Journal of Pathology
J Pathol 2002; **196**: 459–466.
Published online 11 February 2002 in Wiley InterScience (www.interscience.wiley.com). **DOI:** 10.1002/path.1061

Original Paper

Inter-observer variation in the histopathological diagnosis of clinically suspicious pigmented skin lesions

Lieve Brochez[1,2], Evelien Verhaeghe[2,3], Edouard Grosshans[4], Eckhart Haneke[5], Gérald Piérard[6], Dirk Ruiter[7] and Jean-Marie Naeyaert[2]*

[1] *Research Associate, Fund for Scientific Research-Flanders, Belgium*
[2] *Dermatology Department, University Hospital, Ghent, Belgium*
[3] *Scientific Co-worker, Fund for Scientific Research-Flanders, Belgium*
[4] *Dermatology Department, University Hospital, Strasbourg, France*
[5] *Dermatology Department, Academic Teaching Hospital University Düsseldorf, Wuppertal, Germany*
[6] *Dermatopathology Department, University Medical Center Sart Tilman, Liège, Belgium*
[7] *Pathology Department, University Medical Center, Nijmegen, The Netherlands*

*Correspondence to:
Dr J. M. Naeyaert, Department of Dermatology, University Hospital Ghent, De Pintelaan 185, 9000 Ghent, Belgium.
E-mail:
JeanMarie.Naeyaert@rug.ac.be

Received: 19 December 2000
Revised: 10 July 2001
Accepted: 18 November 2001

Abstract

When a biopsy is taken of a suspicious pigmented skin lesion, histological examination is expected to establish the definitive diagnosis. This study evaluated the inter-observer variation of 20 pathologists in the histological diagnosis of a randomly selected set of suspicious pigmented skin lesions (PSLs), by comparing their diagnoses to a reference diagnosis. Overall sensitivity for melanoma was 87%, ranging from 55% to 100% between the observers. Sensitivity was significantly lower for thin (Breslow thickness <1 mm) than for thick melanomas (83% versus 97%, $p=0.005$). Overall melanoma specificity was 94%, ranging from 83% to 100% between observers. Dysplastic naevus was the most important source of false-positive diagnoses, mainly *in situ* melanomas. Positive and negative predictive values in the given test set were 75% and 97%, respectively. In the case of melanoma, there was quite some variation in measured Breslow thickness. This would have led to a different therapeutic approach in 12% of the readings. Some of the variation seemed to be due to a different interpretation of the presence of a co-existent naevus. In 9% (3/35) of the readings, participants did not agree on the presence of ulceration. These results reflect a tendency to overdiagnose mainly thin melanomas in general histopathological practice. They also demonstrate variation in the assessment of major prognostic factors of melanoma.

Keywords: **melanoma; histology; sensitivity; specificity; overdiagnosis**

Introduction

Increased awareness of melanoma among health professionals and the public has led to an increase in the number of biopsies of pigmented skin lesions (PSLs) over time [1,2]. Histological examination is used to predict the biological behaviour of the PSL and to attempt to reach an exact diagnosis. In the case of melanoma, the pathologist evaluates some tumour characteristics which can affect prognosis and patient management, such as the Breslow thickness and the presence or absence of ulceration [3].

Several studies have shown that pathologists do not always agree on the diagnosis and biological behaviour of PSLs [4–11]. The assessment of Breslow thickness and ulceration in melanoma, two major prognostic factors in local disease, has been shown to have good reproducibility [8,12–17], but inter-observer agreement for other tumour characteristics such as Clark's level, histological subtype, and the presence or absence of an associated naevus tends to be poor or intermediate at best [8,12–16]. Most of these studies were performed among expert pathologists who were presented a highly selected set of slides. Moreover, artificial conditions which can induce inter-observer discordance, such as serial sections or lack of any clinical information, were not always taken into account [18].

In the present study, a randomly selected set of biopsies of clinically suspicious PSLs was circulated to a group of general pathologists within a routine practice setting. All participants were sent the original slides and provided with the original clinical information. For each slide, information on slide quality and additional investigations that would be performed in routine practice was acquired.

Materials and methods

Histological slides were selected from the Dermatopathology Department, University Hospital Ghent. All diagnostic slides of PSLs removed because of one or more clinically suspicious features (asymmetry, border irregularity, colour variegation, size of more than 6 mm, elevation, increase in size or darkening of a mole) and received in January, April, July, and October 1996 were selected ($n=66$). Twenty pathologists volunteered to participate. Most of them were

recruited during a medical education session on breast pathology.

