

Neonatal brain MRI: how reliable is the radiologist's eye?

B. Morel^{1,2} · G. Antoni³ · J. P. Teglas³ · I. Bloch² · C. Adamsbaum⁴

Received: 16 July 2015 / Accepted: 6 October 2015 / Published online: 22 October 2015
© Springer-Verlag Berlin Heidelberg 2015

Abstract

Introduction White matter (WM) analysis in neonatal brain magnetic resonance imaging (MRI) is challenging, as demonstrated by the issue of diffuse excessive high signal intensity (DEHSI). We evaluated the reliability of the radiologist's eye in this context.

Methods Three experienced observers graded the WM signal intensity on axial T2-weighted 1.5T images from 60 different premature newborns on 2 occasions 4 weeks apart with a semi-quantitative classification under identical viewing conditions.

Results The intra- and inter-observer correlation coefficients were fair to moderate (Fleiss' kappa between 0.21 and 0.60).

Conclusion This is a serious limitation of which we need to be aware, as it can lead to contradictory conclusions in the challenging context of term-equivalent age brain MRI in premature infants. These results highlight the need for a semiautomatic tool to help in objectively analyzing MRI signal intensity in the neonatal brain.

Keywords Neonatal brain MRI · Signal analysis · DEHSI

Introduction

Perinatal imaging plays a major role in the assessment of human brain development. Magnetic resonance imaging (MRI) is considered the most powerful tool for exploring the anatomy and signal intensity of the neonatal brain. But reliably assessing the shape, volume, and signal intensity of cerebral structures—particularly the white matter (WM)—is a challenge. Unlike in CT imaging, there is no MRI equivalent to the Hounsfield scale to help radiologists quantify and compare signal intensities.

Newborns differ from older populations by their smaller brain size and an inverted white/grey matter contrast on MRI. Brain maturation is a dynamic process that can be charted in vivo via the development of myelination features, which have been described exhaustively using MRI [1, 2]. Grey and white matter components can be analyzed in two complementary ways, based on either morphology or signal intensity. Each radiologist has his or her own way of looking at this exam, as illustrated by the numerous scoring systems for determining the nature and extent of MR imaging abnormalities [3]. The premature newborn brain is a peculiar entity—no longer fetal, but not yet adult. Subjective analysis can be an issue, as demonstrated recently by the debate over diffuse excessive high signal intensity (DEHSI). DEHSI was initially described as regions of “high signal intensity” in the periventricular frontal and parieto-occipital area on T2-weighted images [4, 5], and has been reported in up to 80 % of very premature infants at term-equivalent age [6]. Though its prognostic value is under debate [7–12],

IB and CA contributed equally to this study.

✉ B. Morel
baptistemorelaphp@gmail.com

¹ Pediatric Radiology, A. Trousseau Hospital AHPH, 26 avenue du Dr Arnold Netter, 75012 Paris, France

² LTCI, CNRS, Télécom ParisTech, Université Paris-Saclay, Paris, France

³ INSERM, CESP Centre for Research in Epidemiology and Population Health, U1018, Reproduction and Child Development, F-94807 Villejuif, France

⁴ Faculty of Medicine, Paris Sud University, Pediatric Radiology Department Bicêtre Hospital AHPH, Paris, France



Fig. 1 Visual comparison of the highest signal intensity between circular regions of interest. Comparison 1 was between the periventricular frontal (A) and subcortical frontal (B) white matter. Comparison 2 was between the periventricular frontal (A) and subcortical occipital (C) white matter at term-equivalent age on T2-weighted axial slices

DEHSI are now considered a development-related imaging pattern [8]. This finding can influence medical decisions when associated with other abnormalities such as white matter punctate lesions, in which case it may then have significant ethical implications. Our purpose, however, is to focus on the analysis of white matter signal intensity in T2-weighted MRI to evaluate the reliability of the radiologist's eye, independent of any associated abnormalities.

