Observer reliability in assessing placental maturity by histology

T Y Khong, A Staples, R W Bendon, H M Chambers, S J Gould, S Knowles, S Shen-Schwarz

Abstract

**Aims**—To evaluate the ability of five experienced perinatal pathologists to assess placental maturity reliably by histology.

**Methods**—Twenty-four haematoxylin and eosin slides, six each from placentas of 27, 31, 35, and 39 weeks' gestation, were circulated to five pathologists on three separate occasions. The slides were labelled with the correct or incorrect gestational ages.

**Results**—The mean absolute error over all 360 readings was 2.72 weeks. Only 54% of the slides were assessed within two weeks of the correct gestation. Pathologists tended to overestimate younger gestations and underestimate older gestations. Two, and possibly three, pathologists were influenced by the gestational age stated on the label. One pathologist, who did not appear to be influenced by the label, was more accurate in diagnosing gestation of the placentas than other colleagues.

**Conclusions**—Experienced pathologists can have difficulty in assessing the villous maturity of placentas by histology. They can also be influenced by clinical information provided, such as gestational age. Other observer reliability studies must address the issue of the influence of labelled information on observer variation. A difference in maturation would have to be of a six week magnitude to have a chance of being detected by current methods. This may limit the value of the histological diagnosis of placental dysmaturity as a surrogate marker for uteroplacental ischaemia.


Keywords: Observer reliability, placental pathology, placental maturity, uteroplacental ischaemia.

It has been claimed that an accurate evaluation of placental maturation is important in identifying the cause of fresh stillbirth, antenatal hypoxia, and fetal growth retardation.¹ In general, there may be villous immaturity or accelerated maturation. The former is seen in fetal anaemia due to maternal-fetal rhesus isoimmunisation or twin–twin transfusion, maternal diabetes mellitus, anencephaly, sphihtis, and in some instances of Down syndrome,²³ but is significantly infrequent in preeclamptic pregnancies.⁴ Accelerated villous maturation has been described in preeclampsia⁵ and has been used as a surrogate marker for uteroplacental vascular insufficiency and defective placentation.⁶⁷ It is clear, then, that identification of abnormalities of villous maturation is an important component of histopathological assessment of placental pathology. However, the accuracy of this assessment has not been formally tested: it has been said that “all pathologists can readily recognise placentas of less than 28 weeks or more than 36 weeks’ gestation,” or that “villous immaturity is exceptionally difficult to recognise and assess in placentae from pregnancies of less than 36 weeks’ gestation”.⁶⁸ Between 28 and 36 weeks of gestation, fetal age is difficult to determine by histological placental examination. Thus the utility and validity of villous immaturity or accelerated villous maturity as signs of placental or pregnancy pathology would be compromised if assessment of villous maturity is not accurate or reproducible. This study sets out to evaluate observer reliability in assessing villous maturity.

**Methods**

**SLIDE SELECTION AND LABELLING**

Two blocks each, taken from the centres of functional circulatory units in order to yield an ideal section, were selected from 12 placentas, three each of 27, 31, 35, and 39 weeks’ gestation. These placentas were taken from pregnancies that were uncomplicated by maternal hypertensive disease or diabetes and had resulted in liveborn infants of birth weight appropriate for gestational age, and were normal in every aspect apart from preterm labour in those placentas from 27, 31, and 35 weeks’ gestation. Triplicate haematoxylin and eosin stained sections from these 24 blocks were prepared to provide three sets of identical slides for distribution to the pathologists for assessment. The slides were labelled with the following features. With the exception of the 39 week gestation sections, each of the triplicate slides was labelled with the correct gestational age once and an incorrect gestational age twice: for the 39 week gestational age group, 38 weeks was designated as the “correct” gestational age. The remaining two slides in each triplicate set were incorrectly labelled in reference to this “correct” gestational age. The gestational age in the incorrect slides were chosen in one week increments or decrements to span at least 27 to 39 weeks’ gestation. The order of submission of each set of slides to the pathologist was in ascending gestational age of the labelled slides, that is, disregarding their original correct gestational age. The selection and labelling of the
Assessment of placental maturity by histology

placental slides was performed by one pathologist (TYK) who coordinated distribution of the slides and did not participate in the histological assessment of villous maturity.

