Observer error in histological assessment of marrow hypocellularity

A. MORLEY AND J. BLAKE

From the Department of Medicine, University of Adelaide, Adelaide, Australia

SYNOPSIS Observer error in the assessment of marrow hypocellularity was studied by examining histological sections from 37 mice having normal or reduced marrow cellularity and comparing the estimated cellularity with a previously measured value. For marrows of mild or moderately reduced cellularity the subjective estimate of cellularity was almost valueless both in terms of accuracy and reproducibility. Only severely hypocellular marrows, having a cellularity of less than 17% of normal, could be recognized as being abnormal with some degree of accuracy and reproducibility.

Assessment of overall cellularity is an integral part of examination of the bone marrow in man and is important in the diagnosis of marrow hypoplasia or aplasia. This assessment is usually done either by examination of histological sections prepared from marrow particles obtained at aspiration or by examination of histological sections of a core of bone obtained by needle biopsy or trephine. One recognized source of error in this assessment is sampling error due to the fact that only a minute proportion of marrow is examined. Another, which has received less attention, is the observer error involved in the actual assessment of the degree of cellularity of the specimen being examined. To our knowledge, this latter factor has never been systematically studied. The impression amongst haematologists is that assessment of reduced marrow cellularity is inaccurate and that only moderate or severe degrees of hypoplasia can be recognized. But this impression has never been substantiated since the methodological difficulties are great. Absolute measurement of the total marrow cellularity in man is a cumbersome research procedure (reviewed by Harrison, 1962) and for this reason a study in man comparing a subjective histological assessment to a measured absolute cellularity presents great difficulties. A means of studying the problem, at least for murine bone marrow, became available during the course of an investigation of experimental aplastic anaemia (Morley and Blake, 1974). In this investigation a large number of mice with various degrees of marrow hypo- or aplasia were studied and for many mice a section of femoral marrow and an absolute measure of marrow cellularity were obtained. Thus it became possible to compare a subjective assessment of marrow cellularity with a true assessment and to determine the accuracy and reproducibility of the former.

Methods

The mice studied had had various degrees of marrow hypo- or aplasia produced by previous treatment with four doses of busulphan. They were not sacrificed until at least two months had elapsed since their last dose of busulphan. When sacrificed, the ends of the femur were cut off and the bone was placed in formalin. Histological sections were cut at a thickness of 6 μ. Total tibial cell counts were performed by cutting off the ends of the bone, expressing the cells into saline, suspending them, and counting the resulting suspension. In a number of mice independent counts were performed on the two tibiae in order to assess the degree of variation in cellularity of the marrow between two different sites. Cell counts were not performed on the femur owing to the lesser reproducibility of the measurement of cell counts from this bone.

Three haematologists participated in the study. Each was initially shown 10 normal marrows in order to familiarize himself with the appearance of murine marrow. Assessment of the unknowns was performed ‘blindly’ on two occasions one to two weeks later. On each occasion the observer was first shown two slides corresponding to tibial cell counts in each of the ranges 0-2, 2-4, 4-6, 6-8, 8-10, and...
Observer error in histological assessment of marrow hypocellularity

greater than $10 \times 10^6$ cells/tibia and he then attempted to place a series of randomized slides into these categories.

Results

Previous studies have shown that in the strain of mouse used the mean normal tibial cell count is $12.1 \times 10^6$ million and the value 2 SD below the mean is $8.4 \times 10^6$, but for simplicity the lower limit of normal was taken in the present study as $8.0 \times 10^6$ cells/tibia. In all, 37 marrow specimens were examined histologically and in the categories 0-2, 2-4, 4-6, 6-8, 8-10, and greater than $10 \times 10^6$ cells there were 11, five, four, five, six, and six specimens respectively. Since each specimen was examined on two occasions by three observers the total number of assessments was 222. An example of a histological section of a normal marrow is shown in fig 1, of a moderately hypocellular marrow in fig 2, and of a severely hypocellular marrow in figure 3. The relationship between each individual assessment and the measured value for the corresponding tibia is shown in figure 4. The correlation coefficient between measured and assessed values is a measure of the accuracy of assessment. For all values the coefficient was 0.72; for marrows having cellularity greater than $2 \times 10^6$ it was 0.39, and for marrows less than $2 \times 10^6$ it was 0.95. The correlation coefficient between the two assessments made on the same specimen by the same observer is a measure of the reproducibility of assessment. For all pairs of assessments the coefficient was 0.83; for marrows having a cellularity greater than $2 \times 10^6$ it was 0.61 and for marrows having a cellularity less than $2 \times 10^6$ cells it was 0.95.

In the table the data have been analysed to show the effect of using various 'cut-off' points to separate normality and abnormality, and to show the reciprocal changes in the number of false positives and false negatives that result. The optimal cut-off point would appear to correspond to $4 \times 10^6$ cells per tibia since only below this point were no normal marrows classed as abnormal. However, with this level, which was 33% of the mean for normal marrows and 50% of the lower limit of normal, the sensitivity for detection of hypocellularity was unavoidably low. Thus only nine of 84 (11%) marrows corresponding to tibial counts in the range of $2.8 \times 10^6$ were classed as abnormal whereas 57 of 66 marrows (86%) corresponding to counts of less than $2 \times 10^6$ were so classed. Of the nine normal assessments in the latter group, five were derived from one femur having a corresponding tibial count of $1.7 \times 10^6$ which was the highest cellularity in the group.

