Agreement and disagreement between pathologists in histological diagnosis

H. A. SISSONS

M.D., F.R.C.P., F.R.C.S., F.R.C.Path.

Institute of Orthopaedics, Royal National Orthopaedic Hospital, London

Summary

Studies on differences of observation and interpretation between pathologists in connection with histological diagnosis are rare. The present paper gives a general account of such a study, by members of the Cancer Research Campaign's Bone Tumour Panel, of certain histological features of 'malignant round-cell tumours' of bone. Significant differences were found, not only between the different cases studied, but between the observations of the different pathologists, and between the repeated observations of individual pathologists.

Introduction

That different pathologists sometimes reach different conclusions when examining the same histological sections is well known. When we ask for a 'second opinion' on a section which appears to pose some difficulty, we assume that factors such as knowledge and experience may, either wholly or in part, be responsible. Our successes and failures in prognosis are, in some measure, a check on this assumption. It is interesting, however, that almost no serious work has been carried out on the factors responsible for differences of observation and interpretation between pathologists in connection with histological diagnosis. Among the very few published studies is that of Cocker, Fox and Langley (1968) on the consistency of histological diagnosis of epithelial abnormalities of the cervix uteri, that of Iversen and Sandnes (1971) on the reliability of pathologists in the diagnosis of lymph node biopsies, and that of Lambourne and Lederer (1973) on observer variation in cytological screening for cervical carcinoma.

The difficulties of an investigation of this type are fairly obvious. In the first place, the area of investigation must be defined, and the questions must be chosen with care. The question 'Is the section liver or kidney?' is one where unanimity of opinion might be expected among experienced observers, while 'Is the section a benign or a malignant lymphoid tumour?' might be expected to produce more varied comments. Agreement and disagreement must be defined. Then the observations, preferably repeated observations, must be made under controlled conditions. Finally, appropriate statistical methods must

be employed to enable the correct conclusions to be drawn from the observations.

The purpose of the present paper is to suggest that there is considerable scope for investigation of 'observer variation'—this phrase is used in preference to 'observer error'—in histological diagnosis, and to illustrate this by reference to a study of one particular aspect of bone tumour diagnosis which has recently been carried out by members of the Cancer Research Campaign's Bone Tumour Panel. Dr John Ball, the present writer, and Mr Laurence Freedman are at present preparing a more detailed account for publication.

Malignant round-cell tumours of bone

The point of commencement of the study was the review, by the Bone Tumour Panel, of the cases in their accumulated material where a diagnosis of 'malignant round-cell tumour' had been made. There had been a careful attempt to exclude cases of myeloma or myelomatosis, lymphoma, carcinoma and metastatic neuroblastoma. The basic problem before the Panel was to decide which of the 'malignant round-cell tumours' were primary in bone and which metastatic, in view of the long-standing suggestion (Colville and Willis, 1933; Willis, 1940) that some of them could be metastases from clinically latent neuroblastomas. This still remains, to some extent, an unsolved problem, and the analysis of the present material has not directly helped to solve it. For the purpose of the present investigations, however, our aim has been to study the histological characteristics of the 'malignant round-cell tumours', to see whether they are a homogeneous group or a collection of different histological entities. We particularly had in mind the conventional histological image of 'Ewing's tumour', made up of uniform cells with indistinct cytoplasmic outlines, large pale nuclei, and conspicuous nucleoli, at least as seen in routine paraffin sections stained with haematoxylin and eosin (Fig. 1). We wanted to find out whether tumours of this type existed as a distinct group, and to see how reliable our attempts were to identify them and to separate them from other tumours of the round-cell group (Fig. 2).

Details of study

From the literature, and from the experience of the

FIG. 1. Round-cell tumour made up of uniform cells with indistinct cytoplasmic outlines and large pale nuclei. Paraffin section stained with HE ($\times 390$).

FIG. 2. Another round-cell tumour, made up of cells some of which have more distinct cytoplasmic outlines and darker nuclei which are rather variable in size and shape. Paraffin section stained with HE ($\times 390$).

members of the Bone Tumour Panel, we drew up a list of histological features which we felt might possibly be of diagnostic significance in the context of the present study. These were: cell characteristics, mitoses, intracellular glycogen, star cells, rosettes,

acini, other cell aggregates, giant cells (binucleate), giant cells (multinucleate), ganglion cells, nuclear staining, nuclear pleomorphism, conspicuous eosinophil nucleoli, nuclear size, reticulin pattern, necrosis, fibroblastic septae and calcification.