PATHOLOGICAL ASSESSMENT

Five pathologists participated in the histological assessment of villous maturity. These pathologists are all experienced perinatal pathologists with considerable experience in placental pathology and were practising in three different continents (North America: RWB, SS-S; Europe: SJG; Australia: HMC, SK) and had largely trained in their country of practice, with the exception of SK who trained in Europe and HMC who trained in North America and Australia. The triplicate sets of slides were read in three rounds by the pathologists, with an interval of 8–12 weeks between each round. The pathologists were asked to assess the slides using their usual criteria and state an exact gestational age. Each set of histological assessments accompanied by their slides was received by the co-ordinating pathologist before the next set of slides was issued for assessment.

DATA AND STATISTICAL ANALYSIS

The pathologists’ assessments were analysed by several methods. First, the readings were directly compared to the true gestational age and the accuracy and direction of error noted. Allowance was then made for a predetermined acceptable error of ±2 weeks from the true gestational age and the errors grouped according to absolute differences of 3–6 weeks or 7–10 weeks from the true gestational age. A dummy pathologist was simulated by assessing all the gestations labelled on the slides as correct, and these simulated assessments were included in the analysis. Finally, the assessments were analysed using a $p 	imes q 	imes r$ factorial experimental design with n replicates per cell ($p = 5$ pathologist, $q = 4$ gestational age groups, $r = 3$ time periods, $n = 6$ slides per cell). The errors (in weeks) of each estimate were calculated and analysed using analysis of variance to identify the sources of variation in the error of assessment by the pathologists.

The influence of the labelled gestational age on the pathologist’s assessment was analysed in three steps. Slides were considered as labelled true when the gestational age on the label was within ±2 weeks of the correct gestational age and as labelled false when it was outside this leeway. The proportion of slides correctly and incorrectly diagnosed in the true and false groups was compared using the $\chi^2$ test. This was recomputed after discarding the 10 slides from 39 week pregnancies, since it is impossible to be in error of three or more weeks, in one opposite direction, of a falselabelled slide. Finally, the incorrect slides were examined with regard to whether they were from the true labels or, if from the false labels, whether they were in the direction of or opposite to the false label. The null hypothesis is that incorrect readings will be distributed in a 50:50 split between true and false labelled slides, and distributed 50:50 between in the direction and opposite to the direction among the false labelled slides if labelling of the slides had no influence on the pathological assessment.

Results

Over the 360 readings (of 24 slides by five pathologists on three occasions), there were 40 estimations which were correct to the gestational week, with 176 underestimations and 144 overestimations. The mean absolute error over all pathologists/time periods and gestations (that is, 360 readings) was 2.72 weeks. One pathologist (A) was clearly more erroneous than the remaining four, primarily because of an overestimation of gestational age at gestations less than 33 weeks (an error of 5–17 weeks on average). Allowance for a “correct” reading to be within the range of ±2 weeks and grouping of the errors according to an absolute difference of 3–6 weeks or 7–10 weeks showed a significant difference between pathologists (table 1) and gestations (table 2). At the lower gestational age group, 25–29 weeks, all pathologists overestimated maturity (34 overestimated it by three or more weeks, 1 underestimated it by three or more weeks). However, as gestations increased so too did the tendency to underestimate: at >39 weeks’ gestation, there were no overestimates of three or more weeks and 38 underestimates of three or more weeks. Overall, there were 193 correct predictions (within ±2 weeks), 143 errors of three to six weeks, and 24 errors of seven to 10 weeks. Two thirds of the 24 errors of seven to 10 weeks were due to pathologist A.

Table 1 Accuracy of diagnosis to the gestational group

<table>
<thead>
<tr>
<th>Gestational age</th>
<th>Underestimate</th>
<th>Correct</th>
<th>Overestimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>3-6 weeks</td>
<td>7-10 weeks</td>
<td>2 to +2 weeks</td>
<td>3-6 weeks</td>
</tr>
<tr>
<td>27 weeks</td>
<td>1 0</td>
<td>55 20</td>
<td>23 11</td>
</tr>
<tr>
<td>31 weeks</td>
<td>26 7</td>
<td>30 8</td>
<td>26 8</td>
</tr>
<tr>
<td>35 weeks</td>
<td>23 4</td>
<td>56 9</td>
<td>7 0</td>
</tr>
<tr>
<td>39 weeks</td>
<td>37 1</td>
<td>52 3</td>
<td>0 0</td>
</tr>
<tr>
<td>All</td>
<td>87 5</td>
<td>193 40</td>
<td>56 19</td>
</tr>
</tbody>
</table>

