Proton Magnetic Resonance Spectroscopy Lactate/N-Acetylaspartate within 2 weeks of birth accurately predicts 2-year motor, cognitive and language outcomes in Neonatal Encephalopathy after Therapeutic Hypothermia

Subhabrata Mitra[1] (MB BS, DCH), Giles S Kendall[2] (BSc, MB BS, PhD), Alan Bainbridge[3] (MPhys, PhD), Magdalena Sokolska[3] (MPhys, PhD), Mary Dinan[2] (MA, BSc, DiPH), Cristina Uria-Avellanal[1] (MB BS), David Price[3] (M Phys), Katie Mckinnon[2] (MB BS, BA), Roxanna Gunny[2] (MB BS, FRCR), Angela Huertas-Ceballos[2] (MD, MSc), Xavier Golay[4] (PhD), Nicola J Robertson (MB ChB, PhD)[1,5]

1. Institute for Women's Health, University College London, London, UK
2. Women and Children, University College London NHS Foundation Trust
3. Medical Physics and Engineering, University College London NHS Foundation Trust
4. Institute of Neurology, Queen Square, University College London
5. Division of Neonatology, Sidra Medicine, Doha, Qatar

Corresponding author:
Nicola J Robertson,
Institute for Women's Health
University College London,
74 Huntley Street,
London
WC1E 6HX
E mail: n.robertson@ucl.ac.uk

Original research

Competing interest: none declared

Contribution from Authors: SM, GK, KM, CU recruited the babies and were the clinicians in charge of the clinical safety of the baby during the MRI/MRS study. SM wrote the first draft of the paper. MD assisted with the recruitment and care of the babies undergoing MRI. AB and MS performed the MRS analysis and the MD assessment and interpreted the data. DP did the TBSS analysis and interpreted the data. RG scored the MRI images. XG provided advice on the MRI/MRS analysis. A H-C did the Bayley III assessment of the babies in the study. NJR designed and supervised the study and wrote the final drafts of the paper. All authors have been involved in finalizing the manuscript and have seen and approved the final version.

Abstract (250 WORDS)

Objective: Brain proton (^{1}H) magnetic resonance spectroscopy (MRS) lactate/N-Acetylaspartate (Lac/NAA) peak area ratio is used for prognostication in Neonatal Encephalopathy (NE). At 3-Tesla in NE babies, the objectives were to assess: (i) sensitivity and specificity of basal ganglia and thalamus (BGT) ^{1}H MRS Lac/NAA for prediction of Bayley III outcomes at 2-years using optimized metabolite fitting (Tarquin) with threonine and total NAA; (ii) prediction of motor outcome with diffusion-weighted MRI; iii) BGT Lac/NAA correlation with the NICHD MRI score.

Materials and methods: 55 (16 inborn, 39 outborn) NE infants at 39w+5d (35w+5d-42w+0d) admitted between February 2012 and August 2014 to UCH for therapeutic hypothermia underwent MRI and ^{1}H MRS at 3T on day 2-14 (median day 5). MRIs were scored. Bayley III was assessed at 24 (22-26) months.

Results: Sixteen babies died (1 inborn, 15 outborn); 20, 19 and 21 babies had poor motor, cognitive and language outcomes. Using a threshold of 0.39, sensitivity and specificity of BGT Lac/NAA for 2-year motor outcome was 100% and 97%, cognition 90% and 97% and language 81% and 97% respectively. Sensitivity and specificity for motor outcome of mean diffusivity (MD; threshold 0.001 mm^2/s) up to day 9 was 72% and 39% and fractional anisotropy (FA; threshold 0.198) was 100%, and 94% respectively. Lac/NAA correlated with BGT injury on NICHD scores (2A, 2B, 3).

Conclusions: BGT Lac/NAA on ^{1}H MRS at 3T within 14 days accurately predicts 2-year motor, cognitive and language outcome and may be a marker directing decisions for therapies after cooling.

Background

Intrapartum-related neonatal encephalopathy (NE) is a major health care problem across the world.[1] The incidence of NE in Western Europe and North America is around 1-3/1000 term births depending on the definitions used.[2] Although improvements in care might prevent NE and neonatal death in some cases,[3] many cases cannot be prevented and therapies are limited. Therapeutic hypothermia (HT) initiated within 6h of birth improves outcome, yet despite this therapy 44-53% of infants with NE die or suffer moderate to severe disabilities including cerebral palsy, developmental delay, epilepsy and visual impairment.[4]

Magnetic resonance imaging (MRI) is the imaging modality of choice for assessment of injury pattern, severity and prognostication in NE with an optimal timing between 5 and 14 days. The pattern and severity of the National Institute of Child Health and Human Development (NICHD) Neonatal Research Network (NRN) MRI score[5] correlates with the outcome of death or disability and with disability among survivors. In a nested MRI study in the TOBY cooling trial, major MRI abnormalities in the first 4 weeks after birth predicted death or severe disability at 18 months in cooled babies with a sensitivity of 88% and a specificity 82%.[6] Proton (^{1}H) magnetic resonance spectroscopy (MRS) is increasingly used as an independent quantitative tool for clinical prognostication in NE and has been used in the TOBY-Xenon clinical neuroprotection trial as a surrogate outcome measure.[7] ^{1}H MRS in the neonatal brain has prominent signal peaks related to the presence of N-Acetylasparate (NAA), choline, creatine, and lactate (Lac). Metabolite ratios (the signal amplitude of one metabolite versus another) are typically calculated for voxels positioned in the basal ganglia and thalamus (BGT). In the pre-cooling era, a meta-analysis demonstrated that a Lac/NAA threshold of 0.29 (0.24 to 0.4) at 1.5T had a sensitivity of 82% and specificity 95% for predicting an abnormal outcome[8] and was more accurate than conventional MRI and diffusion weighted imaging (DWI).

