf>
9. Sinha IP, Smyth RL, Williamson PR. Using the Delphi technique to determine which outcomes to measure in clinical trials: recommendations for the future based on a systematic review of existing studies. *PLOS Med*. 2011; 8(1):e1000393. [accessed on 14 June 2016] <http://www.plosmedicine.org/article/info%3Adoi%2F10.1371%2Fjournal.pmed.1000393>. [PubMed: 21283604]
10. Mason R, Schwartz J. Using a Delphi method to develop competencies: the case of domestic violence. *Community Med Health Edu*. 2012; 2:2. [accessed 6 April 2013]doi: 10.4172/jcmhe.1000124
11. Boulkedid R, Abdoul H, Loustau M, Sibony O, Albeti C. Using and reporting the Delphi method for selecting healthcare quality indicators: a systematic review. *PLoS One*. 2011; 6:e20476.doi: 10.1371/journal.pone.0020476 [PubMed: 21694759]
12. Viera AJ, Garrett JM. Understanding inter-observer agreement: the kappa statistic. *Fam Med*. 2005; 37:360. [PubMed: 15883903]
13. Knezevich S, Barnhill RL, Elder DE, et al. Variability in mitotic figures in serial sections of thin melanomas. *J Am Acad Dermatol*. 2014; 71:1204. [PubMed: 25239732]

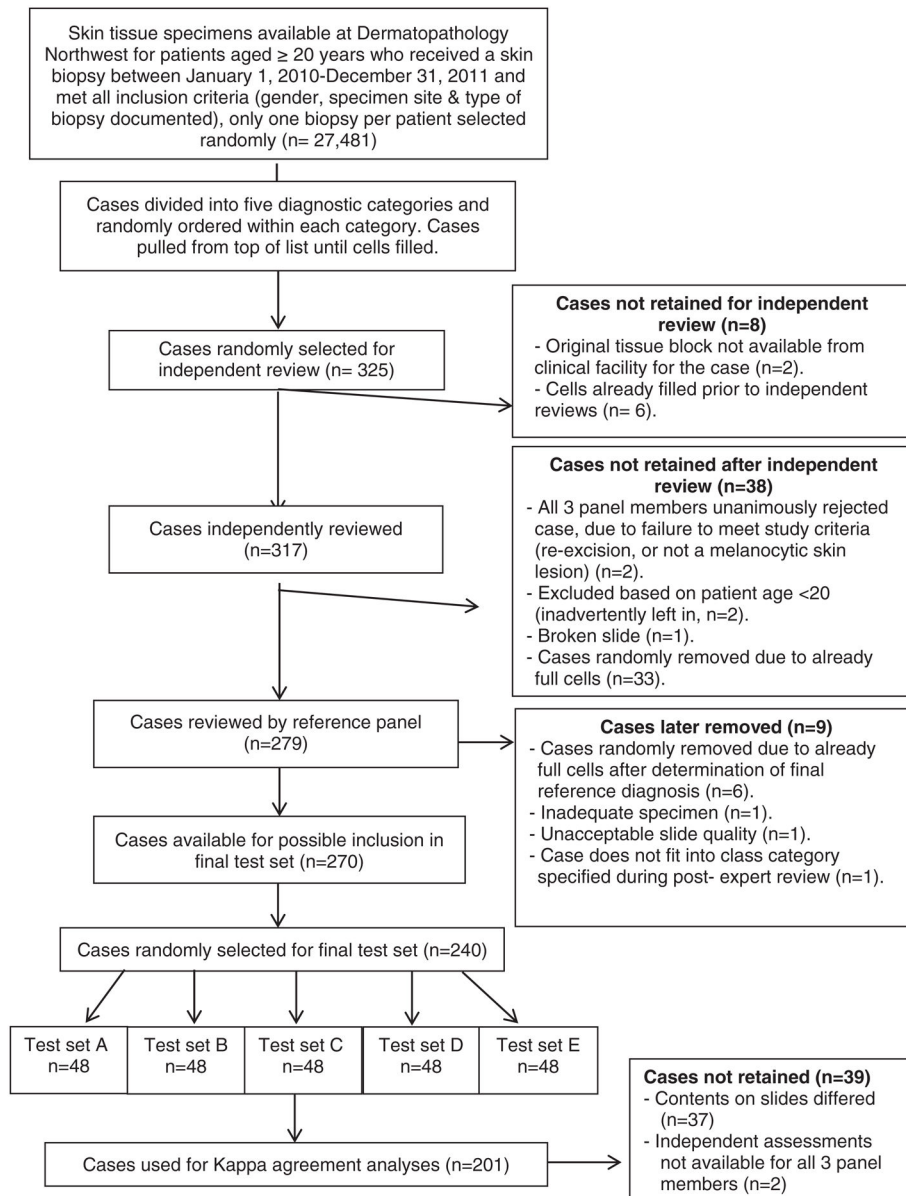


Fig. 1.
Case selection and refinement process.

