Global and mitosis-specific interobserver variation in mitotic count scoring and implications for malignant melanoma staging

Saldanha, Gerald; Ali, Rokiah; Bakshi, Arti; Basiouni, Ahmed; Bishop, Rachael; Colloby, Peter

Published in:
Histopathology

DOI:
10.1111/his.14052

Publication date:
2020

Document Version
Peer reviewed version

DR GERALD S SALDANHA (Orcid ID : 0000-0002-1910-020X)

Article type : Original Article

Global and mitosis-specific interobserver variation in mitotic count scoring and implications for malignant melanoma staging

Gerald Saldanha, [1, 2] Rokiah Ali, [3] Arti Bakshi, [4] Ahmed Basiouni, [5] Rachael Bishop, [2] Peter Colloby, [6] Paul Craig, [7] Philip Da Forno, [2] Sara Edward, [8] Olivia Espinosa de los Monteros, [9] Alan Evans, [10] Lynne Jamieson, [11] Ed Rytina, [12] Mark Bamford [2]

[1] Leicester Cancer Research Centre, University of Leicester. [2] University Hospitals of Leicester NHS Trust. [3] The Rotherham Hospital NHS Foundation Trust. [4] Royal Liverpool University Hospital NHS Trust. [5] Royal United Hospitals Bath NHS Foundation Trust. [6] University Hospitals Birmingham NHS Trust. [7] Gloucestershire Hospitals NHS Foundation Trust. [8] St James's University Hospital, Leeds. [9] Oxford University Hospitals NHS foundation Trust. [10] Ninewells Hospital and Medical School, Dundee. [11] Salford Royal Hospital. [12] Cambridge University Hospitals NHS Foundation Trust.

Conflicts of interest and sources of funding

The authors declare that they have no conflicts of interest.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/HIS.14052

Corresponding author

Gerald Saldanha, Level 3 Sandringham Building, Leicester Royal Infirmary, Leicester LE1 5WW, UK.

Email: gss4@le.ac.uk

Word Count

3,310

Running title

Mitotic count observer variation

Abstract

Aims: Staging is the gold standard for predicting malignant melanoma outcome but changes in its criteria over time indicate ongoing evolution. One notable recent change from the 8th edition of the AJCC staging manual was removal of mitotic count. We explore the extent that this feature is limited by interobserver error in order to find ways to improve its fitness for use should it be revisited in future staging versions.

Methods and Results: In a cohort of 476 patients with melanoma ≤ 1.0 mm, a mitotic count of 0 vs 1 was significant for metastasis-free survival, but not melanoma-specific or overall survival. In 10 melanomas that were 0.9 to 1.0 mm thick, the mitotic count intra-class correlation coefficient for histopathologists was 0.58 (moderate agreement). Uniquely, we also assessed agreement for specific putative mitotic figures, identifying precise reasons why specific mitotic figures qualified for scoring or elimination. A kappa score was 0.54 (moderate agreement). We also gathered data on other staging features. Breslow thickness had an intraclass correlation coefficient of 0.41 (moderate agreement) and there was a systematic difference between histopathologists across cases ($p = 0.04$). Every case had a range that crossed the AJCC8 0.8 mm pT1a/pT1b staging boundary. Ulceration was only identified in 2 out the 10 cases. For ulceration, kappa agreement score was 0.31 (fair).

Conclusion: This study supports the removal of mitotic count from staging but shows that its scoring is substantially affected by interobserver variation, suggesting that more prescriptive guidelines might have a beneficial impact on its prognostic value.

Key words

Malignant melanoma; staging; mitotic count; interobserver agreement

Introduction

Malignant melanoma is an aggressive skin cancer where clinical stage at diagnosis is a key determinant of outcome. Most patients with early stage remain disease-free while outcome is poor for advanced disease. The American Joint Committee on Cancer (AJCC8) and the Union for International Cancer Control (UICC8) version 8 schemes are internationally recognised for staging and there are national guidelines that are used to help histopathologists apply the scheme in a consistent manner and foster its widespread adoption, such as those used in the UK [1] and USA [2].

One important change with the latest 8th version of staging (AJCC8) was the removal of mitotic count as a pT1 staging feature, where it had been used previously to distinguish pT1a from pT1b melanoma. Its value had been called into question [3, 4] and notably tumours classed as pT1b on the basis of mitoses had better outcomes than thicker pT1a melanomas in some analyses [5, 6]. In particular, the validity of the hotspot mitotic counting method has been questioned because there is evidence that the distribution of mitoses in a histological section follows a form of Poisson distribution [7]. This would mean that finding a single mitosis in melanomas with low numbers of mitotic figures might reflect chance rather than biology. Instead of mitotic count, AJCC8 distinguishes non-ulcerated pT1a from pT1b using a Breslow thickness (BT) cut point of < 0.8 mm [8].

The jettisoning of mitotic count shows that the staging scheme remains in a state of evolution in the quest for optimal prediction and begs the question about why prognostic features that have apparent sound biological basis can be deemed unfit for purpose. One issue is whether the guidelines are sufficiently precise for histopathologists to reliably measure a feature while controlling for interobserver discrepancy. Our primary study aim was to determine whether the prognostic value of a single mitosis for pT1 melanoma was diminished because the change between 0 and 1 mitosis per mm^2 was too precise compared to the level of interobserver variation in mitotic count scoring. We sought to analyze this by first assessing whether the 0 versus 1 mitosis cut point had a significant association with outcome in a cohort of pT1 melanoma patients and then to investigate to

what extent those findings might have been affected by poor agreement between observers. In other words, whether "signal" was swamped by "noise". We also gathered data on other staging features, so our secondary research aims were to assess interobserver agreement for these, namely Breslow thickness, ulceration and microsatellitosis. The driving force behind this study was to determine whether guidance accompanying AJCC and UICC staging should be more prescriptive and less open to interpretation.

