Dermoscopic and Histopathologic Diagnosis of Equivocal Melanocytic Skin Lesions

An Interdisciplinary Study on 107 Cases

Gerardo Ferrara, M.D.[1]
Giuseppe Argenziano, M.D.[2]
H. Peter Soyer, M.D.[3]
Rosamaria Corona, M.D.[4]
Francesco Sera, Ph.D.[4]
Bruno Brunetti, M.D.[5]
Lorenzo Cerroni, M.D.[3]
Sergio Chimenti, M.D.[6]
Laila El Shabrawi-Caelen, M.D.[3]
Angela Ferrari, M.D.[8]
Rainer Hofmann-Wellenhof, M.D.[3]
Steven Kaddu, M.D.[3]
Domenico Piccolo, M.D.[7]
Massimiliano Scalvenzi, M.D.[8]
Stefania Staibano, M.D.[9]
Ingrid H. Wolf, M.D.[3]
Gaetano De Rosa, M.D.[9]

[1] Pathologic Anatomy and Histology Service, G. Rummo General Hospital, Benevento, Italy.

[2] Department of Dermatology, Second University of Naples, Naples, Italy.

[3] Department of Dermatology, University of Graz, Graz, Austria.

[4] Clinical Epidemiology Unit, Istituto Dermopatico dell'Immacolata, Rome, Italy.

[5] Division of Dermatology, Santa Maria della Speranza Hospital, Battipaglia, Italy.

[6] Department of Dermatology, University Tor Vergata, Rome, Italy.

[7] Department of Dermatology, University of L'Aquila, L'Aquila, Italy.

[8] Department of Dermatology, University Federico II, Naples, Italy.

[9] Department of Biomorphological and Functional Sciences, Pathology Section, University Federico II, Naples, Italy.

Address for reprints: H. Peter Soyer, M.D., Department of Dermatology, University of Graz, Auenbruggerplatz 8, A-8036 Graz, Austria; Fax: +43-316-385-4957; E-mail: peter.soyer@uni-graz.at

Received August 15, 2001; revision received December 22, 2001; accepted April 8, 2002.

BACKGROUND. Dermoscopy (dermatoscopy, epiluminescence microscopy) is increasingly employed for the preoperative detection of cutaneous melanoma; dermoscopic features of pigmented skin lesions have been previously defined using histopathology as the key to the code. In a preliminary study on 10 cases evaluated by nine dermoscopists and nine histopathologists, the authors experienced that when at least two dermoscopists disagree in evaluating a melanocytic lesion, even histopathologic consultations may give equivocal results.

METHODS. One hundred seven melanocytic skin lesions, consecutively excised because of equivocal clinical and/or dermoscopic features, were retrospectively examined by eight dermoscopists and eight histopathologists; the diagnostic interobserver agreement was calculated by means of the Schouten k statistics. After histopathologic consultations, all 107 lesions underwent unblinded dermoscopic re-evaluation in order to find which dermoscopic features had given rise to histopathologic diagnostic difficulties.

RESULTS. The interobserver ageement was good for both dermoscopy ($k = 0.53$) and histopathology ($k = 0.74$). Out of 48 cases evaluated by the dermoscopists in complete accordance, only 8 (16.7%) received at least one conflicting histopathologic diagnosis. Instead, among the remaining 59 cases with at least one disagreeing dermoscopic diagnosis, 21 (35.6%) received at least one disagreeing histopathologic diagnosis. The unblinded dermoscopic re-evaluation showed that five out of seven lesions with clear-cut regression structures were histopathologically controversial.

CONCLUSIONS. At least for selected and reasonably difficult lesions, a diagnostic discrepancy among formally trained dermoscopists seems to be predictive for a diagnostic disagreement among histopathologists. Lesions showing clear-cut regression structures are prone to give some histopathologic disagreement. ***Cancer* 2002;95:1094–1100.** © *2002 American Cancer Society.*
DOI 10.1002/cncr.10768

KEYWORDS: dermoscopy, dermatoscopy, epiluminescence microscopy, histopathology, melanoma, melanocytic skin lesions, diagnosis, interobserver agreement.

