

Clinical and Dermatoscopic Diagnosis of Malignant Melanoma

Assessed by Expert and Non-expert Groups

HENRIK LORENTZEN¹, KAARE WEISMANN¹, CARSTEN SAND PETERSEN¹, FREDERIK GRØNHØJ LARSEN¹, LENA SECHER¹ and VERA SKØDT²

Departments of ¹Dermatology and ²Pathology, Bispebjerg Hospital, Copenhagen, Denmark

We investigated the nosographic and diagnostic probabilities and likelihood ratios of dermatoscopy in order to evaluate the method's role in decision-making regarding melanoma. Clinical slides and dermatoscopic photos were obtained from 232 patients referred for dermatoscopy. Four dermatoscopy "experts" and 5 "non-experts" assessed the slides. Diagnoses were compared with histopathology. Sensitivity of the clinical assessments was 0.78 vs. 0.69 ("experts" vs. "non-experts"), sensitivity of dermatoscopy assessment was 0.83 vs. 0.69 ($p=0.04$). The expert group demonstrated increased specificity (from 0.89 to 0.94) when applying dermatoscopy compared with clinical assessment alone ($p=0.03$). Positive likelihood ratios were doubled in the "expert group" and the negative likelihood ratios improved 25% with dermatoscopy compared with clinical assessment. Key words: malignant melanoma; dermatoscopy; epiluminiscence microscopy; likelihood ratio; decision-making.

(Accepted January 6, 1999.)

Acta Derm Venereol 1999; 79: 301–304.

Henrik Lorentzen, Department of Dermatology, Bispebjerg University Hospital, DK-2400 Copenhagen NV, Denmark.

The increased incidence of cutaneous malignant melanoma (CMM) has been termed a "melanoma epidemic" (1). The morbidity from CMM has not been accompanied by a corresponding increment in mortality, which is probably due to extended awareness in the general public and amongst physicians. Early excision is pivotal to the cure of this otherwise eventually fatal disease. For this reason early detection is an important key-point. This may be facilitated by dermatoscopy or epiluminiscence microscopy. A dermatoscope is a hand-held skin-surface microscope providing approximately $10\times$ magnification for *in vivo* examination of pigmented skin lesions. With the application of immersion oil the corneal layer of the epidermis becomes translucent, making the different layers of the epidermis and the upper dermis visible. The colours and structures seen with the dermatoscope have been described extensively and associated with histological findings (2–7).

The purpose of the present study was to investigate the diagnostic validity of dermatoscopy applied to an unselected material of pigmented skin elements or elements suspect of CMM observed in patients referred to the outpatients' clinic.

MATERIALS AND METHODS

Between 1994 and 1997 clinical photographs and dermatophotographs (Heine Dermaphot, Heine Optotechnik) of pigmented skin lesions were obtained from 242 patients. All lesions were surgically removed

and histopathologically diagnosed. For this study, 10 cases were considered unfit for evaluation, leaving a total of 232 cases.

Four dermatologists with 4–5 years' daily experience with dermatoscopy (designated the "expert group") and five residents with 1–2 years' interest and formal training in dermatoscopy (designated the "non-expert group") participated in the study.

Clinical photographs showing the lesion at distance and close-up and 1–5 dermatoscopic photographic slides (depending on the size of the lesion) were projected to a screen in a darkened room. Each observer first recorded the clinical diagnosis and then the dermatoscopic diagnosis on an entry form. The observers were not allowed to discuss or modify their clinical assessments after the presentation of the dermatophotos. The observers were familiar with both the ABCD-rule of dermatoscopy proposed by Stolz et al. (6) and Kenet et al.'s risk stratifying algorithm of pigment network features of dermatoscopy (8). The observers were not constrained by either of the rules. The ABCD scores were not used to obtain the diagnoses. Rather a pattern recognition process was intended.

Each case was presented for approximately 2 min, which all observers regarded as sufficient time. The performance of assessment was divided into 3 sessions of 3 h each.

