Is Dermoscopy (Epiluminescence Microscopy) Useful for the Diagnosis of Melanoma?

Results of a Meta-analysis Using Techniques Adapted to the Evaluation of Diagnostic Tests

Marie-Lise Bafounta, MD; Alain Beauchet, MD, PhD; Philippe Aegerter, MD, PhD; Philippe Saiag, MD

**Objective:** To assess, by means of meta-analysis techniques for diagnostic tests, the accuracy of dermoscopic (also known as dermatoscopy and epiluminescence microscopy) diagnosis of melanoma performed by experienced observers vs naked-eye clinical examination.

**Data Sources:** MEDLINE, EMBASE, PASCAL-BIOMED, and BIUM databases were screened through May 31, 2000, without any language restrictions.

**Study Selection:** Original studies were selected when the following criteria were met: spectrum of lesions well described, histologic findings as standard criterion, and calculated or calculable sensitivity and specificity. Eight of 672 retrieved references were retained.

**Data Extraction:** Three investigators extracted data. In case of disagreement, consensus was obtained. Summary receiver operating characteristic curve analysis was used to describe the central tendency of the studies, and to compare dermoscopy and clinical examination.

**Data Synthesis:** Selected studies represented 328 melanomas, mostly less than 0.76 mm thick, and 1865 mostly melanocytic benign pigmented skin lesions. For dermoscopic diagnosis of melanoma, the sensitivity and specificity ranges were 0.75 to 0.96 and 0.79 to 0.98, respectively. Dermoscopy had significantly higher discriminating power than clinical examination, with respective estimated odds ratios of 76 (95% confidence interval, 25-223) and 16 (95% confidence interval, 9-31) \( (P = .008) \), and respective estimated positive likelihood ratios of 9 (95% confidence interval, 5.6-19.0) and 3.7 (95% confidence interval, 2.8-5.3). The roles of the number of lesions analyzed, the percentage of melanoma lesions, the instrument used, and dermoscopic criteria used in each study could not be proved.

**Conclusion:** For experienced users, dermoscopy is more accurate than clinical examination for the diagnosis of melanoma in a pigmented skin lesion. Arch Dermatol. 2001;137:1343-1350

**Morbidity and mortality attributed to skin melanoma have dramatically increased in recent years.** As there is a strong inverse correlation between survival rate and tumor thickness, and no effective therapy exists for advanced melanoma, early diagnosis and excision are essential to reduce morbidity and improve patients’ survival. Dermoscopy (also known as dermoscopy, epiluminescence microscopy, incident light microscopy, and skin-surface microscopy) is a noninvasive technique to examine pigmented anatomic structures of the epidermis, dermoeipidermal junction, and superficial papillary dermis that are not visible to the naked eye. The technique uses in vivo microscopy with fluid applied to the skin to render the stratum corneum more translucent. Various instruments have been used, ranging from widely used, inexpensive, handheld instruments (eg, Dermatoscope [Heine Ltd, Herrshing, Germany] and Episcope [Welch Allyn Inc, Skaneateles Falls, NY]) that provide a fixed magnification of \( \times 10 \), to binocular microscopes and, more recently, to digital videomicroscopy, the last 2 options providing higher magnification.

**For editorial comment see page 1361**

It has been claimed that dermoscopy improves sensitivity (up to 30%) and specificity of melanoma diagnosis compared with clinical diagnosis and thus could lower the excision rate of histologically benign but clinically doubtful nevi. However, there is no consensus on this
METHODS

We followed the guidelines for meta-analyses evaluating diagnostic tests.13,15

DATA SOURCES

Four bibliographic databases were screened through May 31, 2000: (1) the online MEDLINE (National Library of Medicine, Bethesda, Md; http://www.ncbi.nlm.nih.gov/entrez/query.fcgi) from January 1968; (2) Elsevier Science online EMBASE (http://www.ovid.com/products/databases/database_info.cfm) from 1989; (3) PASCAL-BIOMED on CD-ROM (http://www.silverplatter.com/catalog/pbma.htm), which also contains congressional reports, from 1987; and (4) BIUM (Bibliothe`que Universitaire Medica, Paris, France), which contains all French doctoral theses in medicine from 1985.