		Readers					
		1	2	3	4	5	
Section 1	Obs. 1	S_{11}	S_{21}	S_{31}	S_{41}	S_{51}	O_1
	Obs. 2	S_{12}	S_{22}	S_{32}	S_{42}	S_{52}	O_2
Section 2	Obs. 1	S_{13}	S_{23}	S_{33}	S_{43}	S_{53}	O_3
	Obs. 2	S_{14}	S_{24}	S_{34}	S_{44}	S_{54}	O_4
		R_1	R_2	R_3	R_4	R_5	T

FIG. 3. Form of tabulation of results for one histological feature as recorded for one histological block. S_{11} – S_{54} are the individual observations of the five readers, each making two observations on each of two sections. The totals O_1 – O_4 represent the different occasions, the totals R_1 – R_5 represent the different readers, and the overall total T represents the total experience with this particular block.

A preliminary investigation led us to conclude that some questions were unsuitable, in the form in which they were presented, for the cases studied or for the pathologists attempting to use them. Neither star cells nor rosettes, for example, were to be found in the cases under investigation: evaluation of nuclear diameters and mitotic counts led to difficulties which, perhaps regrettably, led to their exclusion from the investigation. From the preliminary investigation, our interest became directed to a limited number of histological features which we regarded as biologically interesting and technically feasible. These were: cell characteristics (syncytial or separate), nuclear staining (dark or pale), nuclear pleomorphism (slight, moderate or marked), conspicuous eosinophil nucleoli (present or absent), reticulin pattern (? around individual cells) and intracellular glycogen (present or absent).

Sections from forty cases of 'malignant round-cell tumour' of bone were prepared and stained with Ehrlich's haematoxylin and eosin, Gordon and Sweet's method for reticulin, and with the PAS technique for glycogen using a diastase-treated section as control. In an attempt to assess the importance of differences between observers, as well as differences between the studied cases, and also differences due to variation ('error') in the observations of individual observers, five members of the Bone Tumour Panel examined the sections from each of the forty cases on four separate occasions, and recorded their findings for each of the listed features. The finding for each feature could be recorded as positive, negative or indeterminate, except for nuclear pleomorphism where a three-grade assessment (+, ++, +++) was used. No attempt was made to provide guidance as to what constituted a positive or a negative finding for any feature, as the investigation was intended to test the assumption that this was common knowledge to people who might be regarded as experts in this

field of pathology. It was also possible to study differences between tissue samples as in some cases more than one histological block was available. Differences between histological sections from the same tissue sample were also studied, except for questions relating to glycogen and reticulin, by preparing two haematoxylin and eosin sections from each paraffin block: these were stained at the same time and under the same conditions.

When the observations were collected, a numerical value (1, 2 or 3) was allotted to each observation. Figure 3 shows the form of tabulation adopted. For each histological block there were twenty observations, in this case the results of the five 'readers' each making two observations each of two sections. The totals O_1 – O_4 represent the different occasions, the totals R_1 – R_5 represent the different readers, and the overall total T represents the total experience with this particular block.

Results

The results, of course, varied from feature to feature, and it is not possible to discuss them in detail in the present paper. Simple inspection of the tabulated figures showed, for some blocks and some features, complete agreement between all readers on all occasions, while for others there were surprising differences both between the observations of different readers and between the repeated observations of a single reader. It was evident that without further clarification of the problem, the examination of a section, on one occasion, by one pathologist, was not a reliable way to answer the particular questions in which we were interested.

We next proceeded to study the results in more detail, using the method of analysis of variance, to determine the significance and the magnitude of differences between patients, between blocks from the same patient, between readers, and due to interactions between these factors. The results for

TABLE 1. Results for reticulin pattern (negative = 1, indeterminate = 2, positive = 3). The significance and magnitude of differences due to each of the factors studied is shown. P, B and R represent differences due to patients, blocks and readers. O, the residual variation, represents differences between the observations of individual readers

Component	Degrees of freedom	Mean square	Variance ratio	Significance	Magnitude of effect
P	30	20.8994	17.65	0.001	0.664
B	15	1.1839	3.64	0.001	0.043
R	4	2.7542	8.46	0.001	0.013
P × R	120	0.6121	1.88	0.01	0.048
B × R	60	0.3256	1.75	0.001	0.035
O	690	0.1859	—	—	0.186
Total	919	—	—	—	0.989

TABLE 2. Results for cell characteristics (separate = 1, indeterminate = 2, syncytial = 3). The significance and magnitude of differences due to each of the factors studied is shown. P, B, R and S represent differences due to patients, blocks, readers and sections. O, the residual variation, represents differences between the observations of individual readers

Component	Degrees of freedom	Mean square	Variance ratio	Significance	Magnitude of effect
P	30	11.9635	6.50	0.001	0.341
B	15	1.8395	2.41	0.01	0.065
R	4	14.3523	31.95	0.001	0.076
P × R	120	0.9470	2.11	0.025	0.084
B × R	60	0.4492	1.94	0.001	0.054
S	46	0.3130	1.35	N.S.	0.008
S × R	184	0.2315	1.33	N.S.	0.029
O	460	0.1739	—	—	0.174
Total	919	—	—	—	0.831

reticulin, where relatively consistent histological observations were made, and for cell characteristics, where less consistent findings were obtained, are shown in Tables 1 and 2.