The biopsy specimens, which were obtained after the clinical and dermatoscopic photographs had been performed, were stained routinely with haematoxylin-eosin and immunostained with HMB-45 and S-100 (DAKO, Denmark) in the case of suspicion of melanoma. One of us (VS) re-evaluated all cases to confirm the pathology diagnosis, which was used as the gold standard in this study.

Statistics

For each observer the diagnoses "CMM" and "other than CMM" were listed in a 2×2 table against the corresponding histopathology diagnoses. A case was termed true positive (TP) if the clinical or dermatoscopic diagnosis and the histopathology diagnosis was melanoma. True negative cases (TN) were cases where the clinical or dermatoscopic as well as the histopathology diagnoses were "other than melanoma". False negative cases (FN) were those with a clinical diagnosis "other than melanoma" and histopathology showing melanoma (i.e. missed melanomas). False positives (FP) were cases with a clinical or dermatoscopic diagnoses of "melanoma" and histopathology demonstrating "other than melanoma".

Sensitivity is the fraction of the histopathologically proven CMM, which was correctly diagnosed by clinical inspection or dermatoscopy, respectively. Sensitivity was calculated as $TP/(TP+FN)$.

Specificity is the fraction of patients not suffering from CMM, which could be ruled out by clinical assessment or dermatoscopy. Specificity was calculated as $TN/(TN+FP)$.

The predictive value of a positive test result, PV_{POS} , is the fraction of the clinical or dermatoscopic melanoma suspicious cases verified by histopathology. PV_{POS} was calculated as $TP/(TP+FP)$.

The predictive value of a negative test result, PV_{NEG} , was the fraction of cases where the clinical or dermatoscopic assessment was "other than CMM", which also had melanoma ruled out by histopathology. PV_{NEG} was calculated as $TN/(TN+FN)$.

The positive likelihood ratio (LR+) is the probability that a patient with a disease (e.g. CMM) has a positive test outcome (e.g. CMM

diagnosed by dermatoscopy) divided by the probability that a patient without the disease has a positive test outcome.

The negative likelihood ratio (LR \div) is the probability that a patient with a disease has a negative test outcome (e.g. CMM ruled out by dermatoscopy) divided by the probability that a patient without the disease has a negative test outcome. Hence, a diagnostic test should have a LR $+$ greater than 1 (the higher the better) and a LR \div less than 1 (the lower the better). The likelihood ratios express the ability of a test to separate a diseased from a non-diseased population and may be used in medical decision-making. To use a test rationally, the test outcome should have a potential of altering the diagnostician's attitude as to whether a patient suffers from a given disease, rather than merely confirming an already established diagnosis. Simel et al. (9) described the relationship: Odds of disease after the test has been performed equals the pretest odds multiplied by the likelihood ratio.

Positive likelihood ratio, LR $+$, is calculated as sensitivity/(1-specificity) and negative likelihood ratio, LR \div , is calculated as (1-sensitivity)/specificity

The 95% confidence interval for LR $+$ (8)=

$$\exp\left(\ln \frac{\text{sensitivity}}{(1-\text{specificity})} \pm 1.96 \cdot \sqrt{\frac{(1-\text{specificity})}{TP} + \frac{\text{specificity}}{FP}}\right)$$

And the 95% confidence interval for LR \div (8)=

$$\exp\left(\ln \frac{(1-\text{sensitivity})}{\text{specificity}} \pm 1.96 \cdot \sqrt{\frac{\text{sensitivity}}{FN} + \frac{1-\text{specificity}}{FP}}\right)$$

For comparison of the diagnostic performances between the "expert" and "non-expert" group a "naive pooling" of data from each group was performed and McNemar's chi-square test was used for comparisons of sensitivity and specificity and the chi-square test for proportions from independent samples was used for comparing predictive values.

