
Article:
Griffiths, P.D. orcid.org/0000-0002-2706-5897, Bradburn, M., Campbell, M.J. orcid.org/0000-0003-3529-2739 et al. (9 more authors) (2017) Change in diagnostic confidence brought about by using in utero MRI for fetal structural brain pathology: analysis of the MERIDIAN cohort. Clinical Radiology, 72 (6). pp. 451-457. ISSN 0009-9260

https://doi.org/10.1016/j.crad.2017.01.009

Change in diagnostic confidence brought about by using *in utero* MR imaging for fetal structural brain pathology: Analysis of the MERIDIAN cohort

Introduction

The 'Magnetic Resonance imaging to enhance the diagnosis of fetal developmental brain abnormalities *in utero'* (MERIDIAN) study is a multi-centre, prospective cohort study designed to evaluate the diagnostic and clinical impact of in utero MR (iuMR) imaging of fetal brain abnormalities[1]. The synoptic overview of the MERIDIAN results has been reported elsewhere[2] and describes improvements in diagnostic accuracy (at least 22%) when iuMR is included in the diagnostic pathway of fetuses with suspected brain abnormalities recognised on ultrasonography (USS). Diagnostic changes were accompanied by major changes in counselling of pregnant women and changes in management.

A further facet of diagnostic impact, diagnostic confidence, was analysed on a descriptive level in the earlier paper and showed encouraging results, with an overall increase in diagnoses made with high confidence by 13% on iuMR compared with USS[2]. There was a 3.5-fold reduction of incorrect diagnoses made with high confidence on iuMR compared with USS and a 2.5-fold reduction of correct diagnoses made with low confidence on iuMR.

In this paper we perform additional analyses on the MERIDIAN cohort to assess whether improvement in diagnostic accuracy was matched by an increase in diagnostic confidence. The cohort was analysed on an individual case basis by three assessments;

a) conventional uncorrected (C_2-C_1%)
b) conventional (C_2-C_1%) with the 'Omary correction'[3]

c) score-based weighted average method described by Ng and Palmer[4,5].

We discuss the relative merits and disadvantages of each technique, including an attempt to provide definitive conclusions about the contribution of iuMR to diagnostic confidence in this field.

Materials and Methods

Ethics

MERIDIAN was performed in accordance with the Medicines for Human Use (Clinical Trials) Regulations 2004 through adherence to the University of Sheffield Clinical Trials Research Unit's (CTRU) standard operating procedures. Ethics approval was obtained for a multi-centre prospective study and written informed consent obtained from all participants prior to entering the study.

The primary results of MERIDIAN have been presented elsewhere[2]. In this paper we provide a detailed assessment of the changes in diagnostic confidence on that same cohort.

Patient Characteristics

Participants were recruited between July 2011 and August 2014. The full recruitment process and numbers of recruits are described in the earlier paper[2] but summarised here. Inclusion criteria were – pregnant women aged ≥16 years whose fetus had any form of brain abnormality detected by USS at a gestational age of ≥18 weeks, had no contraindications to iuMR and consented to the study. Recruitment was based on consecutive selection from 16 fetal medicine (FM) units in the UK with a referral base of 28 million people.

We needed to collect 504 fetuses with complete outcome reference diagnoses (ORD) who had their iuMR study within 2 weeks of the USS based on the power calculation underlying

the entire study[2]. The ORD used in the study was the results of autopsy in cases of termination of pregnancy or spontaneous fetal loss, or the results of clinical neuroimaging performed within the first 6 months of life. As shown in Table 1, we aimed to perform iuMR studies on 720 fetuses with abnormal brain examinations on USS allowing for attrition and predictions of incomplete ORD sets. In practice, we performed iuMR on the 570 fetuses who had ORD and had their iuMR study within 2 weeks of the USS examination. Those fetuses are the basis of the studies reported in this paper.

Imaging studies

No specific requirements were made for the USS technique but all of the USS studies were performed by appropriately trained FM consultants working in the UK National Health Service (NHS). Each fetal brain abnormality recognised on USS was recorded in a tabulated fashion using nomenclature from the "ViewPoint" antenatal ultrasound reporting software (GE Healthcare, Chalfont St Giles UK). The FM consultants were also asked to record their certainty of diagnosis for each brain abnormality using a five-point Likert scale[6]. The descriptions and attributed percentages of diagnostic confidence are: 'Very unsure' - 10% certain, 'Unsure' – 30% certain, 'Equivocal' – 50% certain, 'Confident' – 70% certain, and 'Highly confident' – 90% certain. Some of the subsequent analyses require binary descriptions of confidence and in those situations 10%, 30% and 50% certainties are described as 'low confidence' and 70% and 90% certainties as 'high confidence'.