For reticulin pattern (Table 1), where we attempted to record the presence or absence of a network of reticulin fibres between individual tumour cells or small groups of tumour cells, the difference between patients, blocks and readers are all highly significant. This means that the cases studied really differ from one another in this respect, but that real differences between different samples of tissue from the same case, and between the observations of the different pathologists, also exist. When the magnitude of the effects of these factors is examined, it can be seen that differences between patients is the largest, and is responsible for almost 70% of the total variation. The residual variation, due to differences between the observations of individual readers, is the next largest and is responsible for approximately 20% of the total variation.

For cell characteristics (Table 2), where we attempted to record whether the cells had a separate or a syncytial arrangement, and where we investigated differences between sections from the same histological block, the differences between patients,

blocks and readers are again all significant. For this feature, differences between sections were not significant. When the magnitude of the effects of these factors is examined, it can be seen that although differences between patients is still the largest factor, it is responsible for only about 40% of the total variation.

This type of analysis allowed us to assess the objectivity of our observations in a way that had not hitherto been possible. For the other features, too, we found significant differences between the cases studied, but it was salutary to see how much of the variation was due to other factors. In this respect, the results of our study are rather similar to what has been found in investigations of differences between examiners in marking various types of paper, including those of the final examination in medicine (Hartog and Rhodes, 1936; Bull, 1956). These studies have vividly brought to attention the importance of variation due to differences between examiners, as opposed to differences between examinees, and have encouraged the increasing use of 'multiple-choice' examination questions, where subjective factors in assessment can be eliminated. In our own study, the reasons for the variation in our observations are at present being investigated.

Some of them lie with the observers, others with the histological material, where the presence of contrasting types of tissue in the same tumour appears to be important.

General conclusions

This is not the place to present our detailed results, but it is appropriate to comment briefly on the type of conclusion it is possible to reach with regard to the group of round-cell tumours of bone with which we were concerned.

For each histological feature, we obtained a series of totals (T in Fig. 3), each representing the overall view of the observers with regard to one particular case. We sought to find out whether the results for the different features were related, and we did this by calculating correlation coefficients for the totals for the various pairs of features. We found that there was a high degree of positive correlation between the results for cell characteristics and intracellular glycogen, and a high degree of negative correlation between these two results and those for nuclear staining, nuclear pleomorphism and reticulin pattern (i.e. an association between a syncytial pattern of cell arrangement, the presence of intracellular glycogen, pale nuclear staining, little nuclear pleomorphism, and the absence of an intercellular network of reticulin). When a combined score is obtained by taking into account the results of all these questions, cases at one end of the scale show the associated features itemized above, while those at the other end of the scale show the reverse features. The distribution of scores shows some degree of separation into two groups, although several cases

in the middle do not readily fall into either of these. It is tempting to regard the first of the groups as having the histological features usually associated with Ewing's tumour, and the second group as having those usually associated with reticulosarcoma, but this is going rather beyond the scope of the present investigation.

Our results, however, do indicate a real diversity of histological characteristics in the group of cases studied, and a real association of certain histological features in certain cases. By means of an assessment of 'observer variation', the study has been able to achieve a greater degree of objectivity than could be provided by the uncontrolled observations of a single observer.

References

- BULL, G.M. (1956) An examination of the final examination in medicine. *Lancet*, **ii**, 368.
- COCKER, J., FOX, H. & LANGLEY, F.A. (1968) Consistency in the histological diagnosis of epithelial abnormalities of the cervix uteri. *Journal of Clinical Pathology*, **21**, 67.
- COLVILLE, H.C. & WILLIS, R.A. (1933) Neuroblastoma metastasis in bones, with criticism of Ewing's endothelioma. *American Journal of Pathology*, **9**, 421.
- HARTOG, P. & RHODES, E.C. (1936) *The Marks of Examiners*. Macmillan. London.
- IVERSEN, O.H. & SANDNES, K. (1971) The reliability of pathologists. A study of some cases of lymph node biopsies showing giant follicular hyperplasia or lymphoma. *Acta pathologica microbiologica scandinavica*, **79**, 330.
- LAMBOURNE, A. & LEDERER, H. (1973) Effects of observer variation in population screening for cervical carcinoma. *Journal of Clinical Pathology*, **26**, 564.
- WILLIS, R.A. (1940) Metastatic neuroblastoma in bone presenting the Ewing's syndrome, with a discussion of 'Ewing's Sarcoma'. *American Journal of Pathology*, **16**, 317.