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

CRC Melanoma Pathology Panel

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. These thin or borderline lesions are disproportionately 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. The panel included a histopathologist from each of the seven study areas associated with the CRC “Mole Watcher” health education programme and a dermatopathologist 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, 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, would remain the type of melanoma requiring more active treatment and follow up. As in most other studies of observer variation in melanoma diagnosis, 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
Survey of observer variation in the diagnosis of thin cutaneous malignant melanoma

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. 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)

<table>
<thead>
<tr>
<th>Circulation</th>
<th>Diagnosis</th>
<th>Nationwide sample</th>
<th>Panel</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Kappa</td>
<td>No. of readings</td>
<td>%</td>
</tr>
<tr>
<td>First</td>
<td>Benign naevi with no atypia</td>
<td>0.52</td>
<td>801</td>
</tr>
<tr>
<td></td>
<td>Benign naevi with atypia</td>
<td>0.19</td>
<td>891</td>
</tr>
<tr>
<td></td>
<td>Melanoma</td>
<td>0.60</td>
<td>1558</td>
</tr>
<tr>
<td></td>
<td>Overall Kappa (SE)</td>
<td>0.65 (NA)</td>
<td></td>
</tr>
<tr>
<td>Second</td>
<td>Benign</td>
<td>0.67</td>
<td>925</td>
</tr>
<tr>
<td></td>
<td>MIN + microinvasion</td>
<td>0.51</td>
<td>801</td>
</tr>
<tr>
<td></td>
<td>MIN + microinvasion</td>
<td>0.50</td>
<td>718</td>
</tr>
<tr>
<td></td>
<td>Melanoma (vertical growth phase)</td>
<td>0.76</td>
<td>1245</td>
</tr>
<tr>
<td></td>
<td>Overall Kappa (SE)</td>
<td>0.63 (NA)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Benign</td>
<td>0.67</td>
<td>925</td>
</tr>
<tr>
<td></td>
<td>MIN ± microinvasion</td>
<td>0.62</td>
<td>1519</td>
</tr>
<tr>
<td></td>
<td>Melanoma (vertical growth phase)</td>
<td>0.76</td>
<td>1245</td>
</tr>
<tr>
<td></td>
<td>Overall Kappa (SE)</td>
<td>0.68 (NA)</td>
<td></td>
</tr>
</tbody>
</table>

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

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 has been reported, but another study of atypical naevi reported less satisfactory agreement. 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 observed variation, although the first and last sections were checked for this. Schmoeckel 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. 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, 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.

We are very grateful to the pathologists who participated in the study. We would like to thank the Cancer Research Campaign for funding this research, Richard Trent for preparing the slides, Jeremy Sloane and Joyce Basili for assistance with data collection and processing, and Ruth Ellman for advising the panel.

Members of the CRC Melanoma Panel: M G Cook, Royal Surrey Hospital, Guildford; T J Clarke, Royal Devon and Exeter Hospital, Exeter; S Humphreys, King's College Hospital, London; A Fletcher, Leicester Royal Infirmary; K M McLean, University of Dundee; P F Smith, St Thomas's Hospital, London; A Stevens, Queen's Medical Centre, Nottingham; J M Theaker, Southampton General Hospital; with adviser J Melia of the Cancer Screening Evaluation Unit, Institute of Cancer Research, Sutton, Surrey. Neil Smith sadly died before this paper was submitted for publication.

10 Philips R, Hastings A, Briggs J, Sizer J. Are malignant melanoma time trends explained by changes in histopatho-
Survey of observer variation in the diagnosis of thin cutaneous malignant melanoma

16 Schmoeckel C. How consistent are dermatologists in reading early malignant melanomas and lesions "precur-