Materials and methods

Study population

For Kaplan-Meier analysis of a single mitosis in pT1 melanoma a starting cohort of 970 cutaneous melanoma from the archives of the University Hospitals of Leicester NHS trust was used representing cases diagnosed between January 1st 2004 and December 31st 2011. The cases have been described previously [9]. From these, a set of 476 was identified that were ≤ 1 mm thick, with either 0 or 1 mitoses per mm^2 and without ulceration. For mitotic count agreement analysis we chose a different subset of cases that were between 0.9 and 1.0 mm thick with hotspot mitotic counts in the original pathology report ranging from 0 to 4 mitoses per mm^2, with 2 of each count to yield 10 cases in total.

Histological feature agreement

To analyze mitotic count agreement, all of the original glass slides used by the reporting histopathologists were available to the participants, including deeper levels and extra stains. These were anonymized and analyzed in turn by each of 10 United Kingdom expert dermatopathologists, all of whom were experienced practitioners. The histopathologists were from outside of the coordinating centre, Leicester, where the original report was generated. Each participant reviewed slides independently. Each assessed the features according to his/her own experience and interpretation of published criteria. All participants were allocated a unique code, rendering them anonymous. A simple form based on one published by the Royal College of Histopathologists for their dataset on

melanoma reporting [1] was created and the histopathologists were asked to document for each case: presence of dermal invasion, histological subtype, Breslow thickness, ulceration, mitotic index, lymphovascular invasion, microsatellite metastasis, margin involvement, neurotropic/perineural invasion, growth phase, tumour-infiltrating lymphocytes, regression and T category. The histopathologists were not informed that mitotic count agreement was the focus of the study and they were instructed to report each case as if it was part of their diagnostic workload to minimise bias away from routine practice. Research ethics committee approval was obtained for the use of tissue (IRAS ID 220400).

In order to assess agreement about individual putative mitotic figures, a printed booklet was created from a set of PowerPoint slides containing 20 putative mitotic figures that were photographed at low, medium and high power. This was sent to each participating pathologist with a response sheet. Eight pathologists were available for scoring. The participants were given the instruction, "For each slide that follows please answer the following three questions about the potential mitotic figure near to the red arrow. 1. Is this a mitotic figure? 2. Is the potential mitotic figure in a melanocyte? 3. Is the potential mitotic figure in the dermis?" The PowerPoint file used to generate the booklet is provided as supplementary material (Supplementary_file.pdf).

Statistical analyses

Statistical analysis was performed in R version 3.2.0 [10].The baseline data table for the survival analysis cases was created using the "tableone" package [11]. Survival analyses were performed with the "survival" package [12] and Kaplan Meier plots were generated with the "survminer" package [13]. Time to event analysis was performed for overall survival (OS), melanoma specific survival (MSS) and metastasis-free survival (MFS). The date of diagnosis was the date of primary melanoma sample accession in the pathology database. For OS, failure was death from any cause. For MSS death from melanoma was considered as failure while death from another cause was regarded as censoring. For MFS, the event was the time of first metastasis (either loco-regional or distant). Survival was analysed using the

Kaplan Meier method and log rank tests were used to compare survival curves. Kappa statistics and intraclass correlation coefficients were calculated using the "irr" package [14]. For the intraclass correlation coefficient, a two way model was used for absolute agreement. For multirater kappa statistic, Fleiss's kappa was used. To determine whether systematic differences were present between rater scores, a Friedman test was used. For statistical tests, a two-tailed p value of < 0.05 was regarded as significant.

Results

Prognostic value of a single mitosis in thin melanoma

We first determined whether a single mitosis had prognostic value in pT1 melanoma using a single centre cohort of 476 melanomas. The baseline features are shown in Table 1. This series had a mean survival of 76 months and only 10.5% of cases died, reflecting the good survival for patients with thin melanoma. The median mitotic count was 0 per mm^2. Kaplan Meier plots showed substantial overlap of the 0 versus 1 mitosis mm^2 survival curves for OS and MSS and a log rank test revealed no significant difference. In contrast, for MFS patients with 1 mitosis per mm^2 had worse outcome (Chisq= 14.1 on 1 degree of freedom, $p < 0.001$). It should be noted that there were only 7 events for MSS, which puts a substantial limit on statistical power. These findings are shown in Figure 1A-C. The Kaplan Meier plots represent a univariable analysis of the association of a single mitosis with outcome, but by limiting the analysis to cases with Breslow thickness $\leq 1.0\mathrm{mm}$, the analysis was in essence stratified by thickness. Nevertheless, the mitotic count might still be confounded by thickness even within this stratum of thin melanomas. To assess this for the only significant univariable outcome, MFS, a Cox PH model was fitted to the data entering mitotic count (0 or 1) and Breslow thickness as covariables. The adjusted hazard ratio for mitotic count was 8.11 (confidence interval 2.49 to 26.45), $p = 0.00052$. This confirmed the prognostic value of mitotic count for predicting MFS. Breslow thickness was not significant ($p = 0.23$) with an adjusted hazard ratio of 0.17 (confidence interval 0.01 to 3.13).

The effect of interobserver variation on mitotic count scoring

We next sought to determine whether interobserver variation might dilute the prognostic association of 0 versus 1 mitosis per mm^2 in T1 melanomas, which had previously been used to distinguish T1a and T1b. A series of melanomas was selected that were either 0.9 or 1.0 mm Breslow thickness, comprising two cases each with mitotic counts of 0, 1, 2, 3 and 4 based on the original surgical pathology report, yielding a total of 10 melanomas. In particular, we selected cases at the top of the pT1 Breslow thickness range because we wanted to avoid "micro-invasive" melanomas, i.e. those with such scanty melanoma cells in the dermis that mitotic count agreement would be virtually assured. It could be argued, therefore, that this set of melanomas is enriched for the ones that are most difficult for mitotic count scoring. The features of these melanomas are shown in Figure 2. The mitotic count was independently scored using the hotspot method by 10 United Kingdom expert dermatopathologists. The results are shown in Figure 3A. The ICC for mitotic count was 0.58, indicating moderate agreement [15]. Only 2 of the 10 melanomas showed complete agreement between histopathologists about whether the lesion had 0 versus 1 or more mitoses per mm^2 (case 1 and 10). There was marked variability in the sum of mitotic figures across all 10 cases per histopathologist, with histopathologists 1 to 10 finding 11, 15, 15, 11, 16, 3, 17, 20, 28, and 11 mitoses respectively. Thus, histopathologist 6 found only 3 mitoses across the 10 cases while histopathologist 9 found 28, representing more than 9x difference. A Friedman matched samples non-parametric ANOVA revealed evidence of a systematic difference between dermatopathologists (chi-squared = 25.1, df = 9, p-value = 0.003).