Interest is now moving towards the tertiary phase of injury with therapies that can reduce neuroinflammation and improve regeneration and repair.[9] A robust marker of brain injury and prediction of outcome following HT will be essential to detect babies who are likely to have adverse outcomes despite having had cooling therapy. An increased lactate and reduced NAA on MRS (translating to a high Lac/NAA peak area ratio) suggest brain mitochondrial impairment and impaired oxidative metabolism which may be amenable to late therapies. Unlike mean diffusivity, which pseudonomalizes around day 10 in babies who have been cooled,[10] Lac/NAA remains stable over the first 2 weeks after birth and has potential to be used as a quantitative marker of brain injury after HT. Widespread hospital access to 3T MRI is now possible; a higher field strength as well as improved MRS analysis techniques are likely to further strengthen accuracy of Lac/NAA.

^{1}H MRS measurements of brain Lac have focused on the methyl proton resonance at 1.31 ppm with an optimized long echo time (TE) to reduce contamination with lipid signal.[11] The methyl group of threonine (Thr), an essential amino acid present in the brain, also resonates close by at approximately 1.32 ppm.[12] The proximity of the methyl resonances of Lac and Thr mean that with conventional MRS they are not independently resolvable in *in vivo* spectra.[13] A potential benefit of including threonine, in addition to lactate, in the spectral analysis is improvement in the spectral fit in the region around 1.3ppm in cases where lactate is raised. Similarly, including both NAA and NAA-glutamate (NAAG, an abundant neurotransmitter released from axonal terminals with neuronal activity and hydrolyzed to NAA and glutamate; NAA+NAAG = total NAA (tNAA)) improves fitting in the region around 2.0ppm at field strength of 3T.[14]

The objectives of this study were, in NE babies who have been cooled, to assess: (i) basal ganglia and thalamus (BGT) 1H MRS Lac+Thr/tNAA sensitivity and specificity at 3T for motor, cognitive and language Bayley III outcomes at 2 years; (ii) prediction of motor outcome of Lac+Thr/tNAA versus other MRI biomarkers (MD and FA) before pseudonormalization; iii) BGT Lac+Thr/tNAA correlation with the NICHD MRI score.

Subjects and Methods

Approval was given from the University College London Hospital Research Ethics Committee for this retrospective service improvement project. Anonymized data that is routinely collected in our centre was used and requirement for informed consent was waived.

Infants treated with therapeutic hypothermia for moderate to severe NE in University College London Hospitals between February 2012 and August 2014 were reviewed. Infants with perinatal stroke, major congenital malformations and metabolic or congenital infection were excluded from the study. Neurodevelopmental follow up data at 2 years were available in all 58 term newborn infants who were cooled. Of the 58 infants 3 were excluded as they were part of another multicentre trial protocol. Fifty-five term neonates with a mean (range) gestational age (GA) of 39+5 (35+5 - 42+0) weeks and a birth weight of mean (standard deviation; +sd) 3260 (566) g were included. Thirty-nine infants were inborn and 16 were outborn. These neonates fulfilled criteria for HT based on the National Institute for Clinical Excellence criteria.[15] In 5 deliveries, there was uterine rupture and in 2 deliveries there was placental abruption. The mean +sd arterial gas (cord or within 60 minutes of birth) was 6.9 +0.2 and base deficit -15 +7. Mean Apgar scores 2 +2 at 1 minute, 4 +2 at 5 minutes and 6 +3 at 10 minutes. All neonates required resuscitation at birth. Seven infants had a birthweight <10th centile. All infants received intravenous morphine sulphate (up to 30 microgram/kg/hr) during therapeutic hypothermia. All infants were monitored with continuous video EEG throughout cooling and rewarming. Following clinical or electrographic seizure activity, infants received phenobarbitone as first line (up to 40 mg/kg total), phenytoin as second line (20 mg/kg) and midazolam (bolus followed by infusion) as third line anti-epileptic drug. Seizure burden was not assessed as part of this study.

Magnetic resonance imaging and spectroscopy

Magnetic resonance imaging

Scanning was performed at a median (range) age of 5 (2-14) days. In 47 infants the MRI was performed after the 72h HT and following rewarming to normothermia; in 8 infants the MRI was done during cooling to inform the direction of care. Scanning was performed either during natural sleep or after sedation with a morphine infusion. Most neonates (48) were ventilated during scanning and were transported and studied in a Lammers Medical MR Conditional transport incubator (LMT Medical Systems GmbH, Luebeck, Germany) with gentle head restraint. Throughout the examination, neonates were monitored by using MR-compatible pulse oximetry and electrocardiogram and supervised by an experienced neonatologist trained in clinical MR imaging safety.

MRI was performed on a 3T MR system (Philips Medical Systems, Best, The Netherlands). The scanning protocol included T1-weighted imaging (Inversion-prepared 3D gradient echo readout: TI= 1465ms, TR =17ms, TE = 4.6ms, sagittal slice thickness = 1mm, in-plane resolution = 0.82 x 0.97 mm), T2-weighted imaging (Coronal and axial, Turbo Spin Echo: Echo Train length = 11, TR =10,721ms, TE = 130ms, slice thickness = 3mm, in-plane resolution = 0.50 x 0.52 mm), DTI (32 directions, b = 750, EPI readout: TR = 7500ms, TE = 49ms, slice thickness = 2mm, in-plane resolution = 2.0 x 2.04 mm).

Images were scored using the NICHD NRN MRI[5] score by RG who was blind to the perinatal history and clinical outcome. NICHD NRN MRI score has been validated to predict death or IQ at 6-7 years of age following hypothermia for NE. The score describes pattern 0 (normal MRI), 1A (minimal cerebral lesions), 1B (extensive cerebral lesions), 2A (basal ganglia thalamic, anterior or posterior limb of internal capsule, or watershed infarction), 2B (2A with cerebral lesions) and pattern 3 (hemispheric devastation).