For comparison of likelihood ratios in the expert and the non-expert groups we used the Mann-Whitney U test. $p < 0.05$ was considered significant.

RESULTS

There were 49 melanomas (21%) as diagnosed by histopathology. Fifty percent were pigmented naevi, 7% were basal-cell carcinomas, 7% were blue naevi, 5% were seborrhoeic keratoses and 2% were atypical naevi. The remaining 8% were Spitz naevi, Bowen's disease, sarcoid, nevus spilus, hemangioma, and others.

The dermatoscopy experts assessed almost all cases (98–100%), whereas the non-expert group completed fewer assessments, from 76 to 98%.

The "expert" group achieved a higher sensitivity (0.82) using the dermatoscope compared with their clinical assessment (0.77) $p = 0.03$ (Table I). The "non-expert" group did not improve sensitivity with dermatoscopy. There was a decrease from 0.62 with clinical assessment to 0.58 with dermatoscopy ($p = 0.18$). In their dermatoscopy assessments, the "expert" group achieved 24.6% higher sensitivity than the "non-experts" ($p < 0.0001$) and for the clinical assessments, sensitivity was 15.6% higher in the "expert" group than in the "non-expert" group ($p < 0.001$).

The "expert" and the "non-expert" group increased specificity by dermatoscopy (Table I).

The positive predictive value (Table I) was increased from 0.66 to 0.78 in the expert group with dermatoscopy ($p = 0.008$). It was not increased in the "non-expert" group

Table I. Nosographic and diagnostic probabilities

	"Experts" (n=4)		"Non-experts" (n=5)	
	Clinical	Dermatoscopic	Clinical	Dermatoscopic
Sensitivity	0.77	0.82*	0.62	0.58 n.s.
Specificity	0.89	0.94*	0.88	0.92*
PV+	0.66	0.78**	0.60	0.64 n.s.
PV \div	0.94	0.95 n.s.	0.90	0.89 n.s.

PV+ predictive value of a positive test result.

PV \div predictive value of a negative test result.

* Significant at $p < 0.05$.

** Significant at $p < 0.01$.

n.s. Non-significant.

($p = 0.40$). In clinical assessment there was no difference in the positive predictive value between the "expert" and the "non-expert" group ($p = 0.10$). The difference between the two groups became significant when the performance of dermatoscopy was compared ($p = 0.0041$).

Negative predictive values were not increased by dermatoscopy (Table I).

When data from "experts" and "non-experts" were pooled, a significant increase in positive likelihood ratio was observed when dermatoscopy was performed compared with clinical assessment alone ($p = 0.01$). Due to the small number of observers in each group the difference within each group was not statistically significant ($p < 0.20$). In the "expert" group each observer doubled his positive likelihood ratio (Table II). In the "non-expert" group the increase in positive likelihood ratio was less pronounced. In the "expert" group there was a 25% fall in negative likelihood ratio when dermatoscopy was performed compared with clinical assessment. No improvement was observed in the non-expert group.

Pooling the assessments from all observers, the sensitivity for the following non-CMM lesions was: basal cell carcinoma 0.62 (experts 0.66, non-experts 0.59); naevus pigmentosus 0.74 (experts 0.76, non-experts 0.72); blue naevi 0.86 (experts 0.87, non-experts 0.83); and seborrhoeic keratoses 0.39 (experts 0.40, non-experts 0.28). Specificity was: basal cell carcinoma 0.98; naevus pigmentosus 0.84; blue naevi 0.98; and seborrhoeic keratoses 0.97.

DISCUSSION

The results are in agreement with previous reports on the effect of dermatoscopy on the sensitivity and specificity (Table III). The differences in the nosographic probabilities between the various studies may be caused by differences in the experience with dermatoscopy of the observers, differences in studied cases, differences in study design and differences in the frequency of CMM.