The clinical diagnosis of pigmented skin lesions, including melanoma, is an ongoing medical challenge. Reports assessing diagnostic accuracy based on clinical criteria alone have shown that dermatologists are able to detect melanoma in 65-80% of cases.[1–3] Therefore, 20-35% of melanomas escape from clinical diagnosis. To improve the preoperative diagnosis of melanoma, dermoscopy (dermatoscopy, skin surface microscopy, epiluminescence microscopy-[ELM]) has been increasingly used in clinical practice. Dermoscopy has a sensitivity 10-27% higher than the traditional ABCD clinical criteria for diagnosing melanoma.[4] The smaller the lesion,[5,6] the lower the accuracy of the clinical diagnosis[6] and the greater the diagnostic aid provided by dermoscopy.

Reproducibility in dermoscopic diagnosis has been addressed in a study by Binder et al.,[6] who found that formally trained dermoscopists have only moderate intraobserver and interobserver agreement. Remarkably, even in less experienced hands, reproducibility seems to be improved by the application of a new seven point checklist for dermoscopic diagnosis.[7] Of course, these and other studies about preoperative diagnosis of melanoma have always referred to histopathology as the key to the code for the final diagnosis. By performing clinicopathologic correlations, the underlying structures of dermoscopic criteria have been elaborated.[8–10] However, a major potential problem of this clinicopathologic approach comes from the fact that the histopathologic diagnosis and classification of melanoma is a matter of considerable disagreement.[11–15] Remarkably, there is no single article on dermoscopy reporting the histopathologic criteria used to obtain a standardized histopathologic diagnosis.[4]

In a previous pilot study on 10 melanocytic skin lesions,[15] we showed that when at least two formally trained dermoscopists disagree in evaluating a pigmented skin lesion, histopathologic consultations may also give equivocal results. However, the same study[15] also showed that melanocytic lesions proven to be unclassifiable at dermoscopy were not equally histopathologically controversial, and vice versa. These conflicting observations prompted us to evaluate the interobserver agreement on the dermoscopic and the histopathologic diagnoses in a larger series of melanocytic skin lesions.

MATERIALS AND METHODS

Dermoscopy

The study was implemented by retrieving pigmented skin lesions consecutively examined with dermoscopy by two of the authors (G.A., M.S.) and subsequently excised for diagnostic purposes at the Department of Dermatology of the Federico II University of Naples, Italy, from January to December, 1999. Submission of dermoscopic and histopathologic materials for further diagnostic consultations was authorized by the patients or their guardians. Lesions which proved to be nonmelanocytic on histopathologic examination were excluded from the study.

On the basis of the clinical and dermoscopic features noticed by the first observer (G.A. or M.S.), two classes of lesions were recorded: 1) melanocytic proliferations judged to be equivocal, i.e. lesions with clinical and/or dermoscopic features that did not allow the observers to exclude with certainty a diagnosis of melanoma; and 2) melanocytic proliferations with dermoscopic features consistent with spindle and/or epithelioid cell nevus (Spitz and/or Reed nevus [SECN]), when observed in adults.

The latter category of lesions was also submitted to surgical excision and included in the current study because it has been previously shown that there may be considerable overlap in dermoscopic features between SECN and melanoma.[16]

One hundred seven lesions were included in the study; 37 of these lesions (34.6%) showed a dermoscopic pattern consistent with SECN. Prior to excision, all lesions were photographed in vivo at a fixed magnification of $\times$ 10 with Dermaphot photo equipment (Heine Optotechnik, Herrsching, Germany) after being covered with immersion oil. All lesions were small enough to be entirely encompassed within a single photographic field (up to 1.2 cm).

All 107 35 mm color slides were independently studied on a viewer (Kodak Ektapro 5000 Slide Projector, Kodak Aktiengesellschaft, Germany) by eight dermatologists with at least five years of experience in dermoscopy (G.A., B.B., A.F., R.H.W., D.P., M.S., H.P.S., I.H.W.); they have all been cooperating for several years within a research group working on the diagnosis of melanocytic skin neoplasms. As far as the diagnostic criteria were concerned, these eight observers were asked to refer for diagnosis to previously reported systematic methods (standard pattern analysis,[6,17,18] ABCD rule,[19,20] and seven point checklist[7]).

At the time of their examination, dermoscopists were unaware of the histopathologic diagnosis. They were discouraged from using equivocal or doubtful terms such as “suspicious for. . .” or “atypical nevus”. The diagnostic choices given were: melanoma; congenital melanocytic nevus; acquired melanocytic nevus, type Clark; acquired melanocytic nevus, type Spitz; acquired melanocytic nevus, type Reed; other (dermatofibroma, basal cell carcinoma, etc.).

