Malignant mesothelioma of the pleura: interobserver variability

A Andrion, C Magnani, P G Betta, A Donna, F Mollo, M Scelsi, P Bernardi, M Botta, B Terracini

Abstract

Aims—To assess the consistency of histopathological diagnosis of pleural malignant mesothelioma by estimating interobserver agreement between five pathologists.

Methods—Eighty eight histological sets pertaining to original diagnoses of pleural malignant mesothelioma were reviewed separately by each pathologist. Diagnostic likelihood was graded as definite (A), probable (B), possible (C), improbable (D), and definitely not (E) malignant mesothelioma. The following indexes were estimated: observed proportion of agreement (Po), k statistics and proportion of agreement for “positive” (Ppos) and “negative” (Pneg) diagnoses.

Results—Sixty cases (68-2%) were rated by at least three reviewers as A or B and by none of the others as D or E. Five (5-7%) were rated by at least two reviewers as D or E and by none of the others as A or B. Nine (10-2%) showed a serious disagreement, rated A or B and D or E. Agreement for sets obtained at necropsy/surgery (median kw = 0-57) was similar to that for endoscopic material (median kw = 0-54). Agreement was poor on material obtained by needle biopsy (median kw = 0-21). The median value of Ppos varied between 0-94 (necropsy/surgery) and 0-67 (needle biopsy) and that of Pneg between 0-78 (necropsy/surgery) and 0-34 (unspecified biopsy). Interobserver agreement on histological typing was good overall (median k = 0-59).

Conclusions—Of the original histological diagnoses, 70% were consistently reproduced through panel review. Most indexes of agreement between pathologists ranged from poor (needle biopsy) to moderate (necropsy/surgery). Agreement in confirming malignant mesothelioma was greater than that regarding exclusion of this diagnosis. Of the cases finally considered to have malignant mesothelioma, the reproducibility of histological typing was relatively satisfactory.

(J Clin Pathol 1995;48:856–860)

Keywords: Malignant mesothelioma, pathology panel, interobserver variability.

Assessing the consistency of the histopathological diagnosis of cancer is crucial for validation of clinico-epidemiological studies. Malignant mesothelioma is one of the most difficult tumours to diagnose reliably. There are several reasons for this: it is rare, so that large series are difficult to assemble; both the pleura and peritoneum are frequent sites of metastasis from an unknown primary; and the variety of cellular and architectural forms is almost unique among solid neoplasms. Panels (either formal or informal) have been established for many years in several countries to review pathological material from subjects included in epidemiological studies and to assist individual pathologists to reach their diagnosis. Members of these panels were pioneers in describing observer variation in histopathological reporting. However, only two studies quantified diagnostic reproducibility using appropriate statistical techniques.

Until 1985 an asbestos cement producing plant had been active in the district of Casale Monferrato (population 98 000), northwest Italy. Between 1980 and 1989, an average of nine cases of malignant mesothelioma were diagnosed histologically per annum in this district—that is, 10 times as many as expected in industrialised areas. Only one in three of these individuals had been working at the plant. A hypothesis has been put forward that a number of the other cases may have been caused following exposure to asbestos released from the plant and present in the atmosphere. The size of the series and the novelty of the aetiological hypothesis prompted us to evaluate the reliability of the routine histological diagnosis of malignant mesothelioma and the agreement between a number of expert pathologists.

Methods

A survey of the incidence of malignant mesothelioma in the Casale Monferrato district between 1980 and 1989 revealed that 89 individuals has been diagnosed with malignant mesothelioma of the pleura on histology. Six additional cases diagnosed during the same period and identified in an update of the survey were not included in the present study. Vital status and date of death were obtained from the registry office of the town where the subjects lived at diagnosis or moved thereafter. All subjects but one were successfully traced. Survival periods have been computed, according to the method of Kaplan–Meier, from the date of diagnosis to death, and have been verified up to January 1994.
Malignant mesothelioma of the pleura: interobserver variability

PATHOLOGICAL MATERIAL
For 83 of the 89 patients, the original histopathological material was traced to the seven hospitals (including three teaching hospitals) whose catchment areas are within the Local Health Authority of Casale Monferrato. Of these patients, 30 were men and 33 were women with a mean age of 63.7 (+10.6) years. For five patients, there were two distinct sets of diagnostic slides from two of the hospitals. Thus, the study included 88 sets of slides pertaining to separate diagnoses of malignant mesothelioma which were regarded as individual cases. The original histological material comprised haematoxylin and cosin stained slides. Additional histo- and immunohistochemically stained slides were available for a small number of cases and were not considered further. The source of tissue specimens included necropsy, surgery, endoscopic biopsy, transparietal needle biopsy, and unspecified biopsy for, respectively, 14, 12, 38, nine, and 15 cases. A previous biopsy had been performed during life in only one of the necropsy cases. The histological type at original diagnosis was specified for 39 cases: 16 epithelial, 18 mixed and five fibrosarcomatous.

