A comparison of dermatologists' and primary care physicians' accuracy in diagnosing melanoma

Chen S C, Bravata D M, Weil E, Olkin I

Authors' objectives
To compare the accuracy of dermatologists and primary care physicians (PCPs) in identifying pigmented lesions suggestive of melanoma, and in making the appropriate management decisions to perform a biopsy or to refer the patient to a specialist.

Searching
MEDLINE, EMBASE, and Cancerlit were searched from 1966 to October 1999. The MeSH terms used were reported in the paper. The bibliographies of retrieved papers, the proceedings of the American Academy of Dermatology conferences (1997 to 1999), and the Science Citation Index were also searched. Foreign language papers with English abstracts were included.

Study selection
Study designs of evaluations included in the review
No inclusion criteria for the study design were stated. Both prospective and retrospective studies were included.

Specific interventions included in the review
The diagnostic test assessed was examination by either a dermatologist or PCP to: (1) diagnose melanoma; and/or (2) to take appropriate action according to whether the symptoms indicated that further investigation for melanoma was indicated, i.e. a biopsy should be performed or the patient referred to a specialist (B/R). Studies were excluded if they did not specify the type of physician making the diagnosis.

Reference standard test against which the new test was compared
The reference standard used for diagnostic accuracy (DA) was melanoma identified by biopsy. The reference standard used for B/R was either histologically identified melanoma, or the correct application of standard clinical criteria for referral or biopsy, regardless of the actual pathology, as compiled by a melanoma panel.

Participants included in the review
No inclusion criteria for the participants were stated. By implication they had to be presenting with a lesion. Studies of volunteers with mock lesions cosmetically applied were excluded.

Outcomes assessed in the review
The studies must have assessed sensitivity or specificity in relation to either DA or to B/R accuracy, or to have data available to enable their calculation. Studies were excluded if they did not report data specific to melanoma.

How were decisions on the relevance of primary studies made?
One author reviewed the titles and abstracts for potential relevance. The authors do not state how the papers were selected for the review, or how many of the reviewers performed the selection.

Assessment of study quality
The authors applied a validity assessment based on an existing assessment framework. This evaluated the adequacy of the design and conduct of the studies in relation to sample size, direct comparison of the two groups, the reporting of both sensitivity and specificity, and specific issues relevant to this subject, such as including more than one melanoma and early and late types. The seven criteria used were listed in the paper. The authors do not state how the papers were assessed for validity, or how many of the reviewers performed the validity assessment.
**Data extraction**

Two authors independently extracted the data from the included studies, with any discrepancies resolved by discussion. The sensitivity and specificity were either abstracted or, if not directly reported, were calculated from the data by two reviewers independently. The data required for the calculations were requested from the authors if not published.

The other data extracted included the type of physician, lesion characteristics and, for B/R accuracy, comparison with the ‘gold’ standard.

**Methods of synthesis**

How were the studies combined?

Summary receiver operating characteristic (ROC) curves for dermatologists and for PCPs were constructed from the three studies that reported sufficient data on B/R accuracy. To do this, the outcomes in these studies were standardised into whether or not a malignant lesion was suspected. As only one study reported sufficient data for DA, a summary ROC curve could not be constructed for this outcome. The robustness of the summary ROC curves for B/R was also investigated by assessing the effects of a 10% decrease in PCP sensitivity and specificity.

How were differences between studies investigated?

Heterogeneity was discussed in relation to the outcome measurement used in different studies of B/R and to the definition of dermatologists.

**Results of the review**

Thirty-two studies were included. There were 10 prospective studies, which involved 583 dermatologists and 2,366 PCPs assessing between one and 12 melanomas. Five of these provided data for both diagnostic and B/R accuracy. Seven prospective studies provided data for both dermatologists and PCPs, whilst 25 studies provided data on one group only (2 on PCPs only and the remainder on dermatologists only).

The following results are taken from tables 2 and 3 of individual study results in the article. The summary of the results in the text contains discrepancies when compared with these tables.

**Diagnostic accuracy (DA) - prospective studies only.**

The sensitivity of dermatologists ranged from 0.81 to 1.00 (6 studies); the specificity range was not reported. The sensitivity of PCPs ranged from 0.42 to 1.00 (9 studies), while the specificity was 0.98 (1 study).