Slides were sent to the participants in sets of 6–7. The pathologist was asked to complete a standard protocol (Table 1) for each slide and to return it with the slides within 2 weeks. The protocol was approved by all participants before the start of the study. Each protocol was marked with a code specific for the pathologist. The top of the protocol mentioned the slide number and the clinical information provided on the original pathology request form. Pathologists were asked whether each slide was sufficient to establish a definitive diagnosis. They were then asked to make a diagnosis, selecting one of the proposed diagnostic classes. No discussion on the diagnostic criteria was organized prior to the study, as the purpose was to check variability in a normal routine practice setting. In cases of *in situ* or invasive melanoma, participants were asked to evaluate some additional characteristics. Criteria for classifying histological subtype and Clark's level, and for measuring Breslow thickness, were summarized on the protocol.

One set of six slides was sent twice to each participant to evaluate intra-observer variability. Not all pathologists completed all sets of slides (90% of all expected evaluations); the most frequent reason given was lack of time.

Sensitivity, specificity, and predictive value of a diagnosis were calculated by comparison with a reference diagnosis [19]: this was the initial diagnosis made at the dermatopathology department, if more than 80% of all pathologists agreed, or the diagnosis made by a panel of expert (dermato)pathologists (EG, EH, JMN or DR, GP, JMN). In addition, the panel reviewed all slides where an original diagnosis of melanoma was made. A total of 50 slides were reviewed by the panel. In two slides, the panel did not agree on whether the lesion was malignant or benign and these slides were omitted from the analysis. In another two slides, there was disagreement on the exact diagnostic class of the lesion (common versus atypical naevus and congenital versus common naevus). The other slides consisted of 11 melanomas (nine invasive, one *in situ*, one metastasis), 46 naevi (32 common, seven atypical, three blue, two congenital, two Reed's), and five non-melanocytic lesions.

The chi-squared test was used to test for difference of proportions [20]. The Spearman rank correlation coefficient was used to test a linear correlation between two continuous variables [21]. The sign test was used to test for a trend in a series of categorical values [22].

Results

The mean number of practising years among participating pathologists was 14 years (range 0.5–34 years). The load of PSLs in routine practice ranged from 4 to 50 per week, with a mean of 16.

Overall, 91% of all slides were found convenient for

Table 1. Standard protocol

Pathologist's code XXX
Slide number – clinical information

- Slide reading suitability
  - According to your appreciation, can this slide be evaluated adequately? Yes/no
  - In routine practice, would you perform additional immunohistochemical examinations? Yes/no
  - In routine practice, would you ask for a colleague's opinion or would you consult a textbook? Yes/no
  - Did you ask for a colleague's opinion or did you consult a textbook? Yes/no
- Diagnosis
  - melanoma: *in situ*/invasive/no primary tumour
  - common naevus: junctional/compound/intradermal
  - atypical/dysplastic naevus: junctional/compound
  - congenital naevus
  - blue naevus
  - Spitz naevus
  - pigmented spindle cell naevus (Reed)
  - other: ...
- In cases of diagnosis of primary melanoma
  - complete resection: Yes/no
  - histological subtype: – melanoma *in situ*: – radial growth phase with pagetoid pattern
    – radial growth phase with lentiginous pattern
    – radial growth phase, cannot be classified
    – lentigo maligna melanoma/acral-lentiginous melanoma
    – superficial spreading melanoma
    – nodular melanoma
    – not classifiable
  - presence of ulceration: yes/no
  - Breslow thickness: |. | |mm
  - Clark level: I/II/III/IV/V
  - associated naevus: yes/no

adequate evaluation by the participants (range 77–97%). Reported slide reading suitability was lower in melanoma than non-melanoma lesions [86% versus 92% ($p=0.015$)]. Reasons for poor slide reading suitability were remarks about slide preparation (45%), such as staining and thickness of the sections, and insufficient material to establish a definitive diagnosis (39%). Some other reasons were insufficient clinical information (7%), the need for additional investigations such as immunohistochemistry (5%), and damage to the slide (4%). The panel considered 44 of the 48 slides (92%) adequate for evaluation.

In 46% of the melanoma lesions (range from 0 to 100% among observers) and 12% of the non-melanoma lesions (range from 0 to 58% among observers), the participants would have performed immunohistochemistry in routine practice. Sixty-five per cent of the additional immunohistochemical examinations would have consisted of HMB45 and/or S100. Common leukocyte antigen (CLA), trichrome or vimentin detection would have been performed in another 7% of the slide readings. The panel would have performed additional immunohistochemistry in five of the 48 slides (10%), with S100 and/or HMB45 in 80%.

