



Published in final edited form as:

J Cutan Pathol. 2018 August ; 45(8): 588–596. doi:10.1111/cup.13265.

Influence of variability in assessment of Breslow thickness, mitotic rate and ulceration among US pathologists interpreting invasive melanoma, for the purpose of AJCC staging

Laura Taylor¹, Kyle Hood², Lisa Reisch³, Joann Elmore⁴, Michael Piepkorn^{5,6}, Raymond Barnhill⁷, Stevan Knezevich⁸, Andrea Radick³, and David Elder⁹

¹Department of Dermatology, Hospital of the University of Pennsylvania, Philadelphia, Pennsylvania ²Bureau of Economic Analysis, Suitland, Maryland ³Department of Medicine, University of Washington School of Medicine, Seattle, Washington ⁴David Geffen School of Medicine at UCLA, Los Angeles, California ⁵Division of Dermatology, Department of Medicine, University of Washington School of Medicine, Seattle, Washington ⁶Dermatopathology Northwest, Bellevue, Washington ⁷Department of Pathology, Institut Curie Paris Sciences and Lettres Research University, and Faculty of Medicine University of Paris Descartes, Paris, France ⁸Pathology Associates, Clovis, California ⁹Department of Pathology and Laboratory Medicine, Hospital of the University of Pennsylvania, Philadelphia, Pennsylvania

Abstract

Background: Melanoma staging has depended on depth of invasion (Breslow thickness, BT), mitotic rate (MR) and ulceration. In anticipation of the AJCC's eighth edition, variability in pathologists' assessment of these factors and consequently in tumor staging was assessed.

Methods: One-hundred and fifteen cases of invasive melanoma, established by a consensus panel, were assessed by 187 pathologists. Variation was studied in BT, the detection of mitotic figures, and ulceration. The sources of this variation and its effect on tumor staging are considered.

Results: On average, participant assessments closely approached consensus BT. Greater variation was identified in the classification of mitogenicity, which (like ulceration) upstages a T1 melanoma from T1a to T1b in the seventh but not eighth edition. In cases with a T1a diagnosis by the consensus panel, 15.6% of participants identified one or more mitotic figures (indicative of a false positive); and in cases diagnosed as T1b by the consensus panel, 32.0% of participants failed to find mitotic figures (false negative).

Correspondence David E. Elder, MB ChB, FRCPA, Perelman School of Medicine at the University of Pennsylvania, 6 Founders, 3400 Spruce Street, Philadelphia, PA 19104. david.elder@uphs.upenn.edu.

The views expressed in this article are those of the authors and do not reflect the views of the Bureau of Economic Analysis or Department of Commerce. This research was done solely on the author's personal time and did not employ the use of any government resources.

Conflict of interest

The authors declare no relevant potential conflict of interests.

Conclusion: Variability in the staging of T1 melanoma among pathologists when using the AJCC seventh edition criteria is closely related to the detection of mitotic figures, with BT playing a less prominent role. Decreased variability is expected after implementation of the eighth edition.

Keywords

dermatopathology; melanocytic lesions; melanoma

1 | INTRODUCTION

The incidence of melanoma has been increasing, nearly tripling, between 1975 and 2011, but the reasons for this apparent increase, whether natural or artefactual, have not been elucidated.¹ About 64% of incident cases have a Breslow thickness (BT) of less than 1 mm, and are therefore defined as thin melanomas.² Although thin melanomas are generally associated with good prognoses, 15% of melanoma deaths documented in SEER (Surveillance, Epidemiology, and End Results cancer registry) resulted from thin melanoma metastases.^{2,3} Variation in outcomes in thin melanomas has been well-studied and has greatly influenced the staging system of the American Joint Committee on Cancer (AJCC). Until recently, the 3 key staging factors, which we will term *diagnostic observations*, were Breslow thickness (BT), presence of ulceration and mitotic rate (MR). According to the seventh edition AJCC staging guidelines, T1 lesions have a depth of less than or equal to 1.00 mm with the distinction that category T1a shows neither ulceration nor mitotic figures while T1b shows one or both. The remaining stages (T2-T4) are defined by BT; with the qualifiers of “a” or “b” dependent on the absence or presence of ulceration, respectively. In the eighth edition, stage T1 is defined as having a BT of <0.8 mm, and ulceration is the only stage modifier defining stages T1a and T1b.

Tumor thickness has long been established as the most important histologic predictor of patient outcomes since Breslow’s landmark study.⁴⁻⁶ Breslow demonstrated that lesions less than 0.76 mm in thickness rarely metastasized, and numerous studies have shown that survival is closely related to BT. Mitotic rate has also been well studied as a prognostic indicator, the second most powerful predictor of survival after BT, and was incorporated into the seventh AJCC staging criteria.^{5,7} A significant decrease in survival was found between patients with 0 and those with 1 or more mitotic figures.³ It was recommended that sentinel lymph node (SLN) biopsies be offered to patients with T1b staging due to the 4% decreased survival at 10 years.⁶ Given the AJCC’s newly revised and evidence-based eighth edition, which limits T1 melanomas to a BT of 0.8 rather than 1.0 mm, and removes mitotic rate (MR) as a stage modifier,⁸ it is helpful to assess the effects of removal of MR as it relates to staging using an additional national sample and methodology.

