

Reproducibility and Accuracy of MR Imaging of the Brain after Severe Birth Asphyxia

Phillipe Jouvet, Francis M. Cowan, Philip Cox, Edward Lazda, Mary A. Rutherford, Jonathan Wigglesworth, Huseyin Mehmet, and A. David Edwards

BACKGROUND AND PURPOSE: MR imaging of the brain can be used to detect cerebral damage after suspected hypoxic-ischemic injury. This study examines the reproducibility and accuracy of MR imaging soon after severe birth asphyxia.

METHODS: During a 48-month period, full-term newborn neonates, who died within the first week as a result of severe hypoxic ischemic encephalopathy, were included in the study if they had undergone early (<5 days old) MR imaging and postmortem neuropathologic studies. Two trained observers assessed reproducibility by examining multiple brain regions independently with current criteria and then defining and applying improved criteria. Accuracy of MR findings was tested by comparing the brain regions about which the two imaging raters agreed to those regions about which the two pathologists agreed.

RESULTS: Eight neonates, with a median gestational age of 40 weeks (range, 38–40 weeks) and who suffered severe birth asphyxia, were included in the study. In the reproducibility study, MR imaging agreement was moderate when current criteria were used ($k = .44$). Using the improved criteria, agreement increased considerably ($k = .62$). Much of this improvement was due to limiting the analyses to the posterior limb of the internal capsule, thalamus, parietal cortex, hippocampus, and medulla. The posterior limb of the internal capsule was the most reliable region analyzed. MR imaging agreement was similar to that achieved by two experienced pathologists reviewing the histologic sections ($k = .66$). In the accuracy study, MR imaging abnormality was predictive of pathologic abnormality with a sensitivity of .79 and a positive predictive value of 1.0. The predictive value of a single MR imaging abnormality was .79 (95% confidence interval, .61–.96).

CONCLUSION: Criteria that provide substantial reproducibility and accuracy for the interpretation of MR imaging findings very early after birth asphyxia can be derived.

MR imaging is increasingly used in the assessment of full-term infants with hypoxic ischemic encephalopathy (1–6). Abnormalities occurring early and associated with severe injury of the brain have been described; although imaging findings evolve rapidly after birth, the early findings can be subtle. Even in cases of severe injury, images obtained on the first day may appear normal or edema may ren-

der the interpretation of focal lesions difficult. It is important to know how reliable the imaging features are in this clinical setting because MR imaging is becoming increasingly available and may be used to assign prognosis in combination with other neurophysiologic and clinical tests.

Few data are available regarding either the interobserver variability (reproducibility) of MR images in these circumstances or the correlation between images and the histopathologic appearances of the brain (accuracy). The aim of this study therefore was to examine the reproducibility of MR images obtained from a group of neonates soon after birth asphyxia and to compare MR imaging appearances with neuropathologic findings.

Methods

Patients

Neonates were included in the study if all of the following criteria were fulfilled: gestational age was between 37 and 42 weeks; there was evidence of fetal distress; hypoxic-ischemic

Received May 28, 1998; accepted after revision, March 23, 1999.

This study was supported in part by the Medical Research Council and the Garfield Weston Foundation.

From the Department of Pediatrics (P.J., F.M.C., M.A.R., H.M., A.D.E.), the Department of Histopathology (P.C., E.L., J.W.), and Robert Steiner MRI Unit (M.A.R.), Imperial College School of Medicine, Hammersmith Hospital, Du Cane Road, London W12 OHS, England.

Address reprint requests to A.D. Edwards, FRCP, Department of Paediatrics, Imperial College School of Medicine, Hammersmith Hospital, Du Cane Road, London W12 ONN, England.

encephalopathy was staged grade three according to the criteria of Sarnat and Sarnat (7); MR imaging was performed within 4 days of birth; death occurred within the first week after birth; and neuropathologic examination of the brain was conducted. Informed consent was obtained from the parents for both MR imaging and postmortem examination of the brain.

MR Imaging

Neonates were examined at a median age of 24 hours (range, 12 hours–3 days) during natural sleep or after sedation with chloral hydrate (25–70 mg/kg), as previously described (1). When required, full intensive care, including ventilatory support, was maintained throughout the examination. Body temperature was controlled, and fluid and drug administration was maintained throughout the imaging procedure. All neonates were monitored with pulse oximetry and ECG. All examinations were supervised by at least one pediatrician experienced in both intensive care and MR imaging procedures.