No language restrictions were applied. The MEDLINE search was conducted by successively calling the following keywords: dermatoscopy, dermoscopy, epiluminescence microscopy, diascopy, surface microscopy, incident light microscopy, in combination with melanoma or skin neoplasm (exploded mode), sensitivity, and specificity. Similar strategies were used to search EMBASE and PASCAL-BIOMED. Other sources were the references of the retrieved articles and textbooks on the topic. All of the retrieved references were entered into the bibliography management software EndNote 3.01 (ISI Research Soft, Berkeley, Calif), to facilitate the search for duplicate references.

STUDY SELECTION

Three investigators, 2 dermatologists (M.-L.B. and P.S.) and 1 statistician (A.B.), independently reviewed the studies and filled in preestablished evaluation grids. In case of divergent evaluations, consensus was obtained. We first selected original studies on dermoscopy by reading the titles and abstracts. An initial review evaluated selected studies for the following quality criteria: spectrum of included patients well described (with a spectrum of melanoma lesions and lesions commonly confused with melanoma), blinded comparison of diagnostic tests with histologic findings as the standard criterion, and calculated or calculable statistical values (sensitivity, specificity, and positive and negative likelihood ratios).18,19 A study was held to be of insufficient quality when 1 criterion was absent. Studies of acceptable quality were finally selected when the diagnostic performance of dermoscopy in melanoma was compared with that of naked-eye clinical examination and when dermoscopy was performed by experienced operators. Excluded from this research were reviews; editorials; original articles on the development of dermoscopic criteria for diagnosing PSLs, terminology of dermoscopy, accuracy of diagnosis of a single type of PSL, and accuracy of computer-assisted lesion analysis; and studies not comparing dermoscopy with naked-eye examination.

DATA EXTRACTION

The following data were extracted from selected articles (after translation if not in English or French): setting, lesion-selection criteria, total number and spectrum of lesions analyzed, prevalence of melanoma, Breslow index of the tested melanomas, degree of observer blinding, study type (prospective study or analysis of randomly selected lesions from a database), instrument used, criteria applied for clinical and dermoscopic diagnoses, dermoscopy reproducibility, and percentage of lesions in which no dermoscopic diagnosis could be made. The potential bias of these studies was analyzed.15,18 Dermoscopic criteria for melanoma were classified as either pattern analysis20 (even when slightly different pattern criteria were added to the original publication) or ABCD21 (asymmetry, border, color, and differential structure) or 7FFM22 (7 features for melanoma) criteria. Sensitivity, specificity, and positive and negative likelihood ratios were calculated according to standard formulas. When possible, these values were recalculated from the data given in the articles. When dermoscopy was performed by both nonexpert and expert observers, we selected only the results recorded by skilled operators. When individual clinical or dermoscopic evaluations from multiple observers were

LITERATURE SEARCH

Five hundred sixty-four articles were identified from MEDLINE, 223 from EMBASE (with 138 references found in both databases), 117 from PASCAL-BIOMED (with 96 references also found in MEDLINE or EMBASE), and 2 doctoral theses from the BIUM database. After elimina-
given in a study, we retained only the responses most frequently given by the investigators. When melanoma had been clinically diagnosed according to the ABCDE (asymmetry, border, color, diameter, and evolution) criteria, we required the presence of at least 2 criteria; this gives the best sensitivity with acceptable specificity.21

STATISTICS

Sensitivity and specificity, commonly used in diagnostic tests, rely on a single threshold whose modification to increase sensitivity decreases specificity, and vice versa. When studies use different criteria to define positive and negative test results, the thresholds differ. Even when studies apply the same criteria, the thresholds may not be the same, since interpretation of the test depends on subjective judgment. To take different thresholds into account across studies, we used summary receiver operating characteristic (sROC) curve analysis, proposed by Moses et al,24 to describe the central tendency of the studies, and also to compare the discriminatory powers of naked-eye clinical examination and dermoscopy. The significance of differences was statistically analyzed by applying a Wilcoxon paired-sample test to the parameter Ln(odds ratio), where Ln indicates the neperian logarithm. A P value less than .05 was considered significant. Estimated positive and negative likelihood ratios were calculated from the point of intersection of the sROC curve with the line (1 − specificity) + sensitivity = 1, which slopes from the upper left to the lower right corners.24 At that intersection, sensitivity equals specificity. That point is a global measurement of test efficacy. Statistical analysis was performed with the SAS 6.12 software package (SAS Institute Inc, Cary, NC).