**Biopsy/referral (B/R) accuracy - prospective studies only.**

The sensitivity of dermatologists ranged from 0.82 to 1.00 (5 studies), while the specificity ranged from 0.70 to 0.89 (3 studies). The sensitivity of PCPs ranged from 0.70 to 0.91 (6 studies), while the specificity ranged from 0.51 to 0.87 (4 studies).

The summary ROC curve for B/R accuracy (based on 3 studies) showed the curves for the two groups crossing, so that there was no or insufficient evidence that one group was better than the other. This result was reported not to be robust on the basis that a shift of 10% in one group's results would have yielded different results.

All 4 studies that found sensitivities of 1.00 were based on the assessment of one single melanoma. None of these contributed to the B/R summary ROC curve.

The quality assessments of the prospective studies were summarised; there was a lack of studies with comprehensive data for both groups, as reported above. Fewer than half of the DA and of the B/R studies adequately described the type of lesions shown to participants. A third of the DA studies showed at least one early-stage melanoma, while a third showed at least one late-stage melanoma. Eleven per cent of the DA studies had an appropriate study size, whereas none of the B/R studies did.

**Authors' conclusions**

The currently available data are insufficient to support any policy regarding the use of either a gatekeeper system or
direct access to dermatologists. More well-designed studies are needed to determine either the superiority of dermatologists or the adequacy of PCPs.

**CRD commentary**

The inclusion criteria were very broad in relation to the participants and study design so that, for example, neither physicians nor dermatologists were defined, and there was no minimum standard for study design in terms of the number or range of melanomas assessed, and whether both groups of participants were included. Consequently, a large number of the studies included could not be used in the analysis. The authors pointed out potential biases as a consequence of who were included as dermatologists and as physicians.

The authors carried out a thorough search, which included foreign language papers and some attempt to identify 'grey' literature. Details of the individual study were not fully reported so that it was not possible, for example, to assess the characteristics of the three studies used to construct the summary ROC curve. The validity assessment criteria were relevant, although it might have been helpful to have assessed the study design and conduct in more detail. The review procedures, where reported, involved more than one reviewer, which helps to guard against error and any bias. However, the number of reviewers involved in the study selection process and validity assessment was not reported.

The summary of the results in the text was very poorly reported and did not concur with the numerical results in the tables. Discrepancies of this kind, whilst not apparently significant in relation to the review's conclusions, give some cause for concern in case other aspects of the results have not been thoroughly checked. The method of pooling the results of selected studies using a ROC curve was satisfactory. The authors’ conclusions appeared to be appropriate, and there was a full discussion of the methodological inadequacies of the included studies.

**Implications of the review for practice and research**

Practice: The authors did not state any implications for practice.

Research: Well-designed studies of diagnostic and B/R accuracy are needed that directly compare dermatologists with PCPs in the assessment of a range of stages of melanoma. These should use an adequate sample size and report sufficient data to calculate the sensitivity and specificity for each group. Resident and attending physicians should be evaluated separately.

Outcomes for patients with melanoma should be measured in addition to DA, to help inform policy decisions.

**Funding**

Dermatology Foundation, Health Care Policy Clinical Career Development Award; Veterans Affairs Ambulatory Care Fellowship; Veterans Affairs Health Care System; National Science Foundation, grant number DMS 9626265.

**Bibliographic details**


**PubMedID**

11735713

**Original Paper URL**

http://archderm.ama-assn.org

**Indexing Status**

Subject indexing assigned by NLM

**MeSH**
Clinical Competence; Dermatology /standards; Diagnosis, Differential; Humans; Melanoma /diagnosis /pathology; Primary Health Care /standards; ROC Curve; Research Design; Sensitivity and Specificity; Skin Neoplasms /diagnosis /pathology

**AccessionNumber**
12002000085

**Date bibliographic record published**
30/06/2003

**Date abstract record published**
30/06/2003

**Record Status**
This is a critical abstract of a systematic review that meets the criteria for inclusion on DARE. Each critical abstract contains a brief summary of the review methods, results and conclusions followed by a detailed critical assessment on the reliability of the review and the conclusions drawn.