Agreement for individual mitotic figure recognition

So far, the assessment of mitotic count was on a per melanoma basis. However, in any given melanoma there are several opportunities for disagreement to occur, e.g. histopathologists may choose different slides or different regions within the same slide for counting. Even once a potential mitotic figure is found, a histopathologist must decide if it should be counted based on morphology, position in the section and type of cell. We took

photomicrographs of 20 possible mitotic figures at low, medium and high power and asked histopathologists 3 questions: (1) is the feature a mitotic figure? (2) Is the feature in a melanocyte? (3) Is the feature in the dermis? If the answers to questions 1-3 were yes, then this qualified as a mitotic figure. Figure 4A shows the overall yes/no answer (i.e. all of 1-3 = yes or any of 1-3 = no). The multi-rater kappa for overall yes/no was 0.54, representing moderate agreement. Figure 4B-I reveals exactly how any non-qualifying putative mitotic figures failed for each of the 8 participating histopathologists. Thus, all agreed that putative mitotic figure 3 did not qualify for scoring, yet the reason behind this differed. Histopathologist 4 did not think it was a mitosis, while every other histopathologist did. Histopathologists 3 and 5 thought this feature was in the dermis while none of the others did. These data provide deeper insight into disagreement than analysis merely on a per melanoma basis. Examples of putative mitotic figures are shown in Figure 5, which are taken from the booklet given to each participating pathologist.

Agreement for other AJCC-relevant features

The 10 cases that were selected and specifically tailored towards analysis of mitotic counts were also used to gather data on other AJCC-relevant features. We therefore also asked each histopathologist to score Breslow thickness (mm), ulceration (present/absent) and microscopic satellites (present/absent) for each case. The findings (including mitotic count) are summarised in Table 2 and graphically in Figure 3B and 3C. No histopathologists reported the presence of microscopic satellites for any of the 10 melanomas, representing complete agreement, and this feature was not analysed further. Breslow thickness showed important differences between histopathologists. Five histopathologists called case 3 in-situ melanoma while the other 5 all found the thickness to be at least 0.8 mm. This case featured expansile junctional nests that made it difficult to determine which cells were truly dermal (Figure 6A). This perhaps explains why no measurements fell between 0 mm and 0.8 mm, as histopathologists were forced to make an all-or-none decision, and this would also have influenced mitosis scores, since only dermal mitoses should be counted. One histopathologist also regarded case 8 as in-situ disease, while the other 9 called the

case invasive. The interobserver discrepancies were reflected in an intraclass correlation coefficient of 0.41, indicating moderate agreement for this most important histological prognostic feature. There was also a systematic difference between histopathologists, with some reporting greater Breslow thickness across all cases (Friedman chi-squared = 17.9, df = 9, p-value = 0.037).Thus, histopathologist 3 had a mean Breslow thickness of 0.85 mm while histopathologist 6 had a mean of 0.70. It was also notable that one histopathologist measured case 4 as 1.14 mm, which crosses a T stage boundary from pT1 to pT2. These data additionally show the potential for staging accuracy problems with the AJCC8 cut point of 0.8 mm for pT1a and pT1b, because every single case had a range of scores that crossed this staging boundary. Ulceration was only identified in 2 of 10 cases. Three histopathologists deemed ulceration to be present in case 1 and five in case 3, so even for these cases there was no agreement. All histopathologists agreed that there was no ulceration in the remaining 8 cases. Both of these discrepant cases had issues that might have affected interpretation. There was surface disruption in case 1 (Figure 6B) and there was only a very small focus of epidermal loss in case 3 (Figure 6C and 6D). The overall multirater kappa agreement score was 0.31, indicating fair agreement [15].

Discussion

We found moderate interobserver agreement for hotspot mitotic figure scoring in melanoma cases. To our knowledge, this is the first time that agreement about individual putative mitotic figures has been assessed. Moderate agreement was found again. In addition, we found that a cut point of 0 versus 1 mitoses per mm^2 using the hotspot method had limited prognostic value in 476 thin melanomas, supporting the removal of mitotic hotspot counting from AJCC7, but our findings raise the possibility that the prognostic value was limited because the true importance was overwhelmed by interobserver variation. In addition, we found that interobserver variation for Breslow thickness and ulceration has implications for AJCC8 staging.

Prognostic features related to AJCC staging have been studied previously. Mitotic counts first appeared as a staging feature in 2009, in the 7th AJCC version [16] based on evidence

from various studies [17-19]. The hotspot method was adopted because it was regarded as having better interobserver agreement [20] compared to earlier studies using non-hotspot methods [21-24]. However, more recent studies have found that the hotspot method is also subject to discordance between observers [25-27] while others have found good agreement [28, 29], indicating mixed support for hotspot scoring, but it should be noted that some of these studies compared melanoma prognostic features between original pathology reports and those generated at referral centres, where cases may be more complex. Agreement studies on Breslow thickness have generally found strong agreement [20, 23, 28-31], although in one study 387 out of 588 cases (66%) had a revised Breslow thickness following review at a referral centre [27]. Agreement studies for ulceration have also been performed, usually with kappa scores more than 0.8 [20, 29, 31] indicating excellent agreement, but a few studies showed weaker agreement [23, 30, 32].