Following prenatal USS, participants underwent iuMR at one of six sites[2]. IuMR examinations were performed at 1·5T but it was not possible to match protocols exactly across the sites because different manufacturers' MR systems were used. The absolute requirement was to obtain T2-weighted images of the fetal brain in the three orthogonal

planes using the best ultrafast method available (maximum slice thickness 5mm) and a T1-weighted ultrafast sequence in at least one plane (usually axial). The reporting radiologist was aware of the diagnoses and certainty made by the USS expert before the iuMR study was performed and had access to the clinical USS report. The radiologist was required to comment on each diagnosis made on USS, using 'diagnosis excluded' if they disagreed with an USS finding. Extra anatomical diagnoses were added where appropriate. Each diagnosis was accompanied by a confidence rating using the same Likert scale as the USS assessment. The 'diagnosis excluded' option was attributed a 90% certainty.

Data Handling and analysis

The assessment of diagnostic confidence used in this report is based on the 'dominant diagnosis'. In cases where there was only one anatomical/pathological diagnosis this was straightforward but in cases with more than one diagnosis the independent panels defined the 'dominant diagnosis' as the one most likely to influence prognosis. For the a) conventional uncorrected C_2-C_1% and b) conventional C_2-C_1% with Omary correction analyses, described below the only data required for assessment of diagnostic confidence was the Likert-based percentage certainties from USS and iuMR.

As described below, however, the c) score-based weighted average analysis requires information derived from ORD, which was obtained from the Multidisciplinary Independent Expert Panel (MIEP). The full role of the MIEP is described elsewhere[2], but in summary, the MIEP consisted of three NHS consultants (neuroradiology, fetal medicine, paediatric neurology) from a single centre that did not recruit into MERIDIAN. The panel were given tabulated diagnostic results for each fetus and were blinded to whether it was an USS or an iuMR report. They were asked whether each report agreed with the ORD completely and,

where USS and iuMR disagreed, which one indicated the more severe pathology. The results were subsequently unblinded by staff at Sheffield CTRU. In 7% of cases the MIEP required more information and had access to the full clinical reports and imaging, if necessary, at which point blinding was no longer possible.

Diagnostic confidence tests

a) Conventional uncorrected C_2-C_1%

The pre-test confidence (confidence on USS = C_1%) was subtracted from the post-test confidence (confidence on iuMR = C_2%), therefore (C_2-C_1)%. In accordance with the technique of Ng and Palmer[4,5] the difference was converted to an integer based on the difference in the number of 20% intervals to allow direct comparison with the results of the 'score-based weighted average' analysis described below. For example, if a diagnosis of agenesis of the corpus callosum (ACC) was made with 50% confidence on USS and the same diagnosis was made on iuMR with 90% certainty; C_2-C_1% = 90-50% = +40% which was converted to a score of +2. As such, the integer scores ranged from -4 to +4. Positive values indicate more confident diagnoses made by iuMR, negative values indicate more confident diagnoses made by USS, and zero indicating no difference.

b) Conventional C_2-C_1% with Omary correction

This analysis used the same method as above but applied the 'Omary correction'[3], which is applied in the following situation only;

i. The pre-test confidence (USS) is high (70% or 90%) AND

ii. There is a change in diagnosis post-test

That is, the USS and iuMR reports disagree and in those circumstances the change in diagnostic confidence is calculated as (C_2-[100-C_1])% but the results are otherwise handled in

the same way as the uncorrected conventional data. For example, if a diagnosis of ACC was made with 90% confidence on USS and a diagnosis of holoprosencephaly made on iuMR with 90% confidence the Omary corrected score is $(C_2\text{-}[100\text{-}C_1])\% = 90\text{-}10\% = 80\% = +4$). In any cases where the criteria for the Omary correction are not met then the conventional uncorrected score is unchanged. The rationale for applying the Omary correction and its implications are described in the discussion.

