Reliability of the Histopathologic Diagnosis of Malignant Melanoma in Childhood

Janine Wechsler, MD; Sylvie Bastuji-Garin, MD, PhD; Alain Spatz, MD; Christiane Bailly, MD; Bernard Criber, MD, PhD; Lucile Andrac-Meyer, MD; Beatrice Vergier, MD; Sylvie Fraitag, MD; Olivier Verola, MD; Pierre Wolkenstein, MD, PhD; and the French Cutaneous Cancerology Group

Objective: To assess interrater reliability in the diagnosis of malignant melanoma in children.

Design, Setting, and Participants: We collected 85 slides of melanomas diagnosed in patients younger than 17 years through a network of dermatopathologists and dermatologists. The slides were classified into 3 categories: (1) slides from children with metastatic melanoma; (2) slides from disease-free children with a follow-up of less than 5 years; (3) slides from disease-free children with a follow-up of 5 years or longer. Category 1 was considered the gold standard. Four pairs of expert dermatopathologists reviewed the slides and classified them into melanoma, nevus (including Spitz nevus), or ambiguous tumors.

Intervention: None.

Main Outcome Measure: Concordance between pairs of experts.

Results: For category 1 slides (n=20), the concordance was weak to moderate. For category 2 slides (n=47), the concordance was weak. For category 3 slides (n=18), the concordance was poor to moderate.

Conclusion: This study demonstrates that the reliability of diagnosis of melanoma in childhood is poor, even when submitted to experts.

Arch Dermatol. 2002;138:625-628

MEANOMA remains a rare tumor in childhood: fewer than 0.5% of cases of melanoma occur in children and adolescents. Therefore, there are only small series of melanoma reported in children, and the definition of children varies. Some studies included patients aged from 15 to 20 years, and others included only prepubescent children. The diagnosis of malignant melanoma in children is more difficult than in adults because of unusual clinical and histologic features. Although Spitz nevus is now considered a benign melanocytic tumor, the differential diagnosis with melanoma can be extremely difficult. Misdiagnosis seems to be common, occurring in up to 40% of children eventually found to have the disease. Consequently, some studies included only children with metastatic disease because metastatic spread is the only definite proof of malignancy. However, pathologists faced with melanocytic tumors in children have to decide whether they are malignant. Using metastatic spread as the gold standard for malignancy, we conducted a study to evaluate the difficulties of melanoma diagnosis in childhood. Toward that end, we tested the interrater reliability of the diagnosis of melanoma in childhood among a group of expert dermatopathologists.

RESULTS

The 120 coding sheets (85 childhood cases and 35 unambiguous cases) that were mailed to the 4 pairs of reviewers were completed and returned. The mean ±SD age of patients was 12 ± 4 years in category 1, 12 ± 4 years in category 2, and 11 ± 4 years in category 3 (P=.28). Each pair of experts reached a consensus in classifying the slides into 1 of the 3 categories.

For the 35 unambiguous cases slides, the agreement among the 4 pairs of experts was perfect (κ indexes = 100%). There was no discrepancy between the initial and the review diagnoses. The distribution of diagnosis of childhood cases differed widely among pairs of dermatopathologists (P<.001). Diagnosis of melanoma ranged from 45.9% to 58.8% of the cases, and diagnosis of ambiguous tumor ranged...
**PARTICIPANTS, MATERIALS, AND METHODS**

**EXPERT DERMATOPATHOLOGIST PANEL**

The expert dermatopathologist panel was made up of 8 of us (J.W., A.S., C.B., B.C., L.A.-M., B.V., S.F., and O.V.). Each member of this panel has actively practiced dermatopathology for at least 10 years in a French academic dermatology and/or oncology center and deals with melanoma in his or her daily practice. The members of the panel came from different centers throughout France and were not accustomed to working together. All members have an extensive record of publications in dermatopathology, and 3 of us (J.W., A.S., and C.B.) participated in a previous European study on melanoma in childhood.  

In most cases, the experts were not involved in making the initial diagnosis of the cases included in the present study. Considering that in their practice, dermatopathologists often discuss difficult cases with colleagues, 4 groups of 2 experts each were randomly assigned (groups A, B, C, and D). Each pair reviewed the same 120 slides according to a predetermined method.

**STUDIED CASES**

We included patients younger than 17 years with a diagnosis of melanoma made on the basis of histopathologic examination. One representative slide per case was selected. As control cases, slides of patients with unambiguous tumors, either benign nevi (n=14) or typical adult melanomas (n=21), were collected from the slide collections of the following centers: Hôpital Henri-Mondor, Institut Gustave Roussy, Centre Léon Bérard, Hôpital Saint-Louis, University Hospital of Strasbourg, and Hôpital Necker-Enfants Malades. A total of 120 slides representing 120 lesions (1 slide per lesion) were collected: 85 childhood melanoma cases and 35 unambiguous cases. The study childhood cases were obtained from the following sources: the Cutaneous Cancerology Group of the French Society of Dermatology, Paris, France (n=17); a federation of private pathology laboratories, the CRISAP (Centres de Regroupement Informatique et Statistiques en Anatomie Pathologique), Dijon, France (n=18); Institut Gustave-Roussy (Villejuif, France) (n=9); Centre Léon-Bérard (Lyon, France) (n=21); Hôpital Henri-Mondor (Creteil, France) (n=18); and University Hospital (Strasbourg, France) (n=2). The vast majority of these slides were referred to these participating centers for a review after the diagnosis of melanoma was made. Four of the experts reviewed some of the cases in their centers after the diagnosis of melanoma was made (J.W., 18 cases; A.S., 9 cases; C.B., 21 cases; and B.C., 2 cases).