The level of interobserver discordance that we found for mitotic counting is not surprising if one considers how opportunities for subjectivity can arise, and we speculate that this relates to four questions that each histopathologist must consider: Where in the section(s) should the hotspot count be done? Is a putative mitotic figure genuine? Is the putative mitotic figure in a melanocyte? Is the putative mitotic figure in a dermal location? In our study, there was ample opportunity for histopathologists to perform hotspot counts in different fields to each other because we provided every one of the glass slides used by the original reporting histopathologist, comprising sections from each block and all deeper levels. However, our study did not capture this variable. With regard to deciding if a mitotic figure was dermal, case 3 was informative because it had expansile junctional nests that could be regarded as not at all or partially dermal, necessitating a judgement about both tumour thickness and mitosis location, presumably based on participant interpretation of relevant guidelines, literature and clinical experience. This probably reflected why the same case was either called in-situ or 0.8 mm thick or more, with no intermediate measurements. Cases 1 and 3 provided insight into discrepancy regarding the reporting of ulceration. It seems likely that case 1 necessitated a decision about whether this was true ulceration or a technical artefact/traumatic detachment while case 3

required judgement about whether genuine biological ulceration was present or a tiny erosion due to excoriation. Alternatively, it may not have been seen.

In the light of these observations, we proceeded to explore how discrepancies arose at the level of individual putative mitotic figures in contrast to the global assessment across a whole melanoma. We used a set of photomicrographs of putative mitotic figures and were able to identify more precisely how discrepancy occurred. To our knowledge, this sort of analysis has not been done before in cutaneous melanoma.

Our overall findings suggest that mitotic counting could be improved with more robust guidelines. Such guidelines would need to take into account the fact that mitotic figures in a section follow a form of Poisson distribution[7], and so an average rather than a hotspot count might have better prognostic value. Guidelines should also include clearer advice about default histological locations for scoring mitoses. For example, some modification of the approach described by Vollmer, where the average count across the full breadth of the section containing the deepest cell was used [7]. Agreement might also be improved by limiting the choice of slide, for example by limiting the count to the same glass slide where BT is measured. With regard to expansile superficial nests that could be either junctional, dermal or partly both, it may be important to revisit the concept of variant vertical growth phase [33] and to formulate clear rules about whether or not these should be routinely considered for Breslow thickness measurement (assuming no deeper cells are found) and whether or not to count contained mitotic figures. The medical literature is focused on high-tech molecular biomarkers that typically incur cost and have technical barriers to clinical adoption, but it may be that merely improving the guidelines for scoring simple histological features could be a truly low cost way to enhance melanoma outcome prediction.

In conclusion, this study supports the removal of mitotic count from AJCC7 staging. However, our findings suggest that the value of mitotic counting is substantially affected by interobserver variation and raises the possibility that more prescriptive guidelines for mitotic counting could have a beneficial impact on its prognostic value by improving the

signal to noise ratio. As for assessment of other features, it should be emphasized that this is a biased set of cases designed for the specific purpose of assessing mitotic counts in thicker T1 category melanomas. However, within these limits we additionally found that the core staging features of Breslow thickness and ulceration were open to subjective interpretation. Importantly, these results provide insight into potential problems with AJCC8 staging, namely disagreement about the Breslow thickness-based 0.8 mm cut point for pT1 staging. This has important implications for sentinel lymph node biopsy eligibility.

Acknowledgements

This work was unfunded.

List of online Supporting Information

Supplementary_file.pdf

References

1. Slater DN, Walsh M. Standards and datasets for reporting cancers: Dataset for the histological reporting of primary cutaneous malignant melanomas and regional lymph nodes. London: Royal College of Pathologists, 2014.

2. Smoller BR, Gershenwald JE, Scolyer RA *et al.* Protocol for the examination of specimens from patients with melanoma of the skin. College of American Pathologists, 2017.

3. Helm TN. Mitotic figures and the evolving staging of melanoma. *J. Cutan. Pathol.* 2017;**44**;358-359.

4. Weyers W. "Mitogenicity"-the latest and mosthilarious episode in the slapstick comedy of melanoma management. *Dermatol Pract Concept* 2012;**2**;203a211.

5. Kirkland EB, Zitelli JA. Mitotic rate for thin melanomas: Should a single mitotic figure warrant a sentinel lymph node biopsy? *Dermatol. Surg.* 2014;**40**;937-945.

6. 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**;2199-2205.

7. Vollmer RT. A probabilistic analysis of mitotic counts in melanoma. *Am. J. Clin. Pathol.* 2014;**141**;213-218.

8. Gershenwald JE, Scolyer RA, Hess KR *et al.* Melanoma staging: Evidence-based changes in the american joint committee on cancer eighth edition cancer staging manual. *CA. Cancer J. Clin.* 2017;**67**;472-492.

9. Saldanha G, Yarrow J, Pancholi J *et al.* Breslow density is a novel prognostic feature that adds value to melanoma staging. *Am. J. Surg. Pathol.* 2018.

10. R Core Team. R: A language and environment for statistical computing. . Vienna, Austria: R Foundation for Statistical Computing, 2016.

11. Yoshida K, Bohn J. Tableone: Create 'table 1' to describe baseline characteristics. 2017.

12. Therneau T. A package for survival analysis in s. Version 2.38. 2015.

13. Kassambra A, Kosinsji M. Survminer: Drawing survival curves using 'ggplo2'. 2016.

14. Gamer M, Lemon J, Fellows E, Singh P. Irr: Various coefficients of interrater reliability and agreement. 2012.

15. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;**33**;159-174.

16. Balch CM, Gershenwald JE, Soong SJ *et al.* Final version of 2009 ajcc melanoma staging and classification. *J. Clin. Oncol.* 2009;**27**;6199-6206.

17. Azzola MF, Shaw HM, Thompson JF *et al.* Tumor mitotic rate is a more powerful prognostic indicator than ulceration in patients with primary cutaneous melanoma: An analysis of 3661 patients from a single center. *Cancer* 2003;**97**;1488-1498.

18. Francken AB, Shaw HM, Thompson JF *et al.* The prognostic importance of tumor mitotic rate confirmed in 1317 patients with primary cutaneous melanoma and long follow-up. *Ann. Surg. Oncol.* 2004;**11**;426-433.

19. 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**;1129-1134.

20. Scolyer RA, Shaw HM, Thompson JF *et al.* Interobserver reproducibility of histopathologic prognostic variables in primary cutaneous melanomas. *Am. J. Surg. Pathol.* 2003;**27**;1571-1576.