c) Score-based weighted average analysis

There are well described limitations in both the Conventional and Omary corrected analyses because those methods do not require confirmation that the diagnoses were either correct or incorrect (i.e. no ORD required). The cases described in this paper did have ORD so we are able to use the 'score-based weighted average' method described by Ng and Palmer[4,5]. This approach uses not only diagnostic confidence assessments but also indicators of diagnostic accuracy as supplied by the MIEP namely: was the overall diagnosis correct for iuMR, was the overall diagnosis correct for USS, and which described the most severe pathology. This aspect was combined with a binary assessment of diagnostic confidence as either 'high' or 'low' as described above. An algorithm modified from Ng and Palmer (Figure 1) was used to define a route label for each case and derive a route score ranging from -4 to +4. Zero indicated no change in 'appropriate' confidence, whilst positive values indicate a benefit arising from the introduction of iuMR and negatives indicate iuMR had a detrimental effect on confidence, the larger the number the greater the effect. For example, if USS described ventriculomegaly (VM) with high confidence as the only finding but iuMR diagnosed ACC with high confidence the 'score-based weighted average' result will depend on the ORD. If ACC was confirmed, the route label would be A1 with route score +4, whereas if VM was the ORD the route label E5 gives a route score of -3. This is based on the presumption that

149 the new test (iuMR) is correct.

150 For all three types of analyses the frequency of each integer score was plotted as a bar chart
151 and described in terms of the number of cases in which iuMR reported with greater
152 confidence (positive scores) and the number of cases with reduced confidence (negative
153 scores). The mean and standard deviation of the score and 95% confidence intervals were
154 calculated and one sample t-tests were carried out to test the hypothesis that the expected
155 calculated scores were zero.

156

Results

a) Conventional uncorrected C_2-C_1%

A bar chart of the frequency versus score of the data analyzed by the conventional uncorrected C_2-C_1% method in 570 fetuses is shown in Figure 2a. A difference in confidence levels of the dominant diagnosis of any degree was present in 42% of cases, among which the majority were made with greater confidence following iuMR (32%) rather than USS (10%). The mean difference in confidence on the ordinal -4 to +4 scale was +0·44 in favor of iuMR (95% CI 0.35 to 0.54, p<0·0001; see Table 2).

b) Conventional C_2-C_1% with Omary correction

The criteria for the Omary correction were met in 98/570 cases. A bar chart of the frequency versus score data analyzed by the conventional C_2-C_1% with Omary correction method in 570 fetuses is shown in Figure 2b. A difference in confidence levels of the dominant diagnosis of any degree was present in 52% of all cases, 47% were more confident on iuMR and 5% more confident on USS. The mean difference in confidence was +1·10 in favor of iuMR (95%CI 0.98 to 1.25, p<0·0001; see Table 2).

c) Score-based weighted average analysis

The route labels for the 570 cases included in this study are presented in Table 3. A histogram of the frequency versus score data analyzed by the 'score-based weighted average' method is shown in Figure 2c. A difference in confidence levels of the dominant diagnosis of any degree was present in 38% of all cases and the score was positive in 31% (indicating that an appropriate increase in diagnostic confidence for iuMR) and negative in 7%). The mean difference in confidence was +0·75 (95% CI 0.63 to 0.87, p<0·0001; see Table 2).

Discussion

There are several ways to assess the diagnostic impact of a new imaging method or new application of an existing method but diagnostic accuracy and confidence are central to that process. Our previous report on diagnostic accuracy of iuMR identifying fetal brain pathology demonstrated improvements of at least 22% over USS to 92·4% for fetuses scanned between 18-23 weeks gestation and to 93·5% for fetuses >23 weeks gestation. This, together with the encouraging findings of previous systematic reviews and meta-analyses[7-10], suggests that iuMR increases the accuracy of fetal brain diagnoses compared to USS alone. Whilst some of these systematic reviews and meta-analyses have their limitations, such as a lack of outcome reference data and reporting of the time difference between the examinations the MERIDIAN study was able to overcome these as it was a prospective study designed to address these limitations. Despite the difference in methods between the reviews and our study, all conclude this similar finding; therefore adding weight to the evidence base.