Table 2 Accuracy of diagnosis by pathologists

<table>
<thead>
<tr>
<th>Pathologist</th>
<th>Underestimate</th>
<th>Correct</th>
<th>Overestimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>3-6 weeks</td>
<td>7-10 weeks</td>
<td>2 to +2 weeks</td>
<td>3-6 weeks</td>
</tr>
<tr>
<td>A</td>
<td>11 2</td>
<td>28 5</td>
<td>17 14</td>
</tr>
<tr>
<td>B</td>
<td>28 2</td>
<td>33 10</td>
<td>6 3</td>
</tr>
<tr>
<td>C</td>
<td>17 0</td>
<td>44 14</td>
<td>11 0</td>
</tr>
<tr>
<td>D</td>
<td>19 1</td>
<td>40 9</td>
<td>12 0</td>
</tr>
<tr>
<td>E</td>
<td>12 0</td>
<td>48 2</td>
<td>10 2</td>
</tr>
<tr>
<td>All</td>
<td>87 5</td>
<td>193 40</td>
<td>56 19</td>
</tr>
</tbody>
</table>
The mean value achieved by the dummy pathologist is a measure of the inherent bias in the labelling of the slides, for example for 27 weeks' gestation the slides were labelled on average (as a group) 3.56 weeks above the true value. The standard deviation achieved by the dummy pathologist should indicate a maximum variability tolerance. The interaction of pathologist and gestation was significant (table 4). This implies that the pathologist's performance was not consistent across gestations and each pathologist did not necessarily show similar performance at each gestation. This is most likely to be the result of an influence of the labels on the pathologists' assessments. Pathologists C and D, and possibly A, were significantly influenced by the label of the slides. Their performances on "labelled true" slides were significantly better than their performance on "labelled false" slides, even after discarding the falsely labelled 39 week slides (pathologist C: $\chi^2 = 11.23$, p < 0.005; pathologist D: $\chi^2 = 14.70$, p < 0.001; pathologist $\chi^2 = 4.69$, p < 0.04) while pathologists B and E did not appear to be so influenced (pathologist B: $\chi^2 = 0.61$, NS; pathologist E: $\chi^2 = 0.01$, NS).

Pathologists C and D but not pathologists A, B, or E had a significantly greater number of incorrect assessments in the direction of than in the direction opposite to the false label, confirming an influence of the label on histological assessment with pathologists C and D only (pathologist C: $\chi^2 = 16.38$, p < 0.005; pathologist D: $\chi^2 = 26.12$, p < 0.00001; pathologist A: $\chi^2 = 4.90$, NS; pathologist B: $\chi^2 = 0.31$, NS; pathologist E: $\chi^2 = 0.20$, NS (table 5)).

## Discussion

This study shows that experienced pathologists who report on placentas regularly can have difficulty in dating the maturity of the placenta accurately. Dating the placenta to its exact gestation was achieved on 40 out of a possible 360 occasions. It is unreasonable, however, to expect the placenta to be dated exactly and a tolerance of ±2 weeks would seem acceptable. Although the placentas in this study were selected from pregnancies with known and certain last menstrual dates and early ultrasound scan, there is still the possibility of error in the gestational age from these determinations. Furthermore, there is regional variation in the maturity of the placental villi, depending on their location in relation to the functional lobule of the placenta.

The question not answered in this study is whether the limitations in estimating gestational age are inherent in the biological variation in the placenta, or are an artefact of the techniques used by the pathologists. Most of the large (seven to 10 week) errors were made by two pathologists, suggesting that they may have used less reliable subjective criteria than their colleagues. On the other hand, two other pathologists appeared to have been influenced by the stated gestational age on the slide label. All the pathologists tended to overestimate the lower gestation placentas, particularly pathologist A, and to underestimate the higher gestation placentas, particularly pathologist B. This may have been due to an inherent bias in the design of the study, insofar as the slides of mature placentas could not reasonably be mislabelled beyond 40 weeks' gestation, while the younger placentas similarly were not mislabelled below 24 weeks' gestation. There was difficulty with those placentas of 31 and 35 weeks' gestation. With these placentas, it was possible to mislabel the slides by as much under as over the true gestational age before submission to the pathologists for assessment, and thus the accuracy or otherwise of their assessment was a better reflection of their performance. The right bumping of the mean error for the 31 and 35 week gestational age placentas would suggest that pathologists could estimate these gestations reasonably accurately, but the real situation is that pathologists have difficulty in dating the maturity of the placenta accurately.
Assessment of placental maturity by histology

