Published in final edited form as:

Am J Surg Pathol. 2018 June ; 42(6): 715–725. doi:10.1097/PAS.0000000000001034.

Breslow density is a novel prognostic feature that adds value to melanoma staging

Gerald Saldanha, MB ChB, MRCP(UK), FRCPath, PhD¹, Jeremy Yarrow, BSc¹, Jay Pancholi, BSc¹, Katarina Flatman, BSc¹, Kah Wee Teo, MB ChB¹, Somaia Elsheik, MBBCh, MSc, FRCPath, PhD², Rebecca Harrison, BSc, MB ChB, FRCPath³, Marie O'Riordan, BA, MB ChB³, and Mark Bamford, MB ChB, BSc, FRCPath, DipRCPath³

¹Institute of Advanced Studies, University of Leicester

²Department of Cellular Pathology, Nottingham University hospitals

³Department of Cellular Pathology, University Hospitals of Leicester NHS Trust

Abstract

Histomorphological prognostic biomarkers that can be measured using only an H&E stain are very attractive because they are simple and cheap. We conceived an entirely novel biomarker of this type, the Breslow density (BD), which measures invasive melanoma cell density at the site where Breslow thickness (BT) is measured. This study assessed BD's prognostic value. In this study, BD was measured in 1329 melanoma patients. Measurement accuracy and precision was assessed using intraclass correlation coefficient (ICC). Survival was assessed with a primary end-point of melanoma-specific survival (MSS) and also overall survival (OS) and metastasis-free survival (MFS). We found that BD measurement was accurate compared to gold standard image analysis (ICC 0.84). Precision was excellent for 3 observers with different experience (ICC 0.93) and for an observer using only written instructions (ICC 0.93). BD was a highly significant predictor in multivariable analysis for OS, MSS and MFS (each, $p < 0.001$) and it explained MSS better than BT, but BT and BD together had best explanatory capability. A BD cut point of 65% was trained in 970 melanomas and validated in 359. This cut point showed promise as a novel way to upstage melanoma from T stage "a" to "b". BD was combined with BT to create a targeted burden score. This was validated as an adjunct to AJCC stage. In summary, BD can be measured accurately and precisely. It demonstrated independent prognostic value and explained MSS better than BT alone. Notably, we demonstrated ways that BD could be used with AJCC version 8 staging.

Introduction

Cutaneous malignant melanoma is diagnosed in more than 15,400 people in the UK¹ and 76,000 in the USA². In recent years there have been improvements in the care of advanced melanoma patients with immune checkpoint inhibitor and targeted drugs entering routine care³ and studies now showing benefit with these same agents for adjuvant treatment^{4,5}.

Underpinning all of this is accurate stratification of melanoma patients into risk groups for appropriate care.

The gold standard for risk stratification, up to Jan 2018, has been the American Joint Committee on Cancer (AJCC) version 7 staging system⁶. Staging continues to evolve⁷, reflecting the fact that the TNM system does not perfectly predict the outcome for individuals. The pursuit of new prognostic biomarkers is therefore paramount. It is hoped that prognostic biomarkers will emerge from the revolution in our understanding of melanoma's molecular pathology⁸, but it is easy to forget that there may be simple, cheap, rapid morphological prognostic biomarkers hiding in plain sight on the very same haematoxylin and eosin (H&E) stained sections used to diagnose melanoma. We recently described a biomarker of this type that we called Breslow density (BD)⁹, which can be measured in seconds on the H&E section used for diagnosis and is virtually cost-free. BD is an estimate of the percentage of dermis occupied by melanoma cells at the position where the Breslow thickness (BT) is measured, with the premise that of two prognostically equivalent melanomas the one with the greater BD will have a worse outcome. Because the BD is so simple to measure by a pathologist and requires no additional special equipment, it is particularly amenable to adoption into clinical practice and if it were verified as a *bona fide* prognostic feature, it would be worthy of prioritisation for translation into practice.

The goal of the present study was to expand our proof of concept study for BD⁹. We sought to verify and extend our previous findings on interobserver agreement, to investigate how BD relates to other common prognostic features and to determine whether it has independent prognostic value. Finally, we investigated how BD could be used with AJCC staging.

Methods

Patients

Tissues were identified from melanomas diagnosed at the University Hospitals of Leicester. Eligibility criteria were: primary cutaneous melanoma tissue, diagnosis 2004-2011, invasive disease and resident in Leicestershire. The flow of patients through the study is shown in Figure S1. Samples were assessed sequentially by pathology accession number, thus representing an unbiased retrospective cohort. For patients with multiple primary melanomas, the highest stage melanoma was included or, if they were the same stage, the first diagnosed. From a starting population of 1104, a total of 134 (12%) Leicester cases were excluded to yield 970 cases. Excluded cases are shown in Table S1. A set of 359 primary melanomas diagnosed in 2003-2005 and 2008-2010 from Nottingham University Hospitals was provided with anonymised data allowing determination of stage and melanoma specific survival. This study was granted a favourable opinion from an NHS research ethics committees in Leicester and Nottingham. The Leicester and Nottingham cases are henceforth referred to as training and validation cases respectively.

Breslow density scoring

BD scoring was performed using formalin fixed paraffin embedded (FFPE) slides from the histopathology archives. The original sections used for diagnosis were scored as described⁹.