Materials and methods

Data acquisition

We used axial T2-weighted TSE images from 60 different premature newborns (born between 28 and 29 weeks of

gestation) whose clinical status, transfontanellar ultrasound, and electroencephalogram were all normal. Infants underwent MRI after feeding, swaddling, and placement of ear protection. Parents gave their informed consent for medical use of the MR images. The images were acquired at term-equivalent age (between 39 and 40 weeks of gestation) using routine protocols on a Philips 1.5T Achieva system with an 8-element head coil. The acquisition parameters were as follows: TR=3750 ms; TE=110 ms; Turbo-factor=16. The MRI slices were acquired on a 512×512 pixel matrix, covering the field of view, without interpolation, with a resolution of 2.560 pixels per millimeter (pixel size = 0.39×0.39 mm) and slice thickness = 4 mm. Signal intensity analyses were performed with ImageJ software [13].

Observers

Measurements were performed by three people: two senior radiologists (observer 1 and observer 2, with 3 and 20 years of experience in pediatric neuroimaging, respectively) and one senior computer science researcher (observer 3, with 20 years of experience in medical imaging). The rationale for using this sample of readers was to try to evaluate the role of experience in the specific analysis of white matter. The computer science researcher was used as a “non-clinical” control, who would not be influenced by the clinical context or the particularity of newborn brain MRI contrast.

White matter signal intensity analysis

The three readers compared the relative signal intensity of different circular regions of interest (ROI). In all cases, the ROIs were defined by one of the senior radiologists in the same areas—corresponding to the frontal and occipital periventricular white matter—before each experiment session. The viewing parameters were kept constant for all of the images to avoid any change during or between experiments.

Table 1 Inter-observer Fleiss' kappa agreement for the comparisons 1 and 2 with four class scale

Inter-observer Fleiss' kappa agreement with four class scale	Comparison 1: periventricular frontal white matter vs. subcortical frontal white matter		Comparison 2: periventricular frontal white matter vs. subcortical occipital white matter	
	1st occasion	2nd occasion	1st occasion	2nd occasion
Overall agreement	0.414	0.595	0.604	0.368
Observer 1 / Observer 2	0.25	0.627	0.555	0.222
Observer 1 / Observer 3	0.536	0.631	0.626	0.415
Observer 2 / Observer 3	0.456	0.525	0.635	0.479

Table 2 Intra-observer Fleiss' kappa agreement for the comparisons 1 and 2 with four class scale

Intra-observer Fleiss' kappa agreement with four class scale	Comparison 1: periventricular frontal white matter vs. subcortical frontal white matter		Comparison 2: periventricular frontal white matter vs. subcortical occipital white matter	
	1st occasion	2nd occasion	1st occasion	2nd occasion
Observer 1	0.211		0.213	
Observer 2	0.1		0.301	
Observer 3	0.366		0.402	

Comparison 1 was between the highest signal intensity of the periventricular frontal white matter and the subcortical frontal white matter (Fig. 1). Comparison 2 was between the same highest signal intensity of the periventricular frontal white matter and the subcortical occipital white matter (Fig. 1). These two comparisons were done twice, 4 weeks apart, to test the intra-observer variability.

The following semi-quantitative classification was used:

- 1 = Periventricular frontal white matter displays lower signal intensity than the subcortical frontal white matter.
- 0 = No difference in signal intensity.
- 1 = Periventricular frontal white matter displays slightly higher signal intensity than the subcortical frontal white matter.
- 2 = Periventricular frontal white matter displays much higher intensity signal than the subcortical frontal white matter.

Statistical analysis

Intra- and inter-observer agreements on the white matter signal intensity comparisons were assessed using Fleiss' kappa coefficient, where agreement is considered “fair” if κ is between 0.21 and 0.40, “moderate” if κ is between 0.41 and 0.60, “substantial” if κ is between 0.61 and 0.80, and “almost perfect” if κ is between 0.81 and 1. Statistical analysis was performed using R software [14].