The relevance of diagnostic accuracy in assessing an imaging technology is self-evident but the importance of the diagnostic confidence in imaging examinations is often overlooked and less well studied. Our previous paper provided a simplified descriptive report of changes in diagnostic confidence which, whilst encouraging, required a more robust analysis provided in this paper[2].

All three assessments of diagnostic confidence presented here show statistically significant, positive effects of including iuMR in the diagnostic pathway of assessing the fetus suspected of having a structural brain abnormality. However, there are a number of limitations to the methodologies used herein. Firstly the conventional uncorrected method of assessing diagnostic (C_2-C_1%) is a simple approach but has disadvantages when the diagnoses made on USS and iuMR are in conflict as occurred in over 25% of MERIDIAN cases[2]. For example, if

a diagnosis of isolated VM is made on USS with 90% confidence but the iuMR report is 'ACC' with 90% confidence - the conventional (C_2-C_1%) analysis gives 0 - no change in confidence. This overlooks the discrepancy in information given to the woman and the potential (major) impact on outcome. This is an inevitable result of any analysis that does not use ORD. Furthermore, the conventional (C_2-C_1%) data will tend to underestimate the value of the post-test method (iuMR) if it is more accurate than the pre-test method (USS) and the degree of underestimation is closely related to the difference in diagnostic accuracy between the two methods. There were over 22% more correct diagnoses on iuMR compared with USS in our studies which will lead to a major underestimation of improved diagnostic confidence.

Secondly the Omary correction attempts to counteract the tendency to underestimate improved confidence by the conventional uncorrected (C_2-C_1%) analysis. It is applied only when a high confidence diagnosis (70% or 90%) is made on the pre-test method (USS) and that diagnosis does not agree with the post-test (iuMR) diagnosis. Extending the example from the previous paragraph, the calculation using this correction would be (90-[100-90]) = 80% or +4, the highest score possible in favour of iuMR. This better reflects the diagnostic difference but the assumption that underlies the Omary correction is that the post-test diagnosis (iuMR) is correct. This may be spurious because, again, ORD is not used to confirm this. The closer the diagnostic accuracy of the post-test (iuMR) is to 100% the more appropriate this correction becomes. Our data demonstrates that the overall diagnostic accuracy of iuMR at 18-23 weeks was approximately 92%; therefore analysis using this correction will overestimate the value of iuMR. The data from MERIDIAN shows sizable differences in the degree of improved confidence between the conventional analysis and the Omary corrected data consistent with these concerns.

Ng and Palmer[4,5] make the case that errors will inevitably occur unless ORD is used and their major concerns relate to:

a) An appreciation that diagnoses made on the new test (iuMR) can be incorrect, even if made with high confidence.
b) The consequences of incorrect diagnoses are variable.
c) The introduction of a new test may have detrimental effects on patient management.

They suggest that the 'score-based weighted average' reflects a change in 'appropriate' diagnostic confidence, positive scores indicating improved 'appropriate' diagnostic confidence and negative scores indicating deleterious effects on diagnostic confidence. The algorithm shown in Figure 1 indicates that there are a number of ways iuMR could have deleterious effects on assessment (negative scores), such as;

a) iuMR reports the same, ultimately correct diagnosis with low confidence whilst USS reported the correct pathology with high confidence (route label B5 – route score -1). The implication being that doubt has been placed on the pre-existing high confidence diagnosis that may influence counselling or management.
b) iuMR makes an incorrect diagnosis with high confidence (most of the E and F route labels – route score ranging from -1 to -4).
c) iuMR makes an incorrect diagnosis with low confidence (most of the C and D route labels – route score ranging from -1 to -3).

In contrast, iuMR studies will receive positive scores in situations such as;

a) iuMR makes a correct diagnosis with high confidence that USS got wrong (Route labels A1-A4 – route score +2 to +4)

b) iuMR makes a correct diagnosis with high confidence that USS made with low confidence (Route labels A6 – route score +1)
c) iuMR makes a correct diagnosis with low confidence that USS got wrong (Route labels B1-B4 – route score +1 to +3)

One major advantage of this method is the ability to present large amount of complex data in a relatively accessible fashion and it is likely to represent the most accurate method of presenting true changes in diagnostic accuracy. Application of this method to the MERIDIAN data shows statistically significant improvements of diagnostic confidence.