The BD was obtained by estimating the proportion of dermis that was occupied by invasive melanoma cells, and expressed as a percentage to the nearest 5% except at the extremes, i.e. < 5% or >95%, where the estimate could be to the nearest 1%. A description of BD window construction is provided in Figure 1 and further guidance on BD scoring is provided in Figure S2 and Table S2 along with photomicrographs of melanomas from the study and their associated BD scores in Figure S3. Punch or incisional biopsies of melanoma that did not encompass the entire lesion could provide a biased BD. To minimise this, if the deepest melanoma cell was on the punch/incisional biopsy rather than the re-excision, and the tissue specimen was more than a 10x lens wide, BD was scored as normal. If the deepest melanoma cell was found on a re-excision of such a biopsy, the scorer identified both the deepest cell and the scar corresponding to previous biopsy. If the deepest cell lay more than a 10x lens from the scar, BD was scored as normal because the previous sample would not affect the BD. If the deepest melanoma cell and scar both fell within a single 10x field, both biopsy and excision specimens were reviewed together to make a best interpretation of the maximal amount of dermal cells within 10x field width containing the deepest melanoma cell. A similar approach was adopted in the event that the deepest melanoma cell was found very close to the edge of a punch biopsy. Eleven cases were excluded because of BD could not be measured because the above criteria could not be met (see Figure S1). Where image analysis was performed, slides were scanned with a Hamamatsu Nanozoomer, model C12000-02. Images were annotated with the NDP.view2 version 2.6.13 software. The appropriate dermal area and the outline of each invasive tumour island were drawn manually. The areas of each annotation were exported to a spreadsheet and BD was calculated as the tumour area divided by dermal area x100.

Statistical analysis

Statistical analysis was performed in R version 3.3.110 and RStudio version 0.9911. P < 0.05 was considered significant. Analyses were two tailed using Chi-squared test with continuity correction or one way ANOVA. A Kruskal Wallace test was used for non-normal variables. Intraclass correlation (ICC) and kappa coefficients were calculated for multiple raters¹² using the "irr" package¹³ in R.

For survival analysis, we assessed overall survival (OS), melanoma specific survival (MSS, the primary outcome) and metastasis free survival (MFS). For OS, an event was death with censoring if the patient was alive. For MSS, an event was death from melanoma with censoring if the patient died of another cause or was alive. For MFS, an event was first occurrence of clinical metastasis to a regional or distant site with censoring if the patient was metastasis-free or died prior to metastasis. Analysis was performed with R "survival"¹⁴ and "survminer" packages¹⁵ using Kaplan Meier plots and Cox proportional hazards (CPH) regression. The proportionality assumption was checked by examining plots of scaled Schoenfeld residuals against transformed time and a goodness-of-fit test. Proportionality was not violated.

Statistical prediction modeling strategy

To investigate the value of BD as a novel upstaging feature, we used a split sample approach to discover and then validate a BD cut-point for identifying high risk melanoma. MSS was

the response variable. To investigate the use of BD as an adjunct to AJCC8 staging, we focused on (i) model specification and (ii) model coefficient estimation. Model specification entailed the determination of optimal model predictors, with MSS as the response variable. Given that we were investigating a staging adjunct, AJCC stage was necessarily one of the predictors. We therefore focused on determining how to best enter BD into the prediction model. It is common to assess variable transformations (e.g. logarithmic, quadratic), including the possibility of entering BD as part of a composite variable¹⁶. The most logical composite variable, and one we have investigated previously, was a combination of BD with BT⁹. This combination was considered reasonable because two cases with the same BT may each have a different BD, while two cases with the same BD could have a different BT, and so a composite value might be the best way to capture this information. The composite variable was calculated as $BT \times BD/100$. The division by 100 was merely to shrink the BD scale from 1-99 to 0.01 - 0.99, so that its values were closer in magnitude to typical BT values. We called this composite variable the targeted burden score (TBS) and transformations of this were also investigated. We used a split sample design to validate the optimal model specification using a previously described general approach^{17,18}. For model coefficient estimation, evidence from simulation studies suggest that a bootstrap approach is the optimal way to correct for model overfitting/optimism and is better than split sample validation¹⁹. We used a described bootstrap approach²⁰ using the rms package to perform the analysis²¹. We fitted a model using combined training and validation patients, $n = 1329$ (Sample_0), and determined the CPH model coefficients (Model_0). To determine how optimistic/overfitted Model_0's coefficients were, a bootstrap sample (Sample_1) of size 1329 was drawn from Sample_0 and a new CPH model (Model_1) was fitted to Sample_1. Sample_0 was then used as a test set for Model_1 and the test set calibration slope was calculated. This was performed for a total of 100 bootstrap samples to get a stable estimate of the test set calibration slope. We used the average test set calibration slope as a shrinkage factor for the Model_0 coefficients.

Model comparison was performed using the Akaike information criterion (AIC). The model with the lowest AIC, designated AIC_{min} , is deemed the best and it can then be compared the j^{th} model, designated AIC_j , by calculating an evidence ratio from the formula, $\exp((AIC_{min} - AIC_j)/2)$. The resulting value amounts to how many times stronger the evidence is for AIC_{min} relative to the j^{th} model²².