Magnetic resonance spectroscopy

MRS was performed using PRESS (15x15x15 mm voxel position on the left thalamus / basal ganglia, TR=2288ms, TE=288ms, a dynamic series of 16 spectra were acquired each with 8 averages). The dynamic spectra were summed offline after phase and frequency correction and rejection of motion-corrupted data. MRS analysis was performed using Tarquin.[16] The basis set included Threonine (Thr),[11] but did not include lipids or macromolecules. When the signal around 1.3ppm is high we have observed that the spectral fit in this region often has a significant residual. Including Thr in the basis set typically improves the fit in this region of the spectrum (**Figure 1**). The ratio of lactate (Lac) plus Thr to total NAA (N-acetylaspartate plus N-acetylaspartate glutamate), (Lac+Thr)/tNAA was calculated using the amplitudes of the fitted components. The position of the voxel and representative ^{1}H MRS spectra in normal and abnormal outcomes are shown in **Figure 2.**

Mean Diffusivity and Fractional anisotropy

The DTI volumes were analyzed using tools from the FSL brain imaging software library[17] and the DTI-TK toolbox.[18] Non-brain tissues were first removed using image segmentation software and the FSL-EDDY tool used to correct the DTI volumes for eddy current induced distortions in the image and subject movement. After this the diffusion datasets were inspected and those with severe un-correctable motion artifacts were rejected from further analysis. There were 49 usable datasets. These were then processed using the FSL-FDT toolbox, which contains tools to fit the diffusion tensors from the DTI data and calculate mean diffusivity (MD) and fractional anisotropy

(FA) maps. MD and FA values for the BGT **(Figure 3)** were segmented using the atlas described below.

Regional analysis

Regional MD and FA were computed in BGT using BGT masks generated automatically based on both atlas based probabilistic tissue segmentations of 5 tissue classes[19] and a joint multi-atlas label propagation and fusion of 50 manual segmentations of neonatal brain regions.[20] T2W atlas images were non-linearly registered to T2W image of each neonate, and the probabilistic maps and parcellations were propagated to individual subject space using the Nifty-Reg package.[21] Probability maps were used to segment T2W images by the AdaPT[22] algorithm using Nifty-Seg package.[23] T2W subject images were co-registered to the MD maps using rigid registration, with tissue segmentations and brain parcellations propagated to MD space and resampled using point spread function matching.[24] BGT mask was created by combining propagated parcellations of the thalami, sub-thalamic, caudate and lentiform nuclei.[20] The internal capsule was not defined as part of the segmentation and no additional steps were undertaken to separate this structure. All generated segmentations were checked visually for correct anatomical correspondence. Voxels that were obviously mis-assigned (typically within vessels and less than 10 voxels per dataset) were manually edited if necessary. Finally, MD and FA were quantified within BGT segmentation mask.

Tract based Spatial Statistics (TBSS)

Tract based Spatial Statistics (TBSS) is a method allowing for the voxel-wise statistical analysis of the WM tract anatomy between subjects.[25] It is therefore distinct from the measurement of FA within a specific brain region such as the BGT as described above. TBSS was performed on the corrected DTI volumes using an integration of the DTI-TK diffusion toolbox with the FSL-TBSS pipeline. This aimed to combine spatial normalization provided by DTI-TK together with voxel-wise statistical inference for white matter anatomy provide by TBSS. DTI-TK allows the tensor registration of each subject to a population specific template.

An initial template was formed from 5 well-aligned datasets. This was then refined through three stages of alignment; *rigid*, *affine* and *deformable*. Tensor based registration has been shown to provide an improved alignment of WM tracts compared to using the derived FA maps in the standard TBSS pipeline.[26] A population specific template also may avoid biases associated with templates based on the most typical subject. The template and spatially normalized data sets were then used to create a mean FA skeleton and a 4D FA map respectively. These were introduced into the final stages of the standard TBSS pipeline. This allowed for the voxel-wise analysis of white matter FA between outcome groups (normal versus abnormal motor outcome). The FA

threshold used in this study was 0.15. No additional steps were taken to remove non-white-matter voxels from the FA skeleton.

Clinical data and outcome

Neurodevelopmental outcome was assessed at 3, 6, 12 and 24 months as per local and regional protocols. At 24 (22-26) months, the Bayley Scales of Infant Development 3rd edition was used for assessment (Bayley III).[27] The Bayley III has five scales (Cognitive, Language (Receptive and expressive), Motor (Fine and Gross motor), Social Emotional and Adaptive Behavior). The Bayley III independently assesses cognitive and language skills (previous Bayley editions combined cognitive and language into one mental scale). Cognition's raw score is transformed into a composite score. Language and Motor raw scores are analyzed independently for each subscale or as a composite score. Death or a composite score of <85 on language, motor or cognitive scales at 2 years of age was considered abnormal.

Statistical analysis

The values of (Lac+Thr)/tNAA peak area ratio are not normally distributed in this cohort. Therefore, in order to use parametric statistics on this dataset, (Lac+Thr)/tNAA was transformed to $\log_{10}[(\text{Lac+Thr})/\text{tNAA}]$ prior to analysis.

The relationship between $\log_{10}[(\text{Lac+Thr})/\text{tNAA}]$ and the probability of motor, cognitive and language outcome was assessed using logistic regression modelling to each outcome separately. For each model, the value of $\log_{10}[(\text{Lac+Thr/tNAA})]$ that gave the optimum sensitivity and specificity was chosen for classification of 'abnormal' for values above the threshold. From this value, the optimum threshold of (Lac+Thr)/tNAA peak area ratio for classification of 'abnormal' could then be calculated. The relationship between MD and also white matter FA and the probability of motor outcome was also assessed using logistic regression up to 9 days (before pseudonormalization) and for all scans (up to 14 days).