Stolz et al. (10) reported an almost perfect sensitivity of 98%. Their study included only smaller CMM and 93% of the cases were superficial spreading melanomas or melanomas *in situ*. Pigmented skin lesions other than CMM consisted of melanocytic naevi (junction, compound or dermal). The features in the ABCD-rule proposed by Stolz et al. (6, 10) are especially selected for diagnosing superficial spreading melanomas, where pigment network abnormalities prevail, whereas thick CMM are often associated with gray-blue areas

Table II. Positive ($LR+$) and negative ($LR-$) likelihood ratios. In brackets: 95% confidence intervals. The confidence intervals are not symmetric but rather logarithmically distributed

Observer	LR +		LR -	
	Clinical assessment	Dermatoscopic assessment	Clinical assessment	Dermatoscopic assessment
"Experts"				
I	7.1 (4.8, 11.0)	16.4 (8.9, 30.1)	0.23 (0.09, 0.41)	0.09 (0.04, 0.23)
II	7.8 (4.9, 12.4)	24.6 (11.1, 54.7)	0.25 (0.15, 0.42)	0.19 (0.11, 0.34)
III	5.9 (4.0, 8.8)	6.6 (4.4, 10.1)	0.26 (0.15, 0.44)	0.23 (0.14, 0.41)
IV	9.0 (5.5, 14.7)	18.0 (9.0, 36.1)	0.23 (0.14, 0.41)	0.22 (0.13, 0.37)
"Non-experts"				
1	6.8 (3.9, 12.0)	10.0 (5.4, 18.4)	0.51 (0.38, 0.69)	0.41 (0.29, 0.59)
2	13.0 (6.4, 26.4)	15.9 (7.0, 36.1)	0.30 (0.17, 0.52)	0.36 (0.22, 0.58)
3	4.5 (3.0, 6.8)	4.0 (2.8, 4.7)	0.19 (0.09, 0.41)	0.17 (0.07, 0.39)
4	4.8 (3.0, 7.5)	6.4 (3.9, 10.6)	0.42 (0.28, 0.63)	0.38 (0.25, 0.58)
5	8.5 (4.6, 15.9)	10.4 (5.3, 20.6)	0.45 (0.32, 0.64)	0.45 (0.32, 0.63)

and a vascular pattern (11). Atypical naevi, Spitz naevi and seborrhoeic keratoses are differential diagnoses of CMM to be considered. In this study the false-negative cases (CMM erroneously classified as another diagnosis by dermatoscopy) were misdiagnosed as pigmented naevi/seborrhoeic keratosis (59%), basal cell carcinoma (22%) or atypical naevi (13%). Morton & MacKie (12) likewise reported atypical naevi, seborrhoeic keratoses and basal cell carcinomas as frequent erroneous clinical diagnoses of CMM.

Nachbar et al. (13) performed a prospective study using the dermatoscopic ABCD rule to distinguish melanocytic naevi from CMM and reported a sensitivity of 0.91, which was somewhat higher than the sensitivity obtained by the experts in the present study, the difference, however, was not significant ($p=0.17$).

Nilles et al. (2) excluded non-melanocytic lesions from their study. A score based on the presence and the extent of 8 dermatoscopic elements was used to separate benign melano-

cytic lesions from CMM. Binder et al. (14), Christofolini et al. (15) as in the present study, did not select the referred cases before dermatoscopy. This seems more similar to daily clinical work. The lower sensitivity of dermatoscopy reported by Binder et al. (14) could be due to differences in study design. In contrast to their study, we included macroscopic clinical photographs taken at a distance as well as close-up. The macroscopic clinical slides may have indicated risk factors such as age, gender, and location of the element, skin type and presence of moles or ephelides. The non-melanoma cases in the study by Soyer et al. (16) encompassed melanocytic naevi, basal cell carcinomas, seborrhoeic keratoses and others. The observers in their study were experts and the dermatoscopic ($10\times$ magnification) or epiluminiscence microscopy ($6-40\times$ magnification) were performed directly on the patient rather than using the photo-documentation afterwards. It is not clear whether the observers were blinded to the patients' history as the observers were in our study or if obtaining the medical record was part of the clinical/dermatoscopic examination. Soyer et al. (16) reported a higher sensitivity and a lower specificity than we found. In addition to differences in presented patient cases, this points to a lower threshold for diagnosing melanomas in the study by Soyer et al.