This study adhered to REMARK guidelines²³. A REMARK checklist is given in Table S3.

Results

Breslow density shows strong interobserver agreement

BD is a new biomarker, so evidence of accuracy and precision is fundamental to its adoption into clinical practice. For accuracy, we compared BD scores against a gold standard, image analysis. In 12 cases, the ICC was 0.84, indicating excellent accuracy. For precision, we extended our previous investigation of interobserver agreement⁹ using 24 cases with three observers having variable experience (Figure 2A), demonstrating excellent agreement, ICC 0.93. For BD scoring to be generalised, raters should be able to score with only written instructions. A fourth rater, not involved in the study to this point, measured BD in the same

24 cases using printed instructions (subsequently used to create Figure S2, Table S2 and Figure S3), and blinded to previous scores. The BD scores of this rater had an ICC of 0.93 compared to the mean score of 3 raters, indicating excellent agreement.

Patient characteristics

The characteristics of the 970 training patients are shown (Table 1). The median BT was 0.9 mm. Microscopic satellites were identified in 1.8% of cases and 17% of cases showed ulceration. The BD ranged from 2% to 99% and had a bimodal distribution, and a median BD of 60% and interquartile range of 15% to 85%. Table 1 also shows that BD was associated with known melanoma features, specifically age, melanoma type, BT, ulceration, mitotic count, microscopic satellites and AJCC stage, version 7.

Breslow density has independent prognostic value

The median follow up time for the Training patients was 71 months. BD was classified into low, mid and high categories based on tertiles (25%, 30-75% and 80%), with 315, 317 and 338 in the low, mid and high tertiles. Using Kaplan Meier analysis, BD had a significant effect on outcome. For OS, 5-year survival in descending tertiles was 0.93, 0.84 and 0.67; for MSS, 0.98, 0.94 and 0.80; and for MFS, 0.96, 0.87 and 0.67. The curves were significantly different, $p < 0.001$ for each of OS, MSS and MFS (Figure 2B-D).

To account for confounding, a CPH regression model was fitted to the data to enable adjusted hazard ratios to be calculated. The BD hazard ratio was adjusted for seven recognised and validated prognostic factors²⁴, and results are shown in Table 2, where BD was entered as an integer variable. BD showed highly significant adjusted hazard ratios for each outcome. Table 2 shows the hazard ratio for a 1 percentage point change in BD, but for a more meaningful 10 percentage point change the hazard ratio would be 1.11 for OS and 1.22 for both MSS and MFS. Other well recognised prognostic factors also had significant adjusted hazard ratios, but only BD, BT, ulceration and mitotic count remained significant in multivariable analysis for each of the three outcomes.

Proof of concept, using AJCC version 8, that Breslow density has potential for incorporation into future staging versions

To understand how to use BT and BD together, we assessed how each explained the primary outcome measure, MSS. Three CPH models were created, a full model with BD and BT, a BD-only model and a BT-only model. The AIC for each was 1270.6, 1306.1 and 1318.0 respectively, indicating that the full model was the best and the BT-only model was weakest. The evidence for the full model was 4.96×10^7 times stronger than the BD only model and 1.90×10^{10} times stronger than the BT only model. Thus, a model containing both BD and BT is substantially better than either alone. The summaries of the crude models for BT and BD were shown in Table 2 and the BT plus BD model is shown in supplementary Table S4.

These findings prompted us to evaluate how BT and BD could be used together while maintaining the general principles of TNM staging. To accomplish this, we tested whether BD could be an additional way to revise T stage from "a" to "b" in the new AJCC version 8 (AJCC8) staging system. First, we investigated the optimal cut point for stratification by

setting up a series of CPH regression models, each containing BD and T stage with MSS as the outcome. In each model, BD was entered with a different cut point, from 20% to 80% in 5% increments. The one with the best adjusted hazard ratio was selected, which was BD 65%, with an adjusted hazard ratio of 2.03, 95% confidence interval 1.04 to 3.94, $p = 0.037$. We approximated this finding into a simple clinical rule so that a pathologist could readily apply it in daily practice. Specifically, if BD was two thirds or more then the melanoma should be regarded as high risk. This cut point was validated in 359 independent cases, with median BT 1.1 mm, interquartile range 0.6 mm to 2.1 mm. Mitotic counts were provided as absent, present or brisk, with 129 (36%), 192 (53%) and 38 (11%) cases per group respectively. There were 61 (17%) cases with ulceration and 9 (3%) with microsatellites. For AJCC8, 172 (48%) were IA, 79 (22%) IB, 50 (14%) IIA, 26 (7%) IIB, 23 (6%) IIC and 9 (3%) were stage III. These features were comparable to the training ones, meaning these cases were appropriate for validation. The median follow up was 142 months. The adjusted hazard ratio was 3.11, 95% confidence interval 1.34 to 7.24, $p = 0.008$, thus validating the BD cut point. These data are in Figure 3A. We merged the training and validation cases for a final dataset of 1329 to get a precise estimate of the BD cut point's adjusted hazard ratio, which was 2.44, 95% confidence interval 1.46 to 4.07, $p < 0.001$.