21. Cook MG, Clarke TJ, Humphreys S *et al.* The evaluation of diagnostic and prognostic criteria and the terminology of thin cutaneous malignant melanoma by the crc melanoma pathology panel. *Histopathology* 1996;**28**;497-512.

22. Fuchs E, Tumbar T, Guasch G. Socializing with the neighbors: Stem cells and their niche. *Cell* 2004;**116**;769-778.

23. Heenan PJ, Matz LR, Blackwell JB *et al.* Inter-observer variation between pathologists in the classification of cutaneous malignant melanoma in western australia. *Histopathology* 1984;**8**;717-729.

24. Larsen TE, Little JH, Orell SR, Prade M. International pathologists congruence survey on quantitation of malignant melanoma. *Pathology (Phila).* 1980;**12**;245-253.

25. Albo D, Berger DH, Wang TN, Hu X, Rothman V, Tuszynski GP. Thrombospondin-1 and transforming growth factor-beta l promote breast tumor cell invasion through up-regulation of the plasminogen/plasmin system. *Surgery* 1997;**122**;493-499; discussion 499-500.

26. Garbe C, Eigentler TK, Bauer J *et al.* Mitotic rate in primary melanoma: Interobserver and intraobserver reliability, analyzed using h&e sections and immunohistochemistry. *J Dtsch Dermatol Ges* 2016;**14**;910-915.

27. Patrawala S, Maley A, Greskovich C *et al.* Discordance of histopathologic parameters in cutaneous melanoma: Clinical implications. *J. Am. Acad. Dermatol.* 2016;**74**;75-80.

28. Monshizadeh L, Hanikeri M, Beer TW, Heenan PJ. A critical review of melanoma pathology reports for patients referred to the western australian melanoma advisory service. *Pathology (Phila).* 2012;**44**;441-447.

29. 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**;3969-3975.

30. Eriksson H, Frohm-Nilsson M, Hedblad MA *et al.* Interobserver variability of histopathological prognostic parameters in cutaneous malignant melanoma: Impact on patient management. *Acta Derm. Venereol.* 2013;**93**;411-416.

31. Murali R, Hughes MT, Fitzgerald P, Thompson JF, Scolyer RA. Interobserver variation in the histopathologic reporting of key prognostic parameters, particularly clark level, affects pathologic staging of primary cutaneous melanoma. *Ann. Surg.* 2009;**249**;641-647.

32. Spatz A, Cook MG, Elder DE, Piepkorn M, Ruiter DJ, Barnhill RL. Interobserver reproducibility of ulceration assessment in primary cutaneous melanomas. *Eur. J. Cancer* 2003;**39**;1861-1865.

33. Cook MG, Spatz A, Brocker EB, Ruiter DJ. Identification of histological features associated with metastatic potential in thin (<1.0 mm) cutaneous melanoma with metastases. A study on behalf of the eortc melanoma group. *J. Pathol.* 2002;**197**;188-193.

Table 1. The baseline features of 476 melanomas ≤ 1.0 mm (excluding ulcerated cases).

n = 476	n (%)
Gender = Male (%)	227 (47.7)
Age (mean (sd))	56.40 (16.28)
Time to metastasis (mean (sd))	75.50 (26.17)
Time to death (mean (sd))	76.41 (25.18)
Dead (%)	50 (10.5)
Site (%)	
Acral	17 (3.6)
Head & neck	81 (17.0)
Lower limb	121 (25.4)
Trunk	170 (35.7)
Upper limb	87 (18.3)
Type (%)	
SSMM	400 (84.0)
LMM	42 (8.8)
ALMM	12 (2.5)
NMM	7 (1.5)
Desmoplastic	1 (0.2)
Spitzoid	1 (0.2)
Unknown	13 (2.7)
Breslow (mean (sd))	0.55 (0.20)
Mitotic count (median [IQR])	0 [0, 0]

Table 2. Summary of AJCC-relevant features by case, as assessed by each of 10 UK dermatopathologists

	Case 1	**Case 2**	**Case 3**	**Case 4**	**Case 5**	**Case 6**	**Case 7**	**Case 8**	**Case 9**	**Case 10**
	n = 10	n = 10	n = 10	n = 10	n = 10	n = 10	n = 10	n = 10	n = 10	n = 10
Mitoses (per mm^2)										
Median	3.5	0	0	2	0	1	0	0	4.5	1.5
Minimum	2	0	0	0	0	0	0	0	0	1
Maximum	8	1	3	6	1	2	1	1	11	2
Breslow (mm)										
Median	0.91	0.82	0.40	1.07	0.80	0.80	0.75	0.45	0.88	0.84
Minimum	0.8	0.7	0	0.1	0.75	0.7	0.5	0	0.8	0.75
Maximum	0.97	0.98	0.94	1.14	0.9	0.95	0.9	0.8	0.9	0.95
ulcer										
	3	0	5	0	0	0	0	0	0	0
Present (%)	(30.0)	(0.0)	(50.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
Microsatellite nodule										
	0	0	0	0	0	0	0	0	0	0
Present (%)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)

Figure legends

Figure 1. Kaplan Meier plots of pT1 melanomas with hotspot mitosis count of 0 or 1. Ulcerated melanomas were excluded from the analysis. (A) OS, (B) MSS, (C) MFS.

