

Summary

Abnormal placental calcification is associated with poor pregnancy outcome. The aim of this study was to measure inter- and intra-observer variability in assigning placental grades. Five experienced sonographers independently graded ninety images on two occasions. All technological factors which could affect data reliability and consistency were removed. Substantial variations between individuals' scores were observed. A mean κ -value of 0.34 (range from 0.19 to 0.50) indicated fair inter-observer agreement over the two occasions and only nine of the ninety images were graded the same by all five observers. Intra-observer agreement was moderate, with the mean κ -value 0.52. This study demonstrates that, despite standardised viewing conditions, Grannum grading of the placenta is not a reliable technique even amongst expert observers. The need for new methods to assess placental health to improve neonatal outcomes is required and work is ongoing to develop a software based method using 2D and 3D image datasets.

Introduction

The placenta is the most vital support organ for the developing fetus. Antenatal ultrasound assessment of placental morphology plays an important role in evaluating fetal health, revealing abnormalities such as infarcts and calcification. Currently ultrasound assessment of placental calcification relies on Grannum grading (Grannum et al., 1979), which was initially devised to assist in predicting fetal pulmonary maturity. Early studies demonstrated a significant correlation of placental grade with gestational age, pulmonary maturity, and respiratory distress, with a Grade III placenta at term associated with an absence of subsequent neonatal respiratory distress (Tabsh, 1983; Shah & Graham, 1986).

A degree of placental calcification is normal as the fetus approaches term, however accelerated placental maturation is associated with pregnancy induced hypertension, fetal growth restriction, and fetal distress in labour, all factors which contribute to an increased risk of perinatal morbidity and mortality (Hills et al., 1984, Veena & Sapna, 2000). In their study of 1802 low-risk women McKenna et al (2005) confirmed that detection of a grade III placenta at 36 weeks gestation in a low-risk population assists in identifying the “at-risk” pregnancy. Abnormal placental calcification is also thought to be associated with diabetes and Rhesus incompatibility, with delayed placental maturation (DVM) occurring in these conditions (Clair et al, 1983). Delayed placental maturation has been shown to be significantly associated with pre-gestational diabetes, gestational diabetes and antenatal or intrapartum death.

In a randomised control trial conducted by Proud & Grant (1987) of 2000 women scanned at 32-34 weeks and 34-36 weeks gestation, (where women were divided into one group in which the obstetrician was informed of the placental grade and the other group in which they were not), a grade III placenta was reported in 15% of cases of the women scanned between 34-36 weeks. The presence of a grade III placenta at this stage was found to be significantly associated with low maternal age, nulliparity, being white, maternal smoking, an increased risk of meconium stained liquor, fetal distress in labour, low apgar scores, low birth weight and perinatal death (4% of cases). Whilst the authors do not feel that their results justify routine scanning late in the third trimester they do, however, suggest that their results give good reason to recommend that placental grading should be assessed and reported at all third trimester ultrasound examinations. McKenna et al conducted a larger randomised control trial in 2003. In this study (n=1998) women were randomised at 30 weeks to a control group receiving standard antenatal care or to the study group who also received an ultrasound scan, which included assessment of the placental grade. The results suggest that the introduction of an ultrasound scan at 30-32 weeks' and 36-37 weeks' gestation may reduce the risk of a growth restricted fetus, most likely due to increased antenatal interventions. Bricker et al (2008) in their review of routine late pregnancy ultrasound acknowledge the fact that placental grading may be valuable, but question the reproducibility of reported results. With this in mind they recommend that future research of late pregnancy ultrasound should include evaluation of placental textural assessment.

While clinical studies have shown placental grading to hold promise, its role in clinical practice remains controversial. Perhaps, as Bricker et al (2008) suggest this is due to the subjectivity of this method. McKenna et al (2005) propose that ultrasound assessment for placental grading can be performed in a matter of minutes by sonographers with modest training and skills. Placental grading however may in part contribute to the lack of progress in identifying a high risk fetus in a low risk population as it has been shown to lack objectivity, precision and reproducibility. There is only one previous study examining reproducibility of placental grading (Sau et al., 2004). This study found that while intra-observer agreement in grading placentas was generally good, agreement between observers was only fair for all Grannum grades and poor for grade III. These findings suggest that either Grannum grading is not reproducible or that there is a need for training in those performing grading. In this study, however, as there is no mention of controlled viewing conditions (for example monitor consistency or ambient lighting), variability due to technological factors may have affected the study results. Optimum ambient lighting conditions and consistent presentation of digital ultrasound images are essential for both soft- and hard-copy for visualisation of normal anatomy and subtle pathologies (Wade et al., 2005; Fetterly et al., 2008). The aim of this study was to measure inter- and intra-observer variation in the assessment of placental calcification using placental grading, under strictly controlled viewing conditions.