Results

Four-category classification system

The overall inter-observer Fleiss' kappa agreement was moderate for the first comparison, between the periventricular frontal and subcortical frontal white matter signal (Table 1).

The overall inter-observer Fleiss' kappa agreement was fair to moderate for the second comparison, between the periventricular frontal and subcortical occipital white matter signal (Table 1).

The intra-observer Fleiss' kappa agreement was fair for two observers and poor to fair for the third (Table 2).

Three-category classification system

We further analyzed our results to investigate the impact of distinguishing grades 1 and 2 by merging them into one category. The resulting Fleiss' kappa agreement is shown in Tables 3 and 4.

The overall inter-observer Fleiss' kappa agreement was fair to substantial for the first comparison, between the periventricular frontal and subcortical frontal white matter signal.

The overall inter-observer Fleiss' kappa agreement was fair to substantial for the second comparison, between the periventricular frontal and subcortical occipital white matter signal.

The intra-observer Fleiss' kappa agreement was fair to moderate for two observers and moderate for the third.

Table 3 Inter-observer Fleiss' kappa agreement for the comparisons 1 and 2 with three class scale

Inter-observer Fleiss' kappa agreement with three class scale	Comparison 1: periventricular frontal white matter vs. subcortical frontal white matter		Comparison 2: periventricular frontal white matter vs. subcortical occipital white matter	
	1st occasion	2nd occasion	1st occasion	2nd occasion
Overall agreement	0.344	0.625	0.72	0.407
Observer 1 / Observer 2	0.169	0.667	0.732	0.276
Observer 1 / Observer 3	0.587	0.7	0.732	0.461
Observer 2 / Observer 3	0.155	0.506	0.696	0.494

Table 4 Intra-observer Fleiss' kappa agreement for the comparisons 1 and 2 with three class scale

Intra-observer Fleiss' kappa agreement with three class scale	Comparison 1: periventricular frontal white matter vs. subcortical frontal white matter	Comparison 2: periventricular frontal white matter vs. subcortical occipital white matter
Observer 1	0.439	0.383
Observer 2	0.129	0.411
Observer 3	0.498	0.463

Discussion

Brain MRI provides fine details of the newborn brain using multiple sequences. Recent debate on the significance of DEHSI in the white matter [4, 5] of term-equivalent age premature infants deserves special attention for ethical reasons. In our experience, however, the assessment of such white matter “high signal intensity” is somewhat subjective, and inter-observer variability may be underestimated. While it has been shown that the detectability of low-contrast lesions can be affected by retinal photoreceptor light adaptation [15], to our knowledge there is no published work evaluating observer variability in interpreting newborn brain MRI. The scale chosen for the comparison is similar to that used in previously published studies [11, 16]. Our results show fair to substantial intra- and inter-observer agreements in the analysis of white matter signal intensity. If we simplify our scale by combining grades 1 and 2 to reduce subjectivity, we see a slight improvement in the inter- and intra-observer Fleiss' kappa agreements, as might be expected. But the overall results are similar, highlighting the difficulty of visual intensity signal analysis, whatever the scale used. Though the difficulties in interpreting Kappa statistics are well known, they are still commonly used to quantify inter-observer agreement, even in recent publications [3, 5]. In our study, all experiments were done twice under strictly identical viewing conditions by all observers simultaneously to avoid any bias. We were thus able to study inter- and intra-observer variability and obtain more robust results. One previous study focused specifically on the appearance of DEHSI on different T2-weighted sequences (Fast Spin Echo and Single Shot FSE) [17], and concluded that the appearance of DEHSI on MR images following preterm birth is highly subjective, with slightly low intra- and inter-observer agreement (intraclass correlation of 0.04). Though our results showed better agreement for analysis of WM intensity, this is an important limitation of the visual assessment task, and one we should be aware of. Another difficulty is the subjective identification of the highest WM signal intensity, which might explain the moderate inter-observer agreement. A

great deal of caution is needed when drawing conclusions about WM signal intensity, and these results highlight the need for a semi-automatic tool to make signal intensity analysis in the neonatal brain more objective.