In 49% of the melanoma lesions [range from 9% (1/11) to 100% (11/11)] and in 24% of the non-melanoma lesions [range from 7% (4/54) to 73% (37/51)], participants would have discussed the slide with a colleague or would have consulted a book in routine practice. In 15% of the melanoma evaluations and in 9% of the non-melanoma lesions, they effectively did so. The frequency with which they would seek advice was negatively correlated with their experience (number of practice years multiplied by the load of PSLs seen in everyday practice) in non-melanoma lesions ($r=-0.61$, $p<0.01$). This correlation was not statistically significant in melanoma lesions ($r=-0.40$, $p=0.12$).

Melanoma sensitivity and specificity

Overall sensitivity for melanoma was 87% (range 55–100%) (Table 2). Sensitivity was higher in thick melanomas (Breslow thickness >1 mm) than in thin melanomas [97% versus 83% ($p=0.005$)]. Other suggested diagnoses were dysplastic naevus (12%, with 11% compound type) and common compound naevus (5%) in case of thin melanomas, and Spitz (2%) and dysplastic compound naevus (1%) in case of thick melanomas.

Overall melanoma specificity was 94% (range 83–100%) (Table 2). False-positive melanoma diagnoses were made in dysplastic naevi (46%, with 44% of junction type), blue naevi (30%), common naevi (14%, with 11% of junctional type), Reed's naevus (4%), and in a non-melanocytic lesion (7%) (Table 3). Most false-positive melanoma diagnoses were *in situ*

Table 2. Calculated sensitivity, specificity, and positive and negative predictive values of a melanoma diagnosis. In the event that not all slides were reviewed by a participant, the number of slides for which a protocol was completed is given in parentheses

		Sensitivity	Specificity	Positive predictive value	Negative value
Pathologist	1	100	96	85	100
	2	89 (9)	90 (48)	62	98
	3	82	94	75	96
	4	90 (10)	96 (52)	82	98
	5	55	96	75	91
	6	100	92 (52)	73	100
	7	91	89	63	98
	8	100	96	82	100
	9	91	94 (52)	77	98
	10	89 (9)	98	89	98
	11	82	100 (52)	100	96
	12	100	83	55	100
	13	71 (7)	96 (46)	71	96
	14	67 (3)	100 (22)	100	96
	15	90 (10)	96 (52)	82	98
	16	91	90 (51)	67	98
	17	Missing	100 (5)	Missing	100
	18	82	98	90	96
	19	91	92 (52)	71	98
	20	80 (10)	94	73	96
Total		87	94	75	97

Table 3. Diagnostic classification of benign melanocytic lesions: participating pathologists versus reference diagnosis. The percentages of concordant evaluations are italicized

	Pathologists' diagnoses									
Reference diagnosis	**Melanoma *in situ***	**Invasive melanoma**	**Melanoma metastasis**	**Common naevus**	**Dysplastic naevus**	**Congenital naevus**	**Blue naevus**	**Reeds naevus**	**Spitz**	**Other**
Common naevus ($n=32$, 585 evaluations)	1.0%	0.3%	None	*75.6%*	10.8%	5.1%	None	1.7%	1.7%	3.8% (2.4% naevoid lentigo)
Dysplastic naevus ($n=7$, 125 evaluations)	17.6%	3.2%	None	29.6%	*45.6%*	None	None	4%	None	None
Congenital naevus ($n=2$, 35 evaluations)	None	None	None	48.5%	2.9%	*45.7%*	None	None	None	2.9% (naevus spilus)
Blue naevus ($n=3$, 54 evaluations)	None	1.9%	29.6%	7.4%	None	5.6%	*42.6%*	1.9%	1.9%	9.3% (1.9% halo naevus)
Reeds naevus ($n=2$, 38 evaluations)	2.6%	2.6%	None	10.5%	31.6%	None	None	*47.4%*	5.3%	None

melanomas ($n=29$). Nevertheless, eight invasive melanomas with a Breslow thickness ranging from 0.26 to 2.8 mm were diagnosed. Blue naevus caused confusion with melanoma metastasis (Table 3), especially in one particular slide.

Leaving out those slides that were considered not optimal for evaluation did not change sensitivity and specificity significantly ($p>0.05$). Excluding those readings where a second opinion and/or additional immunohistochemistry would have been asked in normal routine practice did not affect melanoma sensitivity ($p>0.05$), while it increased melanoma specificity from 94% to 98% ($p=0.0001$).