Material and methods

This is a prospective study. With institutional ethical approval and maternal written consent ninety digital images of the placenta were stored from different patients, who attended the fetal assessment unit of a large maternity hospital, between July and November 2007. The hospital has approximately 9,000 deliveries per annum and is the national tertiary referral centre for women with high-risk pregnancies. The main indication for the ultrasound examination was to assess the fetal biophysical profile at 41 weeks and 5 days, as is standard hospital protocol. Two experienced fetal medicine consultants and three full-time midwife sonographers (with experience ranging from 2 to 8 years) independently graded the images, on 2 occasions, as per the classifications devised by Grannum et al (1979). A number of measures were taken to guarantee the reliability of the study. Each image was randomly allocated a number, and the images were presented to the observers in numerical order. Prior to viewing the images the observers reviewed written information on Grannum grading and images representing each grade. They then agreed a consensus definition of the appearance of each grade. This was as follows:

Grade 0: The placental tissue and the basal plate are homogenous without the presence of linear highly reflective foci. The chorionic plate is smooth and well defined.

Grade I: The placental tissue contains a few linear highly reflective areas parallel to the basal plate, which remains unchanged. The chorionic plate presents subtle undulations.

Grade II: The placental tissue contains randomly dispersed echoes and is divided by comma-like reflective structures continuous with the chorionic plate. The marked indentations of the chorionic plate do not reach the basal plate, which is well defined by small linear highly reflective areas.

Grade III: The placental tissue is divided into compartments containing central echo-free areas. The chorionic plate indentations reach the basal plate, which contains almost confluent, very highly reflective areas (Jauniaux, 2003).

Viewing conditions were strictly controlled in order to eliminate any technological factors which could adversely affect the reliability of the results. The 5 observers reviewed the images at the same time in a controlled viewing laboratory. The monitors used for viewing the images were all calibrated to the GSDF (greyscale standard display function) standard to ensure equivalent brightness. Liquid crystal displays (LCDs) are the current electronic display technology for viewing medical digital images. Calibration of the luminance response of each LCD is required to ensure that observer perception of an image is consistent on all displays as the inherent luminance properties of different LCDs can vary considerably (Goo et al., 2004; Fetterly et al., 2008).

The room (ambient) lighting was also tightly controlled. High room luminance can have a damaging effect on image quality by significantly reducing image contrast. This is also the case if no light other than that of the monitor is used. Ambient lighting was adjusted to ensure a reading on all monitors of between 25 and 40 lux, which is the level required to ensure maximum diagnostic accuracy when reporting on images (Chakrabarti et al., 2003; Brennan et al., 2007).

Individual monitors were allocated to each observer. The observers sat apart and viewed the images in silence in order not to influence each other's decisions. The observers were asked to allocate a grade of 0, I, II or III to each image and were instructed to report the highest grade observed. They were then asked to review the same images one week later, with the images presented in a different order, to minimise memory artefacts. Prior to the second review it was again clarified that all the observers were agreed on the grading criteria. Viewing conditions were identical to the first session, with each observer viewing the images on the same monitor as on the first occasion and again all monitors were consistent for brightness and contrast, with the ambient lighting set at 25-40 lux.

Statistical analysis

For each of the two viewing sessions statistical analysis was performed to compare, between each observer, the variability of the grades assigned.

Individual observer comparisons of grades allocated on the two occasions were also evaluated. The degree of inter- and intra-observer agreement was compared using Kappa (κ) analysis, which is a statistical measure of inter-rater reliability. The benefit of using Kappa as a statistical parameter of agreement is that it does not necessitate the assumption of a correct diagnosis. It is also a more robust measure than simple percent agreement calculation since κ takes into account the possibility of agreement occurring by chance. Kappa is expressed through a coefficient ranging from - 1.0 to +1.0. A value of zero indicates an agreement the same as that expected by chance and a coefficient of 1.0 indicates perfect agreement. Landis and Koch, (1977) suggest that poor agreement is indicated by a value less than 0.20, fair agreement if the value is between 0.21 and 0.40, moderate agreement if between 0.41 and 0.60, good agreement between 0.61 and 0.80 and that values above 0.81 indicate excellent agreement.