One-way analysis of variance (ANOVA) was used to compare (i) mean Lac+Thr/tNAA levels in 'abnormal' versus 'normal' outcome groups for each of motor, language and cognitive outcomes, and (ii) mean Lac+Thr/tNAA levels across NICHD MRI classification groups 2A, 2B and 3.

Voxel-wise analysis of white matter FA, performed using TBSS, is presented by color-coding voxels with a significant difference (p<0.05) in FA between outcome groups.

Results

Of the 55 infants included, 16 infants died (one inborn, 15 outborn), 20 infants had abnormal motor outcome, 19 had abnormal cognitive outcome and 21 had abnormal language outcome by 2 years of age **(Table 1).** There were 18 infants who had abnormal outcomes in all three domains (motor, language and cognitive). Two other infants had abnormal motor outcome (total 20) without any cognitive or language problems. One other infant had an abnormal cognitive outcome (total 19), without any other motor or language problem. Three other infants had an abnormal language outcome (total 21) without any motor or cognitive problems.

MRI
NICHD MRI score was 0 in 20 babies, 1A in 7 babies, 1B in 3 babies, 2A in 3 babies, 2B in 7 babies, 3 in 15 babies.

1H MRS
The Lac+Thr/tNAA and $\log_{10}[(Lac+Thr)/tNAA]$ ratios by age at scan are shown in **Supplementary Figure 1.** Using a $\log_{10}[(Lac+Thr)/tNAA]$ cut off threshold of -0.41, the sensitivity and specificity for predicting motor outcome in the Bayley III assessment was 100% and 97% respectively with an AUC of 0.997 **(Figure 4A).** This threshold predicted cognitive outcome with sensitivity of 90% & specificity of 97%, AUC of 0.967 **(Figure 4B)** and language outcome with a sensitivity of 81% & specificity of 97%, AUC of 0.913 **(Figure 4C) (Table 2).** Comparing the group means of the normal and abnormal outcomes grouped by motor, cognitive and language outcomes respectively, there were significant differences in the $\log_{10}[(Lac+Thr)/tNAA]$ ($p<0.0001$ for all comparisons) **(Figure 4D, E, F).** The $\log_{10}[(Lac+Thr)/tNAA]$ threshold of -0.41 equates to an untransformed (Lac+Thr)/tNAA threshold of 0.39.

MD and FA
For all scans up to day 14, BGT MD predicted motor outcome with a sensitivity 65% and specificity 100% with an AUC of 0.769. BGT FA predicted motor outcome with a sensitivity 45% and a specificity 91% with an AUC 0.544 **(Table 2A)**. Some correlation existed between Lac+Thr/tNAA and BGT MD (R^2 0.51) but not with BGT FA ($R^2 = 0.01$).

Limiting the analysis to scans done up to day 9 after birth, before the return of MD to normal (pseudonormalization), BGT MD predicted motor outcome with a sensitivity 72% and specificity 100% with an AUC of 0.819 **(Table 2B, Figure 5A).** BGT FA predicted motor outcome with a sensitivity 39% and a specificity 94% with an AUC 0.507 **(Figure 5B)**. Comparing the means of the normal and abnormal groups for motor outcomes, there was a significant difference between MD

(p=0.001) **(Figure 5C)** but not for FA (p=0.707) **(Figure 5D)**.

Relation between Lac+Thr/tNAA and MRI score

Analysis of variance to compare BGT mean $\log_{10}$[(Lac+Thr)/tNAA] across the different MRI scores showed an increase in BGT mean $\log_{10}$[(Lac+Thr)/tNAA] with increasing BGT severity on the NICHD MRI score from score 2A onwards (pattern 2A, any BGT, anterior limb of internal capsule (ALIC) or posterior limb of internal capsule (PLIC) involvement or watershed infarction noted without any other cerebral lesions; pattern 2B, involvement of either BGT, ALIC or PLIC or area of infarction and additional cerebral lesions; and pattern 3, cerebral hemispheric devastation (**Figure 6**).[5]

TBSS

There were 541 significantly lower FA voxels out of a total of 69851 in the mean FA skeleton in the abnormal compared to the normal motor outcome group. There were no other statistically significant associations. The mean FA map was produced through the registration of all subjects diffusion datasets using DTI-TK**.** The mean FA skeleton was overlaid on the mean FA map and the results of the statistical analysis for Bayley III motor outcome are shown in **Supplementary Figure 2.** Voxels wherein infants with poor motor outcome had lower FA (p<0.05) are shown in red-yellow. On the axial view there were significantly lower voxels in the ALIC and PLIC in the abnormal motor outcome group. On the coronal view, there were significantly lower voxels in optic radiations in the abnormal motor outcome group.

Discussion

This study shows that a ^{1}H MRS spectrum acquired at 3T from the BGT using a long echo time (288ms) for Lac+Thr/tNAA peak area ratio within 14 days of birth in babies with moderate to severe NE who had HT accurately predicts 2-year outcome. Using a threshold of 0.39, BGT Lac+Thr/tNAA peak area ratio predicts motor, cognition and language outcomes (sensitivity and specificity of 100% and 97%, 90% and 97%, 81% and 97% respectively). Normal and abnormal motor, cognitive and language outcome groups had significantly different mean BGT $\log_{10}[(Lac+Thr/tNAA)]$ levels (p < 0.001). BGT $\log_{10}[(Lac+Thr/tNAA)]$ correlated with BGT involvement in the NICHD MRI score (scores 2A, 2B, 3).