Histopathology

All the lesions had been initially sent for routine histopathologic examination to the Pathology Section of the University Federico II of Naples, Italy. Subsequently, one representative hematoxylin and eosin stained specimen from each of the 107 selected lesions was circulated to eight histopathologists with at least five years of experience in dermatopathology (L.C., S.C., G.D.R., L.E.S.C., G.F., S.K., H.P.S., S.S.). These histopathologists have been cooperating for several years (also together with the dermoscopists) within a research group working on the diagnosis of melanocytic skin neoplasms.

Each histopathologist independently diagnosed the whole series of microscopic slides avoiding inconclusive or imprecise diagnostic terms. The diagnoses within the spectrum of melanocytic nevi referred to

the clinicopathologic categories outlined by Ackerman et al.[21] Histopathologic criteria for cutaneous melanoma were standardized according to those listed by Ackerman et al.[21] and by Rosai.[22]

Statistics

For statistical analysis, all the diagnoses were grouped into melanoma vs. nonmelanoma. Given $n = 8$ as the number of observers for both dermoscopy and histopathology, the inter-observer agreement among the $n(n\text{-}1)/2 = 28$ possible pairs of observers was evaluated by using the *kappa* statistics introduced by Cohen.[23] The standard error of *kappa* was calculated according to the method developed by Fleiss et al.[24] The *kappa* statistics for multiple observers was calculated by using the method reported by Schouten.[25] In the current study, in which all eight observers judged the fixed set of 107 lesions, the formula proposed by Schouten[25] gives the same *kappa* values as the procedure proposed by Fleiss.[26] The standard error of *kappa* was calculated by the standard jack knife method.[25]

In evaluating the *kappa* statistics, one must remember that the values range between +1 (perfect agreement) and −1 (perfect disagreement); values greater than 0.75 represent an excellent agreement, values lower than 0.40 a poor agreement, and values between 0.40 and 0.75 a fair to good agreement beyond chance.[27]

A diagnostic score for both dermoscopy and histopathology was calculated for each case by counting the number of ratings as melanoma and the number of ratings as nonmelanoma. Thus, all 107 lesions were divided into three groups on the basis of the dermoscopic diagnostic score as well as into three groups on the basis of the histopathologic diagnostic score (HDS): agreement (dermoscopic agreement, D-A; histopathologic agreement, H-A): cases in which all eight observers gave a unanimous diagnosis; partial agreement (dermoscopic partial agreement, D-PA; histopathologic partial agreement, H-PA): cases with at least one disagreement; disagreement (dermoscopic disagreement D-D; histopathologic disagreement, H-D): cases in which four observers diagnosed the lesion as melanoma and four as nonmelanoma.

Cases with one disagreement and cases with two to three disagreements were placed together into the D-PA and the H-PA groups on the basis of a preliminary statistical analysis, whose results are not itemized herein, but can be summarized as follows: the presence of only one dermoscopic disagreement had the same impact on the degree of histopathologic agreement as the presence of two or three disagreements; cases with only one disagreeing diagnosis were not imputable to the same single observer.

In order to assess the degree of concordance between dermoscopic and histopathologic diagnoses, the sensitivity and specificity of the eight dermoscopists were calculated on the subset of lesions for which unanimous histopathologic diagnoses were made.

Re-evaluation of the Dermoscopic Images

After the histopathologic diagnoses, all the dermoscopic images underwent unblinded re-evaluation by one of the authors (G.A.) to find which, if any, dermoscopic features gave rise to diagnostic difficulties at the histopathologic level.

All 107 lesions were thus classified on the basis of their prevailing dermoscopic pattern, according to the following definitions:

- globular pattern: peripheral rim of globules of either a small (Clark-type) or a large (Spitz-type) size;
- starburst pattern: radial streaks regularly distributed at the periphery of the lesion;
- reticular pattern: widespread pigment network;
- homogeneous pattern: diffuse pigmentation throughout the lesion;
- hypopigmented pattern: diffuse structureless areas of decreased pigmentation;
- peripheral hyperpigmentation: peripheral, asymmetric, hyperpigmented areas;
- regression structures: white scar-like areas and/or pepper-like blue areas within an otherwise dermoscopically common melanocytic lesion; and
- multicomponent patttern: three or more local dermoscopic features.