There was no correlation of melanoma sensitivity and specificity with years of practice or the routine load of PSLs ($p>0.05$). There was a negative correlation between the ability to recognize benign disease (melanoma specificity) and the frequency with which a colleague or a textbook would have been consulted in routine practice ($r=-0.52$, $p<0.001$).

In the set of six slides that was sent twice, six different pathologists switched from melanoma to a benign PSL, or vice versa, in five different slides. The second reading tended to be more in agreement with the reference diagnosis (96% versus 87%; $\chi^2=4.39$, $p=0.04$).

Evaluation of melanoma characteristics

Of the ten primary melanomas (one *in situ*, nine invasive), five were classified by the panel as the superficial spreading type. One was considered lentigo maligna melanoma and three were judged not classifiable. Participants' agreement with the panel was 55%. Agreement on the four superficial spreading melanomas reached 93%, while agreement for the other tumours was only 21%.

The panel of experts reported ulceration in two thick melanomas (Breslow depth of 2.6 and 3.3 mm). Participants did not agree in three of the 35 readings (9%). Overall agreement on the presence or absence of ulceration was 98%.

There was 77% agreement on the evaluation of the presence or absence of an associated naevus. The lowest agreement (24%) was found in the two slides where the panel concluded that an associated naevus was present.

Agreement on Clark's level was highest in level II (81%) and lowest in level III (26%). Agreement for Clark's level IV was 65%. The observers' and panel's observation of the level of invasion was concordant in 59% of all evaluations. In 30% of the evaluations, the level reported by the participants was one level lower (20%) or higher (10%) than that of the panel. In 11% of the evaluations, there were two levels of difference with the panel (all lower). Pathologists tended to underscore Clark's level compared with the panel ($p<0.001$).

The mean difference of Breslow thickness assessed by the pathologist and by the panel was -0.09 mm (SD $=0.52$), but the extremes ranged from -2.40 mm to $+1.67$ mm. The range of differences varied considerably with the observer and with the slide (Figure 1a and 1b). Variation tended to be higher in thicker tumours (Figure 1b). Discordant evaluations of an associated naevus seemed to be at the basis of some of the observed differences in measured Breslow thickness (Figures 1a and 1b). For 12% of all evaluations there would have been different patient management, based on the Breslow thickness assessed by the pathologist or by the panel (cut-off at 1 mm depth).

Diagnostic classification of benign lesions

On 53 PSLs that were considered benign, the panel did not agree on the exact diagnostic class of two. These two slides were omitted from this analysis.

Overall, there was 70% agreement on the diagnostic classification of benign melanocytic lesions. The majority of these lesions were common naevi and the pathologists had a concordant diagnosis (including naevoid lentigo) in 78% of all evaluations (Table 3). In nearly 11%, a diagnosis of dysplastic naevus was made. *In situ* melanoma was diagnosed in five different slides by five different pathologists, and invasive melanoma was diagnosed in two readings (one Breslow missing and the other 0.90 mm).

In 45.6% of their evaluations, pathologists agreed with the reference diagnosis of dysplastic naevus (Table 3). Four of the seven slides with a reference diagnosis of dysplastic naevus gave rise to 22 false-positive diagnoses of *in situ* melanoma. In addition, four false-positive diagnoses of invasive melanoma (Breslow ranging from 0.26 to 0.75 mm) were made (three different slides, three different pathologists).

All three slides of blue naevi caused problems of diagnostic classification (Table 3). A diagnosis of melanoma metastasis was suggested in 16 evaluations. In 13, this was due to one particular slide where the PSL had been excised near the scar of a former melanoma resection. In the latter slide, there was also one false-positive diagnosis of invasive melanoma, Breslow 2.8 mm.

Reed's naevus was recognized in 47.4% of all evaluations. In about one-third, it was diagnosed as dysplastic naevus. There was one diagnosis of melanoma *in situ* and one of invasive melanoma (Breslow 0.77 mm).

Three non-melanocytic lesions were diagnosed as melanocytic by five different pathologists: a dermatofibroma was diagnosed as a desmoplastic naevus and an angioma as a common naevus; an early evolving seborrhoeic keratosis was diagnosed four times as melanoma *in situ*, of which two were classified as Dubreuilh's melanosis. One seborrhoeic keratosis was diagnosed twice as basal cell carcinoma.