While it is known that there is substantial interobserver variability in measurement of the key diagnostic and prognostic observations,⁹⁻¹¹ the specific clinical importance of such variability as we move to using the AJCC eighth edition is unknown. The present study aims to assess and understand sources of discordance in melanoma staging and the potential impact on clinical care. Variability of staging observations—including BT, mitotic figures and ulceration—is compared among participants and in reference to a consensus panel with

particular focus on observations which cross AJCC-defined staging thresholds, leading to differences in the designated stage. Additionally, since guidelines for “thin” melanoma staging in the seventh edition use cutoffs of 1.00/1.01 and 0.75/0.76 mm for BT and 1/mm² for MR, we consider whether rounding preferences prevail in reporting of BT, and whether these might induce participants to cross diagnostic thresholds. In the eighth edition, only a single digit after the decimal point is used.

2 | MATERIAL AND METHODS

A total of 240 melanocytic skin lesion patient cases were reviewed by a consensus panel of 3 experienced dermatopathologists to reach consensus on both diagnosis and treatment recommendation, as previously described.¹² In particular, agreement on mitotic rates was assessed for each case by the 3 panelists and another pathologist at a multi-headed microscope.¹³ Five slide sets of melanocytic skin lesions were then identified (48 patient cases each, for a total of 240 cases).

M-Path study methods have been previously described.^{9,14–17} In brief, 187 pathologists from 10 geographically diverse states enrolled in the M-Path study. This study was approved by the Institutional Review Boards of the University of Washington, Fred Hutchinson Cancer Research Center, Oregon Health & Science University, Rhode Island Hospital, and Dartmouth College. Participating pathologists provided informed consent. Participation in the study required each pathologist to review 1 of the 5 slide sets developed by the consensus panel. Eligible participants were those who had completed their pathology training (residency and/or fellowship), interpreted melanocytic skin biopsies within the previous year, and expected to continue interpreting melanocytic skin lesions for the next 2 years. Participants completed an online survey to gather standardized information on pathologists’ characteristics and clinical experience and then interpreted a slide set of 48 patient cases (22–24 of which were defined as invasive melanomas by the consensus panel), providing their diagnostic interpretations, BT measurements, mitotic counts, presence of ulceration, treatment recommendations, confidence in their assessment and difficulty level for each case into an online histology form (MPATH-Dx).¹⁸

2.1 | Analytical methods

The objective of our analysis was to identify and understand variability in the diagnosis of melanoma. Therefore, we confined the statistical analysis to 115 cases deemed by both the consensus panel and at least 1 participant to be invasive melanoma (seventh edition AJCC stages T1a, T1b or T2+).

Based on the AJCC seventh edition, we first summarized participant and consensus panel assessments separately for each stage (as defined by the consensus panel) and observation of features (BT, MR and ulceration). We report the mean and SD of BT and the distributions of the other 2 measures (MR is divided into 3 groups¹⁹ comparing the distributions of the observations among both the consensus panel and the participants, we determined the most likely reasons that staging varies between participants and the consensus panel: Statistical bias in reporting (differences in means of these observations), or variation in reporting (large

standard deviations in participant assessments). We also assessed the variability of these measures (BT, MR and ulceration).

We then assigned assessments into partitions that are defined by diagnostic criteria; cases are first categorized by the reference seventh edition AJCC stage. Within each reference AJCC stage, participant assessments are placed into subcategories defined by BT: {BT > 1.00 vs BT < 1.01}. Within these subcategories, lesions are further classified by mitotic figures and ulceration. By separating assessments into these categories, we were able to identify the sources of diagnostic discordance.

We next assessed the participant characteristics that are associated with discordance of these diagnostic observations between the consensus panel and participants. We measured discordance according to classifications of BT ($\{>1.00 \text{ vs } <1.01\}$) and MR ($\{<1 \text{ vs } 1\}$), separately; because ulceration is rarely reported in these thin melanomas it was omitted from this analysis). We then correlated discordance with a collection of participant characteristics (years of experience, percent of melanocytic lesions in practice, fellowship training, overall confidence in diagnosing melanocytic lesions), as well as participant-reported assessment-specific characteristics (confidence in diagnosis, desire for second opinion and difficulty of diagnosis). Each of the 2 outcome measures was correlated with each participant characteristic measure using a separate univariate linear regression, with robust standard errors clustered by participant to correct for within-participant correlation of un-observables.

We then assessed whether preferences for rounding affected measures of BT and potentially diagnostic concordance. Participants are asked to report BT to the second decimal place. If a 0 or 5 appear as the second decimal in a large number of assessments, this suggests a preference for rounding: Each digit (0–9) should appear with equal probability (0.10). We determined digit preference by computing the frequencies at which each digit is selected. Because assessments are not independent of each other (assessments may be correlated within participants), we bootstrapped confidence intervals for these frequencies, redrawing at the participant level.

3 | RESULTS

One hundred and eighty-seven participants provided interpretations of 115 cases of invasive melanoma, resulting in a total of 2985 assessments. Characteristics of M-Path study participants have been previously described,⁹ and are also depicted in the Appendix. The means and standard deviations of BT and summaries of the distributions of MR and ulceration are shown separately by seventh edition AJCC stage in Table 1 for the 115 cases and 2985 independent interpretations, with results shown for participants, independent experts and consensus panel results. Means of BT measurements are strikingly similar by stage, (0.47 for participants, 0.46 for experts and 0.44 for the consensus panel; 0.81 for participants, 0.78 for experts and 0.67 for the consensus panel; and 2.18 for participants, 2.23 for experts and 2.18 for the consensus panel; for stages T1a, T1b and T2+, respectively). Participant and expert standard deviations are higher than the consensus panel, the former representing variation both across lesions and among participants/consensus panel, the latter only variation across lesions. For T1a lesions, experts show little variation in

BT not explainable by variation across lesions (0.19 standard deviations of BT in both cases), whereas participants show more variation (0.37). However, variation among experts and among participants is similar for T1b lesions. Neither participants nor experts showed much variation in selecting BT for T2+ lesions beyond that implied by lesional characteristics.