Composition of test set cases according to the MPATH-Dx diagnostic category and patient age

Table 1

MPATH-Dx diagnostic assessment class	Suggested Clinical Management per MPATH-Dx ⁸	Age 20–49	Age 50–64	Age 65+	Total
(I) Nevus/mild atypia	No further treatment required	9	8	8	25
(II) Moderate atypia	Consideration of narrow but complete re-excision or follow-up observation	14	11	11	36
(III) Severe atypia/melanoma <i>in situ</i>	Repeat excision with at least 5 mm margins	20	21	19	60
(IV) T1a melanoma	Wide excision	19	20	19	58
(V) T1b melanoma	Wide excision with additional treatment	20	21	20	61
Totals by age		82	81	77	240

MPATH-Dx, Melanocytic Pathology Assessment Tool and Hierarchy for Diagnosis.

Table 2

Comparison of three-panel member's independent reviews and final consensus interpretations according to MPATH-Dx diagnostic class (n = 201)

	Interpretations from independent reviews					Interpretations from consensus reviews				
	Nevus/mild atypia, n = 22 N (%)	Moderate atypia, n = 33 N (%)	Severe atypia/melanoma <i>in situ</i> , n = 55 N (%)	T1a melanoma, n = 42 N (%)	T1b melanoma, n = 49 N (%)					
Nevus/mild atypia										
0 of 3 pathologists agreed	0	15 (45)	53 (96)	42 (100)	49 (100)					
1 of 3 pathologists agreed	1 (5)	15 (45)	2 (4)	0	0					
2 of 3 pathologists agreed	8 (36)	3 (9)	0	0	0					
All 3 pathologists agreed	13 (59)	0	0	0	0					
Moderate atypia										
0 of 3 pathologists agreed	16 (73)	2 (6)	35 (64)	41 (98)	49 (100)					
1 of 3 pathologists agreed	5 (23)	10 (30)	15 (27)	1 (2)	0					
2 of 3 pathologists agreed	1 (5)	16 (48)	5 (9)	0	0					
All 3 pathologists agreed	0	5 (15)	0	0	0					
Severe atypia/melanoma <i>in situ</i>										
0 of 3 pathologists agreed	19 (86)	16 (48)	1 (2)	24 (57)	48 (98)					
1 of 3 pathologists agreed	3 (14)	13 (39)	8 (15)	13 (31)	1 (2)					
2 of 3 pathologists agreed	0	4 (12)	21 (38)	5 (12)	0 (0)					
All 3 pathologists agreed	0	0	25 (45)	0	0 (0)					
T1a melanoma										
0 of 3 pathologists agreed	22 (100)	33 (100)	42 (76)	0	44 (90)					
1 of 3 pathologists agreed	0	0	13 (24)	5 (12)	5 (10)					
2 of 3 pathologists agreed	0	0	0	17 (40)	0					
All 3 pathologists agreed	0	0	0	20 (48)	0					
T1b melanoma										
0 of 3 pathologists agreed	22 (100)	33 (100)	55 (100)	39 (93)	0					
1 of 3 pathologists agreed	0	0	0	3 (7)	0					
2 of 3 pathologists agreed	0	0	0	0	6 (12)					
All 3 pathologists agreed	0	0	0	0	43 (88)					

MPATH-Dx, Melanocytic Pathology Assessment Tool and Hierarchy for Diagnosis.