With the accessibility of relatively inexpensive and easy-to-handle dermatoscopes, dermatoscopy has moved from an experimental phase to use in daily clinical practice. Outside an experimental setting, skin lesions which are not considered suspicious of malignancy, may not be excised and therefore may not proceed to the golden standard examination: histopathology. For this reason the predictive value of a method is of importance to the clinician. We found high positive predictive values and negative predictive values for experts performing dermatoscopy. In contrast to sensitivity and specificity, the predictive values are dependent on the

Table III. Review of studies performed on sensitivity and specificity of the melanoma diagnosis by dermatoscopy

	Sensitivity	Specificity	n total	n CMM
Stolz et al. (10)	0.98	0.90		
Soyer et al. (16)	0.94	0.82	159	65
Nachbar et al. (13)	0.93	0.91	194	69
Nilles et al. (2)	0.90	0.85	260	72
Christofolini et al. (15)	0.87	0.79	187	33
Binder et al. (14)				
Experts	0.68	0.91		
Non-experts	0.44	0.85	-	-
This study				
"Experts"	0.83	0.94	232	49
"Non-experts"	0.69	0.90	-	-

disease prevalence. Therefore the predictive values cannot be extrapolated to large-scale screening or case finding programs. The high negative predictive values in this study indicate that the risk of misclassifying a CMM as a benign skin element was low. According to Bayes theorem the negative predictive value will increase, whereas the positive predictive value will decrease when a population with a lower prevalence of CMM is investigated. For this reason fewer false negatives (missed melanomas) and more false positives (unnecessary surgery) would be the consequence of applying dermatoscopy to a screening or case finding setting. For example, in a population sample with a melanoma prevalence of approximately 1%, $PV+$ would be 12.26% and $PV- = 99.82\%$.

We intended to simulate an every-day clinical situation, but we did not present patient histories. Patient history is known to improve interobserver homogeneity from poor to substantial agreement on recognition of skin malignancy (17).

Higgins et al. (18) investigated a 7-point checklist including the clinical ABCD rule, that directs patients' and physicians' attention to possible CMM. They found that 70% of benign pigmented skin lesions were erroneously regarded as suspect of melanoma (false positives), whereas 58% of atypical naevi were not considered suspect according to the 7-point rule.

From this data, it may be calculated that the odds that a benign pigmented skin lesion was judged to be suspicious was 2.3 to 1, i.e. in favour of excision. If one of these patients had dermatoscopy performed by an "expert" and the outcome of the examination was "not CMM" the odds of CMM shifted to (negative likelihood ratio times prior odds) 0.42 to 1, i.e. in favour of an expecting attitude. This example demonstrates that dermatoscopy may influence the clinical decision-making.

The positive likelihood ratio was doubled when an "expert" performed dermatoscopy, whereas no benefit was demonstrated for "non-experts". It has previously been shown that "non-experts" exhibit a decline in sensitivity using dermatoscopy, whereas experts gain from using it (14). A brief training session of 9 h resulted in an improvement in non-expert dermatoscopy performance (19). We found a similar tendency supporting the view that formal training is indispensable for correct usage of dermatoscopy.

ACKNOWLEDGMENT

Leo Pharmaceutical Products A/S, Ballerup, Denmark, kindly supported this work. We thank the participants for their cooperation and interest in the study.