Participants observed at least 1 mitotic figure in 15.6% of assessments of the consensus reference T1a lesions (12.3% of assessments with 1 MR and 3.4% with >1 MR), and did not observe a mitotic figure in 32.0% of assessments of reference defined T1b lesions. There was a tendency to undercount mitotic figures relative to the consensus, and counts seemed to vary across participants, especially among T1b lesions. This suggests that the detection of mitotic figures may be an important source of discordance in the diagnosis of melanocytic lesions.

Ulceration was rarely reported by participants and experts when interpreting the reference defined T1a and T1b lesions. It was not a major factor in treatment-oriented staging in either case, as most appeared in T2+ lesions for which SLN biopsy is routinely indicated.

Table 2 shows results from the partitioned data. The table is divided into 3 separate “sub-tables,” 1 for each of the seventh edition AJCC stages (T1a, T1b, T2+) as defined by the reference diagnosis. Rows show 3 strata that are defined by participant assessments: $BT > 1$, $BT = 1$ and all assessments. Within these, columns show the number (and percent of assessments within the seventh edition AJCC stage) of assessments for which participants reported mitoses (none or at least 1), ulceration (yes/no), either or neither. Bins for which participants disagree with the reference panel diagnoses are colored: Overdiagnosis is red, underdiagnosis blue.

This table shows the major sources of misdiagnosis. In 1068 assessments on cases with a reference stage T1a, 175 assessments with $BT < 1.01$ identified ulceration or at least 1 MF. In 17 assessments, BT was measured as being greater than 1, although in 6 of these MF were observed which would have yielded discordance otherwise. Thus, 16.4% of assessments were up-staged to T1b, and 1.0% to T2+. In 908 assessments on cases with reference stage T1b, 252 assessments (27.8%) with participant $BT < 1.01$ also failed to identify mitoses or ulceration. In 1009 assessments on reference T2 + cases, 93 assessments were a $BT < 1.01$; however, only 28 of these assessments (2.8%) yielded discordance as neither mitoses nor ulceration were observed. This table suggests that the major source of discordance in classification of T1 melanoma as T1a or T1b is not measurement of BT, but in mitotic counts. The distributions of BT for each case, ordered by mean assessment is shown in Figure 1. Blue dots represent the median of the reference panel’s observation, while 2 bars represent the second and third quartiles of the data (25th percentile to median and median to 75th percentile). Thus, 50% of the data are contained within the 2 bars, 25% above, and 25% below. The previous 1 mm BT cutoff is shown as a blue line, while the new 0.8 mm BT cutoff line is shown in green. It is clear from these data that while some cases show more variability across participants than others and larger average differences from the reference diagnoses than do others, the participants report BT measurements that largely follow those reported by the expert panel. In particular, it should be noted that in only 3 cases does the

dot appear on the other side of the 1 mm line from the bar, suggesting that crossing this diagnostic threshold is rare when the cutoff is 1.0 mm. When the cutoff is 0.8 mm, there is somewhat greater variation. Specifically, in a total of 5 cases, the experts consider BT to be > 0.8 , and 0.8 is within the bar (indicating that more than 25% of participants consider BT < 0.8); in 4 cases, the opposite is true, where experts consider BT is < 0.8 , and more than 25% of participants consider BT to be > 0.8 .

Approximately 120 participants reported ulceration in the same number of assessments as the consensus, while about 25 reported ulceration in 1 fewer assessment and more than 25 in 1 more assessment. While some participants reported ulceration in more cases than others, the modal participant agreed with the consensus in the number of ulcerations seen. Over all, over-reporting of ulceration is more common than under-reporting, but reports of ulceration were not driven by only a few participants.

Participant characteristics and assessment characteristics that are associated with BT and mitotic rate concordance with the reference panel are shown in Table 3. The coefficients are interpreted as follows: A value of 0.05 would indicate that if predictor variable increases by 1, the participant's rate of agreement with the reference diagnosis increases by 5 percentage points. Positive values indicate greater agreement is associated with higher values of the covariate (eg, a positive coefficient for confidence in assessing melanocytic lesions would indicate that higher confidence in assessing melanocytic lesions is associated with greater agreement with the expert panel).

We focus here mainly on statistically significant coefficients. Participant reporting of more melanocytic lesions seen in their respective practices is associated with better agreement in mitotic figure counts ($P = .004$). Yet number of years interpreting such lesions does not show a pattern. Dermatopathology fellowship training is associated with agreement in mitotic counts ($P < .001$) and BT ($P = .015$). Confidence in interpreting melanocytic lesions as reported by the participant is associated with mitotic figure agreement ($P = .012$). When a participant reports a desire for a second opinion on a particular assessment, agreement is lower for mitotic figures ($P < .001$). When a participant reports in an assessment that the case is difficult to diagnose, agreement in mitotic figure counts is lower ($P < .001$). Finally, confidence in a particular assessment is associated with mitotic figure agreement ($P < .001$). Overall, experience, confidence and training appear to be associated with better agreement with the reference panel. These patterns hold up more often for agreement with mitotic figure counts than for agreement with BT measurements. This is important because most often, it is mitotic counts that are the source of diagnostic discordance, as established above.

Table 4 shows the frequency of participant choices of the second digit in their assessments of BD, along with bootstrapped confidence intervals. Observations in red highlight the second digits with a frequency of use larger than 0.10. Participants were significantly more likely to report BD using second digits of 0 or 5. Digits 0 and 5 are selected substantially more often than expected for all stages of lesions ($P < .01$). Rounding to 0 appears to be less common in thin lesions than in thick lesions.