CONSTRUCTION OF sROC CURVES

A diagnostic test within a single study may have several combinations of sensitivity and specificity values, depending on the threshold value. The graph reporting these values (where the y-axis represents true-positive rate values [TPR, sensitivity] and the x-axis represents false-positive rate values [FPR, 1−specificity]) defines the receiver operating characteristic (ROC) space, and the points within this space are the combinations of the different values. The ROC curve joins the different points and enables assessment of the variation of sensitivity as a function of varying specificity. The closer the curve comes to the upper left-hand corner of the graph, the better the test.

While the ROC curve illustrates data for a diagnostic test coming from a single study, an sROC curve fits data from different studies into the same curve. The method used in this study, proposed by Moses et al,24 follows these steps:

1. Convert each FPR and TPR into its logistic transformation (logit):

   \[ U = \logit(FPR) = \log\left(\frac{FPR}{1-FPR}\right) \]
   \[ V = \logit(TPR) = \log\left(\frac{TPR}{1-TPR}\right) \]

2. For each study, calculate:

   \[ D = V - U = \log(\text{odds ratio}) \]

   where the odds ratio equals \( \frac{[TPR/(1−TPR)]}{[FPR/(1−FPR)]} \) and D measures how well the test discriminates between subjects with and without the disease.

3. Plot each study’s point \((S_i, D_i)\) in an \((S, D)\) space and fit an unweighted least-squares regression line. The linear regression model is:

   \[ D = \alpha + \beta S \]

   The regression coefficient \( \beta \) provides an estimate of the extent to which the Ln(odds ratio) is dependent on the threshold used. If it is near 0, the common odds ratio is given by \( \exp^\alpha \).

4. Reverse-transform (TPR) and plot the sROC curve onto the FPR-vs-TPR graph.

   \[ \text{TPR} = \frac{1 + \exp^{-\alpha - \beta S}}{1 + \exp^{-\alpha - \beta S}} \frac{(1-FPR)/FPR}{(1-FPR)/FPR} \]

5. Compare the 2 diagnostic tests. All the points are plotted for both tests in the \((S, D)\) space, and the regression line is fitted. If \( \beta \) is nonstatistically different from 0, a Wilcoxon paired-samples test is applied to compare the Ln (odds ratio) for clinical examination and dermoscopy.

STUDY SELECTION

Forty-nine studies were subjected to further selection procedures after we read their titles and abstracts. Thirty-four references fulfilled the required quality criteria. Eight studies, all referenced in the MEDLINE database, met all of our selection criteria.7,12,22,25

The study by Stanganelli et al23 was included despite the absence of histologic findings for 92% of the 3372 PSLs assessed, for the following reasons: (1) all lesions clinically or dermoscopically suspected of being melanoma had been subjected to histologic examination, thus allowing the calculation of true- and false-positive evaluations, and (2) all of the included patients were inhabitants of the region covered by the Romagna Cancer Registry, and the entire series of patients with benign clinical or dermoscopic diagnosis and no referral for excision were cross-checked with this registry a median of 31 months after dermoscopic examination. This registry has good indicators of satisfactory levels of registration completeness,25 and we ascertained that it contains the expected number of patients with melanoma,26 as compared with the most recent figures of disease incidence in Italy.27 Only 2 additional patients with melanoma have been found. Thus, only a few patients with melanoma seem to have been missed in that study. As our standard criterion was histologic findings, only verified lesions were considered for the analyses presented in Table 1 (see below), including those found through the cancer registry.

The exclusion of certain studies needs further explanation. In one article, the standard criterion did not allow clear discrimination between melanoma and atypical nevus.28 Although Ascierto et al29 tested 15719 PSLs, the results of clinical and dermoscopic examinations were
**Table 1. Design of the Selected Studies**