REFERENCES

- Burton RC, Coates MS, Hersey P, Roberts G, Chetty MP, Chen S, et al. An analysis of a melanoma epidemic. *Int J Cancer* 1993; 55: 765–770.
- Nilles M, Boedeker R, Schill W. Surface microscopy of naevi and melanomas—clues to melanoma. *Br J Dermatol* 1994; 130: 349–355.
- Stanganelli I, Burrioni M, Rafanelli S, Bucchi L. Intraobserver agreement in interpretation of digital epiluminescence microscopy. *J Am Acad Dermatol* 1995; 33: 584–589.
- Bahmer FA, Fritsch P, Kreutsch J, Pehamberger H, Rohrer C, Schindera I, Smolle J, et al. Meeting report: terminology in surface microscopy. *J Am Acad Dermatol* 1990; 23: 1159–1162.
- Steiner A, Binder M, Schemper M, Wolff K, Pehamberger H. Statistical evaluation of surface microscopy criteria for melanocytic pigmented skin lesions. *J Am Acad Dermatol* 1993; 29: 581–588.
- Stolz W, Braun-Falco O, Bilek P, Landthaler M, Cognetta AB. *Color Atlas of Dermatoscopy*. First English Ed. Oxford: Blackwell Science, 1994.
- Guillod JF, Skaria AM, Solomon D, Saurat J-H. Epiluminescence videomicroscopy: black dots and brown globules revisited by stripping the stratum corneum. *J Am Acad Dermatol* 1997; 36: 371–377.
- Kenet RO, Kang S, Kenet BJ, Fitzpatrick TB, Sober AJ, Barnhill RL. Clinical diagnosis of pigmented lesions using digital epiluminescence microscopy. *Arch Dermatol* 1993; 129: 157–174.
- Simel DL, Samsa GP, Matchar DB. Likelihood ratios with confidence: sample size estimation for diagnostic test studies. *J Clin Epidemiol* 1991; 44: 763–770.
- Stolz W, Rieman A, Cognetta AB, Pillet L, Abmayr W, Hölzel D, et al. ABCD rule of dermatoscopy: a new practical method for early recognition of malignant melanoma. *Eur J Dermatol* 1994; 4: 521–527.
- Argenziano G, Fabbrocini G, Carli P, De Giorgi V, Delfino M. Epiluminescence microscopy: criteria of cutaneous melanoma progression. *J Am Acad Dermatol* 1997; 37: 68–74.
- Morton CA, MacKie RM. Clinical accuracy of the diagnosis of cutaneous malignant melanoma. *Br J Dermatol* 1998; 138: 283–287.
- Nachbar F, Stolz W, Merkle T, Cognetta AB, Vogt T, Landthaler M, et al. The ABCD rule of dermatoscopy. *J Am Acad Dermatol* 1994; 30: 551–559.
- Binder M, Schwarz M, Winkler A, Steiner A, Kaider A, Wolff K, Pehamberger H. Epiluminescence microscopy. A useful tool for the diagnosis of pigmented skin lesions for formally trained dermatologists. *Arch Dermatol* 1995; 131: 286–291.
- Cristofolini M, Zumiani G, Bauer P, Cristofolini P, Boi S, Micciolo R. Dermatoscopy: usefulness in the differential diagnosis of cutaneous pigmentary lesions. *Melanoma Res* 1994; 4: 391–394.
- Soyer HP, Smolle J, Leitinger G, Rieger E, Kerl H. Diagnostic reliability of dermoscopic criteria for detecting malignant melanoma. *Dermatology* 1995; 190: 25–30.
- Whited JD, Horner RD, Hall RP, Simel DL. The influence of history on interobserver agreement for diagnosing actinic keratoses and malignant skin lesions. *J Am Acad Dermatol* 1995; 33: 603–607.
- Higgins EM, Hall P, Todd P, Murthi R, Du Vivier AW. The application of the seven-point check-list in the assessment of benign pigmented lesions. *Clin Exp Dermatol* 1992; 17: 313–315.
- Binder M, Puespoeck-Schwartz M, Steiner A, Kitler H, Muellner M, Wolff K, Pehamberger H. 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.