We next assessed whether the two thirds or more BD cut point might be used to upstage AJCC8 from "a" to "b". We partitioned the melanomas into those comprising T1a, T2a, T3a and T4a and those comprising T1b, T2b, T3b and T4b. Amongst the "a" group, the two thirds cut point was used to categorise them into low and high risk. This is shown graphically in Figure 3B. Our premise was that the "a" high risk Kaplan Meier survival curve should diverge from the "a" low risk curve and approach the "b" curve. The curves for individual T stages are shown (Figure 3C-F). In Figure 3G, each T stage group was merged into composite curves for T1-4a low risk, T1-4a high risk and T1-4b groups.

For T1, 60 of 530 (12%) were upstaged by BD; for T2, 172 of 256 (67%); for T3, 85 of 109 (78%); and for T4, 37 of 45 (82%). Overall, this amounted to 354 of 940 (27%) patients with melanoma being upstaged by BD, with an increasing proportion as T stage increased. The T1a high risk curve sits above both of the other curves but there are very few events to render the plot meaningful. However, this finding suggests that BD may not be of value for T1 melanoma upstaging. The T2a high risk curve sits directly on top of the T2b curve up to the 10 year time point. The T3a high risk curve is similar to the T3b curve as far as the 2 year time point. The T4a high risk curve closely follows the T4b curve to approximately the 3 year time point. Finally, the T1-4a composite high risk curve is closer to the T1-4b curve than the T1-4 low risk curve. Direct comparison of the T1-4a low risk and high risk curves revealed a statistically significant difference ($p < 0.0001$). However, the T1-4a high risk and the T1-4b composite curves were also significantly different ($p = 0.024$), suggesting that the risk of MSS differs for all 3 curves. Overall, these data show that BD has potential for AJCC upstaging from "a" to "b" and emphasises that a combination of absent ulcer and low BD indicates a relatively low risk subset of melanoma cases, i.e. the T1-4a low risk group.

Breslow density can be used as an adjunct for AJCC version 8 stage

As an alternative to integration into staging, we looked whether BD could be an adjunct for AJCC8. We previously explored the concept of a targeted burden score (TBS), entailing the use of BD as a correction factor for BT9. We use the training cases to find the best specification of a multivariable CPH model with MSS as the outcome. Preliminary exploration of BD and TBS (which was defined as $BT \times BD/100$) in a Cox model showed that a combination of AJCC8 and the natural log of TBS ($\ln TBS$) provided the optimal model fit based on AIC values (Table S5). Using this model, AJCC8 was significant, as expected, but so too was $\ln TBS$, $p = 0.01$. The model coefficients and p values are shown in Table S6. To validate this model, we sought to show discrimination and calibration in new cases^{17,18} using only the pre-specified model developed on the training cases, and without re-fitting it to new validation cases. The linear predictor from the training model was applied to each validation case, yielding a prognostic index (PI). To test discrimination, we cut the PI into quantiles to create Kaplan Meier plots. Exploratory analysis revealed that the lower 6 deciles had overlapping curves, in keeping with the fact that melanomas in this study contained a large proportion of low stage, thin cases. Therefore, to best demonstrate discrimination we plotted curves using quantile cut points of 0.6, 0.75 and 0.9, representing risk groups 1-4 respectively. The Kaplan Meier plots showed that the training and validation cases had broadly similar curve characteristics (Figure 4A and B), although group 1 and 2 curves were closer together for validation cases. However, all 4 curves still remained separate, demonstrating evidence of discrimination in the validation cases. To check calibration, the Kaplan Meier curves of the validation cases, representing observed failure rates (Figure 4B), were compared to the curves generated by the CPH model (Figure 4C), representing expected failure rates. This revealed slight mis-calibration, most notably with steeper predicted curves for PI groups 2 and 4. Other measures of validation are shown in Figure 4D. The Gonen and Heller K statistic shows that model fit was retained between training and validation sets. Harrell's c index showed slight loss of discrimination. This was reflected by the smaller hazard ratios for PI groups 2 to 4 in the validation cases. Regression on the PI in training cases shows a calibration slope of 1.0 by design, since the PI was developed using these. Perfect calibration in any validation set would be represented by a calibration slope of 1, and the slope in the validation cases was 0.88 (se 0.12). Importantly, this was not significantly different from the ideal slope of 1.0 ($p = 0.50$), but it does suggest slight over-fitting during model training. This is particularly evident for high risk melanomas, where observed failure rates were fewer than predicted (see PI group 4 curves in Figure 4B and C). Taking all of these analyses together, the trained model showed overall satisfactory validation. We therefore deemed it appropriate to merge the training and validation cases so that the larger combined dataset could be used to derive a final dataset of 1329 in order to produce a CPH model with the best estimate of its coefficients. This final model could suffer from overfitting, so we used a bootstrap approach to assess this. The bootstrap analysis suggested that shrinkage of the coefficients by 2.1% would be required to compensate for overfitting and hence allow for better external validation in future studies. The resultant hazard ratios are shown in Table 3. The final fitted model, with shrunk coefficients, was used to generate Kaplan Meier plot in Figure 4E for risk groups 1 to 4 with cut points as above of 0.6, 0.75 and 0.9. This plot is annotated with 5 and 10 year survival for each risk group and the spread of AJCC8 stages across each PI group is shown in Figure

4F, demonstrating that individual AJCC8 stages are dispersed across risk groups, apart from stage IA. To facilitate tailoring of care, we finally generated a nomogram to predict outcome for individual patients (Figure 4G). This also can be used to allocate a patient to one of the risk groups in Figure 4E.

