A nationwide survey of observer variation in the diagnosis of thin cutaneous malignant melanoma including the MIN terminology

CRC Melanoma Pathology Panel

Members of the CRC Melanoma Pathology Panel
M G Cook
T J Clarke
S Humphreys
A Fletcher
K M McLaren
N P Smith
A Stevens
J M Theaker
with advisor J Melia

Correspondence to:
Dr J Melia, Cancer Screening and Evaluation Unit, Block D, 15 Cotswold Road, Sutton, Surrey SM2 5NG.

Accepted for publication 5 November 1996

Abstract

Aim—To investigate observer variation in the diagnosis of thin cutaneous malignant melanoma and related lesions in a nationwide sample of histopathologists in the UK.

Methods—Out of a random sample of 195 pathologists, 148 (76%) participated in two circulations, the first with 20 slides and the second with 25 slides. The results were compared with those for the Cancer Research Campaign (CRC) Melanoma Pathology Panel, consisting of seven histopathologists and one dermatopathologist, which had developed and evaluated diagnostic criteria.

Results—In the first circulation, when no standardised diagnostic criteria were used, a fair level of agreement was achieved for an overall diagnosis using the categories benign naevi with no atypia, benign naevi with atypia and melanoma (Kappa = 0.45). This was low compared with the agreement of the panel who used agreed criteria (Kappa = 0.75). Moreover, participants in the nationwide survey were more likely to diagnose melanoma and less likely to diagnose benign naevi without atypia than the panel. In the second circulation, when diagnostic criteria and diagrams were used, there was a higher level of agreement for overall diagnosis using the categories benign, melanocytic intraepidermal neoplasia (MIN) with or without microinvasion and melanoma with vertical growth phase, and was the same as that achieved by the panel using the same criteria (Kappa = 0.68).

Conclusions—As the incidence rate of thin melanomas has been increasing in the UK, it is important that standardised diagnostic criteria are used to ensure accurate reporting of incidence and correct management of patients. The use of MIN and the vertical growth phase seemed to be generally acceptable.
(*J Clin Pathol* 1997;50:202–205)

Keywords: melanoma; skin; observer variation; pathology; diagnosis.

In several countries, there has been a notable increase in the incidence rate of cutaneous malignant melanoma and a shift towards diagnosis of lesions with a thin Breslow thickness.[1–3] These thin or borderline lesions are dispropor-

tionately difficult to diagnose accurately and consistently. This has important implications at the individual level for patient management and at a nationwide level for the monitoring of trends in incidence of the disease. In 1991, the Cancer Research Campaign (CRC) Melanoma Pathology Panel was set up to investigate observer variation in the diagnosis of melanoma and to develop criteria to improve the differentiation of thin melanomas from melanocytic lesions showing severe dysplasia.[4] The panel included a histopathologist from each of the seven study areas associated with the CRC "Mole Watcher" health education programme[5] and a dermatopathologist (NS) with a major interest in melanoma. They studied 96 randomly selected lesions from the study areas and found that overall observer variation was reduced with use of diagnostic criteria developed by the panel. Nevertheless, there remained poor concordance in distinguishing benign lesions with atypia from melanoma in situ lesions. These lesions represent a presumed progression and as their clinical management is the same,[6] the panel recommended that both lesions could be termed melanocytic intraepidermal neoplasia (MIN). Melanomas with a vertical growth phase, based on the definitions of Clark *et al*,[7] would remain the type of melanoma requiring more active treatment and follow up. As in most other studies of observer variation of melanoma diagnosis,[8–11] the number of histopathologists involved was small and included those with a special interest in melanoma. Therefore, the panel surveyed histopathologists in the UK to study observer variation in the differential diagnosis of thin melanomas, and to test the acceptability and the effect of diagnostic criteria for MIN on observer variation.

Methods

Histopathologists (n = 250) were randomly selected from the Royal College of Pathologists Membership Directory; 191 were eligible to participate, the remaining 59 being retired, doing other specialised work, ill, emigrated, or untraceable. Four histopathologists were recommended for inclusion because their work was more appropriate than colleagues who had been selected originally. The aim of the first circulation in the survey was to study variation in diagnosis of melanoma without the use of definitions. Twenty slides were chosen from those previously studied by the panel to provide a selection of four benign naevi with no

CRC NATIONAL MELANOMA SURVEY PROFORMA
1st Circulation

Pathologist's Code   Circulation   Slide No.

| 9 | 3 | 1 |

DIAGNOSIS: (please circle one answer)

Benign with no atypia / Benign with atypia / Melanoma

CRC NATIONAL MELANOMA PATHOLOGY SURVEY PROFORMA
2nd Circulation

Pathologist ....

