Pitfalls in using neonatal brain NAA to predict infant development

Peter Lally and colleagues[1] assessed the relationships between N-acetylaspartate (NAA) concentration and cognition, language, and motor scores on the Bayley Scales of Infant and Toddler Development, third edition (Bayley-III). They present conventional MRI and MR spectroscopy data from 223 infants, 190 of whom were followed up after mild (n=37), moderate (n=163), or severe (n=23) neonatal encephalopathy treated with therapeutic hypothermia. Infants were assessed at age 18–24 months. Moderate or severe disability are defined in the Article as composite cognitive and language scores of less than 70 (severe disability) and 70–84 (moderate disability) and any of the following: blindness, hearing impairment, or Gross Motor Function Classification System level of at least 2. The primary outcome was the association between MR biomarkers and an adverse neurodevelopmental outcome, defined as death or moderate or severe disability.

The study's main conclusion is based on the 82 infants in whom thalamic concentration of NAA was measured within 14 days after birth, which the authors maintain "can accurately predict neurodevelopment 2 years after neonatal encephalopathy" and "can be applied to increase the power of neuroprotection trials while reducing their duration".[1,2] However, I argue that these conclusions are far too strong, and that the authors' presentation of the relevant statistics is misleading.

	Bailey-III cognitive or language score	
	<85	≥85
Thalamic N-acetylaspartate concentration, mmol/kg		
<5·5	11	1
>5·5	14	56

Bailey-III cognitive or language scores <85 indicate adverse neurodevelopmental outcomes and scores ≥85 indicate normal neurodevelopment.

Table: **Number of infants categorised by thalamic N-acetylaspartate concentration and Bayley-III scores**

In 70 infants, NAA concentration is greater than 5·5 mmol/kg, and their Bayley-III scores are between the lower levels of 50–70 and upper levels of 120–150. There is no indication that Bayley-III increases with increasing NAA concentration in the plots presented in the Article. Linear regressions give slight increases in the motor and cognitive score plots and a slight decrease in the language score plot, but the standard deviations of residuals are large. It appears impossible, on the basis of NAA concentration alone, to discriminate between neonates who will have poor Bayley-III scores and who might perhaps have cerebral palsy 2 years later, and neonates who will have good Bayley-III scores at 2 years.

I have done my own calculations (table) using the data for the 82 individuals in the Article.[1] The authors claim that 12 of the 82 infants had adverse outcomes 2 years later. From their figures, and using their definition of adverse outcome, I can identify at least 25 infants with adverse outcomes, 14 of them among the 70 infants with NAA concentration greater than 5·5 mmol/kg. Only 12 infants had NAA concentrations of less than 5·5 mmol/kg. 11 of these had very poor Bayley-III scores, ten of which had the lowest score achievable for the test. One infant with NAA concentration of about 4·8 mmol/kg had very high (144) Bayley-III cognitive scores and scores greater than 100 in the other two domains. Thus, at least 20% of the infants with an NAA concentration greater than 5·5 mmol/kg had moderate or severe disability 2 years later.

These numbers definitely do not support the statistical statements in the Article of "area under the curve [AUC] of 0·99 [95% CI 0·94–1·00]; sensitivity 100% [74–100]; specificity 97% [90–100]; n=82". Thus, thalamic NAA concentration acquired soon after birth in infants with neonatal encephalopathy is not a sufficient predictor of neurodevelopment outcome 2 years later. My conclusion is that neurodevelopmental follow-up at age 2 years or later will continue to be necessary to provide a measure of outcome in neonatal trials.

I declare no competing interests.

Lars Walløe
lars.walloe@medisin.uio.no

Division of Physiology, Institute of Basic Medical Sciences, University of Oslo, Oslo NO-0317, Norway

1 Lally PJ, Montaldo P, Oliveira V, et al. Magnetic resonance spectroscopy assessment of brain injury after moderate hypothermia in neonatal encephalopathy: a prospective multicentre cohort study. *Lancet Neurol* 2019; **18:** 35–45.

2 Gunn AJ, Battin M. Towards faster studies of neonatal encephalopathy. *Lancet Neurol* 2019; **18:** 21–22.

Authors' reply

We thank Lars Walløe for his subgroup analysis of the Bayley Scales of Infant and Toddler Development, third edition (Bayley-III) score scatter plot in our study[1]. Unfortunately, there are errors in his digitalisation, data analysis, and interpretation due to a misunderstanding of the primary outcome of our study.

In the MARBLE study, we defined adverse outcome as death or moderate or severe disability at age 18–24 months.[2] We defined severe disability as any of the following: Bayley-III composite cognitive and language scores less than 70, Gross Motor Function Classification System (GMFCS) levels 3–5, hearing impairment requiring hearing aids, or blindness. We defined moderate disability as composite cognitive and language scores of 70–84, and any of the following: GMFCS level 2, hearing impairment without the need for amplification, or a persistent seizure disorder. This is a widely accepted definition and has been used in all major cooling trials.[3] We provided the data for N-acetylaspartate (NAA) concentration in our 2019 Article.[1]

Walløe defined adverse outcome as Bayley-III language or cognitive scores less than 85 and used a different threshold for NAA concentration in his reanalysis. Hence, it is not surprising that he obtained different results.