In the set of benign melanocytic lesions that was read twice by the participants (three common naevi, one dysplastic naevus, and one Reed's naevus), the diagnostic class was changed in 23% of the second

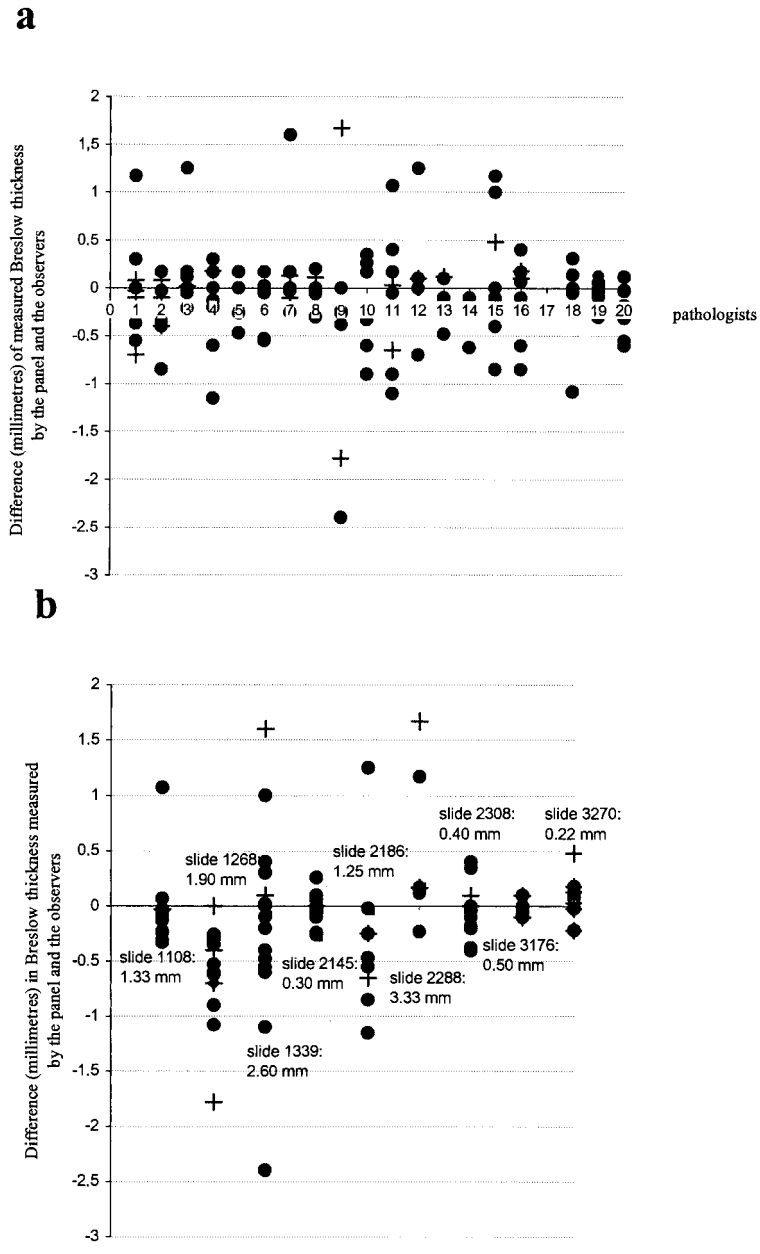

Figure 1. Difference in measured Breslow thickness between the panel and the participating pathologists. (a) Differences for the individual pathologists. (b) Differences for the individual slides. Plus signs mark evaluations where there was a discordant opinion on the presence or absence of an associated naevus between the participating pathologist and the panel

readings. The second reading tended to be more in agreement with the panel's diagnosis, although this was not statistically significant [76% versus 67%, ($p = 0.24$)].

Discussion

In this study, a randomly selected set of biopsies of PSLs with one or more clinically suspicious features was circulated to a group of volunteer general pathologists with a routine practice. Care was taken to avoid conditions which could induce inter-observer variability [18,23].

Overall sensitivity for melanoma was 87%. Sensitivity was significantly lower in thin melanomas (Breslow < 1 mm), where one in six readings resulted in a diagnosis of naevus (mainly dysplastic and mainly of the compound type). Inter-observer variation in thin melanomas, which have shown a considerable increase in the second half of the 20th century [24], has been reported [7,8] and the exact malignant potential of these lesions has been discussed [25]. The concept of radial growth phase (RGP) melanoma, which was proposed by Clark *et al.* in 1984 [26], implies a progression stage of invasive melanoma incapable of metastasis [27]. Inter-observer variation for the recognition of RGP versus vertical growth phase (VGP) in melanoma ranged from fair to good [10,17].