In the 2010 meta-analysis of Lac/NAA peak area ratio, including 32 studies (860 infants) with NE who were not cooled, sensitivity and specificity was 82% and 95% for motor outcome for data acquired between day 1-30.[8] In our study at 3T in babies in the cooling era, BGT Lac+Thr/tNAA had a sensitivity and specificity for predicting motor outcome of 100% and 97% respectively. Therefore, BGT Lac+Thr/tNAA acquired within 14 days of birth and after HT remains a robust predictor of motor outcome in babies who have had HT. For the first time BGT Lac+Thr/tNAA, using a threshold of 0.39, is shown to predict cognitive and language outcomes as well as motor outcomes **(Table 2 and Figure 3)**. The improved accuracy and prediction across other domains may be attributed to the improved spectral fitting with inclusion of Thr, acquiring MRS within 14 days and the use of higher field strengths allowing for a greater spectral separation. The prediction of specific neurodevelopmental domains with BGT Lac+Thr/tNAA further emphasizes the central importance of injury to the BGT in determining outcome across several domains. The influence of BGT on outcome was also seen in the correlation between Lac+Thr/tNAA and the BGT involvement in the NICHD MRI score (pattern 2A, 2B, 3 BGT involvement). The importance of the BGT predicting outcome was observed in the MRI score described by de Vries and colleagues; in their study only the grey matter score was included, suggesting this location most closely determines outcome in NE.[28] The sensitivity and specificity for 2-year neurodevelopmental outcome of this MRI score was 92% and 95% for the retrospective cohort and 42% and 98% for a prospective cohort. There was no ^{1}H MRS included in the prospective cohort, which may explain the reduced accuracy.[28]

[1]H MR Spectra were analyzed using Tarquin, a freely available spectroscopy analysis tool that fits signals in the time-domain using a metabolite basis set. Tarquin is similar to LCModel in that it uses a basis set of metabolite signals to fit to the acquired spectrum. Tarquin has been shown to be robust, to produce similar results to LC Model in healthy volunteers,[16] and to have comparable

11

errors to LC Model using *monte carlo* simulations.[13] It has been used for the quantitative evaluation of brain and brain-tumour metabolites in humans[29-31] and in an animal model.[32]

Threonine (Thr) was included in the basis set for the spectral analysis. Thr is an essential amino acid, which is important for protein synthesis and folding. Foods such as meat, milk, beans and fish are a source of Thr. Thr has been detected in human brain on autopsy[30] and also by ^{1}H MRS.[16] Thr brain levels are known to be affected by protein intake with low protein diets leading to reduced Thr levels in experimental rats.[31] Chronic energy deficiency however was not associated with changes in brain Thr levels.[33] Nevertheless, we have observed that including Thr in the metabolite basis set improves the signal fit in the region around 1.3ppm, especially when the signal in this region is large **(Figure 1).** Lac and Thr are not easily resolvable in *in vivo* spectra[16] and so the sum of Lac+Thr is the most meaningful quantity to report.[34] The inclusion of Thr has only a small effect on the fitting of other metabolites in the basis set other than Lac due to minimal overlap.

Restricting to scans acquired up to day 9 when pseudonormalization of MD occurs in cooled babies with NE,[10] sensitivity and specificity for motor outcomes for MD (threshold $0.001mm^2/s$) was 72% and 100% and FA (threshold 0.198) was 39% and 94%. These values were similar to MD and FA sensitivity and specificity for all scans up to day 14; as only 5 babies were scanned after day 9 these would have had a small influence. DTI analyzed by TBSS has been previously used to show a treatment effect of HT in babies with NE;[35] babies who were not cooled (and likely to have adverse outcomes) had lower FA in the ALIC, PLIC and optic radiations. These findings are similar to the regional differences in our normal and abnormal motor outcome groups. In a subsequent study, Tusor *et al* report that NE infants with unfavorable outcomes after HT had lower FA values ($p < 0.05$) in centrum semiovale, corpus callosum, ALIC, PLIC, fornix, cingulum, cerebral peduncles, optic radiations, and inferior longitudinal fasciculus.[36] It is notable that the current cohort demonstrates a relatively limited distribution of lower FA after HT compared with that reported by Tusor. The reason this discrepancy is not clear as patient populations, acquisition and TBSS methods for the two studies are similar.

There are some limitations to the study. Firstly, the threshold of 0.39 Lac+Thr/tNAA was chosen as the threshold for normal/abnormal outcome for motor, cognitive and language outcomes by assessing the value that gave the optimum sensitivity and specificity on logistic regression. This Lac+Thr/tNAA threshold needs to be validated in another cohort of babies with NE. Secondly, the accuracy of Lac+Thr/tNAA for predicting cognitive and language as well as motor outcome may be influenced by overlap of poor outcome domains within individuals (18 infants had abnormal outcomes in all domains). Thirdly, the highest brain lactate levels were seen in the scans acquired

during the first 4 days; this pattern may have been influenced by earlier scans done in the most severely affected infants to help with clinical management decisions. Lastly, the process of pseudonormalization of MD is a biological continuum and the cut-off of day 9 is somewhat arbitrary and based on a small retrospective study.[10] Variation in pseudonormalization within the study population could therefore have confounded the diffusion weighted MRI analysis (MD).