<table>
<thead>
<tr>
<th>Source</th>
<th>Setting/Study Type</th>
<th>Lesion Selection Criteria</th>
<th>Observers</th>
<th>Method of Clinical Diagnosis of Melanoma</th>
<th>Method of Dermoscopy (Magnification)</th>
<th>Dermoscopic Criteria for Melanoma</th>
</tr>
</thead>
<tbody>
<tr>
<td>Benelli et al&lt;sup&gt;22&lt;/sup&gt;</td>
<td>Dermatologic surgery department/prospective in vivo evaluation</td>
<td>All consecutive PSLs excised during 1 y (No. of patients not given)</td>
<td>2 Dermatologists, not involved in decision whether to excise lesions</td>
<td>ABCDE (≥2 criteria)</td>
<td>Not given</td>
<td>7FFM (score ≥2)</td>
</tr>
<tr>
<td>Carli et al&lt;sup&gt;11&lt;/sup&gt;</td>
<td>Dermatology department/retroactive evaluation of photographs</td>
<td>Not given</td>
<td>15 Dermatologists, all skilled</td>
<td>Not given</td>
<td>Dermaphot (slides taken at ×10 magnification)</td>
<td>Pattern analysis</td>
</tr>
<tr>
<td>Cristofolini et al&lt;sup&gt;30&lt;/sup&gt;</td>
<td>Dermatologic surgery department/prospective in vivo evaluation during a campaign for the early detection of melanoma</td>
<td>All consecutive PSLs excised during 9 mo (700 patients seen), all not excised with ABCDE criteria ≥1</td>
<td>4 Dermatologists, with consensus evaluation</td>
<td>ABCDE criteria (≥2 criteria)</td>
<td>Videomicroscopy (×400)</td>
<td>Pattern analysis</td>
</tr>
<tr>
<td>Dummer et al&lt;sup&gt;6&lt;/sup&gt;</td>
<td>Dermatology department/prospective in vivo evaluation</td>
<td>All consecutive difficult-to-diagnose PSLs seen during 1 y (No. of patients not given)</td>
<td>3 Dermatologists, not involved in decision whether to excise lesions, not involved in clinical diagnosis</td>
<td>ABCDE criteria (threshold not given)</td>
<td>Pattern analysis</td>
<td></td>
</tr>
<tr>
<td>Klahn et al&lt;sup&gt;8&lt;/sup&gt;</td>
<td>Dermatology department/prospective in vivo evaluation</td>
<td>Not given</td>
<td>Not given</td>
<td>Not given</td>
<td>Not given</td>
<td>Not given</td>
</tr>
<tr>
<td>Lorentzen et al&lt;sup&gt;7&lt;/sup&gt;</td>
<td>Dermatology department/retroactive evaluation on photographs</td>
<td>All clinical and dermoscopic photographs of PSLs taken during a 4 y period</td>
<td>4 Expert dermatologists (5 novices discarded)</td>
<td>Not given</td>
<td>Dermaphot (slides taken at ×10)</td>
<td>Pattern analysis</td>
</tr>
<tr>
<td>Soyer et al&lt;sup&gt;12&lt;/sup&gt;</td>
<td>PSL clinic/prospective in vivo evaluation</td>
<td>Not given, but among clinically difficult-to-diagnose PSLs referred by general practitioners or dermatologists</td>
<td>3 Expert dermatologists</td>
<td>Not given</td>
<td>Handheld dermoscope (×10) and stereomicroscope (×6 to ×40)</td>
<td>Pattern analysis</td>
</tr>
<tr>
<td>Stanganelli et al&lt;sup&gt;25&lt;/sup&gt;</td>
<td>Skin cancer clinic/prospective in vivo evaluation</td>
<td>All patients with PSLs recorded in database (see “Results” section)</td>
<td>1 Expert dermatologist</td>
<td>ABCDE (threshold not given)</td>
<td>Stereomicroscope (×6 to ×16) and digital videomicroscopy (×6 to ×40)</td>
<td>Pattern analysis, with digital contrast and enhancement procedures</td>
</tr>
</tbody>
</table>

*PSL indicates pigmented skin lesion. ABCDE are clinical criteria for the diagnosis of melanoma<sup>23</sup>; 7FFM, dermoscopic criteria for the diagnosis of melanoma.<sup>22</sup>†Heine Ltd, Herrshing, Germany.

not clearly reported. The possibility of performing “informal” dermoscopy to support the clinical evaluation, as explained by Nachbar et al,<sup>4</sup> could have biased the clinical diagnosis. In both studies, the choice of lesions to be assessed by the standard criterion (ie, histologic findings) was also dictated by the result of dermoscopy (verification bias). The well-done study by Steiner et al<sup>6</sup> was eliminated because of the unavailability of data to calculate dichotomous outcomes of melanoma or nonmelanoma for dermoscopy, although such data were given for the clinical evaluations. Two studies<sup>17,34</sup> were excluded because dermoscopy was compared with pictures taken without oil at magnifications of ×16 and ×10, respectively, and not with naked-eye examination. The study by Pazzini et al<sup>31</sup> was rejected because the data needed to calculate sensitivity, specificity, and likelihood ratios were not reported, although figures were given for diagnostic accuracy and index of suspicion.

