MRI combined with early clinical variables are excellent outcome predictors for newborn infants undergoing therapeutic hypothermia after perinatal asphyxia

Marianne Thoresen\textsuperscript{a,b,*,}, Sally Jary\textsuperscript{a,}, Lars Walløe\textsuperscript{b}, Mathias Karlsson\textsuperscript{a,c}, Miriam Martinez-Biarge\textsuperscript{a,d}, Ela Chakkarapani\textsuperscript{a}, Frances M Cowan\textsuperscript{a,d}

\textsuperscript{a} Neonatal Neuroscience, Translational Health Sciences, University of Bristol, Bristol, United Kingdom
\textsuperscript{b} Department of Paediatrics, Imperial College London, London, United Kingdom
\textsuperscript{c} Section for Physiology, Institute of Basic Medical Sciences, University of Oslo, Oslo, Norway
\textsuperscript{d} Department of Medical Sciences, Biomedical Structure and Function, Uppsala University, Uppsala, Sweden

\textsuperscript{*} Corresponding author.
\textsuperscript{#} Contributed equally

Abstract

Background: Binary prediction-models for outcome [death, cognition, presence and severity of cerebral palsy (CP)], using MRI and early clinical data applicable for individual outcome prediction have not been developed.

Methods: From Dec 1\textsuperscript{st} 2006 until Dec 31\textsuperscript{st} 2013, we recruited 178 infants into a population-based cohort with moderate or severe hypoxic-ischaemic encephalopathy (HIE) including postnatal collapse (PNC, n = 12) and additional diagnoses (n = 12) using CoolCap/TOBY-trial entry-criteria including depressed amplitude-integrated EEG (aEEG). Early clinical/biochemical variables and MRI scans (median day 8) were obtained in 168 infants. Injury severity was scored for cortex, basal ganglia/thalami (BGT), white matter (WM) and posterior limb of the internal capsule, summating to a total injury score (TIS, range 0–11). Outcome was categorized as adverse or favourable at 18–24 months from Bayley-III domains (cut-off 85) and neurological examination including CP classification.

Findings: HIE and entry-aEEG severity were stable throughout the study. Outcome was favourable in 133/178 infants and adverse in 45/178: 17 died, 28 had low Cognition/Language scores, (including 9 with severe CP and 6 mild); seven had mild CP with favourable cognitive outcome. WMxBGT product scores and TIS were strong outcome predictors, and prediction improved when clinical/biochemical variables were added in binary logistic regression. The Positive Predictive Value for adverse outcome was 88\%, increasing to 95\% after excluding infants with PNC and additional diagnoses. Using WMxBGT in the regression predicted 8 of the 9 children with severe CP.

Interpretation: Binary logistic regression with WMxBGT or TIS and clinical variables gave excellent outcome prediction being 12\% better than single variable cross-tabulation. Our MRI scoring and regression models are readily accessible and deserve investigation in other cohorts for group and individual prediction.

Funding: We thank the National Health Service (NHS) and our Universities and funders in UK and Norway: SPARKS, The Moulton Foundation, The Norwegian Research Council, The Lærdal Foundation for Acute Medicine and charitable donations for their support for cooling therapy.

© 2021 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)
1. Introduction

Therapeutic hypothermia (TH) for perinatal asphyxia was implemented region-wide from 2007 in South-West England in two level-3 hospitals, both key recruiting centres for the first two international randomised controlled trials (RCTs) of TH [1,2]. Following these trials, three days of TH (33.5 °C) remains the only effective intervention for moderate-severe hypoxic-ischaemic encephalopathy (HIE). Cooling for longer (5 days) or deeper (32°C) were ineffective in improving outcomes [3]. TH reduces the severity of brain lesions associated with HIE as determined from magnetic resonance imaging (MRI) [4] and improves outcomes.

Unlike the protocol used in the original RCTs, our centre now cools infants with postnatal asphyxia and additional concurrent diagnoses, [5–7] and 13% of currently cooled infants would not have been recruited to the original TH trials [6].

Additionally, we start passive cooling early rather than keeping body temperature at 37.0°C until commencing active TH. Finally many centres, but not ours, now cool infants with mild HIE and also those of gestational age (GA) <36 weeks; these practices make cooled cohorts different [8,9].

Pre-cooling, neonatal MRI was found to be a particularly powerful predictor of neurodevelopmental outcomes after HIE [10–14]. In cooled infants, studies exploring MRIs predictive value are often small, not accurate enough or require complex MRI assessment [15–20]. Additionally, the changes in clinical practice over years may affect the pattern and severity of injury in cooled cohorts, hence early and readily available predictors of outcome need re-evaluating.

One widely-used MRI scoring system introduced by Rutherford et al., in the nested TOBY study of TH, [4] classifies the severity of injury in the basal ganglia and thalamus (BGT), cortex (CX), white matter (WM) and posterior limb of the internal capsule (PLIC). We summarised these regional scores as a Total Injury Score (TIS) providing a continuum from 0 (all normal) to 11 (maximum lesion load) to quantify injury and predict outcome on a binary basis [21].

We have previously explored traditional clinical/biochemical factors as outcome predictors [22,23]. These are the severity pattern of aEEG, [21,24] the peak LDH (LDHpeak), LDH value at 72h (LDH72h), [25,26] time for plasma lactate to fall below 5 mmol (lactatehrs <5mmol) [21] and the number of inotropic and anticonvulsant drugs used during TH [27] as proxy-markers for hypotension and seizure burden respectively. We confirm in this paper the usefulness of these measures previously proposed for outcome prediction [21,24,26,27].