REVIEWING PROCEDURE
Five pathologists (AA, P-GB, AD, FM, and MS) reviewed the original slides. Experience in pathology practice ranged from 18 to 38 years. One of the authors (CM) coordinated the panel.

In five instances slides were not available and new slides were prepared from the original paraffin wax blocks. Subsequently, all original labels were removed and a new identification number was randomly assigned to each case. As most of the original diagnoses were based on haematoxylin and cosin stained slides, only this stain was submitted to the panel. The histological slides were reviewed by each member of the panel independently. All steps of the study were carried out blindly and the reviewers were unaware of the original pathological and clinical diagnoses, the source of the material, the history of occupational exposure, sex, age, hospital, and diagnoses by other panel members. The panel did not discuss the slides nor were there any consensus meetings. Nevertheless, in the second phase, 29 cases for which there were huge discrepancies were presented to each panel member for a second time along with the anonymous opinions previously expressed by all members. For these cases, only the conclusion of the second review was considered. For each case, each pathologist compiled an entry sheet preliminarily approved by all panellists.

HISTOPATHOLOGICAL DIAGNOSTIC CRITERIA
The reviewers agreed to use widely accepted criteria. In addition, the diagnostic likelihood was graded according to the following categories: A, definite malignant mesothelioma (no doubt as to the histopathological diagnosis); B, probable malignant mesothelioma (reasons for hesitation include lack of differentiation, absence of certain histological criteria, poor quality, etc.); C, possible malignant mesothelioma (the diagnosis cannot be confidently ruled out, but there is insufficient evidence to come to a positive conclusion); D, improbable malignant mesothelioma (probably not a mesothelioma, but the diagnosis cannot be confidently excluded); E, definitely not a malignant mesothelioma; and F, unsatisfactory material, no valid evidence for a diagnosis. For categories A and B, observers were requested to suggest one of the three more frequent morphological forms of malignant mesothelioma—that is, epithelial, mixed and fibrosarcomatous. For categories D and E, the reviewers were requested to indicate an alternative diagnosis. The latter have not been considered further in the present analysis.

STATISTICAL ANALYSIS
Interobserver agreement was evaluated using the following statistical indexes: Po (observed proportion of agreement), \( \kappa \) (observed proportion of agreement corrected for agreement expected by chance) and proportion of agreement for “positive” and “negative” diagnoses (Ppos and Pneg), separately. The \( \kappa \) statistic is a widely used technique for the analysis of agreement when no gold standard is available. It is computed as the ratio between Po and the agreement expected by chance. \( \kappa \) can also be used when observations are grouped according to three or more categories and can be weighted as suggested by Cicchetti. As the extent of disagreement depends on the “distance” between categories, we assumed that the B rating (probable malignant mesothelioma) is closer to A (definite malignant mesothelioma) than to C (possible malignant mesothelioma). A and B categories were regarded as “true” malignant mesothelioma. Similarly, the D (improbable malignant mesothelioma) and E (definitely not malignant mesothelioma) ratings were grouped together. Compared with the weights suggested by Cicchetti, we reduced the difference in weights between categories A and B slightly, and increased the difference between B and C (table 1). Unweighted \( \kappa \) was used for measuring agreement in histological typing. The \( \kappa \) index is influenced not only by the agreement among observers, but also by the prevalence of the disease and by the imbalance of the marginal totals. Its value ranges between +1 and −1, where +1 represents complete agreement, 0 chance agreement and −1 complete

<table>
<thead>
<tr>
<th>Category</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D or E</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1.0</td>
<td>0.9</td>
<td>0.3</td>
<td>0.0</td>
</tr>
<tr>
<td>B</td>
<td>0.9</td>
<td>1.0</td>
<td>0.8</td>
<td>0.2</td>
</tr>
<tr>
<td>C</td>
<td>0.3</td>
<td>0.8</td>
<td>1.0</td>
<td>0.5</td>
</tr>
<tr>
<td>D or E</td>
<td>0.0</td>
<td>0.2</td>
<td>0.6</td>
<td>1.0</td>
</tr>
</tbody>
</table>
disagreement. In fact, the range of variation of both weighted and unweighted $k$ indices is narrower when the marginal totals are imbalanced; this occurs for groups of cases with the same kind of diagnosis such as in the present study. As a summary measure of agreement, we considered, for the 10 possible combinations for two reviewers out of the pool of five, the median and the range of the following indexes: the observed proportion of agreement ($P_o$), weighted $k$ ($kw$) and observed proportions of positive ($Ppos$) and negative ($Pneg$) agreements. In all calculations comparisons rated as inadequate ($F$) by one of the two reviewers forming each pair were excluded.