**INTERNAL AND EXTERNAL VALIDITY OF THE SELECTED STUDIES**

The main characteristics of the 8 retained studies are given in Table 1. All came from dermatology departments, were published between 1993 and 2000, and were set in dermatology clinics or even in specialized PSL clinics. Two studies were performed on images originating from a database, and 6 prospectively recorded dermoscopy results in vivo. As required by our selection criteria, all studies used histologic findings as a standard criterion, but only 4 had histologic findings verified by an external review or derived from a consensus among at least 2 ob-
servers. Although clearly stated only for the 2 studies with image databases, clinical and dermoscopic examinations seem to have been conducted in all studies independent of the standard criterion because of evident temporal relationship. The choice of patients to be assessed by the standard criterion was not entirely independent of the test results, as, in all studies, lesions were excised mainly because of a clinically presumed risk of malignancy or, possibly, patient request. Only Stanganelli et al\textsuperscript{25} collected information aimed at calculating the true false-negative rate (see above). Lesion-selection criteria were absent or almost absent from 2 articles. It was stated in only 4 reports that all consecutive PSLs excised\textsuperscript{4,9} or recorded\textsuperscript{7,23} within a determined period had been included.

To assess the validity of the comparison between clinical and dermoscopic evaluations, we investigated their independence. Only 1\textsuperscript{11} of the 2 retrospective studies clearly stated that the dermoscopic evaluations had been made without knowing the results of the clinical ones, and both claimed that dermoscopy did not influence clinical judgment.\textsuperscript{7,11} For the 6 vivo dermoscopic evaluations, it is highly probable that dermoscopy had been performed after the lesions were examined with the naked eye. Clinical evaluations were clearly recorded before dermoscopy was done in all but 2 of these studies.\textsuperscript{1,2,25}

Criteria for the clinical diagnosis of melanoma were specified in 4 reports, with a threshold value clearly given in only 2. The number of clinically atypical nevi was not clearly stated in most studies. The instruments used differed among the studies. Transparencies taken with a Dermaphot (Heine Ltd) at original \(\times 10\) magnification were analyzed in 2 studies.\textsuperscript{7,11} One study used a handheld monocular dermoscope with \(\times 10\) magnification,\textsuperscript{10} and another used a digital videomicroscope with magnifications up to \(\times 400.9\) One study used both a handheld dermoscope (\(\times 10\)) and a binocular stereomicroscope (\(\times 6\) to \(\times 40\)),\textsuperscript{2} and another one, both a stereomicroscope (\(\times 6\) to \(\times 16\)) and digital videomicroscopy (\(\times 6\) to \(\times 40\))\textsuperscript{25} with digital contrast and enhancement procedures, but none clearly distinguished their results. One study independently assessed dermoscopy (with an unspecified instrument) and high-frequency sonography.\textsuperscript{8} No authors stated that a skin lesion could not be diagnosed by dermoscopy, and only one group\textsuperscript{7} stated that 4% of the photographs were not suitable for evaluation. Various dermoscopic criteria have been used to diagnose melanoma: pattern analysis for 6 studies and 7FFM criteria for 1 study.\textsuperscript{4} Interobserver agreement for dermoscopy was assessed simultaneously in only 1 study.\textsuperscript{11}

The spectrum of lesions examined is given in Table 2. There were 2193 lesions, with 328 melanomas. Sample size varied from 15 to 824 lesions. Melanoma lesions represented 3% to 49% of the excised lesions. Most melanomas were thin (<0.76 mm) in the 4 studies that provided such information. Nonmelanoma lesions represented 1865 lesions. Their histologic findings were detailed in all but 1 study\textsuperscript{25}; they were mainly melanocytic lesions (67%-100% of all nonmelanoma PSLs). No study gave demographic information on included patients.