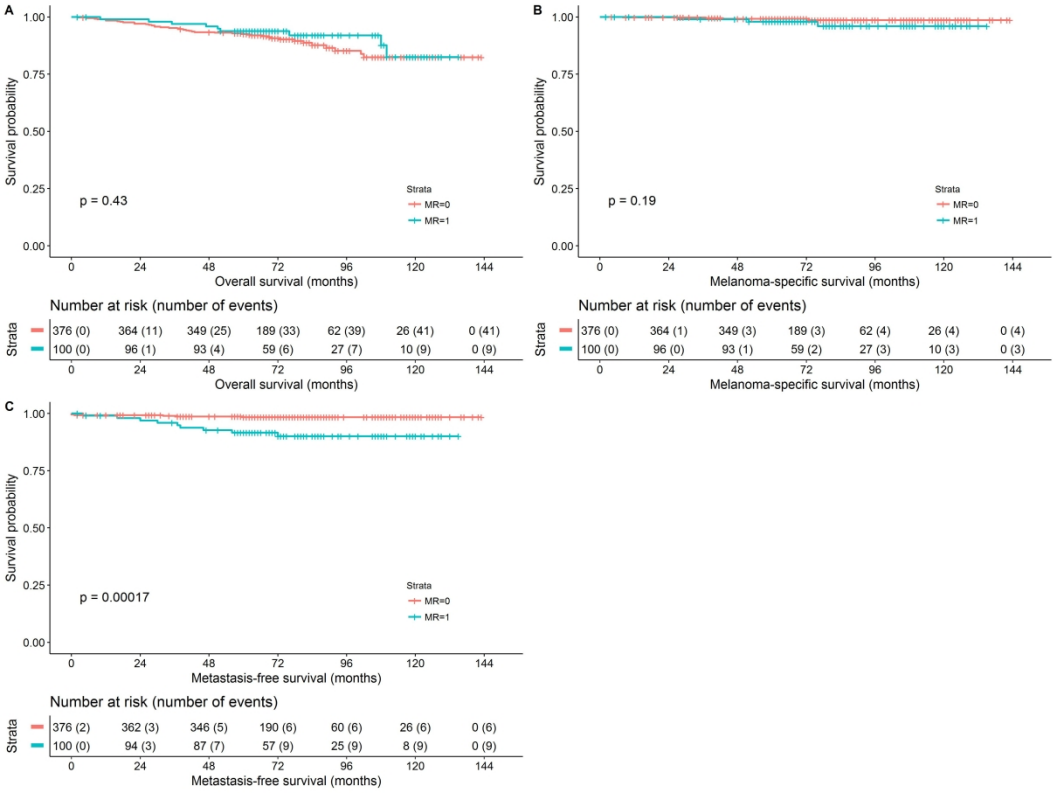
Figure 2. Melanoma cases used for mitotic count agreement analysis. TILs = tumour-infiltrating lymphocytes; LVI = lymphovascular invasion; PNI = Perineural invasion.

Figure 3. Mitotic count, Breslow thickness and ulceration agreement on 10 melanomas for each of 10 UK histopathologists.

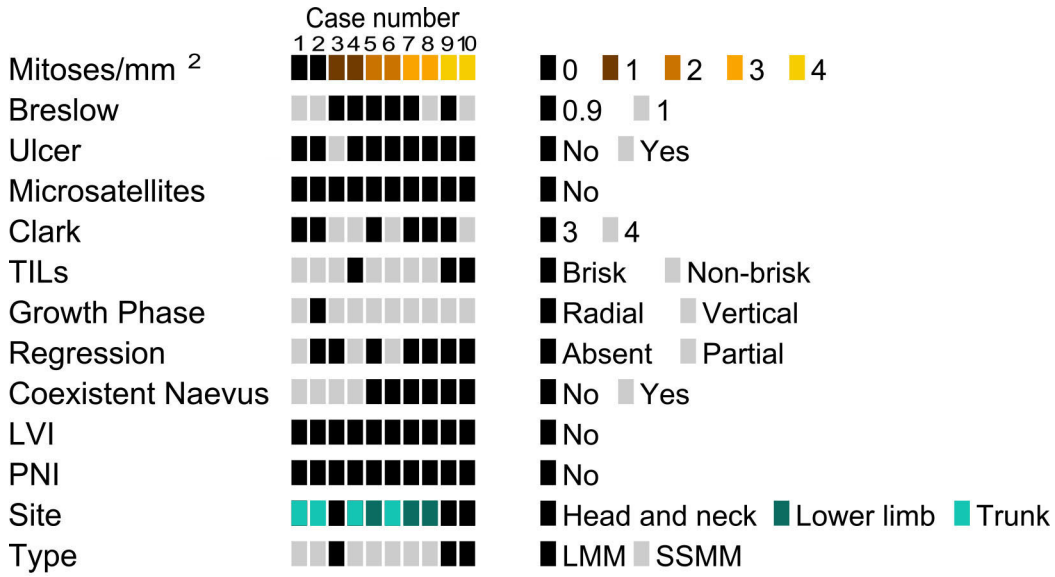
Figure 4. Histopathologist agreement for categorizing potential mitotic figures. (A) Overall yes/no agreement. (B - I) Putative mitotic figures for 8 pathologists, P-1 to P-8.

Figure 5. Examples of putative mitotic figures from three cases, each taken from the booklet used by participating histopathologists for scoring. Case 1 was agreed by all participants to show a mitotic figure. Cases 13 and 16 showed disagreement. The number of participants saying yes to questions 1-3 are shown for each case: Is the feature a mitosis? Is the feature in a melanocyte? Is the feature in the dermis? Only a yes to all 3 questions results in a feature being labelled as a mitotic figure. LP = low power, MP = medium power, HP = high power.