Brain lactate has been observed to persist for months in the brain of babies with adverse outcome after NE;[37 38] this persisting brain lactate is associated with abnormal MRI and brain alkalosis.[38] Mechanisms leading to persisting raised brain lactate include impaired mitochondrial function and oxidative phosphorylation leading to an increase in anaerobic glycolysis, an altered redox state, the presence of phagocytic cells, gliosis, loss of perfusion autoregulation and/or altered buffering mechanisms. This persisting abnormal brain metabolism may reflect tertiary brain injury[9] and be a marker of ongoing injury, which may be a target for therapies. NAA is the second most abundant amino acid-related metabolite in the brain after glutamate. Both NAA and NAAG synthesis in neurons is ATP-dependent and hypoxia-ischemia initiates a decline in NAA/NAAG within hours. Much of the NAA synthesized in the neuron is transported to oligodendrocytes where it is a substrate source for myelin synthesis; it is likely that decreases in NAA following NE will lead to a reduced myelination.[14] In adult stroke, NAA levels can predict the fate of an acute ischemic lesion with more accuracy than DWI.[39 40] The dissociation between NAA and DWI after hypoxia-ischemia has been well described; restoration of ADC occurred despite the subsequent infarction of tissue whereas NAA levels continued to show a decline in the same area.[41] Our findings in babies with NE concur with the experience in adult stroke of better accuracy of ^{1}H MRS in predicting outcome compared to MD and FA. We previously described the central importance of NAA in predicting outcome: absolute concentrations of NAA [NAA] in babies with NE were more accurate than peak area ratios of Lac/NAA and could discriminate control, normal/mild, and severe/fatal outcome groups during the first 4 days after birth.[42] Time constraints, however, make collection of data for absolute concentration impractical in clinical MRI/MRS studies. Moreover, a significant advantage of peak-area ratios eg with Lac+Thr/tNAA, is that they depend on both metabolite T2s and concentrations, both of which are pathologically modulated and hence injury severity prediction is improved.[42]

In conclusion, we have shown in 55 babies with moderate to severe NE who have undergone HT, that BGT Lac+Thr/tNAA acquired at 3T within 14 days after birth predicts Bayley III motor, cognitive and language outcomes with a high degree of accuracy; the central importance of the BGT in prediction of all three domains was clear. Optimized spectral fitting with Tarquin and higher field strength may contribute to the improved accuracy compared to previous studies. Lac+Thr/tNAA correlated with MRI scores of BGT injury. Lac+Thr/tNAA is a robust marker of

13

neurodevelopmental outcome on which decisions for adjunct therapies in the subacute phase after HT might be based upon.

What is already known on this topic

1. Lac/NAA peak area ratio acquired in the first 30 days in non-cooled babies with NE predicts motor outcome with a sensitivity of 82% and specificity of 95% and is more accurate and stable than mean diffusivity

2. MRS is easy to acquire on modern MRI scanners. Peak area ratios are simple calculate and require only a single acquisition.

3. Improved metabolite peak definition may be seen at 3T versus 1.5T

What this study adds

1. BGT Lac+Thr/tNAA acquired at 3T within 14 days after birth in babies who have been cooled predicts Bayley III motor, cognitive and language 2-year outcomes with a high degree of accuracy

2. Higher field strength (3T) and inclusion of threonine with lactate in the metabolite basis set may contribute to better prediction

3. BGT Lac+Thr/tNAA had improved sensitivity and specificity for motor outcomes than MD and FA during the period up to day 9 before MD pseudonormalization

References

1. Lee A, Kozuki N, Blencowe H, et al. Intrapartum-related neonatal encephalopathy incidence and impairment at regional and global levels for 2010 with trends from 1990. *Pediatr Res* 2013;74(Suppl 1):50-72.

2. Kurinczuk J, White-Koning M, Badawi N. Epidemiology of neonatal encephalopathy and hypoxic-ischaemic encephalopathy. *Early Hum Dev* 2010;86(6):329-38.

3. NPEU. MBRRACE-UK: Mothers and Babies: Reducing Risk from Audits and confidential Enquiries across the UK https://www.npeu.ox.ac.uk/mbrrace-uk2017 [

4. Jacobs S, Berg M, Hunt R, et al. Cooling for newborns with hypoxic ischaemic encephalopathy. *Cochrane Database Sys Rev* 2013;1:CD003311. doi: 10.1002/14651858.CD003311.pub3.

5. Shankaran S, Barnes P, Hintz S, et al. Brain injury following trial of hypothermia for neonatal hypoxic-ischaemic encephalopathy. *Arch Dis Child Fetal Neonatal Ed*;97(6):F398-404.

6. Rutherford M, Ramenghi L, Edwards A, et al. Assessment of brain tissue injury after moderate hypothermia in neonates with hypoxic-ischaemic encephalopathy: a nested substudy of a randomised controlled trial. *Lancet Neurol* 2010;9(1):39-45.

7. Azzopardi D, Robertson N, Bainbridge A, et al. Moderate hypothermia within 6 h of birth plus inhaled xenon versus moderate hypothermia alone after birth asphyxia (TOBY-Xe): a proof-of-concept, open-label, randomised controlled trial. *Lancet Neurol* 2015;pii: S1474-4422(15)00347-6.

8. Thayyil S, Chandrasekaran M, Taylor A, et al. Cerebral magnetic resonance biomarkers in neonatal encephalopathy: a meta-analysis. *Pediatrics* 2010;125(2):e382-95.

9. Fleiss B, Gressens P. Tertiary mechanisms of brain damage: a new hope for treatment of cerebral palsy? *Lancet Neurol* 2012;11(6):556-66.

10. Bednarek N, Mathur A, Inder T, et al. Impact of therapeutic hypothermia on MRI diffusion changes in neonatal encephalopathy. *Neurology* 2012;78(18):1420-7.

11. Kreis R, Hofmann L, Kuhlmann B, et al. Brain metabolite composition during early human brain development as measured by quantitative in vivo 1H magnetic resonance spectroscopy. *Magn Reson Med* 2002;48(6):949-58.

12. Govindaraju V, Young K, Maudsley A. Proton NMR chemical shifts and coupling constants for brain metabolites. *NMR Biomed* 2000;13(3):129-53.

13. Choi C, Coupland N, Kalra S, et al. Proton spectral editing for discrimination of lactate and threonine 1.31 ppm resonances in human brain in vivo. *Magn Reson Med* 2006;56(3):660-5.

14. Igarashi H, Suzuki Y, Huber V, et al. N-acetylaspartate decrease in acute stage of ischaemic stroke: a perspective from experimental and clinical studies. *Magn Reson Med Sci* 2015;14(1):13-24.