Results

Overall agreement over the two occasions between observers was fair, with a mean κ -value of 0.34 with a range 0.19 to 0.50. Only nine of the ninety images (ten percent) were graded the same by all five observers, five of these grade III.

Inter-observer agreement – First image assessment

Results from the first image review session demonstrate fair agreement between observers, with a mean κ value of 0.31 (range 0.19 to 0.43). Eight

out of ten of the observer groupings showed fair agreement, and there was moderate agreement between one of the observer groupings only, with a κ value of 0.43 (observers B and C). There was poor agreement between observers B and E (κ value 0.19). The κ values for inter-observer variability for the first image review session are represented in Figure 1.

Inter-observer agreement – Second image assessment

There was very little difference in the inter-observer variability at the second image review session. The mean κ value was slightly better at 0.37 (fair agreement), with values ranging from 0.29 to 0.50. None of the observer groupings had poor agreement and again only one observer grouping (A and D) showing moderate agreement (κ value 0.55). Observers B and C, who had moderate agreement in the first session, had only fair agreement in the second. Figure 2 represents inter-observer variability for the second image review session.

Intra-observer agreement

Intra-observer agreement (represented in Figure 3) had a moderate mean κ -value of 0.52, with individual comparisons ranging from 0.45 to 0.66. There was good agreement of placental grading between the two image review sessions for observer B only (κ -value 0.66), with the other four observers showing moderate agreement. There was generally only one grade

difference between observations, however fourteen of the ninety images had a difference of two grade points between observations, and one image was assigned grades 0, I, II and III. This image is shown in Figure 4.

Agreement Grannum grade III

The results shown in this study highlight the obvious difficulty in any sonographer, including the authors, determining what constitutes a grade III placenta. However agreement on which placentas were assigned Grannum grade III is perhaps the most significant as this grade has potential implications for clinical outcome. In an attempt to ascertain the level of agreement between observers on a grade III placenta the authors chose the images assigned grade III by observer B as this observer has the best intra-observer agreement (κ value 0.66). Observer B graded twenty one images as III, observer A agreed in 7 cases, observer C in 16 cases, and observers D and E in twelve cases.

Discussion

The low level of agreement in this study between observers in assessing placental calcification highlights the subjectivity of placental grading. This occurred even though every effort was made to ensure consistency both in terms of agreed consensus prior to each session on the classifications to be used for Grannum grading (1979), and the strictly controlled viewing conditions that were employed. Monitor luminance and ambient light,

factors which can adversely impact on image display and therefore soft-copy viewing conditions were within the recommended levels and were consistent for all PCs for each session (Wade et al., 2005). These rigorous controls guaranteed that the only variable in this study was the observer.

All observers were trained to a very high standard in obstetric ultrasound, the three midwife sonographers working full-time in this area and the two consultants highly experienced in the field of fetal medicine, overseeing this busy national tertiary referral centre for high risk pregnancies. At both image review sessions inter-observer agreement was generally only fair and indeed at the first review of the images there was poor agreement between observers B and C. Sixteen percent of the images had a difference of two grade points in the grades assigned over the two sessions, and there was complete agreement on a grade III placenta in only five of the ninety images. Intra-observer agreement was slightly better, with four of the observers showing moderate agreement.

One of the observers remarked that it was difficult to report on still images, particularly on some of the images in relation to distinguishing the basal plate from the chorionic plate. Any difficulties in distinguishing the different placental layers, however, should have been encountered to the same degree by all five observers. Also it must be noted that in some centres ultrasound images are not reported on by the sonographer performing the live scan, but are in fact reported on at a later stage, with assessment of still images by a fetal medicine specialist or radiologist. Nevertheless real time ultrasound or the review of video clips is far superior to reviewing still images.

The results from this study are disappointing and in fact less favourable than those reported in the previous study by Sau et al (2004), even though in our study all variables, bar the observer, were excluded. The authors acknowledge however that the controlled viewing environment under which the images were reviewed, and the use of still images, is not truly reflective of clinical practice. A further study is planned in an uncontrolled environment, using real time imaging, which will be more reflective of normal clinical practice. Nonetheless this study strongly confirms the view that there is a lack of objectivity, precision and reproducibility associated with grading of the placenta. This was particularly borne out by the fact that one image was assigned grades 0, I, II and III.