Figure 2 shows the distribution of BD, truncated on the left at 0.65 and on the right at 1.1. The height of each bar represents the percent of assessments for which the participant measured a BD equal to the value of the horizontal axis, limited to cases in which participants reported BD between 0.65 and 1.1. For example, in 7% of assessments, participants selected 0.75, while about in 1% 0.76 was selected. In nearly 10% of cases, participants selected 1.00, while in only 1% was 1.01 selected. Large spikes can be observed at each measure ending in a 0 digit or a 5 digit. This visual information confirms the intuition of the results in Table 4.

4 | DISCUSSION

The seventh edition AJCC staging guidelines for invasive melanomas were based on BT, mitotic rate and ulceration. With the introduction of the newly revised eighth edition guidelines, it is important to assess how these revisions and the changes in status of individual factors (eg, mitotic rate and BT) will influence staging agreement consistency. This is of great importance since treatment recommendations are based on the designated stage, and errors in staging could have severe consequences for patients. As shown in the literature (and the present study) most of this discordance occurs within the category of thin melanomas. This is particularly problematic since the treatment guidelines for these lesions may differ. For example, the presence of a T2+ lesion indicates the patient should receive sentinel lymph node biopsy. However, seventh edition AJCC guidelines recommend the sentinel lymph node biopsy for T1b lesions but not for T1a, with the only distinction between these lesions being ulceration or 1 (subsequently revised in NCCN guidelines to “multiple”)^{19,20} mitotic figure (s).

Our study identifies significant variability in the diagnostic observations reported by participants, in particular with regard to mitotic rate. This variability affects melanoma staging in a substantial number of cases, with upstaging of up to 18% of reference T1a assessments and down-staging of up to 28% of reference T1b assessments (Table 2). The most important source of variability in staging appears to be in the recognition of mitotic figures. In contrast, despite some variability among observers, the variance of BT rarely crosses staging thresholds. When the BT does cross a threshold, the resulting variance would usually only affect indicated treatments in cases where pathologists also differ in observations of ulceration and/or mitotic rate. There is somewhat greater variance around the 0.8 mm threshold adopted in the eighth edition. It is possible that the 1.0 mm threshold, being well known to the pathology readers, has been measured with greater attention to precision, suggesting that the same attention might be paid to the 0.8 mm threshold in the future.

Agreement between the reference panel and participants in reporting of BT and mitotic figures is correlated with participant and assessment characteristics. Concordance in reporting of mitotic rates is positively affected by the number of lesions seen in practice, dermatopathology fellowship training and confidence in assessing melanocytic lesions, as well as high assessment-specific confidence, and negatively associated with a high perception of diagnostic difficulty, low assessment-specific confidence and the desire for a second opinion ($P < .01$ in all cases). Discordance in measurement of BT shows a similar

trend, although magnitudes and significance levels are not as strong. Because of the importance of mitotic figures in staging discordance, it is likely that any or all of these pathologist-specific and assessment-specific factors may influence classification and staging and thus treatment recommendations. Although experience and training are correlated with concordance, in cases where high discordance exists the pathologist was significantly more likely to send the case for a second opinion (this is observed in Table 3), and therefore this may not represent the discordance rates in practice.

Participants show a clear tendency to report BT with numbers that end in a 0 or 5 in the hundreds place ($P < .01$), indicating a preference for rounding. This tendency is stronger, however, in thick melanomas ($BT > 1.00$), suggesting that participants may be paying closer attention when diagnostic thresholds affect treatment. Even though BT was not necessarily the cause of a large degree of discordance in the present study, these results suggest that care should be exercised in any redefinition of diagnostic cutoffs.

The present study shows that variation in the reporting of mitotic rate is the principal factor responsible for discordance in the staging of T1 melanoma using seventh edition AJCC system, and that this variation appears reduced with the eighth edition system. However, there is greater variation in staging around the new cutoff of 0.8 mm. This variance might be attributed to reviewer characteristics, such as paying closer attention to known staging cutoff values. However, future studies should assess in greater detail how observer backgrounds or practices may account for this discordance. For example, does a dermatopathology-trained reviewer have a keener eye or a more disciplined/standardized approach to the search? Machine learning and image analysis systems are also being developed, and antibodies are available to assist in the detection of mitotically active cells,^{19,20} although these methods have not yet been perfected and not yet adopted by consensus groups like the AJCC.

The issue of the identification of mitotic figures becomes much less important in staging decisions in the new eighth edition AJCC system.⁸ Nevertheless, mitotic figures are of importance not only in staging but also in diagnosis of melanocytic tumors and in estimation of prognosis²¹, so that many of the above considerations remain applicable.

Most of the discordance between the consensus panel and participants in staging of invasive melanoma in the AJCC 7th Edition system occurs within the category of thin melanomas (Table 5). The major source of discordance is up-staging of T1a lesions to T1b, or downstaging of T1b lesions to T1a, based on disagreement about mitotic figures, while variation in BT does not result in substantial staging discordance. Disagreement in the recognition of mitotic figures is related to the confidence, experience, and training of the participant performing the assessment. These issues are mitigated in the eighth edition staging system. Substantial preferences for rounding appear to be present in BT measurements, suggesting that staging thresholds should be chosen carefully. This issue may also be mitigated in the new system, wherein BT is rounded to 1 number after the decimal point, instead of 2 as in the seventh edition system. Further research gathering new data examining variability in eighth edition staging factors is recommended.