In the present study, melanoma specificity was 94%, increasing to 98% excluding the readings where additional advice or immunohistochemistry was not

performed while it would have been in normal routine practice. The negative correlation between specificity and the frequency with which a colleague or textbook would have been consulted in routine practice could indicate a tendency to overdiagnose melanoma in case of doubt about the biological behaviour of the lesion. A tendency to overdiagnose thin melanomas has been reported [7].

Dysplastic naevi of the junctional type were the PSLs most often misdiagnosed as melanoma, mainly melanoma *in situ*. Cook *et al.* suggested the use of the MIN (melanocytic intraepidermal neoplasia) terminology, by analogy with CIN (cervical intraepithelial neoplasia), because of the difficulty in discriminating between (severely) dysplastic naevi and *in situ*/RGP melanoma [7,10]. However, dysplastic naevi and *in situ*/RGP melanoma may have a different clinical significance. While the latter are considered progression stages of malignant disease [26], the clinical significance of naevi with histological dysplasia is not established and its prevalence in the normal population has been estimated from 10% up to 47% [28,29]. In contrast, clinically atypical naevi are a marker for increased melanoma susceptibility and are possible precursor lesions of melanoma [30–33]. There is a poor correlation between clinical atypia and the presence of histological dysplasia [33]. Studies on the recognition of naevi with histological dysplasia have shown divergent results [34]. In this study, less than 50% of the dysplastic naevi were classified as such. However, histopathological criteria for diagnosing dysplastic naevi have been shown to be reproducible [35,36] and problems seem to lie mainly at the extremes of the spectrum (dysplastic naevi with severe dysplasia versus melanoma *in situ*, dysplastic naevi with mild dysplasia versus common naevus) [5,35].

Despite the relatively high specificity for melanoma, there is an important number of false-positive melanoma diagnoses, because the majority of the PSLs presented to the pathologist are benign: in 6% of the readings of 53 benign PSLs, the pathologist did not agree with the innocuous nature of the lesion, resulting in 57 false melanoma diagnoses, consisting mainly of *in situ* melanomas and thin invasive melanomas. This important 'burden' of false-positive diagnoses is also reflected by the positive predictive value in the current study setting (Table 2): overall, one-quarter of all melanoma diagnoses were made in benign lesions. On the other hand, the negative predictive value in the current test setting was high (97%), indicating that the risk for a patient having melanoma in the event of a negative diagnosis is low. It is likely that in general histology practice the ratio of benign/malignant lesions is even higher than in this experimental setting (4.8 : 1), which would further reduce the calculated positive predictive value of a melanoma diagnosis.

In the set of six slides that was sent twice, the second reading tended to be more in agreement with the reference diagnosis, indicating a learning effect and/or habituation to the specific slide processing characteristics.

Even among experts, there may be disagreement on the estimated biological behaviour of PSL [9,11] and the only certainty about the benign/malignant nature of a lesion is the eventual outcome for the patient.

Thickness measured according to Breslow is a major prognostic factor in local disease, on which further management is often based [3]. Despite the fact that this is an objective, reproducible measurement [8,12–14,16,17], this study indicates that there is still an important variation in routine practice, with possible therapeutic implications for the patient. Errors inherent in the measuring instrument may be expected to produce variations of 0.1–0.5 mm [12,14]. Other variations may be due to different interpretations of the deepest invasive tumour cell [12,14]. In this study, some of the differences in measured Breslow thickness between the participants and the panel seemed to be related to a different interpretation of the presence or absence of a co-existent naevus. Larger variations may be attributable to errors in the conversion of micrometre units into millimetres [12,14]. Finally, although we presumed that participating pathologists used the same measuring technique and criteria (described on the protocol), some may use other techniques in routine practice [37]. Clark's level has been reported to be less reproducible than Breslow thickness [8,12–14,17]. This is probably inherent in the fact that it is a categorical variable, forcing the pathologist to choose one category [12,13]. The lowest agreement has been observed for Clark level III. This has been attributed to the difficulty in assessing the interface between papillary and reticular dermis [12,13].

Ulceration is the second major independent prognostic factor in local disease, which will be incorporated in the new pTNM staging system [3]. In this study, the presence of ulceration was missed in two melanoma cases by three pathologists. Although several studies reported good reproducibility of this characteristic [8,13,16], Heenan *et al.* reported that inter-observer congruence was lower than one might expect. They suggested that this might be related to the recognition of small foci of ulceration (less than 3 mm wide), of which the prognostic significance is questionable [13]. However, the areas of ulceration in this study were clearly wider, so that discordant observations probably pertain to inaccuracy of the observer.