Acknowledgments The authors thank N. Friedman for English language assistance and N. Girard for her friendly and contributive support.

Compliance with ethical standards We declare that all human studies have been approved by the local Ethics Committee and have therefore been performed in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki and its later amendments. We declare that all patients gave informed consent prior to inclusion in this study.

Conflict of interest We declare that we have no conflict of interest.

References

- Girard N, Raybaud C, du Lac P (1991) MRI study of brain myelination. *J Neuroradiol* 18:291–307
- Barkovich AJ (1998) MR of the normal neonatal brain: assessment of deep structures. *AJNR Am J Neuroradiol* 19:1397–1403
- Kidokoro H, Neil JJ, Inder TE (2013) New MR imaging assessment tool to define brain abnormalities in very preterm infants at term. *AJNR Am J Neuroradiol* 34:2208–2214
- Counsell SJ, Shen Y, Boardman JP et al (2006) Axial and radial diffusivity in preterm infants who have diffuse white matter changes on magnetic resonance imaging at term-equivalent age. *Pediatrics* 117:376–386
- Counsell SJ, Allsop JM, Harrison MC et al (2003) Diffusion-weighted imaging of the brain in preterm infants with focal and diffuse white matter abnormality. *Pediatrics* 112:1–7
- Dyet LE, Kennea N, Counsell SJ et al (2006) Natural history of brain lesions in extremely preterm infants studied with serial magnetic resonance imaging from birth and neurodevelopmental assessment. *Pediatrics* 118:536–548
- Jeon TY, Kim JH, Yoo S-Y et al (2012) Neurodevelopmental outcomes in preterm infants: comparison of infants with and without diffuse excessive high signal intensity on MR images at near-term-equivalent age. *Radiology* 263:518–526
- de Bruïne FT, van den Berg-Huysmans AA, Leijser LM et al (2011) Clinical implications of MR imaging findings in the white matter in very preterm infants: a 2-year follow-up study. *Radiology* 261:899–906
- Woodward LJ, Anderson PJ, Austin NC et al (2006) Neonatal MRI to predict neurodevelopmental outcomes in preterm infants. *N Engl J Med* 355:685–694

10. Parikh NA, He L, Bonfante-Mejia E et al (2013) Automatically quantified diffuse excessive high signal intensity on MRI predicts cognitive development in preterm infants. *Pediatr Neurol* 49:424–430
11. Leitner Y, Weinstein M, Myers V et al (2014) Diffuse excessive high signal intensity in low-risk preterm infants at term-equivalent age does not predict outcome at 1 year: a prospective study. *Neuroradiology* 56(8):669–678
12. Van't Hooft J, van der Lee JH, Opmeer BC et al (2015) Predicting developmental outcomes in premature infants by term equivalent MRI: systematic review and meta-analysis. *Syst Rev* 4:71
13. Schneider CA, Rasband WS, Eliceiri KW (2012) NIH Image to ImageJ: 25 years of image analysis. *Nat Meth* 9:671–675
14. Team RC (2012) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0
15. Baxter B, Ravindra H, Normann RA (1982) Changes in lesion detectability caused by light adaptation in retinal photoreceptors. *Investig Radiol* 17:394–401
16. Calloni S, Cinnante C, Bassi L et al (2015) Neurodevelopmental outcome at 36 months in very low birth weight premature infants with MR diffuse excessive high signal intensity (DEHSI) of cerebral white matter. *Radiol Med (Torino)* 1–8
17. Hart AR, Smith MF, Rigby AS et al (2010) Appearances of diffuse excessive high signal intensity (DEHSI) on MR imaging following preterm birth. *Pediatr Radiol* 40:1390–1396