RESULTS

The 107 selected lesions belonged to 105 white patients (40 males, 65 females) aged 11 to 86 years (mean, 34.11 years). The most common locations were the buttocks (50 out of 107) and the trunk (47 out of 107); mean diameter was 0.8 cm (range, 0.3 to 1.2 cm).

Dermoscopic Diagnoses

The distribution by group (melanoma vs. nonmelanoma) of the dermoscopic diagnoses made by the eight observers is shown in Table 1. Out of 856 dermoscopic diagnoses, 325 (37.96%) were given as melanoma. The extremes were observer A, who diagnosed only 26 lesions as melanoma, and observer B, who diagnosed 60 lesions as melanoma. Of the possible 28 pairs of dermoscopists, 26 showed a *kappa* value ranging between 0.40 and 0.75. For two pairs, namely, B/E and B/H, the *kappa* values were 0.39 and 0.33, respectively. The Schouten kappa value for all eight observers was 0.53 (95% confidence interval [CI]: 0.44-0.62).

Table 2 shows the distribution of the 107 lesions according to the degree of diagnostic agreement and the prevailing diagnosis.

TABLE 1
Distribution of Dermoscopic and Histopathologic Diagnoses by Diagnostic Groups (M vs. nM)

	M		nM	
Dermoscopist	***n***	**%**	***n***	**%**
A	26	24	81	76
B	60	56	47	44
C	42	39	65	61
D	27	25	80	75
E	38	36	69	64
F	44	41	63	59
G	49	46	58	54
H	39	36	68	64
	M		**nM**	
Histopathologist	***n***	**%**	***n***	**%**
1	28	26	79	74
2	41	38	66	62
3	34	32	73	68
4	30	28	77	72
5	42	39	65	61
6	29	27	78	73
7	36	34	71	66
8	34	32	73	68

M: melanoma; nM: nonmelanoma.

TABLE 2
Distribution of 107 Melanocytic Skin Lesions According to the Diagnostic Score Groups

		HDS GROUPS					
		H-A		H-PA		H-D	
DDS GROUPS		**M**	**nM**	**M**	**nM**	**Undefined**	**Total**
D-A	M	11		4			15
	nM		29	**1**[a]	2	1	33
D-PA	M	7	**5**[b]	3		1	16
	nM	**1**[a]	21	**1**[a]	7	4	34
D-D	Undefined	1	3	3	2		9
Total		20	58	12	11	6	107

HDS: histopathologic diagnostic score; D-A/H-A: dermoscopic/histopathologic agreement (all diagnoses in agreement); D-PA/H-PA: dermoscopic/histopathologic partial agreement (one to three disagreeing diagnoses); D-D/H-D: dermoscopic/histopathologic disagreement (no prevailing diagnosis); DDS: dermoscopic diagnostic score; M: melanoma; nM: nonmelanoma.

[a] Putative dermoscopic false negative.

[b] Putative dermoscopic false positives.

The major discrepancies among dermoscopic and histopathologic diagnoses are highlighted in bold.

The D-A group included 48 cases (44.9%), 33 of which were unanimously diagnosed as nonmelanoma and 15 as melanoma. The D-PA group included 50 cases (46.7%), 34 of which were diagnosed as nonmelanoma and 16 cases as melanoma by al least five out of eight observers. The D-D group included nine cases (8.4%), since each had been diagnosed as melanoma by four observers and as nonmelanoma by the other four.

Patients' age and lesion locations did not significantly vary across the three groups of lesions (data not shown).

Histopathologic Diagnoses

The distribution by group (melanoma vs. nonmelanoma) of the histopathologic diagnoses made by the eight histopathology observers is shown in Table 1.

The total number of histopathologic diagnoses of melanoma (273 out of 856; 31.89%) was slightly lower than the number of dermoscopic diagnoses of melanoma (325 out of 856; 37.96%). The extremes were histopathologist 1, who diagnosed 27 lesions as melanoma, and histopathologist 5, who diagnosed 42 cases as melanoma.

Among the 28 possible pairs of histopathologists, 14 showed a *kappa* value of 0.40 to 0.75. The other 14 pairs showed an excellent agreement, their *kappa* value ranging from 0.75 to 0.86. The Schouten *kappa* value was 0.74 (95% CI: 0.66-0.83).