In conclusion, the increased presentation of borderline lesions seems to have complicated the histopathological diagnosis of PSLs. On the one hand, a proportion of the melanomas removed in an early phase of tumour progression may not be recognized; many of these lesions, however, may not display clinically aggressive behaviour [27,38]. On the other hand, there may also be an ongoing overdiagnosis of (especially thin) melanomas, which may be further stimulated by increased awareness and the pressure not to miss such an important diagnosis as melanoma.

It is difficult to estimate the extent, and hence the

cost, both at the community and at the individual level, of melanoma under- and over-diagnosis in general histopathological practice. In this study, 13% of the readings were reclassified from benign to malignant and 2–6% were reclassified in the other direction. Since melanoma is a relatively infrequent cancer in most European countries, it is likely that patients receiving an inappropriate diagnosis of melanoma outnumber those in whom melanoma is missed. It also implies that the histological diagnosis of a benign lesion is probably more reliable than that of a malignant lesion (with a high negative predictive value and a lower positive predictive value).

These phenomena are inherent in the fact that histological diagnosis is based on the interpretation of a visual image, which is subjective to some extent [9,39]. New diagnostic techniques for detecting genetic or functional changes specific for the malignant transformation of melanocytes would bring significant progress in the classification of borderline lesions.

Acknowledgements

We thank D. Debacquer, PhD, Biostatistician (University Ghent), for advice about data analysis. This study was supported by the Fund for Scientific Research-Flanders (grant F7801).