MR imaging investigation was performed using a 1-T HPQ system (Picker International, Inc., Cleveland, OH). The images were obtained using age-related inversion recovery (3600/30 [TR/TE]; inversion time, 700) and T1- (860/20) and T2-weighted (3400/120) spin-echo sequences in the transverse plane.

MR Imaging Reproducibility Study

The protocol for the reproducibility study was as follows. A first review of MR images (including inversion recovery, T1-weighted, and T2-weighted images) was made separately by two observers experienced in the interpretation of MR images of neonates with hypoxic ischemic encephalopathy who were unaware of the neuropathologic results. The following areas were independently coded as either normal, abnormal, or not assessable according to criteria then used at our institution: the posterior limb of internal capsule, the thalamus, the parietal cortex, the occipital cortex, the frontal cortex, parietal white matter, frontal white matter, occipital white matter, the hippocampus, the dentate nucleus, and the medulla. These results were classified as the "training set." A first statistical analysis was conducted to define interobserver variability. After reviewing the level of agreement in the training set, practice was reviewed. Improved criteria for normality and abnormality in each brain area were defined. Brain regions in which clear criteria could not be determined were excluded from further analysis. The improved criteria were tested on a large number of other images for a period of 2 months. The original study images were reassessed by the same observers using the revised criteria, and a second statistical analysis of interobserver variability was performed.

MR Imaging Accuracy Study

The protocol for the accuracy study was as follows. A neuropathologic examination of the brain was conducted for each neonate. The brain was removed intact at the time of postmortem examination and was suspended in 10% formalin for at least 1 month before sectioning. Coronal cuts were made at 1-cm intervals. Representative blocks from the parietal cortex, subcortical parietal white matter, thalami and basal ganglia, and hippocampus were taken and processed to paraffin wax. Five-micrometer-thick sections were stained with hematoxylin and eosin. Independent histologic examinations were performed separately by two perinatal pathologists experienced in the assessment of the neonate brain. Both were unaware of the MR imaging findings. The brain areas for each patient were coded as abnormal, normal, or not assessable according to the presence or absence of eosinophilic neurons, nuclear pyknosis, karyorrhexis, gliosis, macrophages, hemorrhages, or white matter edema. The following brain areas were examined: the posterior limb of internal capsule, thalamus, parietal cortex,

parietal white matter, hippocampus, dentate nucleus of cerebellum, and medulla. Statistical analysis was conducted to define interobserver variability. Areas about which both MR imaging observers agreed regarding abnormality or normality using the revised criteria were selected. Areas about which there was disagreement between reviewers or that one observer thought had an image that was not assessable were rejected. A similar selection of agreed upon cases was made from the pathologic data. The agreement between MR imaging diagnoses and pathologic findings was determined.

Statistical Methods

Interobserver variability was assessed by the calculation of unweighted kappa statistics. The interpretation of kappa statistics is presented by Landis and Koch (8): $k > .81$, agreement almost perfect; .61–.8, substantial agreement; .41–.6, moderate agreement; .21–.4, fair agreement, and 0–.2, poor agreement.

The predictive ability of MR imaging for histopathologic appearances was assessed by calculation of sensitivity/specificity and positive/negative predictive values. Because of the problems of interpreting these values in small samples without correction for disease prevalence, the data were analyzed further using a Bayesian approach (9). To determine the ability of a single MR imaging diagnosis to predict the histopathologic appearances, the overall predictive *P* value was calculated. A beta density curve, which estimates the probability density of prediction correctness, was constructed according to the method presented by Berry (9). From the mean, fifth, and 95th centiles of this distribution, the predictive probability of the test, together with the 95% confidence intervals for that value, were calculated. The interpretation of predictive probability is that it is the probability that the next assessment made using the revised MR imaging criteria for the study of a neonate suffering birth asphyxia will correctly predict the histopathologic appearances of that brain region. Predictive probability can be regarded as analogous to positive predictive value. Calculations of kappa statistics were made using Stata statistical software (Statacorp, College Station, TX).