Assessment

We are not aware of any previous studies on observer reliability that have addressed the influence of information given with the specimen. This is important since histopathologists, like clinicians, rely on all available clinical information and cues,19 and may well be misled by erroneous or distracting information. Examples include dermatoses, which may evoke a limited range of pathological features, or previous cytology smear/biopsy/colposcopic findings in squamous cervical lesions. In the context of this study, the assumed gestational age of the pregnancy may be inaccurate because of the recording of the last menstrual period or absence of an early ultrasound scan. Two, and possibly three, pathologists were influenced by the labelled gestational age on the slides. It is likely that the pathologists compromised between their estimated gestational age and that given on the label when the discrepancy was large, but accepted that given on the label as correct when the discrepancy was not significantly obvious to them. It is likely that if the pathologists had trusted their own judgements, a fair better performance would have ensued. In the study design, the slides were labelled in one week incremental steps in an attempt to elucidate the resolution limit of the pathologists. In the event, this could not be analysed because of the confounding effects of label influence and pathologists’ errors.

This study may be somewhat artificial in that the assessment of maturity was based on one slide, whereas in practice the pathologist would assess maturity, and hence dysmaturity, from more than one slide from each placenta. Nonetheless, an inescapable conclusion from this study is that pathologists are poor at determining the term of placentas using conventional light microscopy. This has several implications. If dysmaturity—evident in villous maturity is either advanced or delayed, or not within the spectrum of normal growth and development of the villous tree—is to be an objective and reproducible diagnostic description, then the use of morphometry may be of value in assessing variation within a sample and in selecting the best discriminating variables. Secondly, the use of dysmaturity, and more specifically accelerated villous maturation, as a surrogate for uteroplacental vascular insufficiency is questionable given that its presence may be irreproducible or unrecognised. This is particularly important given that the more gross errors, even allowing for a ±2 week tolerance, were found among the 31 and 35 week placentas, which would account for many of the placentas from preclamptic pregnancies which may be delivered early for maternal or fetal reasons. It is also obvious that, even if pathologists can apply agreed and consistent criteria for assessing placental maturity and consequently diagnose abnormalities of villous maturity, this may be of limited utility. This is because there are no agreed published guidelines on how to interpret the pathological and chronological maturity to be made. In this study, although the pathologists were successful in estimating the gestational age within two weeks in only 193 of 360 placentas (54%), they correctly estimated the gestation within six weeks in 336 out of 360 placentas.

This result suggests that for the usual diagnostic question, that is, whether intrauterine growth retardation was due to uteroplacental ischaemia, a difference in expected maturation would have to be of a six week magnitude to have a (93%) chance of being detected. For example, in a case of growth retardation, a 30 week infant should have at least 36 week placent morphology to diagnose uteroplacental ischaemia on the basis of accelerated maturation. Another implication is that advanced maturation cannot be diagnosed after 34–36 weeks’ gestation.

Audit of observer reliability has become a standard approach in many areas of tissue diagnosis and, if conventional placental pathology is to be genuinely meaningful, perinatal pathologists must assess the reliability of their diagnoses.20 It would be of interest to test objectively whether the pathologists are limited in the accuracy of placental diagnoses that incorporate maturational estimates in patients with known disease, for example, maternal hypertensive disease or poorly controlled diabetes.

We thank the technical staff at Queen Victoria Hospital for their patient preparation of the sections and Ms Ann Harmer and Miss Lynne Smith for secretarial assistance.

References

1 Naeve RL. Functionally important disorders of the placenta, umbilical cord, and fetal membranes. Hum Pathol 1987; 18:609-12.
Observer reliability in assessing placental maturity by histology.

T Y Khong, A Staples, R W Bendon, H M Chambers, S J Gould, S Knowles and S Shen-Schwarz

doi: 10.1136/jcp.48.5.420

Updated information and services can be found at:
http://jcp.bmj.com/content/48/5/420

These include:

Email alerting service
Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

Notes

To request permissions go to:
http://group.bmj.com/group/rights-licensing/permissions

To order reprints go to:
http://journals.bmj.com/cgi/reprintform

To subscribe to BMJ go to:
http://group.bmj.com/subscribe/