Observer reliability in assessing villitis of unknown aetiology

T Y Khong, A Staples, L Moore, R W Byard

Abstract

Aims: To evaluate observer variation in diagnosis of villitis of unknown aetiology. Methods: Fifty haematoxylin and eosin stained sections were circulated to three pathologists who were asked to assess if villitis was present. These slides, with an additional 20, were recirculated and reassessed by the same pathologists. Results: Intra-observer agreement was 84-7% (range 74—92%) and interobserver agreement was 81%. A joint review by the three pathologists revealed that sources of differences included the overlooking of isolated single or small numbers of affected villi, the difficulty in assessing stromal cellularity close to infarcted parenchyma, and apparent stromal hypercellularity in immature villi. Conclusions: Experienced pathologists can show a significant interobserver variation in assessing villitis of unknown aetiology. Future studies on villitis of unknown aetiology should address the problem of observer reproducibility of diagnosis.

The diagnosis of villitis of unknown aetiology has implications for patients because of its significant association with intrauterine growth retardation, pre-eclampsia, and stillbirth. It may recur with successive pregnancies with similar associations. It has also been suggested that villitis of unknown aetiology may be a morphological expression of a maternal immune attack against placental tissues. Despite the clinical and experimental importance of villitis of unknown aetiology, the accuracy of its diagnosis has not been questioned. In discussing the apparent geographic distribution of the lesion with the reported incidences ranging from 6% (United States of America), 7-6% (Australia), 9% (Canada), 13.6% (United Kingdom), 18% (United States of America) and 33.8% (Argentina), Knox and Fox considered several factors that could account for these differences. These included the possibility that true variations exist among the populations studied, that the diagnosis of villitis may be influenced by the number of placental blocks taken, and that the incidence may have been influenced by the study size. They also felt that recognition of minor degrees of villitis may be difficult and that distinguishing between a mild inflammatory infiltrate and villous stromal hypercellularity is highly subjective. This study was undertaken to examine observer variation in the diagnosis of villitis.

Methods

A list of placentas was derived by scanning the files of the Department of Pathology, Queen Victoria Hospital. This comprised 41 cases from pregnancies with complications not known to be associated with a high incidence of villitis (such as twin pregnancy, preterm delivery), 33 cases from pregnancies with complications known to be associated with a high incidence of villitis (pre-eclampsia or intrauterine growth retardation), and 29 cases of villitis diagnosed by various pathologists. From this list, 50 cases were randomly selected by a non-participant in the study. One haematoxylin and eosin stained histologic section from each case, identified by number only, was circulated to the three pathologists. The pathologists were asked to assess whether villitis was present in each of the sections. No criteria for the diagnosis of villitis were given to the pathologists. At the end of the first round of assessment of the 50 cases, an additional 20 cases were included to act as subterfuge slides in assessing intraobserver variation, and all 70 slides were renumbered and circulated for assessment of the presence of villitis.

The results were analysed in three different ways. First, a direct measure of agreement between two observers was calculated as the percentage of slides on which they agreed, and this was calculated for each set of viewing. Second, a £ test was used to analyse the results, assuming a majority rule decision of the first set of 50 sections as being the “correct” answer. Finally, by introducing “time” as the second factor between the first and second viewings by the pathologists, a two factor experiment—a P x Q dichotomous factorial model with replication—is simulated and analysis of variance was performed on this model.

All 70 sections were reviewed jointly by the three pathologists at the end of the study and a consensus opinion on each section was obtained. Sources of disagreement were noted.

Results

The intraobserver agreement rate was 92%,
74%, and 88% for pathologists A, B, and C respectively (four, 13, and six differences of diagnosis of villitis over two sets of 50 slides each).

When agreement between the pathologists was measured by the percentage of slides on which they agreed, agreement between pathologists A and B was 82% (first round), 72% (second round), and 77% (overall); between pathologists A and C it was 86%, 90%, and 88%; and between pathologists B and C it was 84%, 70%, and 77%, indicating that the greatest agreement was between pathologists A and C. The average interobserver agreement was 81%. Assuming that a majority rule principle of correctness is applied to the first viewing of the 50 slides, the interobserver agreement rates between pathologist A and pathologists B and C were 92% (first round of 50 slides), 96% (second round of 50 slides), and 94% (overall) respectively; between pathologist B and pathologists A and C it was 90%, 72%, and 81%, respectively and between pathologist C and pathologists A and B it was 94%, 90%, and 92%, respectively. Analysis of the data, assuming that a majority rule principle applied to the second viewing, yielded almost similar results: overall agreement rates for pathologists A, B, and C were 94%, 82%, and 92%, respectively. These results again indicate that there was a statistical difference among the three observers (p = 0.005; \( x^2 = 30.7 \)) and that pathologist B was different from the other two (p = 0.005; \( x^2 = 20.6 \)).

Analysis of variance (ANOVA) of the simulated two factor experiment indicates that there was a significant difference between and within pathologists at the p = 0.05 level (table).

The conjoint viewing indicated that there were 32 sections without villitis and 18 sections with villitis present. Agreement was reached among the three pathologists over both the independent viewings as well as the third conjoint viewing on 27 slides negative for villitis and two slides positive for villitis. When the opinions of the first and second sets of viewings were compared with the consensus opinion, the agreement rates for pathologist A were 88% (first round), 88% (second round), and 88% (overall); for pathologist B these were 76%, 70%, and 73%, and for pathologist C these were 86%, 86%, and 86%, respectively.