As explained in the Article, although NAA concentration had a positive association with Bayley-III composite cognitive (r^2=0·42), language (r^2=0·55), and motor (r^2=0·45) scores, the relationship tailed off at higher values. Thus, it is not surprising that Walløe's subgroup analysis of babies with normal Bayley scores did not find a relationship with NAA concentration. Although brain injury reduces NAA concentration, higher than normal NAA concentration does not indicate advanced development. In fact, this might be a sign of a degenerative disease such as Canavan's disease.[4] It is neither required nor desirable for any surrogate biomarker to have a linear relationship with Bayley scores, as Bayley scores at age 2 years do not have a linear correlation with later childhood outcomes.[5]

The key innovation of the MARBLE study is the development of a cross-platform MR spectroscopy technique, which can be used across all common 3T MR scanners in clinical use. We validated the prognostic accuracy of this MR sequence in the prospective multicountry study on MR biomarkers (MARBLE), over a 6-year period. These data suggest that thalamic NAA concentration has a good prognostic accuracy (AUC 0·99 [95% CI 0·94–1·00]), far higher than amplitude-integrated electroencephalography 0·73 (0·65–0·79) and conventional MRI 0·82 (0·76–0·87).

Our previous meta-analysis, involving more than 800 babies, had already shown that thalamic MR spectroscopy biomarkers had far higher prognostic accuracy than conventional MRI.[6] However, none of these studies used MR sequences that were harmonised across commonly used clinical scanners. Without such sequences, MR spectroscopy loses its potential as a powerful quantitative tool, because the results from two different scanners cannot be directly compared. The MARBLE sequences allow direct comparison of quantitative MR spectroscopy across sites and make these measurements accessible on common clinical 3T MR scanners.

We are not suggesting that MR spectroscopy or any surrogate biomarker should replace neurodevelopmental and cognitive assessments in childhood. Clinical outcome (death or disability) should remain the gold standard in phase 3 clinical trials and for health policy decisions. However, MR spectroscopy biomarkers should be used in phase 2 clinical trials of new neuroprotective drugs and to provide a yes or no decision before proceeding to an expensive and large phase 3 clinical trial. Phase 2 trials using thalamic NAA concentration can be completed rapidly and would need to recruit fewer than 100 babies, whereas an adequately powered phase 3 trial with a clinical endpoint would need to recruit more than 600 babies.

MR spectroscopy thalamic NAA concentration acquired using the MARBLE sequences has the potential to produce a paradigm shift in the way we do neonatal neuroprotection trials. Funding bodies might be able to ensure that only therapies that improve neonatal MR biomarkers are funded for further assessment in phase 3 trials—this optimises public resources and maximises the likelihood of translating effective therapies into clinical practice.

We declare no competing interests.

*Paolo Montaldo, Ravi Swamy, Paul Bassett, Peter J Lally, Seetha Shankaran, *Sudhin Thayyil*
s.thayyil@imperial.ac.uk

Centre for Perinatal Neuroscience, Imperial College London, London W12 0HS, UK (PM, RS, PJL, ST); StatsConsultancy, Amersham, UK (PB); Wayne State University, Detroit, MI, USA (SS)

1 Lally PJ, Montaldo P, Oliveira V, et al. Magnetic resonance spectroscopy assessment of brain injury after moderate hypothermia in neonatal encephalopathy: a prospective multicentre cohort study. *Lancet Neurol* 2019; **18:** 35–45.

2 Lally PJ, Pauliah S, Montaldo P, et al. Magnetic Resonance Biomarkers in Neonatal Encephalopathy (MARBLE): a prospective multicountry study. *BMJ Open* 2015; **5:** e008912.

3 Shankaran S, Laptook AR, Pappas A, et al. Effect of depth and duration of cooling on death or disability at age 18 months among neonates with hypoxic-ischemic encephalopathy a randomized clinical trial. *JAMA* 2017; **318:** 57–67.

4 Moffett JR, Ross B, Arun P, Madhavarao CN, Namboodiri AM. N-Acetylaspartate in the CNS: from neurodiagnostics to neurobiology. *Prog Neurobiol* 2007; **81:** 89–131.

5 Burakevych N, McKinlay CJ, Alsweiler JM, Wouldes TA, Harding JE. Bayley-III motor scale and neurological examination at 2 years do not predict motor skills at 4·5 years. *Dev Med Child Neurol* 2017; **59:** 216–23.

6 Thayyil S, Chandrasekaran M, Taylor A, et al. Cerebral magnetic resonance biomarkers in neonatal encephalopathy: a meta-analysis. *Pediatrics* 2010; **125:** e382–95.