Results

During a 48-month period (November 1, 1992–November 1, 1996), eight neonates with severe hypoxic ischemic encephalopathy, who underwent early MR imaging (Fig 1) and were admitted to the neonatal intensive care unit of the Hammersmith Hospital, London, were included in both the reproducibility and accuracy studies. The patients' characteristics are shown in Table 1. Hypoxic-ischemic injury occurred during delivery, except in Patient 4, who experienced a sudden unexplained cardiac arrest at 9 hours of age. The neurologic prognosis of these eight newborns was extremely poor, and owing to clinical, biologic, electroencephalographic, MR imaging, and MR spectroscopic findings, active treatments were withdrawn after discussion with parents.

For all of the neonates, the MR imaging abnormalities that were documented were bilateral. There were no unilateral or markedly asymmetrical changes.

MR Imaging Reproducibility Study

The training set showed moderate agreement in interpretation of the images between observers,

TABLE 1: Characteristics of the newborns included in the accuracy and reproducibility study of early MR brain imaging

Case	GA (weeks)	BW (g)	Cord pH	Apgar Score (1 & 5 min)	EEG Activity the First 24 Hours	Age of Imaging (days)	Age at Death (days)	Age at Autopsy (days)
1	40	3230	6.90	1/2	Burst suppression	2	2	9
2	39	3175	NA	0/0	isoelectric	1	5	8
3	38	3205	NA	0/2	fits	1	1	6
4	40	3270	NA	3/9*	NA	2	4	5
5	40	3770	NA	0/4	NA	1	2	4
6	42	3270	6.90	1/3	NA	3	5	7
7	39	3390	7.20	3/2	isoelectric	0	1	4
8	41	3845	NA	2/5	suppressed	1	3	6

Note.—Case 4 had a cardiac arrest at 3 hours of age. GA, gestational age; BW, birth weight; NA, not available.

TABLE 2: MR imaging criteria used to compare early MR imaging and pathology findings in 8 infants with severe hypoxic-ischaemic encephalopathy

Area	Level	Abnormal Criteria
PLIC	Low ventricular level angled through the basal ganglia and thalami (Figure 1A)	Low signal throughout on T1-WSE or IR on either or both sides of the brain, or high signal < 1/3 PLIC on IR either or both sides (Figure 1B)
Cortex	Around the central fissure	Excessive high signal on T1-WSE or IR on either or both sides of the central fissure or excessive low signal on T2-WSE on either or both sides of the central fissure
Thalamus	Low ventricular level angled through the basal ganglia and thalami (Figure 1A)	Excessive high signal and/or medial extension of the normal high signal in the region of the ventro-lateral nucleus of the thalamus on T1-WSE or IR images (Figure 1B).
Brain stem	Pons and medulla	Regions of low SI on T1-WSE or IR (as low or lower than the temporal white matter) or loss of the normal high signal of at least one of the myelinated tracts or nuclei: medial lemniscus, medial longitudinal fasciculus, inferior colliculus, pyramidal tracts.
Hippocampus	Temporal horn	Diffuse high or low signal on T1-WSE or IR

Note.—The images were obtained on a Picker 1.0T HPQ system using inversion recovery (IR) and T1- and T2-weighted spin echo (T1-WSE, T2-WSE) sequences in the transverse plane. PLIC signifies posterior limb of internal capsule.

with an agreement in 64.8% of the cases ($k = .44$). The agreement for each area was as follows: six of eight for the posterior limb of internal capsule, six of eight for the thalamus, eight of eight for the parietal cortex, four of eight for the occipital cortex, four of eight for the frontal cortex, six of eight for the parietal white matter, five of eight for the occipital white matter, five of eight for the frontal white matter, five of eight for the hippocampus, two of eight for the dentate nucleus, and six of eight for the medulla.