ACKNOWLEDGEMENTS

Supported by the National Cancer Institute (R01CA151306 and R01CA201376) The funding agency had no role in the study design; in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the article for publication. We thank the study participants for their commitment to improving clinical care in dermatopathology.

Funding information

National Cancer Institute, Grant/Award Numbers: R01CA151306R01CA201376, R01CA201376, R01CA151306

APPENDIX

TABLE A1

Self-reported characteristics of M-Path study pathologists who completed the baseline survey ($N=187$)

Physician characteristics	<i>N</i> (%)
Demographics	
Age (years)	
<40	31 (16.6%)
40–49	56 (29.9%)
50–59	63 (33.7%)
60	37 (19.8%)
Gender	
Female	73 (39.0%)
Male	114 (61.0%)
Training and experience	
Affiliation with academic medical center	
No	134 (71.7%)
Yes, adjunct/affiliated	34 (18.2%)
Yes, primary appointment	19 (10.2%)
Residency	
Anatomic/clinical pathology	168 (89.8%)
Dermatology	15 (8.0%)
Both dermatology and anatomic/clinical pathology	4 (2.1%)
Training	
Board certified or fellowship trained in dermatopathology ^a	74 (39.6%)
Other board certification of fellowship training ^b	113 (60.4%)
Years interpreting melanocytic skin lesions	
<5	29 (15.5%)
5–9	45 (24.1%)
10–19	57 (30.5%)
20	56 (29.9%)
Percent of caseload interpreting melanocytic skin lesions	
<10%	79 (42.2%)
10–24%	72 (38.5%)
25–49%	28 (15.0%)

Physician characteristics	N (%)
50%	8 (4.3%)
Average number of melanoma cases (melanoma in situ and invasive melanoma) interpreted per month	
<5	82 (43.9%)
5–9	47 (25.1%)
10	58 (31.0%)
Average number of benign melanocytic skin lesions interpreted per month	
<25	54 (28.9%)
25–49	32 (17.1%)
50–149	51 (27.3%)
150	50 (26.7%)
Considered an expert in melanocytic skin lesions by colleagues	
No	108 (57.8%)
Yes	79 (42.2%)
Feelings/thoughts about interpreting melanocytic skin lesions	
In general, how challenging do you find melanocytic skin lesions to interpret?	
Challenging	179 (95.7%)
Easy	8 (4.3%)
Interpreting melanocytic skin lesions makes me more nervous that other types of pathology	
Agree	129 (69.0%)
Disagree	58 (31.0%)
In general, how confident are you in your assessments of melanocytic skin lesions?	
Confident	161 (86.1%)
Not confident	26 (13.9%)

^aThis category consists of physicians with single or multiple fellowships that include dermatopathology. Also includes physicians with single or multiple board certifications that include dermatopathology.

^bOther includes fellowships or board certifications in surgical pathology, cytopathology, hematopathology, etc.

REFERENCES

1. Howlader N, Noone AM, Krapcho M, et al. SEER Cancer Statistics Review, 1975–2013 Bethesda, MD: National Cancer Institute; 2016.
2. Gimotty P, Botbyl J, Soong SJ, Guerry D. A population-based validation of the American Joint Committee on Cancer Melanoma Staging System. *J Clin Oncol* 2005;23(31):8065–8075. [PubMed: 16258105]
3. Gimotty PA, Elder DE, Fraker DL, et al. Identification of high-risk patients among those diagnosed with thin cutaneous melanomas. *J Clin Oncol* 2007;25(9):1129–1134. [PubMed: 17369575]
4. Thickness Breslow A., cross-sectional areas and depth of invasion in the prognosis of cutaneous melanoma. *Ann Surg* 1970;172(5): 902–908. [PubMed: 5477666]
5. Balch CM, Gershenwald JE, Soong SJ, et al. Final version of 2009 AJCC melanoma staging and classification. *J Clin Oncol* 2009;27(36): 6199–6206. [PubMed: 19917835]
6. Piris A, Mihm MC, Jr, Duncan LM. AJCC melanoma staging update: impact on dermatopathology practice and patient management. *J Cutan Pathol* 2011;38(5):394–400. [PubMed: 21385199]
7. Thompson JF, Soong SJ, Balch CM, et al. Prognostic significance of mitotic rate in localized primary cutaneous melanoma: an analysis of patients in the multi-institutional American joint committee on Cancer melanoma staging database. *J Clin Oncol* 2011;29(16):2011.