## Results

The distribution of cases in each category of diagnostic likelihood, according to each of the five reviewers, is presented in table 2. The proportion of cases rated as definite (A) or probable (B) malignant mesothelioma ranged from 60-3 to 77-3%, while the proportion rated as improbable (D) or definitely not (E) malignant mesothelioma ranged from 2-3 to 12-5%. The intermediate category (C) ranged from 11-4 to 23-9%, and the proportion of inadequate material (F) varied between 3-4 and 11-4%.

The number of cases, categorised according to the number of ratings, confirming (A or B) or excluding (D or E) a diagnosis of malignant mesothelioma is presented in table 3. Sixty (68-2%) cases were rated by at least three reviewers as A or B and by none of the others as D or E. Five (5-7%) were rated by at least two reviewers as D or E and by none of the others as A or B. Nine (10-2%) showed a serious disagreement, rated A or B and D or E. The remaining 14 (15-9%) cases were rated by the majority as C or F.

All subjects but one died within five years of the original histological diagnosis. Survival times ranged from one to 64 months. When the necropsy cases were excluded, 49 cases were rated by most of the reviewers as A or B. Seven of these were alive two years after original diagnosis; this was the case for two of 21 subjects (necropsy cases excluded) for which there was some disagreement.

Table 4 summarises agreement between reviewers according to the source of pathological material. Agreement on slides obtained at necropsy/surgery (median $kw=0.57$) was similar to that for endoscopical material (median $kw=0.54$), but was poor for needle biopsy specimens (median $kw=0.21$). Similar findings were observed for $Pneg$, which was highest for sets obtained at necropsy or surgery. Compared with $Ppos$, absolute $Pneg$ was lower and the median ranged between 0.78 (necropsy/surgery) and 0.34 (unspecified biopsy).

As mentioned earlier, overall and pairwise calculations of agreement by histological type were based only on cases rated by the reviewers as A or B; absolute numbers are given in table 5. The proportion of cases categorised as epithelial ranged between 54-2 and 69-1%, as mixed between 20-3 and 39-6%, and as fibrosarcomatous between 3-4 and 10-9%. A diagnosis of unspecified histological type largely originated from one observer. The median and range of the indexes of agreement for histological type computed for all possible pairwise combinations are given in table 6. The relative frequency of each histological type was also computed (data not shown). Material obtained at necropsy or surgery was categorised as mixed or fibrosarcomatous (41% mixed and 15% fibrosarcomatous) more often than that obtained on biopsy (22% and 5%, respectively).
Malignant mesothelioma of the pleura: interobserver variability

Table 6  Median values of indexes of agreement for histological type computed for all possible pairwise combinations for a definitive diagnosis (A or B) of malignant mesothelioma with respect to the source of the pathological material

<table>
<thead>
<tr>
<th>Source of material</th>
<th>Po</th>
<th>k (range)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>0.80</td>
<td>0.59 (0.51-0.77)</td>
</tr>
<tr>
<td>Necropsy + surgery</td>
<td>0.74</td>
<td>0.59 (0.40-0.88)</td>
</tr>
<tr>
<td>Biopsy endoscopic</td>
<td>0.88</td>
<td>0.66 (0.67-0.97)</td>
</tr>
<tr>
<td>Needle + unspecified</td>
<td>0.86</td>
<td>0.66 (0.66-0.90)</td>
</tr>
</tbody>
</table>

Discussion

Earlier investigations of observer variability in the histopathological diagnosis of malignant mesothelioma suggested that there is a large subjective component.5 In recent years confidence in the histopathological diagnosis of malignant mesothelioma has grown.6 7 8 9 10 These diagnoses, however, included heterogeneous cases with a putative diagnosis of malignant mesothelioma submitted to a panel for histopathological consultation,11 12 while others included malignant mesotheliomas diagnosed as such by the original pathologist.13 14 Moreover, determining the contribution of each diagnostic approach (routine histology, histochemistry, immunohistochemistry, electron microscopy, etc.) to the final diagnosis is difficult.