The image interpretation protocol constructed after consideration of the training set excluded six of the 11 areas assessed in the first part of the study because no consensus was obtained. Areas excluded from this analysis were the frontal cortex, occipital cortex, dentate nucleus, and frontal, parietal, and occipital white matter (Table 2). When images were reassessed by the same observers, interobserver variability was reduced, with agreement in 80% of the cases ($k = .62$) (Table 3). Individual area agreement was as follows: eight of eight for the posterior limb of the internal capsule, five of eight for the thalamus, six of eight for the parietal cortex, six of eight for the hippocampus, and six of eight for the medulla. Further analysis showed

TABLE 3: Interobserver variation in imaging interpretation

		Observer 1		
		Normal	Not Assessable	Abnormal
Observer 2	Normal	7	0	4
	Not assessable	0	4	1
	Abnormal	3	0	21

Note.—The assessment included the following areas of 8 newborns with hypoxia-ischaemia injury: posterior limb of internal capsule, thalamus, parietal cortex, hippocampus, and medulla. The agreement was 80% ($k = 0.62$).

that if the data from the training set were reanalyzed after exclusion of areas in which defined criteria could not be produced, there was a significant increase in agreement to 77% ($k = .58$).

MR Imaging Accuracy Study

The pathologists agreed in 84% of the cases ($k = .66$). When only those brain regions defined in the revised image interpretation protocol and examined in the second MR imaging review were

FIG 1. MR images.

A, Normal control participant, aged 2 days. Inversion recovery sequence (3800/30; inversion time, 950) shows normal appearances with high signal intensity, consistent with myelin within the posterior half of the posterior limb of the internal capsule.

B, Neonate with severe birth asphyxia, aged 2 days. No high signal from myelin in the posterior limb of the internal capsule occurs.

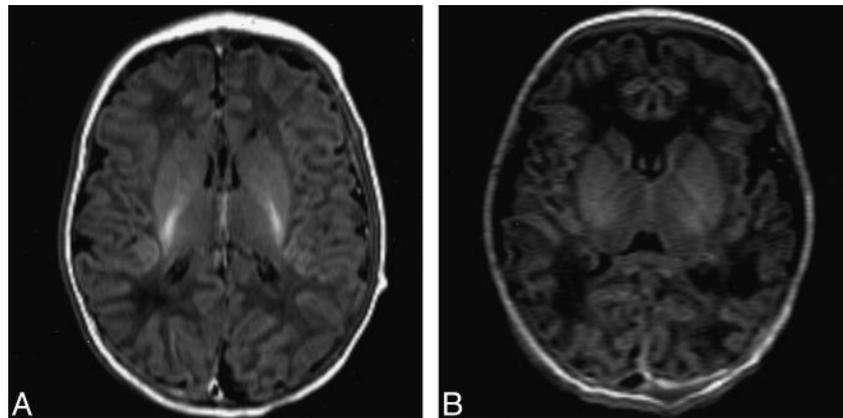


TABLE 4: Accuracy of MR imaging for prediction of neuropathological findings after birth asphyxia

		MR Findings	
		Normal	Abnormal
Neuropathology	Normal	0	0
	Abnormal	4	15

considered, interobserver variability between the pathologists was not significantly changed, with agreement in 85% of the cases ($k = .64$). Individual area agreement was as follows: four of eight for the posterior limb of internal capsule, eight of eight for the thalamus, eight of eight for the parietal cortex, seven of eight for the hippocampus, and six of eight for the medulla. Pathologic findings in the thalamus, parietal cortex, hippocampus, and medulla were eosinophilic neurons, nuclear pyknosis and karyorrhexis, macrophage infiltration, and gliosis. Pathologic findings in the posterior limb of the internal capsule were edema in all cases (Fig 2), combined with karyorrhexis in four cases and astrocyte hypertrophy in two cases. Hemorrhage was observed in the medulla in one case.

Both MR imaging raters agreed about 28 of 40 brain regions regarding normality or abnormality. The neuropathologists disagreed about nine of those 28 brain regions. Thus, a total of 19 brain regions, which included data from all subjects, were available for direct comparison. Agreement about MR imaging and pathologic assessment of different brain regions was reached as follows: posterior limb of internal capsule, four of four cases; thalamus, two of four cases; parietal cortex, five of five cases; hippocampus, two of two cases; and medulla, two of four cases.

Thus, abnormal MR imaging findings and abnormal pathologic findings were observed in a total of 15 regions and normal MR imaging findings and abnormal pathologic findings in four regions (Table 4). There were no histologically normal areas in the regions examined. The sensitivity rating of the ability of MR imaging to be predictive for pathologic abnormality was .79, and the positive predictive

value was 1. The negative predictive value and specificity rating could not be calculated. The predictive probability was .79, with 95% confidence intervals of .61 to .96.