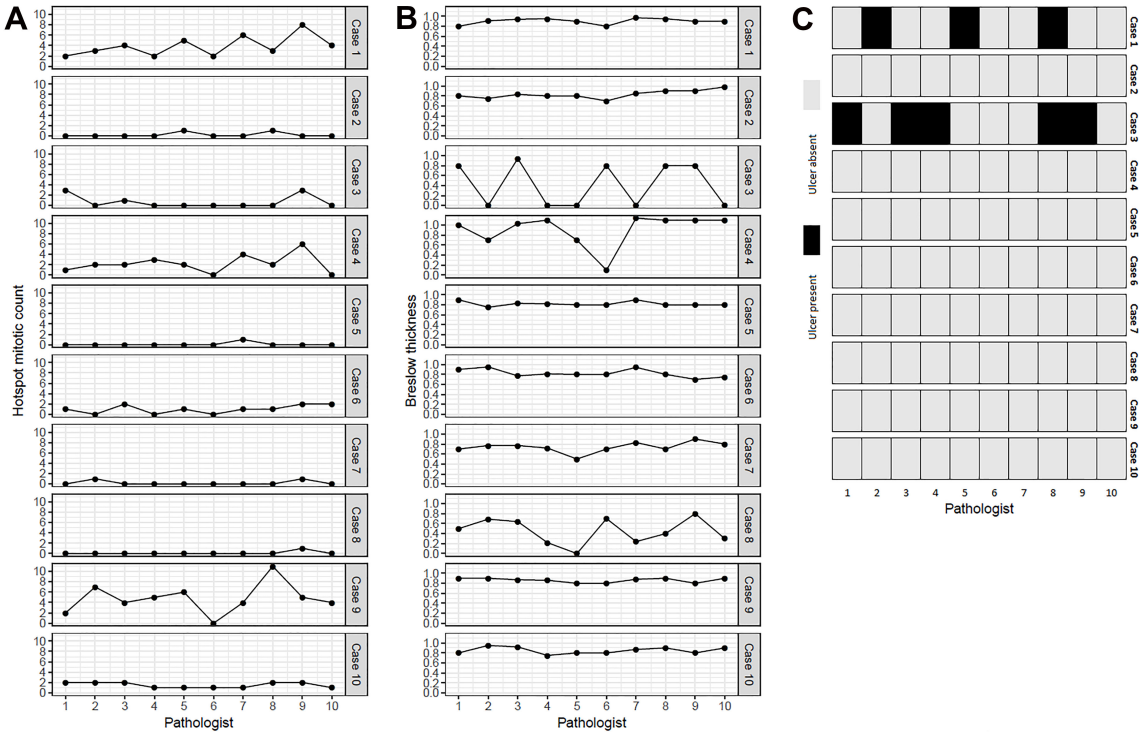
Figure 6. Discrepant cases for AJCC8 staging features. Case 3 had epidermal acanthosis making assessment of mitotic figure location in junctional versus invasive cells difficult and also causing problems for measuring Breslow thickness (A). Case 1 caused disagreement about ulceration because of surface disruption at arrows (B). Case 3 had only a very focal area of epidermal loss at arrows (C), with magnified view of dashed box (C).