8. Gershenwald JE, Scolyer RA, Hess KR, et al. Melanoma staging: evidence-based changes in the American joint committee on Cancer (AJCC) eighth edition Cancer staging manual. *CA Cancer J Clin* 2017; 67:472–492. [PubMed: 29028110]
9. 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. [PubMed: 28659278]
10. Elder DE. Pathology of melanoma. *Surg Oncol Clin N Am* 2015;24(2): 229–237. [PubMed: 25769708]
11. Niebling MG, Haydu LE, Karim RZ, Thompson JF, Scolyer RA. Reproducibility of AJCC staging parameters in primary cutaneous melanoma: an analysis of 4,924 cases. *Ann Surg Oncol* 2013; 20(12):3969–3975. [PubMed: 23851608]
12. 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. [PubMed: 27247109]
13. Knezevich SR, Barnhill RL, Elder DE, et al. Variability in mitotic figures in serial sections of thin melanomas. *J Am Acad Dermatol* 2014;71(6): 1204–1211. [PubMed: 25239732]
14. Carney PA, Frederick PD, Reisch LM, et al. How concerns and experiences with medical malpractice affect dermatopathologists' perceptions of their diagnostic practices when interpreting cutaneous melanocytic lesions. *J Am Acad Dermatol* 2016;74(2):317–324.e8. [PubMed: 26559597]
15. Onega T, Weaver D, Geller B, et al. Digitized whole slides for breast pathology interpretation: current practices and perceptions. *J Digit Imaging* 2014;27(5):642–648. [PubMed: 24682769]
16. Lee K, Peacock SW, Zhao G, et al. Variation among pathologists' treatment suggestions for melanocytic lesions: a survey of pathologists. *J Am Acad Derm* 2016;76(1):121. [PubMed: 27692732]
17. Zhao G, Lee K, Peacock S, et al. The utilization of Spitz-related nomenclature in the histological interpretation of cutaneous melanocytic lesions by practicing pathologists: Results from the M-Path Study. *J Cutan Pathol* 2017;44(1):5–14. [PubMed: 27686456]
18. 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. [PubMed: 24176521]
19. Andres C, Andres-Belloni B, Hein R, et al. iDermatoPath - a novel software tool for mitosis detection in H&E-stained tissue sections of malignant melanoma. *J Eur Acad Dermatol Venereol* 2017;31(7): 1137–1147. [PubMed: 28107565]
20. Ottmann K Detection of mitotic figures in thin melanomas-- immunohistochemistry does not replace the careful search for mitotic figures in hematoxylin-eosin stain. *J Am Acad Dermatol* 2015;73(4): 637–644. [PubMed: 26278815]
21. Valdebran M, Elbendary A, Chaitanya Arudra SK, Torres KM, Elattar I, Elston DM. Nuclear and cytoplasmic features in the diagnosis of banal nevi, Spitz nevi, and melanoma. *J Am Acad Dermatol* 2016;75(5): 1032–1037.e8. [PubMed: 27542585]

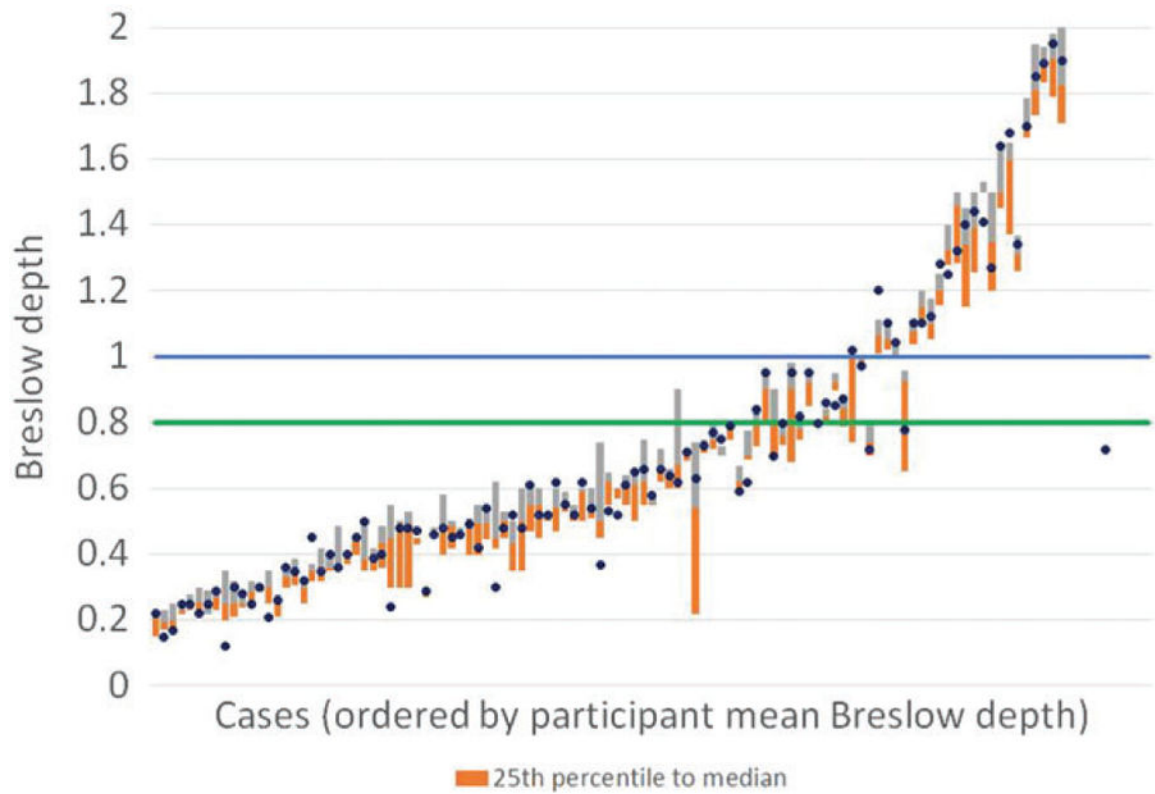


FIGURE 1.

The orange bar extends from the 25th percentile to the median, while the grey bar extends from the median up to the 75th percentile. The blue dots represent the medians of the Breslow depths reported by the members of the expert panel for each case. The blue line represents the AJCC seventh edition T1/T2 cutoff at 1.00. The green line represents the AJCC eighth edition T1/T2 cutoff at 0.8 mm. For ease of interpretation, the figure is truncated at a Breslow depth of 2.00