Discussion

This study provides data regarding the accuracy and reproducibility of MR imaging when used to assess cerebral damage soon after birth asphyxia in term neonates. The application of a Bayesian approach allowed the 95% confidence limits of predictions to be calculated. These data provide a more secure basis for the interpretation of early MR images of neonates suspected of suffering birth asphyxia.

Reproducibility in the initial MR training set was moderate, reflecting that a clear definition of normal and abnormal criteria was required for most of the areas. After review and redefinition of an imaging protocol by consensus, the validation set showed substantial agreement. This improvement in interobserver variability was mainly due to restricting the areas to those for which a protocol could be defined.

The overall agreement between MR imaging interpreters was similar to that achieved in histopathologic examinations. The main source of disagreement between pathologists was difficulty in the interpretation of edema in the posterior limb of the internal capsule. The assessment of edema is difficult because tissue may be affected by fixation or processing artifact. Additional special staining, such as Luxol fast blue, might have improved the histologic diagnosis of edema in these specimens.

When MR imaging data about which there was agreement were compared with pathology results about which the pathologists agreed, the overall probability of MR imaging correctly predicting histologic results was .79 (95% confidence interval, .61–.99). This suggests that when experienced observers apply the defined criteria to a defined brain region, a reasonable probability that they will be correct exists. Because, however, the positive predictive value was 100% (all errors were due to the failure of normal MR imaging to be predictive for

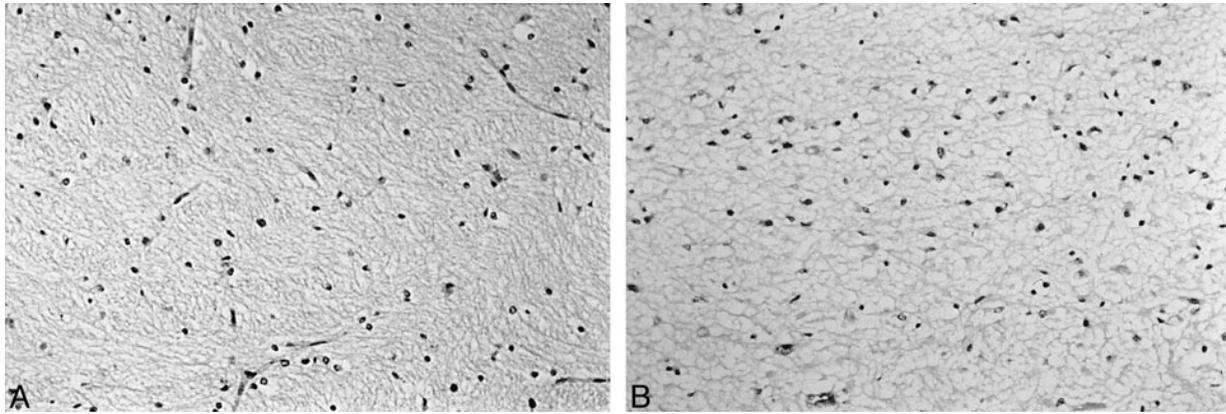


FIG 2. Histologic findings of the posterior limb of the internal capsule sections stained with hematoxylin and eosin (original magnification, $\times 300$).

A, Normal histologic findings of the posterior limb of the internal capsule.

B, Abnormal histologic findings of the posterior limb of the internal capsule show marked interstitial edema and nuclear pyknosis of many of the glial cells.

neuropathologic abnormalities), when an MR imaging abnormality is found, it is highly predictive of tissue damage.

Nevertheless, it is perhaps surprising that more areas of the brain did not look consistently abnormal on the images, considering the widespread nature of the histopathologic evidence of damage. One explanation of the failure of MR imaging to be predictive of tissue damage is that the median patient age at the time of imaging was 1 day and the median patient age at the time of death was 2.5 days. Serial imaging in neonates with hypoxic-ischemic encephalopathy has shown that abnormal signal intensities become more obvious during the first week after birth. We did not use proton density-weighted or diffusion-weighted images in this study, and it is possible that the addition of such sequences may have yielded more abnormal findings during the first few days after the injury. Proton density-weighted images have been shown to depict early "cortical edema" (10), but a comparative study is needed to identify accurately histologic abnormalities.