Conclusion

Placental grading is subjective both in terms of the correct image of the placenta that should be acquired and how that image is assessed for calcification. At this stage it is evident that a protocol must be developed to ensure that the correct parameters are used for optimal image acquisition and that a more objective method of assessing placental function is required. One possible option is digital analysis. Initial work on a new 2D ultrasound imaging software tool developed in the School of Medicine and Medical Sciences, University College Dublin has shown promising results (Ryan et al., 2008). Further research in this area is vital if progress is to be made in utilising the placenta as part of the assessment of fetal health.

References

- Brennan PC, McEntee M, Evanoff M, Phillips P, O'Connor WT, Manning DJ. 2007. Ambient Lighting: Effect of Illumination on Soft-Copy Viewing of Radiographs of the Wrist. *American Journal of Radiology* 188:W177-W180.
- Bricker L, Neilson JP, Dowswell T. 2008. Routine ultrasound in late pregnancy (after 24 weeks gestation). [1]Chakrabarti K, Kaczmarek RV, Thomas JA, Romanyukha A. 2003. Effect of Room Illuminance on Monitor Black Level Liminance and Monitor Calibration. *Journal of Diagnostic Imaging* 16:350-355.
- Clair MR, Rosenberg E, Tempkin D, Androtti RF, Bowie JD. 1983. Placental grading in the complicated or high-risk pregnancy. *Journal of Ultrasound in Medicine* 2:297-301.
- Fetterly KA, Blume HR, Flynn MJ, Samei E. 2008. Introduction to Greyscale Calibration and Related Aspects of Medical Imaging Grade Liquid Crystal Displays. *Journal of Diagnostic Imaging* 21:193-207.
- Goo JM, Choi JY, Im JG, Lee HJ, Chung MJ, Han D, Park SH, Kim JH, Nam SH. 2004. Effect of Monitor Luminance and Ambient Light on Observer Performance in Soft-Copy Reading of Digital Chest Radiographs. *Radiology* 232:762-766.
- Grannum PAT, Berkowitz RL, Hobbins J. 1979. The ultrasonic changes in the maturing placenta and their relation to fetal pulmonic maturity. *American Journal of Obstetrics and Gynecology* 133: 915-922.

- Hills D, Irwin GAL, Tuck S, Baim R. 1984. Distribution of Placental Grade in High-Risk Gravidas. *American Journal of Radiology* 143:1011-1013.
- Jauniaux E. 2003. The Placenta. In: Dewbury K, Meire H, Cosgrave D, Farrant P, editors, *Ultrasound in Obstetrics and Gynaecology*. London. Churchill Livingstone; pp 529-555.
- Landis JR, Koch GG. 1977. The measurement of observer agreement for categorical data. *Biometrics* 33:159-174.
- McKenna D, Tharmaratnam S, Mahsud S, Dornan J. 2005. Ultrasonic evidence of placental calcification at 36 weeks' gestation: maternal and fetal outcomes. *Acta Obstetrica et Gynecologica Scandinavica* 8:7-10.
- McKenna D, Tharmaratnam S, Mahsud S, Bailie C, Harpe A, Dornan J. 2003. A Randomised Trial Using Ultrasound to Identify the High-Risk Fetus in a Low-Risk Population. *Obstetrics & Gynecology* 101:626-632.
- Proud J, Grant AM. 1987. Third trimester placental grading by ultrasonography as a test of fetal wellbeing. *British Medical Journal* 294:1641-1644.
- Ryan JT, McAuliffe F, Higgins M, Stanton M, Brennan PC. 2008. A novel software-based technique for quantifying placental calcifications and infarctions from ultrasound. *Proceedings of SPIE-Optical Society for Optical Engineering* 6290:1-8.
- Sau A, Seed P, Langford K. 2004. Intraobserver and interobserver variation in the sonographic grading of placental maturity. *Ultrasound in Obstetrics and Gynecology* 23:374-377

Shah YG, Graham D. 1986. Relationship of placental grade to fetal pulmonary maturity and respiratory distress syndrome. American Journal of Perinatology 3:53-5.

Tabsh KM. 1983. Correlation of real-time ultrasonic placental grading with amniotic fluid lecithin/sphingomyelin ratio. American Journal of Obstetrics and Gynecology 145:504-508.

Veena A, Sapna J. 2000. Placental Grading and its Correlation with Foetal outcome. Journal of Obstetrics and Gynecology of India 50:59-62.

Wade C, Brennan PC, McEntee MF. 2005. An Assessment of the Performance and Quality Control Procedures of PACS Workstation Monitors used in Irish Radiology Departments. Proceedings of SPIE 574:244-255.