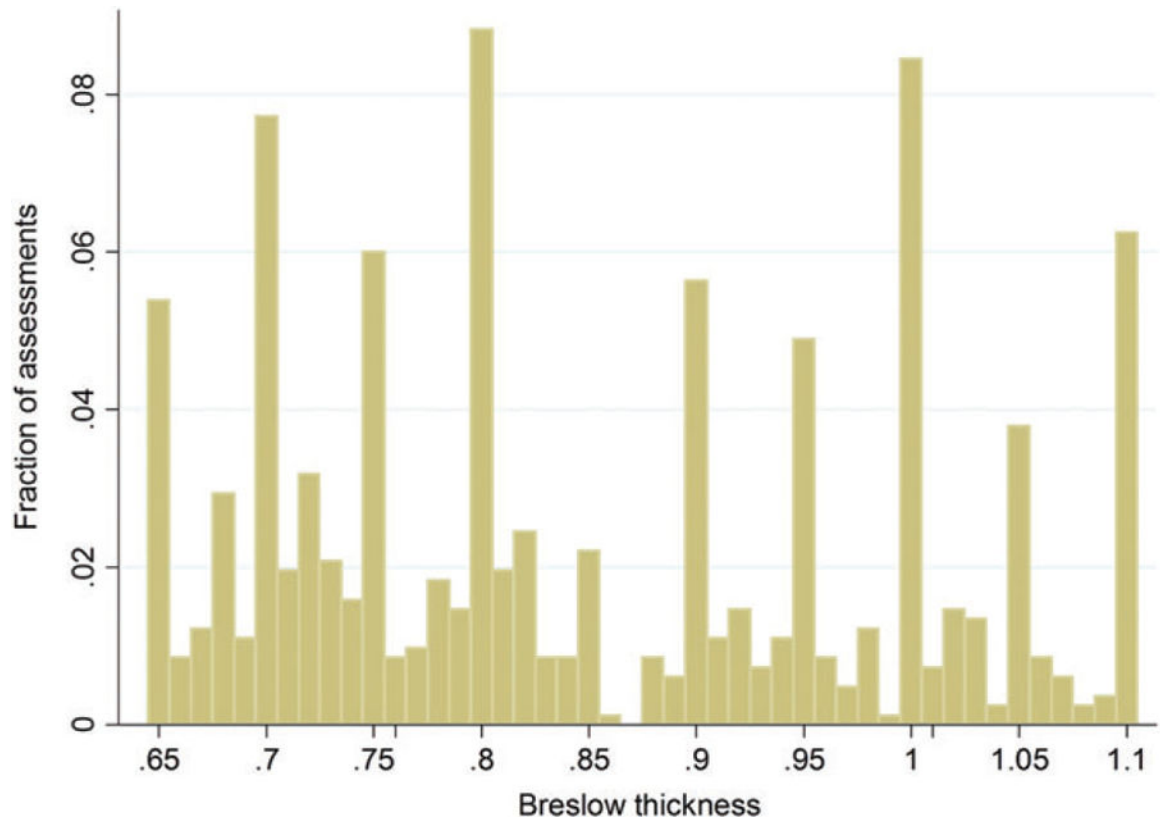


FIGURE 2.

The height of each bar represents the fraction of assessments for which the participant measured a Breslow depth equal to the value of the horizontal axis, limited to cases in which participants reported Breslow depth between 0.65 and 1.1

Summary of participant and expert reporting of Breslow depth, mitotic figure counts and the presence of ulceration. Interpretations provided by 187 US pathologists on 115 cases (N = 2985 total independent interpretations)

TABLE 1

Breslow thickness (BT)									
	Participant			Expert			Consensus		
	Mean	St. Dev.	N (interpretations)	Mean	St. Dev.	N (interpretations)	Mean	St. Dev.	N (cases)
Consensus stage (AJCC) ^a									
T1a	0.47	0.37	1068	0.46	0.19	131	0.44	0.19	54
T1b	0.81	0.67	908	0.78	0.60	86	0.67	0.16	30
T2+	2.18	1.46	1009	2.23	1.40	90	2.18	1.39	31
Mitotic rate (MR)									
	Participant			Expert			Consensus		
	MR = 0	MR = 1	MR > 1	MR = 0	MR = 1	MR > 1	MR = 0	MR = 1	MR > 1
Consensus stage (AJCC) ^a									
T1a	84.4%	12.3%	3.4%	92.4%	6.1%	1.5%	—	—	—
T1b	32.0%	35.0%	32.9%	16.3%	43.0%	40.7%	—	50.0%	50.0%
T2+	14.8%	22.5%	62.7%	3.3%	22.2%	74.4%	3.2%	25.8%	71.0%
Ulceration									
	Participant			Expert			Consensus		
	None	Observed		None	Observed		None	Observed	
Consensus stage (AJCC) ^a									
T1a	98.3%	1.7%		99.2%	0.8%		—	—	—
T1b	94.6%	5.4%		93.0%	7.0%		96.7%	3.3%	
T2+	70.7%	29.3%		75.0%	25.0%		74.2%	25.8%	

^aConsensus stage based on the American Joint Committee on Cancer seventh edition cancer staging manual.

Sources of misdiagnosis for lesions categorized as invasive melanoma both by the reference panel and the participant. Underdiagnosis by participant diagnosis as compared to the reference panel is indicated by the color blue and overdiagnosis is indicated by the color red^a