In the present study 60-3 to 77-3% of the original diagnoses were classified as definite or probable malignant mesothelioma by the reviewers (table 2). Validation of positive cases was based on satisfactory agreement (60/88 (68-2%) in this study). Previous comparable estimates of the proportion of diagnoses of malignant mesothelioma validated by panels (six and three pathologists, respectively) were 59/119 (49-6%) cases in Canadian study4 and 44/67 (65-7%) cases in a Dutch study.7 The proportion of unconfirmed diagnoses (possible malignant mesothelioma) was also similar, ranging from 13-4 to 15-9%. Two other studies estimated agreement in the histological diagnosis of malignant mesothelioma. A panel of three pathologists6 reviewed slides from 22 cases of childhood mesothelioma and confirmed 10 as true malignant mesothelioma. In a study from Italy22 a panel of three pathologists confirmed 91 (87-5%) of 104 original diagnoses of malignant mesothelioma.

Overall, agreement for a diagnosis of malignant mesothelioma seems to be less satisfactory than that estimated in similar studies (the design of which, admittedly, was somewhat heterogeneous) of other cancers, such as those of the breast, prostate, bowel, and lung.23-25 As it has been suggested that a comprehensive clinico-pathological study is the best approach to diagnosing malignant mesothelioma,3 19 this lack of agreement may be related just to the process used to review the histopathological material. However, in their series McDonald et al4 demonstrated that the availability of clinical information and an account of the gross pathology only raised the proportion of positive diagnoses slightly compared with histology alone (53-9% v 49-6%). In our study all patients but one died within five years of the initial diagnosis. When necropsy cases were excluded, only 15% of patients for whom the diagnosis of malignant mesothelioma was confirmed (categories A and B) were alive two years after the original diagnosis. This is compatible with survival rates reported for malignant mesothelioma.26 27 Nevertheless, low survival rates (9% at two years) were also found for those patients with a disputed diagnosis.

A postmortem examination has been regarded as essential for a definite diagnosis of malignant mesothelioma. However, in the present study, of the 14 necropsy cases originally designated malignant mesothelioma, only 10 were confirmed by the reviewers. Necropsy, therefore, does not provide a conclusive diagnosis.28 Some reports mention ancillary techniques (histochemistry, immunohistochemistry, electron microscopy) for final assessment, but do not give figures for their relative contribution to diagnostic validation.29 30 Recent reports suggest that several immunohistochemical markers may enhance diagnostic accuracy because of their greater frequency in adenocarcinoma of the lung than in malignant mesothelioma.30 Preliminary results on the cases presented here indicate that the systematic use of immunohistochemistry may increase diagnostic consistency,30 confirming the results of a Danish study.31

The data were corrected for agreement expected by chance.11 10 Agreement, irrespective of the source of the material, was reasonably good (median kw = 0-46); it was slightly more satisfactory for samples taken at autopsy/surgery (median kw = 0-58) and endoscopy (median kw = 0-57). However, agreement was very poor for samples obtained via needle biopsy or from an unspecified source (median kw = 0-21 and 0-30, respectively). Only two previous studies of diagnostic agreement for malignant mesothelioma used more sophisticated statistical techniques than those used in the present study: the first showed that one in three observers pairs had an unweighted k value as high as 0-40; in the other, which had a different design,4 interobserver agreement was poor (k values were between 0-23 and 0-39).32

In the present series, in analyses broken down by pairs of reviewers, there was only one instance of good agreement (kw = 0-69). For all remaining pairs of reviewers, values indicative of fair to moderate agreement were obtained.

The epithelial form was the most frequently diagnosed histological type, followed by the mixed and fibrosarcomatous forms. This was not the case in other series, in which the most frequently diagnosed histological type varied widely.10 Interobserver agreement was good for confirmed cases (categories A and B) of malignant mesothelioma, irrespective of the source of the sample (median k values ranged from 0-50 to 0-64).

In conclusion, of the original histological diagnoses, 70% were consistently reproduced through panel review. Most indexes of agreement between pathologists ranged from poor (needle biopsy) to moderate (necropsy/surgery). Agreement in confirming malignant mesothelioma was greater than that regarding exclusion of this diagnosis. Of the cases finally...
reproducibility of histological typing was relatively satisfactory.

The authors thank Drs G P Borgo, G Zaidan and D Bellingeri for their valuable assistance. This study was supported by a grant from the Associazione Italiana per la Ricerca sul Cancro (AIRC) and by the EEC Europe Against Cancer Programme.