To classify childhood case slides, the following clinical data were collected in each case: age at diagnosis, presence or absence of metastatic spread and/or death, and date of last follow-up if the patient was disease free. Metastatic melanoma was defined by nodal metastasis and/or distant metastasis. The slides were then classified by the 2 non-pathologists (S.B.G. and P.W.) into the following categories: category 1—slides from children with metastatic melanoma (n=20); category 2—slides from disease-free children with a follow-up of less than 5 years (n=47); and category 3—slides from disease-free children with a follow-up of 5 years or longer (n=18). A qualified person not involved in the study blinded these 120 slides, both childhood cases and unambiguous cases.

**REVIEW PROCEDURE**

The same set of 120 slides was sent to each pair of dermatopathologists (pairs A, B, C, and D) with 1 coding sheet per slide. There was no attempt to reach a consensus about the diagnosis of melanoma in childhood before the study. Our purpose was to follow the routine procedure for reading slides. Nevertheless, slides were sent to the experts without any clinical information. Each pair of experts was asked to classify each lesion as melanoma, nevus, or ambiguous tumors (ie, impossible to classify as benign or malignant). The 2 members of each pair had to agree on a diagnosis by joint analysis using a multiheaded microscope. If no agreement could be reached by the 2 members of a pair of experts, this disagreement had to be specifically notified.

**STATISTICAL ANALYSIS**

Unless a discrepancy between the 2 members of a pair occurred, the statistical unit of the analysis was the answer of each pair of experts. We first analyzed the diagnosis of experts on unambiguous adult case slides as controls by evaluating the 2 × 2 agreement between pairs of experts in paired comparisons. The agreement between the reviewers’ diagnoses and the initial diagnoses were also analyzed.

For the childhood case slides, we first analyzed the univariate distribution of the pairs of dermatopathologists’ diagnoses and tested, using the χ² test, whether the percentage of each diagnosis was different. We then evaluated the agreement between the pairs of experts in paired comparisons. The mean percentages of agreement were also compared among the 3 diagnosis categories by analysis of variance.

The agreement was evaluated by the percentage of concordance. The κ indexes and their 95% confidence intervals (when strength was sufficient) were calculated to evaluate the degree of agreement by taking into account the random concordance. All κ indexes have values approaching zero if agreement is due only to chance and can assume either positive or negative values. A value of 100% indicates perfect agreement. Concordance is considered excellent if the κ value is between 100% and 81%; good, between 80% and 61%; moderate, between 60% and 41%; weak, between 40% and 21%; poor, between 20% and 0%; and bad, lower than 0%.24

<table>
<thead>
<tr>
<th>Table 1 and Table 2</th>
<th>Table 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>from 15.3% to 35.3%</td>
<td>Table 3 summarizes the analysis of the agreement between pairs of experts in the 3 categories of slides. The concordance was weak to moderate in category 1 slides, weak in category 2 slides, and poor to moderate in category 3 slides. The mean κ index</td>
</tr>
</tbody>
</table>

©2002 American Medical Association. All rights reserved.
was slightly higher in cases with metastatic spread than in the other categories, but the difference was not significant.

The distribution of the diagnoses in each category of slides is summarized in Table 4. The experts diagnosed melanoma for category 1 slides more often than for slides from categories 2 and 3 (\(P = .04; \chi^2\) test).

### Table 1. Distribution of Dermatopathologist Expert Diagnoses on 85 Case Slides

<table>
<thead>
<tr>
<th>Pair of Experts</th>
<th>Diagnosis</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Melanoma</td>
<td>Nevi</td>
</tr>
<tr>
<td>A</td>
<td>39 (45.9)</td>
<td>16 (18.8)</td>
</tr>
<tr>
<td>B</td>
<td>46 (54.1)</td>
<td>17 (20.3)</td>
</tr>
<tr>
<td>C</td>
<td>50 (58.8)</td>
<td>22 (25.9)</td>
</tr>
<tr>
<td>D</td>
<td>43 (50.6)</td>
<td>17 (20.3)</td>
</tr>
</tbody>
</table>

*The distributions of diagnoses were significantly different \((P < .001; \chi^2 = 30.9)\). All data are number (percentage) of cases.