15. NICE. National Institute for Clinical Excellence: Therapeutic hypothermia with intracorporeal temperature monitoring for hypoxic perinatal brain injury: guidance. *NICE guidelines: interventional procedures* 2010;http://www.nice.org.uk/nicemedia/live/11315/48809/48809.pdf(IPG 347):ISBN 978-1-84936-251-1.

16. Wilson M, Reynolds G, Kauppinen R, et al. A constrained least-squares approach to the automated quantitation of in vivo ^{1}H magnetic resonance spectroscopy data. *magn Reson Med* 2011;65(1):1-12.

17. Jenkinson M, Beckmann T, Behrens S, et al. FSL. *Neuroimage* 2012;62:782-90.
18. Zhang H, Avants B, Yushkevich P, et al. High-dimensional spatial normalization of diffusion tensor images improves the detection of white matter differences in amyotrophic lateral sclerosis,. *IEEE Transactions on Medical Imaging*;26(11):1585-97.
19. Kuklisova-Murgasova M, Aljabar P, Srinivasan L, et al. A dynamic 4D probabilistic atlas of the developing brain. *Neuroimage* 2011;54(4):2750-63.
20. Gousias I, Edwards A, Rutherford M, et al. Magnetic resonance imaging of the newborn brain: Manual segmen- tation of labelled atlases in term-born and preterm infants. *Neuroimage* 2012;62(3):1499-509.
21. http://sourceforge.net/projects/niftyreg.
22. Cardoso M, Melbourne A, Kendall G, et al. Adapt: An adaptive preterm segmentation algorithm for neonatal brain MRI. *Neuroimage* 2013;65:97-108.
23. http://sourceforge.net/projects/niftyseg/. [
24. Cardoso M, Modat M, Vercauteren T, et al. Scale Factor Point Spread Function Matching: Beyond Aliasing in Image Resampling. Medical Image Computing and Computer-Assisted Intervention -- MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9,2015:675-83.
25. Smith S, Jenkinson M, Johansen-Berg H, et al. Tract-based spatial statistics: voxelwise analysis of multi-subject diffusion data. *Neuroimage* 2006;31(4):1487-505.
26. Bach M, Laun F, Leemans A, et al. Methodological considerations on tract-based spatial statistics (TBSS). *Neuroimage* 2014;100:358-69.
27. San BNBSoIaTDre, Assessment; ATH. 2006
28. Weeke L, Groenendaal F, Mudigonda K, et al. A Novel Magnetic Resonance Imaging Score Predicts Neurodevelopmental Outcome After Perinatal Asphyxia and Therapeutic Hypothermia. *J Pediatr* 2018;192:33–40.e2.
29. Zarinabad N, Abernethy L, Avula S, et al. Application of pattern recognition techniques for classification of pediatric brain tumors by in vivo 3T 1 H-MR spectroscopy-A multi-center study. *Magn Reson Med* 2018;79(4):2359-66.
30. Vicente J, Fuster-Garcia E, Tortajada S, et al. Accurate classification of childhood brain tumours by in vivo ^{1}H MRS - a multi-centre study. *Eur J Cancer* 2013;49(3):658-67.
31. Scott J, Underwood J, Garvey L, et al. A comparison of two post-processing analysis methods to quantify cerebral metabolites measured via proton magnetic resonance spectroscopy in HIV disease. *Br J Radiology* 2016;89(1060):20150979.
32. Doblas S, He T, Saunders D, et al. In vivo characterization of several rodent glioma models by 1H MRS. *NMR Biomed* 2012;25(4):685-94.
33. Kaladhar M, Rao B. Experimental protein and energy deficiencies:effects on brain free amino acid composition in rats. *Br J Nutr* 1977;38:141-44.
34. Hofmann L, Slotboom J, Jung B, et al. Quantitative 1H-magnetic resonance spectroscopy of human brain: Influence of composition and parameterization of the basis set in linear combination model-fitting. *Magn Reson in Med* 2002;48:440-53.
35. Porter EJ, Counsell SJ, Edwards AD, et al. Tract-based spatial statistics of magnetic resonance images to assess disease and treatment effects in perinatal asphyxial encephalopathy. *Pediatr Res* 2010;68(3):205-9.
36. Tusor N, Wusthoff C, Smee N, et al. Prediction of neurodevelopmental outcome after hypoxic-ischemic encephalopathy treated with hypothermia by diffusion tensor imaging analyzed using tract-based spatial statistics. *Pediatr Res* 2012;72(1):63-9.
37. Robertson N, Cox I, Cowan F, et al. Cerebral intracellular lactic alkalosis persisting months after neonatal encephalopathy measured by magnetic resonance spectroscopy. *Pediatr Res* 1999;46(3):287-96.

38. Robertson N, Cowan F, Cox I, et al. Brain alkaline intracellular pH after neonatal encephalopathy. *Ann Neurol* 2002;52(6):732-42.
39. Loh P, Butcher K, Parsons M, et al. Apparent diffusion coefficient thresholds do not predict the response to acute stroke thrombolysis. *Stroke* 2005;36:2626-31.
40. Rivers C, Wardlaw J, Armitage P, et al. Do acute diffusion and pefusion weighted MRI lesions identify final infarct volume in ischemic stroke? *Stroke* 2006;37:98-104.
41. Kikuchi B. Mechanism of secondary deterioration after transient focal brain ischemia:1H MRS study. *Niigata Med J* 2008;122:271-79.
42. Cheong J, Cady E, Penrice J, et al. Proton MR spectroscopy in neonates with perinatal cerebral hypoxic-ischemic injury: metabolite peak-area ratios, relaxation times, and absolute concentrations. *AJNR Am J Neuroradiol* 2006;27(7):1546-54.