Year/circulation | 9 | 4 | 2 |

Slide Number

OVERALL DIAGNOSIS   (Circle one only)

Benign   1

MIN with no microinvasion   2

MIN with microinvasion   3

MELANOMA - vertical growth phase
(MIN may or may not be present)   4

Figure 1   CRC National Melanoma Survey proformas.

atypia, five benign naevi with atypia, one borderline benign/melanoma lesion, five melanomas with thickness <0.76 mm, and five melanomas with a thickness ranging from 0.76 to 2.77 mm according to the panel's majority diagnosis. The aim of the second circulation was to study variation in diagnosis using the panel's definition of MIN[4] and a diagram to illustrate the diagnoses. Nineteen slides were used from the first circulation so that comparisons could be made and six new slides from the panel's previous study were selected to ensure that each of the diagnostic categories were represented by several lesions. Five lesions were benign, six MIN with no microinvasion, four MIN with microinvasion, two MIN with microinvasion which might also be interpreted as melanoma with vertical growth phase, and eight melanoma with vertical growth phase according to the panel's diagnosis. Seventeen consecutive sections were cut from one representative block from each lesion and they were checked by one pathologist (MC) to ensure that the lesion was present in all slides. It took 12 months to circulate the 17 sets of

slides among the 195 pathologists for the first circulation, and 18 months to circulate the same set to each pathologist in the second circulation. During the second circulation, one set was broken in the post and another lost so some pathologists had to read an adjacent set. Individual results were kept confidential and pathologists were asked not to discuss their readings with colleagues who had participated. Pathologists were told the age of the patient and the anatomical site of the lesion but no other details. A standard proforma was used to collect the data (fig 1). Of the 195 eligible pathologists, 148 (76%) participated in the both circulations. The data were collected and analysed by the Cancer Screening Evaluation Unit. The interobserver agreement for diagnosis was studied using Kappa statistics (K) to assess the measurement of overall agreement adjusted for the agreement expected by chance.[12] The percentage of readings by each reader which are the same as the majority or most frequent value of each slide was also used to compare results between observers.

Results
In the first circulation, the overall level of agreement using three categories of diagnosis (benign naevi with no atypia, benign naevi with atypia and melanoma) was lower for the nationwide sample (K = 0.45) than for the panel (K = 0.75) (table 1). The level of agreement was poorest for benign naevi with atypia in both groups of pathologists. Only 42 (28%) of the nationwide sample had a percentage agreement ≥80% with their majority diagnoses. The majority diagnosis for each lesion by the study sample agreed with that determined by the panel for 15 of the 20 lesions. However, there was wide variation in the distribution of diagnoses made by the nationwide sample for each lesion, and, overall, the panel was significantly more likely to diagnose benign lesions with no atypia (46%) and less likely to diagnose melanoma (24%) than the nationwide sample (25% and 48%, respectively, p < 0.001; table 1).

In the second circulation, the overall level of agreement using four categories of diagnosis (benign, MIN with no microinvasion, MIN with microinvasion, and melanoma with vertical growth phase) was similar for the nationwide sample (K = 0.63) and the panel (K = 0.64). Using three categories (benign, MIN with and without microinvasion, and melanoma with vertical growth phase) slightly increased the level of agreement to K = 0.68 in both groups. It was not possible to calculate a standard error for the nationwide sample because of a small number of missing readings (<15 in each circulation). Of the nationwide sample, 66% (97) had a percentage agreement ≥80% with their majority diagnoses. The majority diagnosis by the study sample for each lesion was the same as that made by the CRC panel for 21 of the 25 lesions. The variation in diagnoses by the study sample was smaller than in the first circulation, and the distribution of diagnoses recorded by the nationwide sample was very similar to that recorded by the panel with 25% benign, 41% MIN and 34%

Table 1   Level of agreement (Kappa) for the diagnosis of slides read by the nationwide sample (148 readers) and by the panel (eight readers)

		Nationwide sample			Panel		
Circulation	Diagnosis	Kappa	No. of readings	%	Kappa	No. of readings	%
First	Benign naevi with no atypia	0.52	801	25	0.76	74	46
	Benign naevi with atypia	0.19	891	27	0.66	47	29
	Melanoma	0.60	1558	48	0.83	39	24
	Overall Kappa (SE)	0.45 (NA)			0.75 (0.02)		
Second	Benign	0.67	925	25	0.68	41	20
	MIN + no microinvasion	0.51	801	22	0.57	48	24
	MIN + microinvasion	0.50	718	19	0.53	42	21
	Melanoma (vertical growth phase)	0.76	1245	34	0.76	69	35
	Overall Kappa (SE)	0.63 (NA)			0.64 (0.02)		
	Benign	0.67	925	25	0.68	41	20
	MIN ± microinvasion	0.62	1519	41	0.60	90	45
	Melanoma (vertical growth phase)	0.76	1245	34	0.76	69	35
	Overall Kappa (SE)	0.68 (NA)			0.68 (0.03)		

NA = not applicable because a few readings were mising from some pathologists.

melanoma with vertical growth phase (table 1). The results did not differ when the four recommended pathologists were excluded. Results for analyses with the 19 slides common to both circulations were similar to those presented in table 1. Analyses with the full set of 195 pathologists who took part in the first circulation gave similar results to those for the 148 pathologists who took part in both circulations.