### RESULTS OF SELECTED STUDIES

Results of the 8 studies are summarized in Table 2. The respective sensitivity, specificity, and positive and negative likelihood ratio ranges were as follows: 0.50 to 0.94, 0.55 to 0.89, 1.91 to 10.66, and 0.61 to 0.08 for the clinical diagnosis of melanoma; and 0.75 to 0.96, 0.79 to 0.98, 4.21 to 76.62, and 0.28 to 0.04 for the dermoscopic diagnosis of melanoma. Variations of specificity, sensitivity, and positive and negative likelihood ratios for the diagnoses of melanoma after dermoscopy were as follows: \(-0.05\) to \(+0.30\), 0.00 to \(+0.35\), 0.00 to \(+65.96\), and \(-0.34\) to 0.00, respectively.

According to the results of 5 studies, dermoscopy had higher sensitivity than nondermoscopic diagnosis,\textsuperscript{7,9,11,25} but only 2 of them demonstrated statistical significance.\textsuperscript{7,25} Because individual data were not detailed in the 3 remaining reports, we were unable to perform post hoc statistical tests. In the only study that addressed that issue, the improvement of sensitivity was more pronounced for melanomas less than 0.76 mm thick, as compared with thicker ones.\textsuperscript{9} In all of these studies, specificity and positive and negative likelihood ratios were also better. Four of these studies used high-magnification dermoscopy, with either Dermaphot or videomicroscopy. The remaining one provided no information on the instrument used.

For 2 studies, in which the clinical evaluation already had achieved high sensitivity and specificity, no difference was found between clinical and dermoscopic assessments.\textsuperscript{10,12} For another study, in which the clinical evaluation had high sensitivity but low specificity, dermoscopy had comparable sensitivity but better specificity and positive and negative likelihood ratios.\textsuperscript{22}

### STATISTICS

The sROC curves of clinical and dermoscopic evaluations of melanoma (Figure) indicate that dermoscopy has significantly higher discriminatory power, with estimated odds ratios of 76 (95% confidence interval [CI], 25-223) vs 16 (95% CI, 9-31) for naked-eye examination \((P = .008)\). The weighted least squares, the robust-resistant line, and other methods proposed by Moses et al\textsuperscript{22} gave similar values. For dermoscopy and naked-eye examination, respectively, the estimated positive likelihood ratios were 9 (95% CI, 5.6-19.0) and 3.7 (95% CI, 2.8-5.3), and the estimated negative likelihood ratios, 0.11 (95% CI, 0.05-0.18) and 0.27 (95% CI, 0.19-0.36). Exploratory analyses were performed to investigate whether diagnostic accuracy differed in subgroups defined by the characteristics of the patients and diagnostic tests. No relationship could be established between dermoscopic diagnostic sensitivity and the number of lesions analyzed, the percentage of melanoma lesions, or the dermoscope or dermoscopic criteria used.

### COMMENT

Our results showed that, for dermatologists working within specialized clinics and experienced in dermoscopy, this technique has higher discriminatory power than...
As for all meta-analyses, one important limitation is publication bias, with small studies on dermoscopy or with negative results not having been submitted or accepted for publication. We tried to limit this handicap by conducting a comprehensive search in 4 different databases, including 2 databases potentially covering studies unpublished in the peer-reviewed medical literature. This strategy proved useful, because overlaps between these databases were small. The selection of relevant studies followed strictly established guidelines, with extensive evaluation of the design and validity of selected studies. However, because we retained only studies that met predetermined quality criteria, numerous others were rejected.

mate of diagnostic accuracy of a test among primary studies is not simply the calculation of the mean sensitivity and the mean specificity of this test. Such an approach would be inappropriate, because it is likely that different studies used different explicit or implicit thresholds, so that a primary study with high sensitivity may have low specificity, and vice versa. That is why it is recommended, for each primary study, to plot a scattergram of the true-positive rate (sensitivity) against the false-positive rate (1 – specificity), which are also the axes used for an sROC curve. The odds ratio of each primary study can be combined, allowing the 95% CI to be calculated and statistical comparisons between different diagnostic tests to be made. Furthermore, it is possible to calculate estimated positive and negative likelihood ratios for dermoscopy for the detection of melanoma, which are more useful for clinicians than odds ratios. As dermoscopy achieved likelihood ratios far from 1, this technique seemed an accurate diagnostic test. It should be noted that values greater than 10 or less than 0.1 are considered to generate large and often conclusive changes from pretest to posttest probability, and values of 5 to 10 and 0.1 to 0.2, moderate but substantial shifts of these probabilities. Thus, this technique gives precise information on the confidence clinicians may have in a test to detect or exclude a disease, ie, with dermoscopy of a PSL to exclude melanoma and to decide not to excise a clinically doubtful lesion. However, this technique is complex, particularly when subgroup analysis or modeling is done. Therefore, someone with methodologic expertise should be included in the research team, as in our study.