Tables

Gestational age (weeks + days)	39+5 (35+5 – 42 weeks)
Birth weight (grams)	3260±566
Sex (Male: Female)	23:22
Arterial cord gas pH	6.9±0.2
Arterial cord base deficit	-15±7
Apgar score at 1 min	2±2
Apgar score at 5 min	4±2
Apgar score at 10 min	6±3
Inborn: Outborn infants	16:39
Age at MRI (days)	5 (2-14)

Table 1. Patient characteristics. Gestational age is presented as mean (range), age at MRI as median (range) while other parameters are presented as mean±SD.

Bayley III outcome at 2 years	Cut off / Threshold	Normal outcome at 2 years (n)	Abnormal outcome at 2 years (n)	Sensitivity	Specificity	AUC
Lac+Thr/tNAA vs Motor	0.39	35	20	100%	97%	0.997
Lac+Thr/tNAA vs Cognitive	0.39	33	19	90%	97%	0.967
Lac+Thr/tNAA vs Language	0.39	30	21	81%	97%	0.913
Mean diffusivity (MD) (mm^2/s) vs Motor	0.001	35	20	65%	100%	0.769
Fractional Anisotropy (FA) vs Motor	0.198	35	20	45%	91%	0.544

Table 2A. The logistic regression model for all 55 subjects scanned between 2-14 days after birth. Log_{10}(Lac+Thr/tNAA) was used to determine the diagnostic accuracy of a Lac+Thr/tNAA threshold of 0.39 for motor, cognitive and language outcomes at 2 years along with mean diffusivity (MD) and fractional anisotropy (FA).

Bayley III outcome at 2 years	Cut off / Threshold	Normal outcome at 2 years (n)	Abnormal outcome at 2 years (n)	Sensitivity	Specificity	AUC
Lac+Thr/tNAA vs Motor	0.39	31	19	100%	97%	0.997
Lac+Thr/tNAA vs Cognitive	0.39	31	18	94%	97%	0.987
Lac+Thr/tNAA vs Language	0.39	28	20	85%	96%	0.935
Mean diffusivity (MD) (mm^2/s) vs Motor	0.001	32	18	72%	100%	0.819
Fractional Anisotropy vs Motor	0.198	28	20	39%	94%	0.507

Table 2B. The logistic regression model for subjects scanned between day 2-9 (n=50) after birth. Log_{10}(Lac+Thr/tNAA) was used to determine the diagnostic accuracy of a Lac+Thr/tNAA threshold of 0.39 for motor, cognitive and language outcomes at 2 years along with mean diffusivity (MD) and fractional anisotropy (FA).

Figure Legends

Figure 1: (A) Representative fitting of the 1H MR spectrum using Tarquin and a basis metabolite set that includes both lactate and threonine. The "residual" once the fitting has occurred is minimal (circle). **(B)** Removing threonine from the basis metabolite set leads to a large "residual" to the fit (circle), especially if the lactate signal in the region of 1.3ppm is elevated.

Figure 2: Position of the thalamic ^{1}H MRS voxel for spectral acquisition (sagittal A and axial B) from a neonate with normal outcome (Lac+Thr/tNAA 0.19) (C) and a neonate with poor outcome (Lac+Thr/tNAA 0.62) (D) after HT for NE.

Figure 3: ROIs for MD and FA in BGT were defined automatically based on both an atlas based probabilistic tissue segmentation and on a joint multi-atlas label propagation and fusion of 50 neonatal brain regions.

Figure 4: Logistic regression analysis of $\log_{10}$Lac+Thr/tNAA for outcome prediction. Using a Lac+Thr/tNAA cut off threshold of 0.39 (equating to -0.41 for Log 10 Lac+Thr/tNAA), TARQUIN analysis gave 1 false positive **(A)** for predicting poor motor outcome at 2 years (sensitivity 100%, specificity 97%, AUC 0.997). Analysis for cognitive outcome gave 1 false positive and 2 false negative (sensitivity 90%, specificity 97%, AUC 0.967 **(B)**. Analysis for language outcome gave 1 false positive and 4 false negatives (sensitivity 81%, specificity 97%, AUC 0.913 **(C).** The normal and abnormal group means of $\log_{10}$Lac+Thr/tNAA were significantly different for motor **(D)**, cognitive **(E)** and language **(F)** domains (p<0.0001).

Figure 5: Logistic regression analysis of mean diffusivity (MD) (mm^2/s) and fractional anisotropy (FA) restricted to scans acquired day 1-9 before pseudonormalization for motor outcome prediction only (50 of 55 scans). MD threshold of 0.001mm^2/s gave a sensitivity of 72%, specificity 100%, AUC 0.819) (A); FA threshold of 0.198 gave a sensitivity 39%, specificity 94%, AUC 0.507) (B). The normal and abnormal group motor outcome means of MD were significantly different (p<0.001) (C); for FA there was no difference (p=0.707) (D).

Figure 6: BGT Lac+Thr/tNAA correlated with BGT involvement in NICHD MRI score - scores 2A,2B,3. The BGT injury on MRI is indicated by the dotted box. Analysis of variance to compare BGT mean $\log_{10}$[(Lac + Thr)/tNAA] revealed significant differences between score 2A and B (p=0.0003) and score 3 and 2B (p=0.001).

Supplementary Figure 1: Lac+T/tNAA (left) and $\log_{10}[(Lac + Thr)/tNAA]$ (right) peak area ratio in all NE babies versus age of scan up to day 14.

Supplementary Figure 2: Tract based spatial statistics (TBSS) analysis was performed with the mean FA skeleton overlaid in green on the mean fractional anisotropy map. Axial view at PLIC (A), midline sagittal view(B), coronal view at level of PLIC and splenium (C). Voxels wherein infants with poor motor outcome had lower FA ($p<0.05$) are shown in red-yellow. The axial view indicated lower voxels in PLIC and ALIC. The coronal view revealed lower voxels in optic radiations.