Discussion

We show that BD measurement is both accurate and precise and it retains independent prognostic value even when adjusted for well recognised prognostic covariables. BD explains MSS better than BT alone, but BT and BD are best used together and we explored two ways to achieve this. First, we showed that BD could be a novel way to upstage T stage from "a" to "b". Second, we showed that BD and BT can be combined into a score and this score can be used as an adjunct to AJCC8.

A key aspect of BD is its practicality, because it can be measured in seconds using the same H&E stain used for diagnosis and its only cost is the extra pathologist time, which per case is negligible. Many other prognostic biomarkers have been investigated, including molecular ones such as DecisionDx25. However, molecular biomarkers typically require specialist techniques and sufficient, surplus tissue. They also have greater financial cost.

The idea of BD was conceived by us, so the only prior study was our own proof of concept investigation⁹. BD and BT likely hold prognostic value because each is a surrogate for invasive tumour burden. BT is a one dimensional quantitative measurement, while BD is two dimensional semi-quantitative measurement that compliments BT. The idea of density bears similarity to Clark's concept of growth phase²⁶, in that confluent nodules of tumour cells equate to vertical growth phase and metastatic potential, and one could also speculate BD might correlate with the degree of paracrine signaling, which is likely to be important for melanoma growth, invasion and metastasis²⁷. Thus, BD's prognostic value may relate to both invasive tumour burden and intercellular crosstalk. Aside from BD, few studies investigating invasive burden have been performed for melanoma^{28, 29} and these have generally been small and some lacked correlation with outcome³⁰.

BD is a semi-quantitative and subjective. Nevertheless, we demonstrate its accuracy and precision, regardless of rater experience, and that precision is achievable even with written instruction alone. In future, it is possible that image analysis could become the routine way of assessing BD. However, we believe that the requirement for imaging could provide a barrier to the widespread adoption of BD scoring across pathology departments, and that the very simplicity of the scoring system that we describe is what makes it attractive for translation.

The size of this study allowed us to ensure that the two thirds or more cut point for T stage upstaging could be validated in new cases. In future, it will be important to address whether our discovered cut point is actually applicable to individual T stages. However, it will not be possible to integrate BD directly into AJCC8 in future because the staging system criteria are already established. This study therefore serves only as proof of concept for the idea of incorporating BD into future staging systems beyond AJCC8. As an alternative we looked at

ways of using BD as a staging adjunct. Exploratory analysis identified a combination of BT and BD, which we called the TBS, as the best adjunct for AJCC8. This makes sense if one considers how the BD is measured, namely the BD measurement window is always the same width but its depth varies because it is the same as the BT. Thus, two melanomas with a BD of 50% will not have comparable tumour cell numbers in the BD window if one has a BT of 1.0 mm melanoma and the other 4.0 mm. The TBS, which multiplies BD and BT, mitigates this problem and acts as a crude approximation to tumour burden in the BD measurement window. The PI that was derived from the TBS and AJCC8 could be stratified into four risk groups that refined the prognostic value of AJCC8 stage. For example, melanomas in PI risk group 1 had 5 year MSS of 98% and included patients with AJCC8 stage IA, IB and IIA. This group could potentially have less frequent follow up and/or be spared sentinel node biopsy, while simultaneously ensuring that only high risk patients actually have this procedure. In contrast, prognostic group 4 are high risk cases with 5 year MSS of 56%, and so these could be candidates for adjuvant treatment. Thus, the combination of TBS and stage has potential for creating personalised follow up and treatment plans.

In summary, we show that BD is a novel histomorphological prognostic feature with potential clinical value. Further studies will be needed for true external validation in other centres. Such studies would be made more feasible if BD was to be routinely reported by histopathologists, which would be likely to occur if it was adopted as a standard part of the pathology report, e.g. if BD were recommended by the College of American Pathologists (in the USA) and Royal College of Pathologists (in the UK) as a feature in their pathology report guidelines.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Conflicts of interest and sources of funding

None declared. Research was supported by the British Skin Foundation, the Hope Foundation, the Pathological Society of Great Britain & Ireland and Cancer Research UK.

References

1. [Accessed July 20, 2017] Skin cancer statistics. Cancer Research UK. 2017. Available at: <http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/skin-cancer#heading-Zero>
2. Glazer AM, Winkelmann RR, Farberg AS, et al. Analysis of trends in uS melanoma incidence and mortality. JAMA Dermatology. 2016
3. Volpe VO, Klufas DM, Hegde U, et al. The new paradigm of systemic therapies for metastatic melanoma. Journal of the American Academy of Dermatology. 2017; 77:356–368. [PubMed: 28711086]
4. Weber J, Mandala M, Del Vecchio M, et al. Adjuvant nivolumab versus ipilimumab in resected stage III or IV melanoma. The New England Journal of Medicine. 2017
5. Long GV, Hauschild A, Santinami M, et al. Adjuvant dabrafenib plus trametinib in stage III BRAF-mutated melanoma. The New England Journal of Medicine. 2017
6. Balch CM, Gershenwald JE, Soong S-j, et al. Final version of 2009 AJCC melanoma staging and classification. Journal of Clinical Oncology. 2009; 27:6199–6206. [PubMed: 19917835]