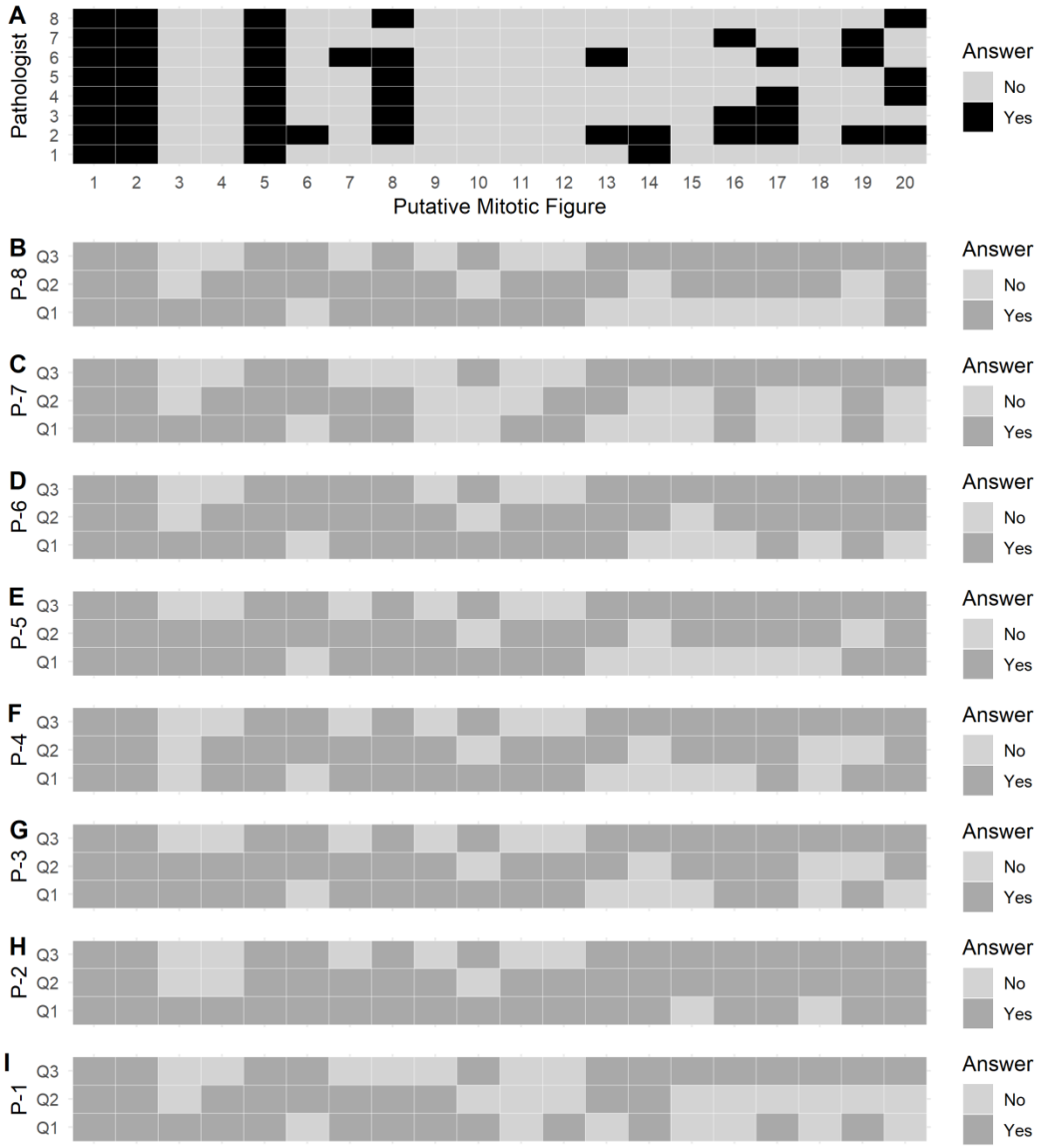
his_14052_f1.jpg



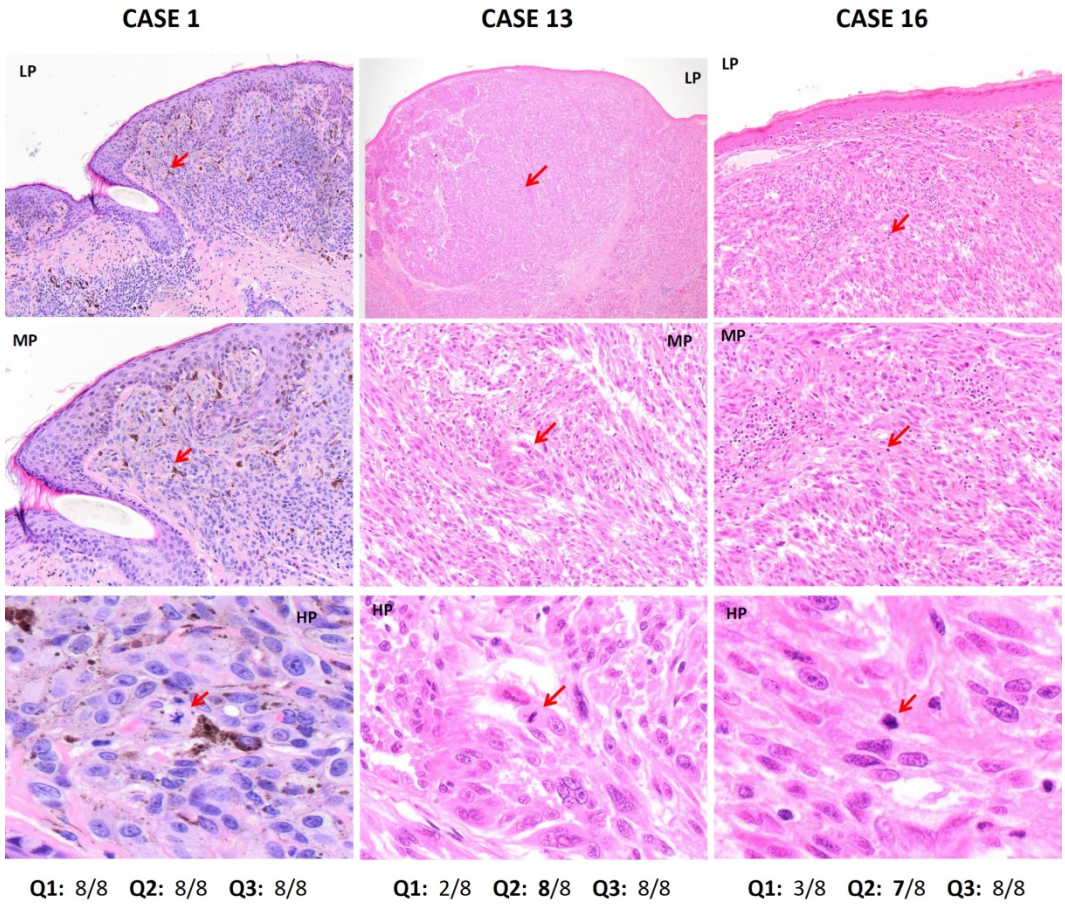
his_14052_f2.jpg



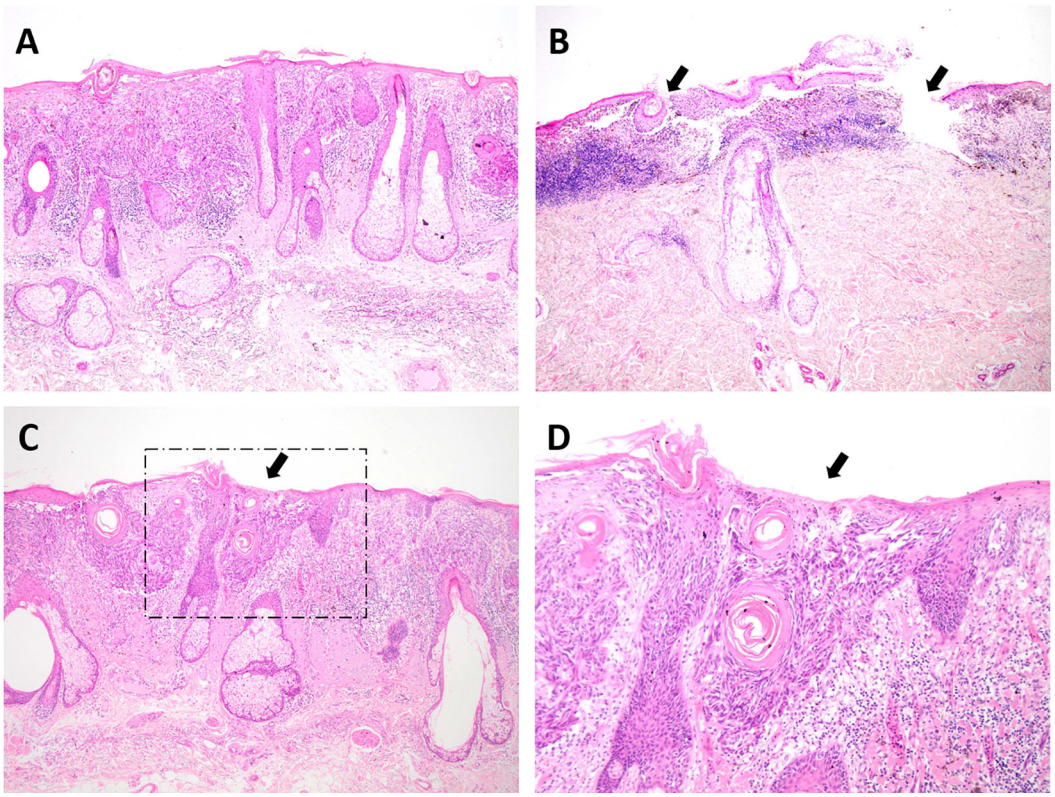
his_14052_f3.jpg



his_14052_f4.png



his_14052_f5.jpg



his_14052_f6.jpg