Tibial counts were performed on both tibiae in 15 mice. The means of the two counts ranged from $1.4 \times 16.8 \times 10^6$ cells per tibia and the correlation coefficient between the two counts was 0.93.

Discussion

The bone marrow of the mouse differs from that of

Fig 1  Histological appearance ($\times$ 50) of a normal murine marrow having tibial cellularity of $11.1 \times 10^6$ cells.
Fig 2  Histological appearance ($\times$ 50) of a moderately hypocellular murine marrow having tibial cellularity of $3.6 \times 10^6$ cells.

Fig 3  Histological appearance ($\times$ 50) of a severely hypocellular murine marrow having tibial cellularity of $0.6 \times 10^6$ cells.
Observer error in histological assessment of marrow hypocellularity

Fig 4  Relationship between measured and assessed values. The line indicates the cut-off point for discriminating between normality and abnormality if it is desired not to class normal marrows as abnormal.

man in that the amount of fat contained therein is much less. However, there is no reason to believe that this factor should make recognition of hypocellularity in the marrow more difficult. A decrease in the number of cells within the rigid bony cavity of either mouse or man must be associated with an increase in the volume of fat or of blood within sinusoids which on section should be recognized as an increase in the area of fat spaces or of sinusoids. Thus the result of the present study should be translatable in broad terms into the human situation, particularly when it is considered that the amount of marrow available for assessment, namely, most of one femur, represents a far larger proportion of the total marrow of the mouse than does the amount of marrow examined even in a trephine section in man.

The high correlation observed between the cellularity of the two tibiae of individual mice suggests that there was little variation in cellularity between different areas of the marrow, at least when areas of the order of size of a whole bone were being considered. If so, this suggests that the measured estimates of tibial cellularity correlated closely with the true values for femoral cellularity. The amount of variability in results of assessment of sections of femoral marrow therefore indicates that there was great error in the assessment of marrow cellularity both in terms of accuracy and reproducibility. The correlation coefficient (r) between the measured and assessed value is an index of accuracy. Thus the correlation coefficient of 0·38 for the marrows with cellularity greater than $2 \times 10^6$ indicates very low accuracy of assessment for such marrows, since by calculating the value for $(1 - r^2)$ it can be shown that 86% of the variance in the assessed value was due to some factor(s) other than linear regression of the assessed on the true value. The correlation between the two values for the one section assessed by the one observer provides a measure of intraobserver reproducibility and again the value of 0·61 for marrows having cellularity greater than $2 \times 10^6$ indicates that this reproducibility was low, since 63% of the variance of the assessment was due to intraobserver variation. Only for marrows corresponding to tibial counts of less than $2 \times 10^6$ did assessment become both accurate and reliable as indicated by the correlation coefficients of 0·95 between measured and assessed value and 0·95 between the two assessments by the same observer.

The usual problem in clinical practice is not to grade cellularity as finely as was attempted in the present study but to decide whether the marrow is normal or hypocellular, although often an attempt is made to grade the latter as mild, moderate, or severe. The method by which haematologists arrive at their cut-off point between normal and hypocellularity is uncertain, but probably this is done by reconciling (a) the subjective definition of the lower limit of normal of a series of marrows which on other grounds are believed probably to be of normal cel-

<table>
<thead>
<tr>
<th>Cellularity</th>
<th>Abnormal</th>
<th>Normal</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>True Positive</td>
<td>False Negative</td>
</tr>
<tr>
<td>$&lt;8 \times 10^4$</td>
<td>87%</td>
<td>13%</td>
</tr>
<tr>
<td>$&lt;6 \times 10^4$</td>
<td>67%</td>
<td>33%</td>
</tr>
<tr>
<td>$&lt;4 \times 10^4$</td>
<td>44%</td>
<td>56%</td>
</tr>
</tbody>
</table>

Table  Effect of different levels of cellularity for distinguishing between normality and abnormality.

1An abnormal assessment has been regarded as positive, a normal assessment has been regarded as negative.
lularity, and (b) a subjective definition of the upper limit of abnormality for those marrows believed probably to be hypocellular on the basis of peripheral blood counts. Clearly this process is subjective, must lead to an unknown number of false positives and false negatives, and must lead to somewhat different cut-off points for different haematologists. The results of the present study, as shown in fig 4 and table I, indicated a great overlap between the assessment of normal and abnormal marrows, with a high percentage of false positive and/or false negative assessments. If the results can be extrapolated to man they suggest that, if a cut-off point corresponding to $6 \times 10^6$ cells were used, then many normal marrows would be classed as abnormal and, if a cut-off point corresponding to $4 \times 10^6$ cells were used, nearly all cases of mild or moderate hypocellularity would be missed. With the latter cut-off point the only marrows which could be reliably recognized as abnormal would be those corresponding to tibial counts of less than $2 \times 10^6$ cells in the mouse. Such marrows would have a count of less than 17% of normal and would appear 'empty' or almost 'empty' on histological section. Even within this group, reliability might be less than ideal for marrows having the higher cellularity since in the present study the marrow having the highest cellularity in the $0.2 \times 10^6$ group was classified as normal on five of six occasions.

This study was supported by grants from the Michell and Anti-Cancer Foundations of the University of Adelaide. We thank Dr R. J. Kimber and Dr D. Handley for taking part in the study and the Institute of Medical and Veterinary Science, Adelaide, for help with histological sections.

References