The H-A group included 78 cases (72.9%), 58 of which were unanimously diagnosed as nonmelanoma and 20 as melanoma (see Table 2). Out of the 20 cases of melanoma, 12 were diagnosed as invasive lesions, and 8 as in situ melanomas.

The H-PA group included 23 cases (21.5%), 11 of which were diagnosed as nonmelanoma and 12 as melanoma by at least five out of eight histopathologists. Of these 12 cases, putative in situ and invasive lesions were equally represented. Finally, six cases (5.6%) were assigned to the H-D group, since each of them had been diagnosed as melanoma by four pathologists and as nonmelanoma by the other four. Among these cases belonging to the H-D group, two lesions showed a dermal, possibly invasive component, whereas the other four were completely intraepidermal.

Patients' age and lesion locations did not significantly vary across the three groups of lesions (data not shown).

Comparison between the Dermoscopic and the Histopathologic Diagnoses

Table 2 shows the distribution of the 107 lesions according to the degree of diagnostic concordance and the prevailing diagnosis simultaneously for both dermoscopy and histopathology.

Within the D-A group,15 lesions were diagnosed as melanoma; 11 out of these 15 cases were also within the H-A group with a confirmed histopathologic diagnosis of melanoma. The other four cases were found in the H-PA group, with a prevailing diagnosis of mel-

TABLE 3
Distribution of 107 Melanocytic Skin Lesions According to the Diagnostic Score Groups

	DDS groups						
	D-A		D-PA + D-D		Total		McNemar test P value
HDS groups	n	%	n	%	n	%	
H-A	40	51.3	38	48.7	78	100	
H-PA + H-D	8	27.6	21	72.4	29	100	
Total	48	44.9	59	55.1	107	100	0.000

DDS: dermoscopic diagnostic score; D–A/H–A: dermoscopic/histopathologic agreement (all diagnoses in agreement); D–PA/H–PA: dermoscopic/histopathologic partial agreement (one to three disagreeing diagnoses); D–D/H–D: dermoscopic/histopathologic disagreement (no prevailing diagnosis); HDS: histopathologic diagnostic score.

TABLE 4
Schouten *Kappa* According to Dermoscopic and Histopathologic Diagnostic Score Groups

	Agreement on histopathologic diagnosis				
DDS groups	*n*	k	SE	95% CI	z test *P* value
D-A	48	0.87	0.05	(0.78–0.96)	
D-PA + D-D	59	0.64	0.07	(0.51–0.77)	0,002
	Agreement on dermoscopic diagnosis				
HDS groups	*n*	k	SE	95% CI	z test *P* value
H-A	78	0.58	0.05	(0.48–0.69)	
H-PA + H-D	29	0.38	0.09	(0.20–0.56)	0,013

DDS: dermoscopic diagnostic score; SE: standard error; CI: confidence interval; D–A/H–A: dermoscopic/ histopathologic agreement (all diagnoses in agreement); D–PA/H–PA: dermoscopic/ histopathologic partial agreement (one to three disagreeing diagnosis); D–D/H–D: dermoscopic/histopathologic disagreement (no prevailing diagnosis).

TABLE 5
Distribution of 107 Melanocytic Skin Lesions According to the Prevailing Dermoscopic Pattern and the Prevailing Histopathologic Diagnosis

	H-A		H-PA		H-D	
Prevailing dermoscopic pattern	M	nM	M	nM	Undefined	Total
Globular	1	14	2	4	2	23
Reticular		15	1	3		19
Starburst		9		3		12
Homogeneous		7				7
Hypopigmented	1	3	2			6
Peripheral Hyperpigmentation	1	3	1			5
Regression		2		2	3	7
Multicomponent	17	5	5		1	28

H–A: histopathologic agreement (all diagnoses in agreement); H–PA: histopathologic partial agreement (one to three disagreeing diagnoses); H–D: histopathologic disagreement (no prevailing diagnosis); M: melanoma; nM: nonmelanoma.

anoma. By contrast, group D-PA encompassed 16 lesions with a prevailing dermoscopic diagnosis of melanoma; five out of these 16 cases were rated as nonmelanoma by all pathologists. An even greater variability in histopathologic diagnosis was observed for nine lesions belonging to the D-D group.