In this paper we have used the 'score-based weighted average' method only to assess and describe the entire MERIDIAN cohort, it can also be conveniently used to evaluate subgroups in order to provide targeted information. For example subsequent analysis will evaluate the role of iuMR in the three commonest anatomical subgroups that were referred into the MERIDIAN study, namely isolated VM, ACC and abnormalities of the posterior fossa. In addition, the 'score-based weighted average' method provides opportunities to study the possible effect of experience of the radiological reports in relation to appropriate increases in diagnostic confidence.

In conclusion, we have presented three analyses of change in diagnostic confidence on an individual case basis, all of which show major improvements when iuMR is included in the diagnostic pathway. We have described the relative strengths and weaknesses of the methods used but we believe that the 'score-based weighted average' method has considerable advantages and should be used as part of any assessment of diagnostic confidence in studies when ORD is available'.

1. Griffiths PD, the MERIDIAN study group. Protocol 11PRT/2491: Magnetic Resonance Imaging to Enhance the Diagnosis of Fetal Developmental Brain Abnormalities in Utero (MERIDIAN) (ISRCTN27626961). The Lancet http://www.thelancet.com/protocol-reviews/11PRT-2491
2. Griffiths PD, Bradburn M, Campbell MJ, Cooper CL, Graham R, Jarvis D, Kilby MD, Mason G, Mooney C, Robson SC, Wailoo A. Use of MRI in the diagnosis of fetal brain abnormalities in utero (MERIDIAN): a multicentre, prospective cohort study. The Lancet 2016; doi: 10.1016/S0140-6736(16)31723-8
3. Reference blinded for review
4. Omary RA, Kaplan PA, Dussault RG, et al. The impact of ankle radiographs on the diagnosis and management of acute ankle injuries. Academic Radiology 1996; 3(9):758–765
5. Ng CS, Palmer CR. Analysis of diagnostic confidence and diagnostic accuracy: a unified framework. The British Journal of Radiology 2007;80(951):152–160.
6. Ng CS, Palmer CR. Analysis of diagnostic confidence: application to data from a prospective randomized controlled trial of CT for acute abdominal pain. Acta Radiol. 2010;51(4):368–74
7. Likert R. A technique for the measurement of attitudes. Archives of Psychology 1932; 22(140):1–55
8. Mundy L, Hiller J, Braunack-Mayer A, Merlin T. MRI for the detection of foetal abnormalities. Adelaide: Adelaide Health Technology Assessment (AHTA). 2007;15(6)
9. Rossi AC, and Prefumo F. Additional value of fetal magnetic resonance imaging in the prenatal diagnosis of central nervous system anomalies: a systematic review of the literature. Ultrasound Obstet Gynecol 2014; 44(4)**:**388–93

10. Van Doorn M, Oude Rengerink K, Newsum EA, Reneman L, Majoie CB, Pajkrt E. Added value of fetal MRI in foetuses with suspected brain abnormalities on neurosonography: a systematic review and meta-analysis. J Matern Fetal Neonatal Med 2015; 29(18):2949-61
11. Jarvis D, Mooney C, Cohen J et al. A systematic review and meta-analysis to determine the contribution of mr imaging to the diagnosis of foetal brain abnormalities In Utero. Eur Radiol 2016; DOI 10.1007/s00330-016-4563-4

Figures and Table Legends

Figure 1. The algorithm used to provide the score-based weighted average data used to assess changes in appropriate confidence (modified from Ng and Palmer[4]).

Figure 2. Assessments of change in diagnostic confidence between USS and iuMR reports using three methods; a) Conventional uncorrected C_2-C_1% method, b) Conventional C_2-C_1% with Omary correction, and c) score-based weighted average method.

Table 1. Predicted and actual recruitment numbers of fetuses into the MERIDIAN study. The three analyses of diagnostic confidence reported in this paper are based on the 570 fetuses who had the iuMR performed within 2 weeks of USS and had complete outcome reference data.

Table 2. Changes in diagnostic confidence using the route score method in the first row and differences in confidence of diagnoses with and without Omary correction in the second and third rows respectively

Table 3. Frequencies of each Route Label and their relevant Route Score

Acknowledgments

This project was funded by the National Institute for Health Research Health Technology Assessment Programme (project number 09/06/01). The views and opinions expressed therein are those of the authors and do not necessarily reflect those of the Health Technology Assessment Programme, NIHR, NHS or the Department of Health.

338

339