As other studies have shown that dermoscopy increased the diagnostic ability only when performed by formally trained operators, dermatologists should make every effort to master this useful tool.

These conclusions are more favorable for dermoscopy than those of Mayer's systematic review. One explanation is that she selected only 2 studies on this issue, whereas we retained 5 additional studies published later and another study published in German.

Moreover, we were able to obtain a quantitative summary of the diagnostic efficacy of dermoscopy by applying meta-analysis methods adapted for the evaluation of diagnostic tests. To the best of our knowledge, we have used this approach for the first time in dermatology, as a MEDLINE search (data not shown) did not find any meta-analysis for diagnosis of skin diseases. Indeed, meta-analyses are more commonly conducted to examine treatments or prognosis. In fact, obtaining a summary estimate of diagnostic accuracy of a test among primary studies is not simply the calculation of the mean sensitivity and the mean specificity of this test. Such an approach would be inappropriate, because it is likely that different studies used different explicit or implicit thresholds, so that a primary study with high sensitivity may have low specificity, and vice versa. That is why it is recommended, for each primary study, to plot a scattergram of the true-positive rate (sensitivity) against the false-positive rate (1 – specificity), which are also the axes used for an sROC curve. The odds ratio of each primary study can be combined, allowing the 95% CI to be calculated and statistical comparisons between different diagnostic tests to be made. Furthermore, it is possible to calculate estimated positive and negative likelihood ratios for dermoscopy for the detection of melanoma, which are more useful for clinicians than odds ratios. As dermoscopy achieved likelihood ratios far from 1, this technique seemed an accurate diagnostic test. It should be noted that values greater than 10 or less than 0.1 are considered to generate large and often conclusive changes from pretest to posttest probability, and values of 5 to 10 and 0.1 to 0.2, moderate but substantial shifts of these probabilities. Thus, this technique gives precise information on the confidence clinicians may have in a test to detect or exclude a disease, ie, with dermoscopy of a PSL to exclude melanoma and to decide not to excise a clinically doubtful lesion. However, this technique is complex, particularly when subgroup analysis or modeling is done. Therefore, someone with methodologic expertise should be included in the research team, as in our study.

As other studies have shown that dermoscopy increased the diagnostic ability only when performed by formally trained operators, dermatologists should make every effort to master this useful tool.

These conclusions are more favorable for dermoscopy than those of Mayer's systematic review. One explanation is that she selected only 2 studies on this issue, whereas we retained 5 additional studies published later and another study published in German.