Tables 3 and 4 summarize the above-mentioned comparative data. It is worth noting that, among the 48 cases belonging to the D-A group, only eight (16.7%) received at least one disagreeing histopathologic diagnosis. Thus, the histopathologic agreement on this subset of lesions was excellent, with a Schouten *kappa* value of 0.87 (95% CI: 0.78-0.96; Table 4). By contrast, among the other 59 cases belonging to the D-PA and D-D groups, 21 (35.6%) showed some histopathologic controversy (Table 3), with a Schouten kappa value of 0.64 (95% CI: 0.51-0.77; Table 4). The difference between these calculated kappa values was statistically significant (z test *P* value = 0.002; Table 4).

By examining the agreement among dermoscopists by degree of histolopathologic diagnostic certainty, we found that out of 78 cases belonging to the H-A group, 38 (48.7%) received at least one disagreeing dermoscopic diagnosis (Table 3), with a Schouten kappa value for dermoscopic diagnosis of 0.58 (Table 4). Conversely, out of 29 cases belonging to the H-PA and H-D groups, 21 (72.4%) showed some dermoscopic controversy (Table 3), with a kappa value for dermoscopic diagnosis of 0.38 (95% CI: 0.20-0.56). The difference between these values of kappa was statistically significant (z test *P* value = 0.013).

Finally, as an overall measure of the dermoscopists' performance, a mean sensitivity of 88% (range, 75-100%) and a mean specificity of 81% (range, 67-98%) were calculated on the subset of 78 lesions for which a unanimous histopathologic diagnosis was made.

Dermoscopic Patterns

Table 5 shows the prevailing dermoscopic patterns (as defined in the Materials and Methods section)

matched with the histopathologic diagnostic score groups.

The most frequent dermoscopic pattern was ascribed to the multicomponent type (28 out of 107, 26.16%). This pattern was strongly associated with a histopathologic diagnosis of melanoma in complete agreement (17 out of 28 cases). On the other hand, the homogeneous pattern, although not commonly found, was always associated with a histopathologic diagnosis of nonmelanoma in complete concordance.

Regression structures were detected in seven lesions with dermoscopy: a prevailing dermoscopic diagnosis of nonmelanoma was made in four cases, but three out of these four cases were within the H-D score group. A prevailing dermoscopic diagnosis of melanoma was made only in one case, but in this case histopathologic consultations gave a unanimous diagnosis of nonmelanoma. Overall, five of these lesions with regression structures showed some histopathologic controversy.

DISCUSSION

Few studies have addressed the issue of interobserver diagnostic reproducibility on melanocytic skin lesions;[6,13–15] however, only one of these studies was carried out in a combined setting of dermoscopy and histopathology.[15] The chance-corrected interobserver agreement seems to be acceptable both for dermoscopy (*kappa* = 0.47 according to Binder et al.[6]) and for histopathology (*kappa* = 0.61 according to Corona et al;[13] *kappa*=0.50 according to Farmer et al.[14]). However, when viewed critically, these data indicate that the degrees of concordance among histopathologists and among dermoscopists are by far less than excellent, suggesting a considerable variability in both dermoscopic and histopathologic diagnosis of melanoma even when made by experts in the field.

In light of a previous pilot study,[15] we issued the hypothesis that the low accuracy of dermoscopic diagnosis due to selected subsets of melanocytic skin lesions could be accounted for by the relatively low reproducibility of the histopathologic diagnosis, especially in difficult lesions. This hypothesis was supported by the finding that when at least two formally trained dermoscopists disagreed in evaluating a pigmented skin lesion, histopathologic consultations might also give controversial results. However, because of the small numbers of lesions (only ten) examined in the above-mentioned study, we decided to test this hypothesis on a larger number of melanocytic skin lesions.

The current study was carried out by dermoscopists and by histopathologists who had been cooperating for some years in the field of melanocytic skin lesions. Therefore, an excellent interobserver agreement was expected. Interestingly, this was found to be almost true only for histopathology. The *kappa* value of 0.74 obtained in this series was nearly excellent and fairly higher than the value previously reported in other published studies.[13,14] By contrast, the agreement among dermoscopists on the same set of lesions was by far less than excellent (*kappa* = 0.59). A likely explanation for the higher variability of dermoscopic diagnoses is that the lesions included in the current study had been consecutively submitted to surgical excision because of their equivocal clinical and/or dermoscopic features. Therefore, the dermoscopic interobserver agreement was not excellent, but, for a selected set of reasonably difficult lesions, the *kappa* value we found among dermoscopists was even higher than the one previously found by Binder et al.[6] in a series of consecutive (nonselected) cases. Moreover the dermoscopic diagnostic discrepancies found in the present series had no clinical consequence, since all the examined lesions had been already submitted to surgical excision.