TABLE 2

		Participant assessments reporting							
		Ulceration	Mitoses	Either	Neither	Ulceration	Mitoses	Either	Neither
Total assessments		N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)
2.1: Reference diagnosis: T1a^b									
Participant Breslow depth < 1.01	1051	17 (1.6%)	163 (15.3%)	175 (16.4%)	876 (82.0%)				
Participant Breslow depth > 1.00	17	1 (0.1%)	6 (0.6%)	6 (0.6%)	11 (1.0%)				
All Breslow depths (total)	1068	18 (1.7%)	169 (15.8%)	181 (16.9%)	887 (83.1%)				
2.2: Reference diagnosis: T1b^c									
Total assessments		N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)
Participant Breslow depth < 1.01	835	43 (4.7%)	566 (62.3%)	583 (64.2%)	252 (27.8%)				
Participant Breslow depth > 1.00	73	6 (0.7%)	55 (6.1%)	55 (6.1%)	18 (2.0%)				
All Breslow depths (total)	908	49 (5.4%)	621 (68.4%)	638 (70.3%)	270 (29.7%)				
2.3: Reference diagnosis: T2+^d									
Total assessments		N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)
Participant Breslow depth < 1.01	93	15 (1.5%)	63 (6.2%)	65 (6.4%)	28 (2.8%)				
Participant Breslow depth > 1.00	916	271 (26.9%)	809 (80.2%)	821 (81.4%)	95 (9.4%)				
All Breslow depths (total)	1009	286 (28.3%)	872 (86.4%)	886 (87.8%)	123 (12.2%)				

^aFor each assessment in which participants and experts selected a diagnosis of invasive melanoma, this table summarizes Breslow depth, mitotic counts and ulceration reported by participants. Three panels show consensus invasive melanoma stages (T1a, T1b and T2+). Blue numbers indicate underdiagnosis, in which the reference diagnosis indicates SLN biopsy (per AJCC 7e) but the participant diagnosis does not. Red numbers indicate overdiagnosis, in which the reference diagnosis indicates no SLN biopsy but the participant diagnosis does.

^bFor a consensus diagnosis T1a, 100% of assessments have consensus BD < 1.01 and neither mitoses nor ulceration.

^cFor a reference diagnosis of T1b, 100% of assessments have consensus BD < 1.01 and either mitoses or ulceration.

^dFor a consensus diagnosis of T2+, 100% of assessments have reference BD > 1.00.

TABLE 3

Agreement with reference panel, covariate analysis (*P* values in parentheses)^a

Pathologist characteristics	Agreement—participant reporting same range in measure as reference panel	
	Mitotic rate agreement	Breslow thickness agreement
	MR < 1, MR = 1 ^b	BT > 1.00, BT < 1.01 ^c
Percent of melanocytic lesions in practice (larger = higher percentage)	0.0684 (0.004)	0.0118 (0.255)
Years interpreting melanocytic lesions (larger = more years)	-0.0362 (0.085)	0.00378 (0.693)
Fellowship (1 = dermpath, 0 = other/none)	0.0723 (0.000)	0.0227 (0.015)
Confidence in interpreting melanocytic lesions (larger = more confident)	0.0600 (0.012)	0.0132 (0.206)
Ask for second opinion (1 = yes, 0 = no)	-0.0799 (0.000)	-0.0148 (0.132)
Level of diagnostic difficulty (larger = more difficult)	-0.0895 (0.000)	-0.00429 (0.642)
Confidence in assessment (larger = more confident)	0.0906 (0.000)	0.0163 (0.073)

^a*P* value in parentheses (standard errors are clustered by participant ID). Each cell represents a separate model. "Agreement" refers to agreement with the reference panel.

^b"Mitotic rate agreement" is defined as 1 if both the participants and consensus identified either no mitotic figures or at least 1, and 0 otherwise.

^c"Breslow thickness agreement" is defined as 1 if both the participants and consensus reported Breslow depth above 1 or below 1, and 0 otherwise.

TABLE 4

Frequencies of digit choice, rounded to hundredths by stage^a

Second digit	T1a		T1b		T2+	
	Mean	Conf. interval	Mean	Conf. interval	Mean	Conf. interval
0	0.312	(0.270,0.354)	0.331	(0.294,0.369)	0.512	(0.468,0.557)
1	0.051	(0.038,0.065)	0.064	(0.048,0.080)	0.034	(0.020,0.048)
2	0.090	(0.069,0.110)	0.097	(0.075,0.118)	0.058	(0.042,0.075)
3	0.051	(0.038,0.065)	0.045	(0.030,0.060)	0.049	(0.034,0.063)
4	0.051	(0.035,0.068)	0.057	(0.041,0.073)	0.046	(0.033,0.058)
5	0.206	(0.178,0.234)	0.216	(0.187,0.244)	0.169	(0.144,0.195)
6	0.056	(0.041,0.071)	0.056	(0.039,0.074)	0.025	(0.014,0.036)
7	0.054	(0.042,0.067)	0.044	(0.030,0.058)	0.034	(0.021,0.047)
8	0.086	(0.070,0.102)	0.053	(0.035,0.071)	0.057	(0.042,0.073)
9	0.041	(0.029,0.054)	0.036	(0.023,0.049)	0.016	(0.007,0.024)

Digits 0 and 5 are highlighted in boxes

Red indicates digits with choice frequencies significantly greater than 10%

^a95% confidence intervals are reported.

Comparison of mitotic counts between the consensus panel and combined participant assessments

TABLE 5

	Participant assessments			Participant agreement		
	MR < 1	MR = 1	MR > 1	MR < 1	MR = 1	MR > 1
Combined						
Consensus						
MR = 0	1069	149	37	85.2%	11.9%	2.9%
MR = 1	418	246	148	51.5%	30.3%	18.2%
MR > 1	292	332	786	20.7%	23.5%	55.7%
Multiple mitoses						
Reference: "Multiple"						
MR = {0,1}	1719	266	266	86.6%		13.4%
MR 2	624	837	837	42.7%		57.3%
Single mitosis						
Reference: Current AJCC^a						
MR = 0	1069	248	248	81.2%		18.8%
MR 1	564	1565	1565	26.5%		73.5%

^aConsensus stage based on the American Joint Committee on Cancer seventh edition cancer staging manual.