Kappa coefficients: Pair 1 – 0.54 (absolute agreement – 63.7%); Pair 2 – 0.59 (absolute agreement – 67.2%); Pair 3 – 0.61 (absolute agreement – 70.2%); 3-way overall – 0.58 (95% CI: 0.52–0.64).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2 shows the full extent of agreement and where disagreements lie among the reference diagnosis panel members, ranging from no agreement to complete agreement according to diagnostic classes. The diagonal cells shaded in grey indicate agreement within each diagnostic category and the unshaded cells indicate where levels of agreement occurred across categories rather than within categories. Thus, for 95% of the cases with a reference consensus diagnosis of nevus/mild atypia cases, two or all three panel members interpreted those cases as such during their initial independent reviews, and none of them interpreted those cases as T1a or T1b melanoma. The majority of pathologists on the panel (two of three or all three of three) were also in agreement with the consensus diagnosis for severe atypia/melanoma *in situ* cases (83%), T1a melanoma cases (88%), and T1b melanoma cases (100%). The lowest level of agreement was for moderate atypia, where for 63% of cases the majority of panel members agreed that the case was moderate atypia in their initial interpretations. Interpretations for this category tended to cross over into severe atypia/melanoma *in situ*.

Table 3
 Level of agreement and types of disagreements from independent review according to MPATH-Dx lexicon (n = 239)

	Reference consensus diagnosis				
	Nevus/mild atypia (n = 25)	Moderate atypia (n = 35)	Severe atypia/melanoma <i>in situ</i> (n = 60)	T1a melanoma (n = 58)	T1b melanoma (n = 61)
All three pathologists agreed – n = 115, n (%)	14 (56.0%)	5 (14.3%)	27(45.0%)	23 (39.7%)	46 (75.4%)
The three pathologists did not agree – n = 124, n (%)	11 (44.0%)*	30 (85.7%) [‡]	33 (55.0%) [‡]	35 (60.3%) [§]	15 (24.6%) ^{//}
Types of disagreement					
Different terms/nomenclature	1	0	1	0	1
Threshold differences	8	27	24	18	3
Missed finding	0	1	2	10	6
Contents on slides differ	2	2	4	8	7
Philosophical/conceptual differences	2	3	8	4	0

MPATH-Dx, Melanocytic Pathology Assessment Tool and Hierarchy for Diagnosis.

* One case had more than 1 type of disagreement.

[‡] Three cases had more than 1 type of disagreement.

[§] Seven cases had more than 1 type of disagreement.

[§] Four cases had more than 1 type of disagreement.

^{//} Two cases had more than 1 type of disagreement.

Table 4
Resolution of disagreements according to final diagnoses and interpretive category

	How disagreements were resolved				
	Re-review with discussion	Consider treatment recommendations	Use of existing literature	Reviewing additional slides for case	Choosing a provisional/ borderline diagnosis to capture diverse opinions
Nevus/mild atypia cases (n = 11 cases; 13 types of disagreement) Reasons for disagreement					
Different terms/nomenclature	1	0	0	0	0
Threshold differences	8	0	0	0	1
Missed finding	0	0	0	0	0
Contents on slides differ	2	0	0	2	0
Philosophical/conceptual differences	2	0	1	0	1
Totals	13	0	1	2	2
Resolution disagreements for moderate atypia cases (n = 30 cases; 33 types of disagreement) Reasons for disagreement					
Different terms/nomenclature	0	0	0	0	0
Threshold differences	27	5	3	2	9
Missed finding	1	0	0	0	0
Contents on slides differ	2	0	0	1	1
Philosophical/conceptual differences	3	0	0	1	1
Totals	33	5	3	4	11
Severe atypia/melanoma <i>in situ</i> cases (n= 33 cases; 39 types of disagreement) Reasons for disagreement					
Different terms/nomenclature	1	0	0	0	0
Threshold differences	24	0	4	1	15
Missed finding	2	0	0	0	0
Contents on slides differ	4	0	0	2	3
Philosophical/conceptual differences	8	0	1	1	5
Totals	39	0	5	4	23
T1a melanoma cases (n = 35 cases; 40 types of disagreement) Reasons for disagreement					
Different terms/nomenclature	0	0	0	0	0
Threshold differences	18	0	1	3	4
Missed finding	10	0	0	1	1
Contents on slides differ	8	0	1	8	0

	How disagreements were resolved				
	Re-review with discussion	Consider treatment recommendations	Use of existing literature	Reviewing additional slides for case	Choosing a provisional/ borderline diagnosis to capture diverse opinions
Philosophical/conceptual differences	4	0	1	1	1
Totals	40	0	3	13	6
T1b melanoma cases (n = 15 cases; 17 types of disagreement) Reasons for disagreement					
Different terms/nomenclature	1	0	0	1	0
Threshold differences	3	0	1	1	0
Missed finding	6	0	1	3	0
Contents on slides differ	7	0	1	7	0
Philosophical/conceptual differences	0	0	0	0	0
Totals	17	0	3	12	0