In the current study, a diagnosis of melanoma was made in 325 out of 856 dermoscopy observations (38%), and in only a slightly lower number of histopathologic observations (273 out of 856; 32%). This finding is quite surprising when considering that dermoscopy was expected to overdiagnose a potentially fatal neoplasm such as melanoma.

A further observation in the current data, made by examining the dermoscopic diagnoses in relation to the prevailing histopathologic diagnoses (as shown in Table 2), showed that dermoscopy yielded only five possible false positive cases and three possible false negative cases. The current results show that in expert hands dermoscopy carries a high diagnostic specificity. However, it should be kept in mind that its main challenge is to reach a sensitivity value close to 100% (no false negative cases), with an acceptable degree of specificity.[7]

The definition of the three diagnostic score groups highlighted some noteworthy issues. Histopathologically, putatively in situ and invasive cases of melanoma showed no preferential distribution among the HDS groups. This finding was unexpected and in contrast with the results of previous studies, which underlined that thin or in situ lesions were prone to give the greatest diagnostic troubles.[13,15] The presence of several spitzoid and nevoid lesions, which are a well-recognized source of diagnostic problems,[28,29] in the current series might explain the apparently paradoxical finding of highly controversial lesions with often large dermal components.

The elucidation of the relationship between the reproducibility of dermoscopic and histopathologic diagnosis of pigmented lesions is the main result of

the current study. As shown in Table 4, in the group of cases in which at least one dermoscopic diagnostic disagreement was encountered, the *kappa* value for histopathology decreased from 0.87 to 0.64. The same effect was observed also when examining the *kappa* values for dermoscopy matched with the degree of histopathologic diagnostic concordance: in the group of lesions with at least one histopathologic diagnostic disagreement, the *kappa* value for dermoscopy of decreased from 0.58 to 0.38 (Table 4). This means that in the current set of data there was a close relationship between dermoscopic and histopathologic diagnostic difficulties, although this might not be true in individual cases in daily clinical practice.

One might suppose that the question of whether difficult cases for dermoscopy are still difficult for histopathology has no clinical consequence, all the doubtful lesions being excised after dermoscopy.

However, our re-evaluation of the prevailing dermoscopic patterns showed that lesions with clear-cut regression structures, which might not be excised solely on the basis of these dermoscopic features, often carry great histopathologic controversy. Should our observation be confirmed by larger studies, the dermoscopic evidence of regression could be a criterion to excise the lesion and particularly to mention this troublesome feature to the histopathologist. The knowledge of a dermoscopic pattern which might be predictive of a histopathologically difficult melanocytic skin lesion should be an alert to a meticulous clinicopathologic correlation and to the request of a dermoscopic and histopathologic second opinion, and perhaps first, by using telecommunication via the Internet.[15]