7. 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: A Cancer Journal for Clinicians. 2017
8. Lo JA, Fisher DE. The melanoma revolution: From uV carcinogenesis to a new era in therapeutics. Science. 2014; 346:945–949. [PubMed: 25414302]
9. Rashed H, Flatman K, Bamford M, et al. Breslow density is a novel prognostic feature in cutaneous malignant melanoma. Histopathology. 2016
10. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2015. Available at: <https://www.R-project.org/>
11. RStudio Team. RStudio: Integrated Development for R. Boston, MA: RStudio, Inc; 2015. (2015). Available at <http://www.rstudio.com/>
12. Shrout PE, Fleiss JL. Intraclass correlations: Uses in assessing rater reliability. Psychological Bulletin. 1979; 86:420–428. [PubMed: 18839484]
13. Gamer M, Lemon J, Fellows E, Singh P. Irr: Various coefficients of interrater reliability and agreement. 2012. Available at: <https://CRAN.R-project.org/package=irr>
14. Therneau T. A package for survival analysis in s, version 2.38. 2015. Available at: <http://CRAN.R-project.org/package=survival>
15. Kassambara A, Kosinski M. Survminer: Drawing survival curves using 'ggplo2'. 2016. Available at: <https://CRAN.R-project.org/package=survminer>
16. Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. European Heart Journal. 2014; 35:1925–1931. [PubMed: 24898551]
17. Royston P, Altman DG. External validation of a cox prognostic model: Principles and methods. BMC Medical Research Methodology. 2013; 13:33. [PubMed: 23496923]
18. Rahman MS, Ambler G, Choodari-Oskooei B, et al. Review and evaluation of performance measures for survival prediction models in external validation settings. BMC Medical Research Methodology. 2017; 17:60. [PubMed: 28420338]
19. Babyak MA. What you see may not be what you get: a brief, nontechnical introduction to overfitting in regression-type models. Psychosom Med. 2004; 66:411–21. [PubMed: 15184705]
20. Steyerberg EW. Overfitting and Optimism in Prediction Models Clinical prediction models. A practical approach to development, validation and updating. Steyerberg EW, editor Springer Science+Business Media, LLC; 2009. 83–100.
21. Harrell FE. Rms: Regression modeling strategies. 2017. Available at: <https://CRAN.R-project.org/package=rms>
22. Burnham KP, Anderson DR, Huyvaert KP. AIC model selection and multimodel inference in behavioral ecology: Some background, observations, and comparisons. Behav Ecol Sociobiol. 2011; 65:23–25.
23. McShane LM, Altman DG, Sauerbrei W, et al. REporting recommendations for tumour mARKer prognostic studies (REMARK). Br J Cancer. 2005; 93:387–91. [PubMed: 16106245]
24. Gould-Rothberg BE, Rimm DL. Construction and analysis of multi-parameter prognostic models for melanoma outcome. Methods Mol Biol. 2014; 1102:227–258. [PubMed: 24258982]
25. Berger AC, Davidson RS, Poitras JK, et al. Clinical impact of a 31-gene expression profile test for cutaneous melanoma in 156 prospectively and consecutively tested patients. Current Medical Research and Opinion. 2016; 32:1599–1604. [PubMed: 27210115]
26. Clark WH, Elder DE, Guerry D, et al. A study of tumor progression: The precursor lesions of superficial spreading and nodular melanoma. Human Pathology. 1984; 15:1147–1165. [PubMed: 6500548]
27. Wang JX, Fukunaga-Kalabis M, Herlyn M. Crosstalk in skin: Melanocytes, keratinocytes, stem cells, and melanoma. Journal of Cell Communication and Signaling. 2016; 10:191–196. [PubMed: 27553358]
28. Friedman RJ, Rigel DS, Kopf AW, et al. Volume of malignant melanoma is superior to thickness as a prognostic indicator. preliminary observation. Dermatologic Clinics. 1991; 9:643–648. [PubMed: 1934639]

29. Bahmer FA, Hantirah S, Baum HP. Rapid and unbiased estimation of the volume of cutaneous malignant melanoma using cavalieri's principle. *The American Journal of Dermatopathology*. 1996; 18:159–164. [PubMed: 8739991]
30. Voss B, Wilop S, Jonas S, et al. Tumor volume as a prognostic factor in resectable malignant melanoma. *Dermatology (Basel, Switzerland)*. 2014; 228:66–70.

Figure 1.

Illustration of the Breslow density window. The BD was the estimated percentage of available dermis within a specified window (brown lines) occupied by invasive melanoma cells (blue circles). The window was an imagined crude rectangle, approximate width 2mm and height extending to the deepest invasive melanoma cell (yellow circle). The window was adjusted laterally to maximise the BD, with the proviso that the window always had to contain the deepest melanoma cell. The top of the window followed the base of the epidermis, passing beneath junctional nests. The bottom of the window replicated any major surface undulations as illustrated, but minor undulations were ignored. Major structures traversing the dermis such as hair follicles (illustrated) were not considered to be part of the available dermis. On many microscopes, the x10 objective is approximately 2 mm in diameter and this was helpful for defining the lateral edges of the window. Once the window was mentally constructed, the BD was then estimated to the nearest 5%, except below 5% and above 95%, where nearest 1% was used.