References

1. Marks R, Jolley D, McCormack C, Dorevitch AP. Who removes pigmented skin lesions? *J Am Acad Dermatol* 1997; **36**: 721–726.
2. McWilliam LJ, Knox F, Wilkinson N, Oogarah P. Performance of skin biopsies by general practitioners. *Br Med J* 1991; **303**: 1177–1179.
3. Balch CM, Buzaid A, Atkins MB *et al*., for the AJCC Melanoma Staging Committee. A new American Joint Committee on cancer staging system for cutaneous melanoma. *Cancer* 2000; **88**: 1484–1491.
4. Duray PH, DerSimonian R, Barnhill R, *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**: 741–749.
5. Duncan LM, Berwick M, Bruijn JA, Byers R, Mihm MC, Barnhill RL. Histopathologic recognition and grading of dysplastic melanocytic nevi: an interobserver agreement study. *J Invest Dermatol* 1993; **100**: 318S–321S.
6. Piepkorn MW, Barnhill RL, Cannon-Albright LA, *et al*. A multi-observer, population-based analysis of histologic dysplasia in melanocytic nevi. *J Am Acad Dermatol* 1994; **30**: 707–714.
7. 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.
8. Corona R, Mele A, Amini M, *et al*. Interobserver variability on the histopathologic diagnosis of cutaneous melanoma and other pigmented skin lesions. *J Clin Oncol* 1996; **14**: 1218–1223.
9. Farmer ER, Gonin R, Hanna MP. Discordance in the histopathologic diagnosis of melanoma and melanocytic nevi between expert pathologists. *Hum Pathol* 1996; **27**: 528–531.
10. CRC Melanoma Pathology Panel. A nationwide survey of observer variation in the diagnosis of thin cutaneous malignant melanoma including the MIN terminology. *J Clin Pathol* 1997; **50**: 202–205.
11. Barnhill RL, Argenyi ZB, From L, *et al*. Atypical Spitz nevi/tumors: lack of consensus for diagnosis, discrimination from melanoma and prediction of outcome. *Hum Pathol* 1999; **30**: 513–520.
12. Prade M, Sancho-Garnier H, Cesarini JP, Cochran A – EORTC Melanoma Pathologist's Group. Difficulties encountered in the application of Clark classification and the Breslow thickness measurement in cutaneous malignant melanoma. *Int J Cancer* 1980; **26**: 159–163.
13. 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.
14. 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**: 245–250.
15. Krieger N, Hiatt RA, Sagebiel RW, Clark WH, Mihm MC. Inter-observer variability among pathologists' evaluation of malignant melanoma: effects upon an analytic study. *J Clin Epidemiol* 1994; **47**: 897–902.
16. Lock-Andersen J, Hou-Jensen K, Hansen JP, Jensen NK, Sogaard H, Andersen PK. Observer variation in the histological classification of cutaneous malignant melanoma. *Scand J Plast Reconstr Surg Hand Surg* 1995; **29**: 141–148.
17. McDermott NC, Hayes DP, Al-Sader MH, *et al*. Identification of vertical growth phase in malignant melanoma; a study of interobserver agreement. *Am J Clin Pathol* 1998; **110**: 753–757.
18. Cramer SF. Interobserver variability in dermatopathology. *Arch Dermatol* 1997; **133**: 1033–1036.
19. Riegelman RK, Hirsch RP. *Studying a Study and Testing a Test: How to Read the Medical Literature*. Little, Brown and Company: Boston (Mass.), 1991; 165–174.
20. Kirkwood BR. The chi-squared test for contingency tables. In *Essentials of Medical Statistics*. Blackwell Scientific Publications: Oxford, 1988; 87–93.
21. Kirkwood BR. Non-parametric methods. In *Essentials of Medical Statistics*. Blackwell Scientific Publications: Oxford, 1988; 151–152.
22. Kirkwood BR. Non-parametric methods. In *Essentials of Medical Statistics*. Blackwell Scientific Publications: Oxford, 1988; 147–151.
23. Cramer SF, Roth LM, Ulbright TM, *et al*. The mystique of the mistake. With proposed standards for validating proficiency tests in anatomic pathology. *Am J Clin Pathol* 1991; **96**: 774–777.
24. MacKie RM, Dole D, Hunter JAA, *et al*. Cutaneous malignant melanoma in Scotland: incidence, survival, and mortality, 1979–94. *Br Med J* 1997; **315**: 1117–1121.
25. Swerlick RA, Chen S. The melanoma epidemic. Is increased surveillance the solution or the problem? *Arch Dermatol* 1996; **132**: 881–884.
26. Clark WH, Elder DE, Guerry D, Epstein MN, Greene MH, Van Horn M. A study of tumor progression: the precursor lesions of superficial spreading and nodular melanoma. *Hum Pathol* 1984; **15**: 1147–1165.
27. Guerry D, Synnestvedt M, Elder DE, Schultz D. Lessons from tumor progression: the invasive radial growth phase of melanoma is common, incapable of metastasis, and indolent. *J Invest Dermatol* 1993; **100**: 342S–345S.
28. Steijlen PM, Bergman W, Hermans J, Scheffer E, Van Vloten WA, Ruiter DJ. The efficacy of histopathological criteria required for diagnosing dysplastic naevi. *Histopathology* 1988; **12**: 289–300.
29. Piepkorn M. A hypothesis incorporating the histologic characteristics of dysplastic nevi into the normal biological development of melanocytic nevi. *Arch Dermatol* 1990; **126**: 514–518.
30. Grob JJ, Andrac L, Romano MH, *et al*. Dysplastic naevus in non-familial melanoma. A clinicopathological study of 101 cases. *Br J Dermatol* 1988; **118**: 745–752.
31. Klein LJ, Barr RJ. Histologic atypia in clinically benign nevi. *J Am Acad Dermatol* 1990; **22**: 275–282.
32. Roush GC, Dubin N, Barnhill RL. Prediction of histologic melanocytic dysplasia from clinical observation. *J Am Acad Dermatol* 1993; **29**: 555–562.
33. Meyer LJ, Piepkorn M, Goldgar DE, *et al*. Interobserver

concordance in discriminating clinical atypia of melanocytic nevi, and correlations with histologic atypia. *J Am Acad Dermatol* 1996; **34**: 618–625.
34. Murphy GF, Mihm M. Recognition and evaluation of cytological dysplasia in acquired melanocytic nevi. *Hum Pathol* 1999; **30**: 506–512.
35. Clemente C, Cochran AJ, Elder DE, *et al*. Histopathologic diagnosis of dysplastic nevi: concordance among pathologists convened by the World Health Organization Melanoma Programme. *Hum Pathol* 1991; **22**: 313–319.
36. de Wit PEJ, van't Hof-Grootenboer B, Ruiter DJ, *et al*. Validity of the histopathological criteria used for diagnosing dysplastic naevi. *Eur J Cancer* 1993; **29**: 831–839.
37. Calder CJ, Campbell AP, Plastow SR. Measurement techniques for melanoma: a statistical comparison. *J Clin Pathol* 1990; **43**: 922–923.
38. Månsson-Brahme E, Carstensen J, Erhardt K, Lagerlof B, Ringborg U, Rutqvist LE. Prognostic factors in thin cutaneous malignant melanoma. *Cancer* 1994; **73**: 2324–2332.
39. McCalmont TH. Melanoma and melanoma *in situ*: build a better diagnosis through architecture. *Semin Cutaneous Medicine Surg* 1997; **16**: 97–107.