Moreover, we were able to obtain a quantitative summary of the diagnostic efficacy of dermoscopy by applying meta-analysis methods adapted for the evaluation of diagnostic tests. To the best of our knowledge, we have used this approach for the first time in dermatology, as a MEDLINE search (data not shown) did not find any meta-analysis for diagnosis of skin diseases. Indeed, meta-analyses are more commonly conducted to examine treatments or prognosis. In fact, obtaining a summary estimate of diagnostic accuracy of a test among primary studies is not simply the calculation of the mean sensitivity and the mean specificity of this test. Such an approach would be inappropriate, because it is likely that different studies used different explicit or implicit thresholds, so that a primary study with high sensitivity may have low specificity, and vice versa. That is why it is recommended, for each primary study, to plot a scattergram of the true-positive rate (sensitivity) against the false-positive rate (1 – specificity), which are also the axes used for an sROC curve. The odds ratio of each primary study can be combined, allowing the 95% CI to be calculated and statistical comparisons between different diagnostic tests to be made. Furthermore, it is possible to calculate estimated positive and negative likelihood ratios for dermoscopy for the detection of melanoma, which are more useful for clinicians than odds ratios. As dermoscopy achieved likelihood ratios far from 1, this technique seemed an accurate diagnostic test. It should be noted that values greater than 10 or less than 0.1 are considered to generate large and often conclusive changes from pretest to posttest probability, and values of 5 to 10 and 0.1 to 0.2, moderate but substantial shifts of these probabilities. Thus, this technique gives precise information on the confidence clinicians may have in a test to detect or exclude a disease, ie, with dermoscopy of a PSL to exclude melanoma and to decide not to excise a clinically doubtful lesion. However, this technique is complex, particularly when subgroup analysis or modeling is done. Therefore, someone with methodologic expertise should be included in the research team, as in our study.
Classifying PSLs, only a few studies commented on how nevertheless obvious. First, although pathologists may vary in general dermatology practice remains unknown. As dermatoscopy is a subjective interpretation of images, a high threshold of the observer to conclude that melanoma is low-magnification, inexpensive dermoscope was not shown to be significantly superior to clinical examination in the only study to test that instrument. All other studies investigated instruments with higher magnification rates, such as stereomicroscope, videomicroscope, or the less expensive Dermaphot, which dramatically increases the magnification of the lesion by the projection on a screen, but is not suitable to give immediate information. Concerning the dermoscopic criteria for melanoma used, only one study used structured explicit criteria such as 7FFM or ABCD, which have been reported to have increased sensitivity and specificity for detecting melanoma. Most studies examined a distorted spectrum of PSLs, with far too many melanoma and too few nonmelanocytic PSLs, such as seborrheic keratoses, as compared with their frequency in usual dermatology consultant practices. As dermoscopy is a subjective interpretation of images, a high number of melanomas in the sample may modify the threshold of the observer to conclude that melanoma is present. Finally, as all selected studies were conducted in clinics with expertise on PSLs, the place of dermoscopy in general dermatology practice remains unknown.

Some limitations of the selected studies are, nevertheless, obvious. First, although pathologists may vary in classifying PSLs, only a few studies commented on how the standard criterion was reached. Second, in some selected studies, naked-eye examination and dermoscopy were not performed totally independently, and dermoscopy could have “improved” the accuracy of the clinical diagnosis proposed. Third, the impact of the verification bias (mainly clinically equivocal lesions were excised) present in most studies must be considered. Although Stanganeli et al gave reassuring indirect information that issue, such studies may overestimate the sensitivity of dermoscopy by missing some false-negative cases. Notably, the clinical threshold for considering a lesion sufficiently suspect to be excised seemed to have been set low (at least 2 ABCDE criteria) in the 2 studies that gave the results. Fourth, data concerning intraobserver and interobserver variability were absent or sparse in most selected studies. However, satisfactory reproducibility of the dermoscopy results was shown in most studies specifically addressing that issue.

Nevertheless, several points enhance the value of dermoscopy in detecting melanoma. First, its superiority to clinical examination was demonstrated despite the high sensitivity and specificity of the clinical diagnosis of melanoma reported in most selected studies, as compared with figures usually published. This finding may be explained by the expertise in PSLs of most investigators, rather than by the high percentage of melanomas among the lesions examined. Second, mostly highly equivocal (ie, mainly or exclusively melanocytic) benign lesions and thin, early melanomas, presumably difficult to diagnose but highly curable by complete excision, were examined. In this situation, sensitivity and specificity of the technique are expected to decrease. However, although dermoscopy seems to have all the qualities required for a good diagnostic test—accuracy, absence of adverse effect, target disorder dangerous if left undiagnosed, and effective treatment of the disease when diagnosed early—its usefulness, which means providing information beyond that otherwise available and inciting a change of patient treatment, remains to be proved.

To provide better evidence of the usefulness of dermoscopy, future research should address the following issues. Studies should be undertaken to analyze, in the standard dermatology practice, a more representative mixture of melanoma and benign PSLs, as well as studies aimed to lower the number of unknown false-negative results. The value of the handheld low-magnification dermatoscope needs to be assessed. However, the most important issue would be to prove in a large patient population that, with dermoscopy, it is possible to reduce the number of excision biopsies performed for nonmelanoma PSLs while increasing the number of thin melanomas excised or, at least, without increasing the number of melanomas missed.

Accepted for publication April 30, 2001.

The authors received no outside support for this study.

We thank Janet Jacobson for editing our English and Marie-Dominique Allen for her help in searching computerized databases.

Corresponding author: Philippe Saiag, MD, Service de Dermatologie, Hôpital Ambroise Paré, 9, av Charles-de-Gaulle, 92104 Boulogne CEDEX, France (e-mail: dermatom@apr.ap-hop-paris.fr).

REFERENCES