**Discussion**

The histological aspects of villitis have been well documented,\(^1\)\(^3\) the sine qua non for diagnosis being an inflammatory infiltrate within the substance of the chorionic villi. Despite this seemingly simple feature, this study has shown that there can be a significant observer error rate in the diagnosis or otherwise of the lesion. Initially this may seem an alarming finding, but studies on observer variation in the diagnosis of various pathological lesions or in clinical decision making have shown similar trends with significant intra- and interobserver variation.\(^{12-17}\) In this instance, it may be a reflection of the subjectiveness of the diagnosis as suggested by Knox and Fox.\(^3\) It can be argued that the higher variation rate of pathologist B is a reflection of the perceived need, for the purposes of this study, to assign a diagnosis of villitis or otherwise to what might be borderline cases of hypercellular stroma v true villitis. We did not use a visual linear analogue scale to measure the degree of confidence of each diagnosis by the pathologists but this may have revealed this to be a source of difficulty in assessing villitis.

When the slides were reviewed jointly by the three pathologists, it became obvious that there were several sources of potential difficulty other than that associated with the difficulty of distinguishing between villous hypercellularity or infiltration with inflammatory cells. First, it was easy to overlook lesions when only one villus was inflamed, and yet this finding would have qualified as villitis. Second, often the inflamed villi were close to, but not contiguous with, areas of infarction and it was therefore difficult to decide sometimes if the infarmination was a reaction to the infarction. Published studies do not provide guidelines for the distance an inflamed villus has to be from a focus of infarction before it becomes villitis of unknown aetiology. This is an especially difficult problem in pregnancies complicated by pre-eclampsia or intrauterine growth retardation, where the lesions of villitis and infarction often coexist; more so when the infarctons are very extensive. Third, the difficulty of assessing the clinical importance of stromal cellularity is illustrated further by the relatively larger immature villi with prominent resident Hofbauer cells (macrophages) in immature placentas. This is a recurrent problem because those very pregnancies with which villitis is associated may have their gestational duration reduced by medical intervention for maternal or fetal indications. Finally, it is possible that the pathologists' background may have contributed to the observed differences in geographic incidence of villitis of unknown aetiology. We had limited our study to three pathologists of contemporaneous seniority, with an interest in perinatal/paediatric pathology, who had trained in different continents (Australia, Europe, and North America). It probably is, however, invalid to infer too much regarding training and early career influences and their effects on observer variation.

We used percentage agreement as a measure of agreement between observers because

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>Sum of squares</th>
<th>Degrees of freedom</th>
<th>Mean square error</th>
<th>F ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pathologists</td>
<td>388.67</td>
<td>1</td>
<td>44.33</td>
<td>4.491*</td>
</tr>
<tr>
<td>Time</td>
<td>403.33</td>
<td>1</td>
<td>403.33</td>
<td>4.061*</td>
</tr>
<tr>
<td>Pathologists x time</td>
<td>6066.66</td>
<td>2</td>
<td>3033.33</td>
<td>3.0730*</td>
</tr>
<tr>
<td>Within cell</td>
<td>29,020.00</td>
<td>29</td>
<td>0.9871</td>
<td></td>
</tr>
</tbody>
</table>

*Indicates significance at 95% level.
it is more easily understood than \( \kappa \) statistics,\(^1\) and because there was no "gold standard" or "correct diagnosis" in this series. In this respect the study lacked the strength in the number of observers for statistical significance because the assumption of a majority rule of correctness means that a minority decision (one out of three pathologists) could easily change into a majority decision if one of the remaining two pathologists were to interpret the slide differently. It is also pertinent to note that a majority rule may not be the "right answer": on four occasions in this study, five readings by the three pathologists over the two microscopic sessions gave a "wrong answer" with respect to that of the unanimous opinion.

If non-subjective criteria were available to indicate the presence or absence of villitis, the accuracy of diagnosis by the three observers could easily be resolved. Recent studies\(^{16-20}\) suggest that immunostaining for HLA-DR by syncytiotrophoblast could identify villitis lesions. However, this begs the question of how the lesion was originally defined. In common with other studies that rely on immunohistological characterisation of morphological entities, those studies relied on conventional light microscopy of haematoxylin and eosin stained sections for defining the lesions of villitis.\(^{16-20}\) In the context of diagnostic anatomical pathology, it is not always feasible to perform immunostaining of about four sections from each placenta, and diagnosis may depend ultimately on conventional light microscopy with the attendant observer errors that this study has demonstrated. Furthermore, immunohistology can create an additional source of observer variation insofar as immunostaining may be equivocal or else conflict with the prior diagnosis by conventional light microscopy.\(^{21}\)

In summary, this study has shown that there can be significant observer variation in the diagnosis of villitis of unknown aetiology. It is important, therefore, that future studies on villitis should address the problem of observer reproducibility of diagnosis.

We thank Ms Christine Mott for randomising the placental cases and Ms Rosa Katsikeros for preparing the haematoxylin and eosin sections. Part of this paper was presented at the 165th Meeting of the Pathological Society of Great Britain and Ireland in July 1992.