Figure 2.

Breslow density precision and effect on survival. Interobserver agreement for Breslow density, where each numbered box is a separate melanoma case and rater 1 is a senior trainee with no specialist dermatopathology training, rater 2 is a dermatopathologist with extensive experience of reporting melanoma cases and rater 3 is a trained medical student (panel A). Kaplan Meier plots for Breslow density tertiles showing OS, MSS and MFS (panels B-D).

Figure 3.

BD as a TNM upstaging feature. A forest plot of BD cut points is shown, with each cut point value and above specifying a high risk group; suboptimal training cut points are black, the optimal training cut point is red and the validation of the red cut point is blue (A). The strategy for creating risk groups is shown, with colour-coding to match the respective Kaplan Meier curves (B). Kaplan Meier plots to assess ability of BD to upstage T stage from "a" to "b", with T1, T2, T3 and T4 plots (C-F respectively). Kaplan Meier plot of composite curves for T1-4 (G).

Figure 4.

Kaplan Meier plot of training case risk groups (A). Kaplan Meier plot of validation case risk groups after applying the model fitted on training cases (B). Survival curves for PI groups in validation cases based on predictions from the trained CPH model (C). Measures of discrimination and calibration in Leicester (training) and Nottingham (validation) cases (D). Kaplan Meier plots of PI groups derived from the final CPH model using combined training and validation cases (E). Cross tabulation of AJCC8 stage and PI groups for combined training and validation cases (F). Nomogram to predict outcome for individuals. The top row

is used to allocate points for each of an individual patient's TBS (transformed to natural log) and AJCC8 stage. Summation of these two scores yields the total points in the fourth row, which can be used to either read off a 5 year survival estimate from the 6th row or allocate the patient to one of the four risk groups shown in Figure 4E. For completeness, the corresponding value of the associated linear predictor from the CPH model is also shown at the bottom (G).

Table 1
Baseline features of Leicester patients.

	Overall	Lower Tertile	Middle Tertile	Upper Tertile	p
n	970	315	317	338	
Sex = Male (%)	458 (47.2)	155 (49.2)	147 (46.4)	156 (46.2)	0.7
Age (mean (sd))	59.88 (16.43)	57.74 (16.17)	58.80 (16.37)	62.88 (16.32)	<0.001
Site (%)					0.7
Head & neck	182 (18.8)	60 (19.0)	53 (16.7)	69 (20.4)	
Upper limb	174 (17.9)	52 (16.5)	59 (18.6)	63 (18.6)	
Trunk	317 (32.7)	112 (35.6)	106 (33.4)	99 (29.3)	
Lower limb	248 (25.6)	74 (23.5)	86 (27.1)	88 (26.0)	
Acral	49 (5.1)	17 (5.4)	13 (4.1)	19 (5.6)	
Type (%)					<0.001
SSMM	659 (67.9)	256 (81.3)	233 (73.5)	170 (50.3)	
LMM	70 (7.2)	38 (12.1)	14 (4.4)	18 (5.3)	
NMM	162 (16.7)	3 (1.0)	41 (12.9)	118 (34.9)	
ALMM	29 (3.0)	10 (3.2)	7 (2.2)	12 (3.6)	
Desmoplastic	16 (1.6)	3 (1.0)	9 (2.8)	4 (1.2)	
Unknown	34 (3.5)	5 (1.6)	13 (4.1)	16 (4.7)	
Breslow (median [IQR])	0.90 [0.50, 2.00]	0.50 [0.35, 0.60]	0.90 [0.70, 1.70]	2.10 [1.20, 4.30]	<0.001
Ulcer = Yes (%)	172 (17.7)	4 (1.3)	39 (12.3)	129 (38.2)	<0.001
Mitotic count (median [IQR])	1.00 [0.00, 3.00]	0.00 [0.00, 0.00]	1.00 [0.00, 2.00]	3.00 [1.00, 8.00]	<0.001
Microsatellites = Yes (%)	17 (1.8)	0 (0.0)	2 (0.6)	15 (4.4)	<0.001
AJCC7 (%)					<0.001
IA	376 (38.8)	268 (85.1)	93 (29.3)	15 (4.4)	
IB	326 (33.6)	43 (13.7)	151 (47.6)	132 (39.1)	
IIA	95 (9.8)	2 (0.6)	36 (11.4)	57 (16.9)	
IIB	78 (8.0)	2 (0.6)	20 (6.3)	56 (16.6)	
IIC	78 (8.0)	0 (0.0)	15 (4.7)	63 (18.6)	
III	17 (1.8)	0 (0.0)	2 (0.6)	15 (4.4)	

Table 2
Cox regression model for Breslow density and recognised prognostic features.