Discussion

This is the first large survey of histopathologists in the United Kingdom investigating the level of agreement for the diagnosis of thin melanomas. Concordance was poor without the use of standardised diagnostic criteria (K = 0.45). In other studies, good agreement for the diagnosis of melanoma[13 14] has been reported, but another study of atypical naevi reported less satisfactory agreement.[15] The results will depend on the experience of the pathologists and the type of lesions selected. In this study, the sample of pathologists are considered to be representative of those nationwide as they were a random selection with a good response rate of 75% to both circulations. The lesions were chosen for their diagnostic difficulty and it is also possible that some variation between sections could have contributed to the observer variation, although the first and last sections were checked for this. Schmoeckel[16] reported inconsistency in the terminology of "early" or thin melanomas used by histopathologists with a special interest in melanocytic lesions. He predicted that there must be even greater disagreement between less specialised pathologists, as found in this study. The panel chose to investigate the use of definitions of MIN and the vertical growth phase in the second circulation because they have been reported to have high prognostic value.[7] It was not feasible to conduct other circulations because the response rate would certainly have dropped to an unacceptable level. The use of MIN was acceptable to the nationwide sample and ensured a level of agreement (K = 0.63) similar to that of the panel. The level of agreement is not high for either the panel or the

nationwide sample, but the study was conducted using a difficult set of borderline lesions. A higher level of agreement could be expected with a routine case mix seen by most pathologists. The use of MIN merits further evaluation on a larger sample of lesions.

As indicated by others,[4 16] there is clearly a need for the use of standardised criteria to reduce observer variability in the diagnosis of thin melanomas. This is important for the monitoring of trends in incidence of melanoma and patient management.

1 MacKie RM, Hole D. Audit of public education campaign to encourage earlier detection of malignant melanoma. BMJ 1992;304:1012–15.
2 MacLennan R, Green AC, McLeod GRC, Martin NG. Increasing incidence of cutaneous melanoma in Queensland, Australia. J Natl Cancer Inst 1992;84:1427–33.
3 Thorn M, Ponten F, Bergstrom R, Sparen P, Adami HO. Trends in tumour characteristics and survival of malignant melanoma 1960–84: a population-based study in Sweden. Br J Cancer 1995;70:743–8.
4 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.
5 Melia J. Early detection of cutaneous malignant melanoma in Britain. Int J Epidemiol 1995;24(Suppl 1):S39–44.
6 Elder D, Clark W, Elenitsas R, Guerry D, Halpern A. The early and intermediate precursor lesions of tumor progression in the melanocytic system. Common acquired nevi and atypical (dysplastic) nevi. Semin Diag Pathol 1993;10:18–35.
7 Clark WHJ, Elder DE, Guerry IVD, Bloggs J. Model predicting survival in Stage I melanoma based on tumour progression. J Natl Cancer Inst 1989;81:1893–904.
8 Colloby PS, West KP, Fletcher A. Observer variation in the measurement of Breslow depth and Clark's level in thin cutaneous malignant melanoma. J Pathol 1991;163:245–50.
9 Heenan PJ, Matz LR, Blackwell JB, Kelsall GR, Singh A, ten Seldam RE, et al. Inter-observer variation between pathologists in the classification of cutaneous malignant melanoma in Western Australia. Histopathology 1984;8:717–29.
10 Philipp R, Hastings A, Briggs J, Sizer J. Are malignant melanoma time trends explained by changes in histopatho-

logical criteria for classifying pigmented skin lesions? *J Epidemiol Community Health* 1987;**42**:14–16.

11 van der Esch EP, Muir CS, Nectoux J, Macfarlane G, Maisonneuve P, Bharucha H, *et al.* Temporal change in diagnostic criteria as a cause of the increase of malignant melanoma over time is unlikely. *Int J Cancer* 1991;**47**:483–90.

12 Fleiss JL. *Statistical methods for rates and proportions.* New York: John Wiley, 1981.

13 Duray PH, DerSimonian R, Barnhill R, Stenn K, Ernstoff MS, Fine J, *et al.* An analysis of interobserver recognition of the histopathologic features of dysplastic nevi from a mixed group of nevomelanocytic lesions. *J Am Acad Dermatol* 1992;**27**:741–9.

14 de Wit PE, van't Hof-Grootenboer B, Ruiter DJ, Bondi R, Brocker EB, Cesarini JP, *et al.* Validity of the histopathological criteria used for diagnosing dysplastic naevi. *Eur J Cancer* 1993;**29A**:831–9.

15 Hastrup N, Clemmensen OJ, Spaun E, Sondergaard K. Dysplastic Naevus: Histological criteria and their interobserver reproducibility. *Histopathology* 1994;**24**:503–9.

16 Schmoeckel C. How consistent are dermatologists in reading early malignant melanomas and lesions "precursor" to them? *Am J Dermatopathol* 1984;**6**:13–24.