REFERENCES

1. Grin CM, Kopf AW, Welkovich B, Bart RS, Levenstein MJ. Accuracy in the clinical diagnosis of malignant melanoma. *Arch Dermatol.* 1990;126:763-766.
2. Miller M, Ackerman AB. How accurate are dermatologists in the diagnosis of melanoma? Degree of accuracy and implications. *Arch Dermatol.* 1992;128:559-560.
3. Wolf IH, Smolle J, Soyer HP, Kerl H. Sensitivity in the clinical diagnosis of malignant melanoma. *Melanoma Res.* 1998;8: 425-429.
4. Mayer J. Systematic review of the diagnostic accuracy of dermatoscopy in detecting malignant melanoma. *Med J Aust.* 1997;167:206-210.
5. Steiner A, Pehamberger H, Wolff K. In vivo epiluminescence microscopy of pigmented skin lesions. 2. Diagnosis of small pigmented skin lesions and early detection of malignant melanoma. *J Am Acad Dermatol.* 1987;17:584-591.
6. Binder N, Schwarz N, Winkler A, et al. Epiluminescence microscopy. A useful tool for the diagnosis of pigmented lesions for formally trained dermatologists. *Arch Dermatol.* 1995;131:286-291.
7. Argenziano G, Fabbrocini G, Carli P, De Giorgi V, Sammarco E, Delfino M. Epiluminescence microscopy for the diagnosis of doubtful melanocytic skin lesions. Comparison of the ABCD rule of dermatoscopy and a new 7-point checklist based on pattern analysis. *Arch Dermatol.* 1998;134:1563-1570.
8. Yadav S, Vossaert KA, Kopf AW, Silverman M, Grin-Jorgensen C. Histopathologic correlates of structures seen on dermoscopy (epiluminescence microscopy). *Am J Dermatopathol.* 1993;15:297-305.
9. Soyer HP, Smolle J, Hoedl S, Pachernegg H, Kerl H. Surface microscopy: a new approach to the diagnosis of cutaneous pigmented tumors. *Am J Dermatopathol.*1989;11:1-10.
10. Soyer HP, Kenet RO, Wolf IH, Kenet BJ, Cerroni L. Clinicopathological correlation of pigmented skin lesions using dermoscopy. *Eur J Dermatol.* 2000;10:22-28.
11. Krieger N, Hiatt RA, Sagebiel RW, Clark WH Jr., Mihm MC Jr. Inter-observer variability among pathologists' evaluation of malignant melanoma: effects upon an analytic study. *J Clin Epidemiol.* 1994;47:897-902.
12. De Wit PEJ, van't Hof-Grootenboer B, Ruiter DJ, et al. Validity of the histopathological criteria used for diagnosing displastic naevi. *Eur J Cancer.* 1993;29A:831-839.
13. 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.
14. 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.
15. Ferrara G, Argenziano G, Soyer HP, et al. Histopathologic interobserver agreement on the diagnosis of melanocytic skin lesions with equivocal dermoscopic features: a pilot study. *Tumori.* 2000;86:445-449.
16. Argenziano G, Scalvenzi M, Staibano S, et al. Dermatoscopic pitfalls in differentiating pigmented Spitz naevi from cutaneous melanomas. *Br J Dermatol.* 1999;141:788-793.
17. Pehamberger H, Steiner A, Wolff K. In vivo epiluminescence microscopy of pigmented skin lesions. I. Pattern analysis of pigmented skin lesions. *J Am Acad Dermatol.* 1987;17:571-583.
18. Binder M, Puespoeck-Schwarz M, Steiner A, Kittler H, Muellner M, Wolff K. Epiluminescence microscopy of small pigmented skin lesions: short term formal training improves the diagnostic performance of dermatologists. *J Am Acad Dermatol.* 1997;36:197-202.
19. Stolz W, Riemann A, Cognetta AB, et al. ABCD rule of dermatoscopy: a new practical method for early recognition of malignant melanoma. *Eur J Dermatol.* 1994;4:521-527.
20. Nachbar F, Stolz W, Merkle T, et al. The ABCD rule of dermatoscopy. *J Am Acad Dermatol.* 1994;30:551-559.
21. Ackerman AB, Cerroni L, Kerl H. Pitfalls in the diagnosis of malignant melanoma. Philadelphia: Lea & Febiger, 1994.
22. Rosai J. Skin. Tumors and tumorlike conditions. In: Rosai J, editor. Ackerman's surgical pathology, 8th ed. St Louis: CV Mosby, 1996:138-166.
23. Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas.* 1960;20:37-46.
24. Fleiss JL, Cohen J, Everitt BS. Large sample standard errors of kappa and weighted kappa. *Psychol Bull.* 1969;72:323-327.
25. Schouten HJA. Nominal scale agreement among observers. *Psychometrika.* 1986;51:453-466
26. Fleiss JL. Measuring nominal scale agreement among many raters. *Psychol Bull.* 1971;76:378-382.
27. Fleiss JL. Statistical methods for rates and proportions, 2nd ed. New York: Wiley, 1981.
28. Walsh N, Crotty K, Palmer Am, McCarthy S. Spitz nevus versus spitzoid melanoma: an evaluation of the current distinguishing histopathologic criteria. *Hum Pathol.* 1998;29:1105-1112.
29. McNutt NS. "Triggered trap": nevoid malignant melanoma. *Semin Diagn Pathol.* 1998;15:203-209