A further problem arose, which is common to all studies comparing images to postmortem data. Neonates who die almost always have severe tissue injury. In our study, there were relatively few brain regions in which there was no histopathologic injury; no region available for the accuracy study was microscopically normal. This may introduce some bias into the results, which should properly be applied only to neonates suffering severe asphyxia.

Because the accuracy of MR imaging was assessed by comparing MR images, about which both raters agreed, with pathologic data, about which both pathologists agreed, the number of brain regions available for comparison was considerably reduced. This was necessary, however, to avoid problems due to reproducibility. In calculating sensitivity, specificity, and predictive values, we also made the assumption that all of the regions compared were rated independently. This assumption

might be questioned, and the summary statistical measures should be used with appropriate caution; however, the results of the analysis provide a useful indication of the accuracy of MR imaging.

Even in very severe birth asphyxia, total agreement between MR imaging assessors was reached only concerning the posterior limb of the internal capsule. This shows the difficulty of assessing brain damage of neonates very soon after birth asphyxia by conventional MR imaging, but confirms our previous finding that the posterior limb of the internal capsule is a particularly useful region for defining the prognosis of brain damage after birth asphyxia (11). Although the loss of normal signal is consistent with the presence of edema identified by histologic examination, it also suggests the loss of myelin. Unfortunately, we were not able to use myelin staining in this study to confirm this, although we have noted that in some neonates with less severe asphyxia, the normal signal from myelin may return within weeks (11). In this previous study, we did note that the internal capsule could look normal on day 1 and then become abnormal during 24 to 48 hours. If this sign, which is easy to detect because it involves a reversal of the normal tissue signal, is still evolving during the first days after birth, then it is not surprising that in other areas, where the abnormal MR findings are exaggerations of the normal signal, complete agreement is more difficult.

Conclusion

MR imaging of the brain performed within the first days after severe hypoxic-ischemic injury has substantial reproducibility and accuracy if precise assessment criteria are applied. Interpretations are most reproducible if confined to the posterior limb of the internal capsule, the thalamus, the parietal cortex, the hippocampus, and the medulla.

Acknowledgment

We received invaluable help from Professor Graeme Bydder.

References

1. Rutherford M, Pennock J, Schwieso J, Cowan F, Dubowitz L. **Hypoxic ischaemic encephalopathy. Early magnetic resonance imaging findings and their evolution.** *Neuropediatrics* 1995;26:183-191
2. Byrne P, Welch R, Johnson M, Darrah J, Piper M. **Serial magnetic resonance imaging in neonatal hypoxic-ischaemic encephalopathy.** *J Pediatr* 1990;117:694-700
3. Rollins N, Morris M, Evans D, Perlman J. **The role of early MR in the evaluation of the term infant with seizures.** *AJNR Am J Neuroradiol* 1994;15:239-248
4. Kuenzle C, Baenziger O, Martin E, et al. **Prognostic value of early MR imaging in term infants with severe perinatal asphyxia.** *Neuropediatrics* 1994;25:191-200
5. Barkovich A. **MR and CT evaluation of profound neonatal and infantile asphyxia.** *AJNR Am J Neuroradiol* 1992;13:959-972
6. Rivkin MJ. **Hypoxic-ischemic brain injury in the term newborn. Neuropathology, clinical aspects and neuroimaging.** *Clin Perinatol* 1997;24:607-625
7. Sarnat H, Sarnat M. **Neonatal encephalopathy following fetal distress.** *Arch Neurol* 1976;33:696-705
8. Landis J, Koch G. **The measurement of observer agreement of categorical data.** *Biometrics* 1977;33:159-174
9. Berry DA. **Statistics: A Bayesian Approach.** Belmont: Duxbury; 1997:196-232
10. Barkovich AJ, Hajnal BL, Vigneron D, et al. **Prediction of neuromotor outcome in perinatal asphyxia. Evaluation of MR scoring systems.** *AJNR Am J Neuroradiol* 1998;19:143-149
11. Rutherford MA, Pennock J, Counsell S, et al. **Abnormal magnetic resonance signal in the internal capsule predicts poor outcome in infants in the hypoxic-ischaemic encephalopathy.** *Pediatrics* 1998;102:323-328