### Table 2. Degree of Agreement Between Dermatopathologist Expert Pairs’ Diagnoses

<table>
<thead>
<tr>
<th>Expert Pairs</th>
<th>Observed Agreement, %</th>
<th>(\kappa) Index (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A-B</td>
<td>54</td>
<td>26 (10-43)</td>
</tr>
<tr>
<td>A-C</td>
<td>61</td>
<td>38 (22-55)</td>
</tr>
<tr>
<td>A-D</td>
<td>66</td>
<td>46 (29-62)</td>
</tr>
<tr>
<td>B-C</td>
<td>62</td>
<td>36 (20-55)</td>
</tr>
<tr>
<td>B-D</td>
<td>62</td>
<td>38 (21-55)</td>
</tr>
<tr>
<td>C-D</td>
<td>62</td>
<td>38 (21-55)</td>
</tr>
</tbody>
</table>

*The distributions of dermatopathologists’ diagnoses were statistically different \((\chi^2 = 30.9; P < .001)\). CI indicates confidence interval.

### Table 3. Degree of Agreement Between Pairs of Experts According to the Different Categories of Case Slides

<table>
<thead>
<tr>
<th>Pairing of Expert Pairs</th>
<th>Category 1† (n = 20)</th>
<th>Category 2‡ (n = 47)</th>
<th>Category 3§ (n = 18)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Observed Agreement</td>
<td>(\kappa) Index</td>
<td>Observed Agreement</td>
</tr>
<tr>
<td>A-B</td>
<td>70</td>
<td>28</td>
<td>51</td>
</tr>
<tr>
<td>A-C</td>
<td>90</td>
<td>56</td>
<td>58</td>
</tr>
<tr>
<td>A-D</td>
<td>85</td>
<td>48</td>
<td>58</td>
</tr>
<tr>
<td>B-C</td>
<td>70</td>
<td>25</td>
<td>57</td>
</tr>
<tr>
<td>B-D</td>
<td>75</td>
<td>43</td>
<td>60</td>
</tr>
<tr>
<td>C-D</td>
<td>80</td>
<td>26</td>
<td>60</td>
</tr>
<tr>
<td>Mean ± SD (\kappa) index</td>
<td>. . .</td>
<td>37.7 ± 13.1</td>
<td>. . .</td>
</tr>
</tbody>
</table>

*All data, including \(\kappa\) indexes, are percentages. Ellipses indicate not applicable.
†Patients with metastatic carcinoma.
‡Donor with metastatic carcinoma.
§Disease-free patients with follow-up of less than 5 years.
¶Disease-free patients with follow-up of 5 years or longer.
||No difference was evidenced between the group means \((F = 0.49; P = .62)\).
tion confirms, the challenge of making the diagnosis of melanoma in childhood; it may not be easily diagnosed even by a panel of expert pathologists. Our study confirms the poor reliability of this diagnosis. The concordance among pairs of experts was moderate to weak in the group of documented metastatic tumors even though the diagnosis of melanoma was made more frequently in this category. The concordance among experts was even lower in the other cases.

This is the first study to evaluate the interrater reliability of the histopathologic diagnosis of childhood melanomas. To improve diagnosis reliability, we need to identify those morphologic criteria that will allow us to separate the metastatic cases from the benign or ambiguous tumors. Diagnostic criteria should be documented.27 Nevertheless, it may not be possible to develop definitive diagnostic criteria because there are many overlapping features with benign nevi and Spitz nevi, even in adult melanoma. Unless clearly delineated criteria are identified, progress is unlikely.

Accepted for publication January 22, 2002.

This study was supported by a grant from Vaincre le Mélanome (SANOFI), Paris, France (Dr Wolkenstein).

This study was presented in part at the Journées Dermatologiques de Paris, Paris, December 1-4, 1999.

We would like to thank L. Thomas, MD, Service de Dermatologie, Hôtel-Dieu, Lyon; M. Delaunay, MD, *Unité de Dermatologie-Cancérologie, Hôpital Saint André, Bordeaux; P. Bioulac-Sage, MD, Département d’Anatomie Pathologique, Hôpital Pellegrin, Bordeaux; B. Guillot, MD, *Service de Dermatologie, Hôpital Saint-Eloi, Montpellier; C. Labrèze, MD, Dermatologie Pédiatrique, Hôpital Pellegrin, Bordeaux; F. Truchetet, MD, *Hôpital Beauregard, Thionville; F. Boitier, MD, *Service de Dermatologie, Hôpital Henri-Mondor, Créteil; P. Joly, MD, *Service de Dermatologie, Hôpital Charles Nicolle, Rouen; C. Lok, MD, *Service de Dermatologie, Hôpital Sud, Amiens; B. Dreno, MD, *Clinique Dermatologique, Hôtel-Dieu, Nantes; G. Lorette, MD, *Service de Dermatologie, CHU Trouseau, Tours; J. J. Grob, MD, *Service de Dermatologie, Hôpital Sainte-Marguerite, Marseille; and la Fédération des CRISAP. (Asterisk indicates that the person is a member of the French Cutaneous Cancerology Group).

Corresponding author: Pierre Wolkenstein, MD, PhD, Department of Dermatology, Henri-Mondor Hospital, F-94010 Créteil CEDEX, France (e-mail: pierre.wolkenstein@hmn.ap-hop-paris.fr).

REFERENCES