	Crude					Adjusted		
Variable	Total	Event	HR	2.5 % to 97.5 %	p value	HR	2.5 % to 97.5 %	p value
OS								
Breslow density	53.76 (±33.87)	72.32 (±27.96)	1.02	1.02 to 1.03	< 0.001	1.01	1.01 to 1.02	< 0.001
Age	59.88 (±16.43)	70.87 (±14.97)	1.06	1.05 to 1.08	< 0.001	1.05	1.04 to 1.06	< 0.001
Sex					< 0.001			0.007
Female	512	100 (19.53%)	1	ref		1	ref	
Male	458	137 (29.91%)	1.73	1.33 to 2.24		1.48	1.12 to 1.96	
Site					0.005			0.09
Central	499	140 (28.06%)	1	ref		1	ref	
Peripheral	471	97 (20.59%)	0.69	0.53 to 0.89		0.79	0.59 to 1.04	
Breslow	1.95 (±2.88)	3.92 (±4.42)	1.15	1.13 to 1.17	< 0.001	1.06	1.03 to 1.10	< 0.001
Ulcer					< 0.001			0.001
No	798	138 (17.29%)	1	ref		1	ref	
Yes	172	99 (57.56%)	4.8	3.70 to 6.23		1.69	1.23 to 2.33	
Mitoses / mm²	2.80 (±5.09)	6.15 (±7.35)	1.09	1.07 to 1.10	< 0.001	1.03	1.01 to 1.04	0.003
Microsatellites					< 0.001			0.1
No	953	222 (23.29%)	1	ref		1	ref	
Yes	17	15 (88.24%)	7.94	4.69 to 13.45		1.59	0.91 to 2.79	
MSS								
Breslow density	53.76 (±33.87)	80.18 (±21.10)	1.04	1.03 to 1.05	< 0.001	1.02	1.01 to 1.03	< 0.001
Age	59.88 (±16.43)	64.25 (±14.75)	1.03	1.01 to 1.04	< 0.001	1.01	0.99 to 1.02	0.26
Sex					0.005			0.11
Female	512	44 (8.59%)	1	ref		1	ref	
Male	458	62 (13.54%)	1.75	1.19 to 2.58		1.41	0.92 to 2.15	
Site					0.03			0.19
Central	499	64 (12.83%)	1	ref		1	ref	
Peripheral	471	42 (8.92%)	0.66	0.44 to 0.97		0.75	0.49 to 1.15	
Breslow	1.95 (±2.88)	4.86 (±4.70)	1.16	1.14 to 1.19	< 0.001	1.09	1.05 to 1.13	< 0.001
Ulcer					< 0.001			0.005
No	798	52 (6.52%)	1	ref		1	ref	
Yes	172	54 (31.40%)	6.76	4.60 to 9.93		1.97	1.23 to 3.16	
Mitoses / mm²	2.80 (±5.09)	8.07 (±8.18)	1.1	1.09 to 1.12	< 0.001	1.04	1.02 to 1.06	< 0.001
Microsatellites					< 0.001			0.56
No	953	99 (10.39%)	1	ref		1	ref	
Yes	17	7 (41.18%)	7.74	3.58 to 16.76		1.28	0.57 to 2.87	
MFS								
Breslow density	53.76 (±33.87)	78.95 (±21.93)	1.03	1.03 to 1.04	< 0.001	1.02	1.01 to 1.03	< 0.001
Age	59.88 (±16.43)	64.28 (±15.28)	1.02	1.01 to 1.03	< 0.001	1	0.99 to 1.01	0.9

Variable	Total	Event	Crude			Adjusted		
			HR	2.5 % to 97.5 %	p value	HR	2.5 % to 97.5 %	p value
sex					0.003			0.01
Female	512	73 (14.26%)	1	ref		1	ref	
Male	458	96 (20.96%)	1.59	1.18 to 2.16		1.54	1.11 to 2.15	
Site					0.4			0.69
Central	499	91 (18.24%)	1	ref		1	ref	
Peripheral	471	78 (16.56%)	0.88	0.65 to 1.19		0.94	0.67 to 1.30	
Breslow	1.95 (±2.88)	4.46 (±4.21)	1.14	1.12 to 1.16	< 0.001	1.08	1.04 to 1.11	< 0.001
Ulcer					< 0.001			0.01
No	798	90 (11.28%)	1	ref		1	ref	
Yes	172	79 (45.93%)	5.8	4.27 to 7.86		1.64	1.12 to 2.40	
Mitoses / mm ²	2.80 (±5.09)	7.47 (±8.05)	1.13	1.11 to 1.15	< 0.001	1.06	1.04 to 1.09	< 0.001
Microsatellites					< 0.001			0.15
No	953	158 (16.58%)	1	ref		1	ref	
Yes	17	11 (64.71%)	9.52	5.13 to 17.66		1.64	0.84 to 3.18	

Table 3
Final Cox proportional hazards model for creating a prognostic index.

	Adjusted HR	95% CI	p value	Optimised HR**
lnTBS *	1.62	1.28 - 2.04	< 0.001	1.60
AJCC8				
IB	1.63	0.80 - 3.32	0.18	1.61
IIA	2.59	1.18 - 5.70	0.02	2.54
IIB	3.43	1.35 - 8.70	0.01	3.34
IIC	5.14	2.09 - 12.63	< 0.001	4.96
III	9.21	3.48 - 24.42	< 0.001	8.79

* Natural log Breslow.

** Based on shrunk parameters following bootstrap analysis of model optimism.