In this population-based cohort study with a wide range of infants, with HIE including those with comorbidities or additional diagnoses more representative of current cooling practices we aim to test:

1. The predictive ability of different MRI scoring combinations, with and without clinical/biochemical markers, for binary favourable or adverse outcome (including the presence and severity of CP and death).
2. Whether this predictive ability is worse in infants who were cooled with diagnoses outside the original TH trial entry criteria.

2. Methods

2.1. Cohort

This was a prospective, population-based cohort study of infants born at ≥36 weeks GA with signs of moderate or severe neonatal encephalopathy, within the southwest region of the United Kingdom, who received 72h of TH in St Michael’s Hospital, Bristol, one of the two regional tertiary cooling centres. This hospital accepts infants for TH on alternate days. Thus, the included infants are approximately a 50% random selection of those needing TH. The regional population is 2.6 million with 30,000 deliveries annually and 1.7/1000 term deliveries undergo TH. We run a 24/7 retrieval service administering TH en route. All referring hospitals have aEEG monitoring and servo-controlled cooling equipment allowing remote aEEG assessment and early TH [28].

2.2. Data collection

With research ethics approval and waiver consent (CH/2009/3091) to collect an ongoing (until April 1st 2023) anonymised database, we analysed anonymised data from 178 infants treated with TH during the seven years between December 2006 and December 2013. All authors had access to the data. A flowchart with the outcome of the study cohort is shown in Supplementary Figure 1. Maternal and neonatal demographic data, clinical imaging and outcome...
information are given in Supplementary Table 1. The criteria for starting TH and a list of descriptive variables explored as explanatory variables, output variables and details of handling missing data are all shown in Supplementary Table 2. The annual distribution of precooling aEEG pattern is shown in Supplementary Figure 2

2.3. Therapeutic hypothermia

Infants with clinical signs of moderate or severe neonatal encephalopathy within an hour of birth, or who suffered a PNC within 48h [29] of birth (fulfilling the same entry criteria except Apgar score), had their external heating turned off just after resuscitation, initiating passive cooling. Most had no clear etiology for the collapse, though two were hypotensive in relation to large sub-galeal haemorrhage and five had a transiently low blood sugar; none had positive blood or other cultures. Their post-collapse blood pH was not different from the typically presenting HIE infants and their need for inotropic and anticonvulsant medication was also similar. Of the PNC infants, one infant died, two had a poor outcome without CP and one had a good outcome but CP GMFCS level 1. The remaining nine infants had Bayley-III scores > 100 in all domains and very low MRI injury scores. Details of the infants with additional diagnoses are given in Supplementary Table 1.

Rectal temperature was monitored continuously and hypothermia always avoided. Active cooling [servo-controlled whole-body cooling (Criticool, Mennen Medical)] was started within 6h of birth or PNC once TH entry criteria had been fulfilled as per the CoolCap and TOBY trials, [1,2] and summarised in the NICE guidelines [30] plus additional criteria defined in the Bristol cooling management protocol [28].

TH was continued for 72h at 33.5 °C followed by 6h of rewarming at a rate of 0.5 °C/h to 36.5 °C. Infants were intubated before active cooling until the end of rewarming and received morphine during TH for comfort and pain reduction [26,31]. aEEG monitoring was started within 1h of birth and continued until after rewarming for assessing background activity, seizure load and anticonvulsant medication effects. Clinical and electrical seizures were treated according to an escalating protocol and the number of anticonvulsants given recorded [27] as was the number of different inotropes needed to maintain mean arterial blood pressure > 45 mmHg.

2.4. Indices of Deprivation

Maternal postcode at time of birth was used to classify all infants using the English indices of deprivation [32] which give a summed score (deprivation score, DS) for income, employment, education, housing, health, disability and crime (1 most deprived; 10 least deprived). The scoring and degree of high resolution postcode data is unique to England and not applicable generally. This Deprivation Index is used in an example in Supplementary Document 1.

2.5. MRI protocol

Neonatal MRI scans were obtained in 168 infants (Supplementary Figure 1), post-feed usually without sedation. Hearing protection was used and infant movement limited by a vacuum mattress. We monitored heart rate, core temperature, transcutaneous oxygen saturation and respiratory rate. All infants had at least T1-weighted imaging in axial and sagittal planes and axial T2-weighted imaging; most had diffusion-weighted imaging (DWI) and more recently susceptibility-weighted imaging and MR venogram. All scans were reported clinically and again later by one author (FC), aware of the infants GA and postnatal age at scan but no other clinical details or outcome and this scoring was used in the analyses. Note was made of scan quality, anatomical development and maturation, evidence of recently acquired and longstanding lesions, subdural haemorrhage and venous thrombosis. Median (IQR) age at scan was 8 (6.5–9) days. DWI information was used when appropriate (26/168 infants were scanned < 6d). Injury patterns were mostly consistent with perinatal hypoxia-ischaemia [33].

2.6. MR scoring

Scans were scored according to Rutherford et al. [4] for evidence of BGT, WM and CX injury (each on a scale of 0–3), and PLIC signal (scale 0–2), the higher number indicating more severe abnormality. We then calculated the TIS (range 0–11) for each infant [21].

A subset of 52 scans was reviewed and re-scored by FC for intra-observer reliability and scored independently by a second assessor (MMB) for inter-observer reliability. These scans, which encompassed the full TIS range, included scans from the early (2007–2009) and late (2012–2013) data collection periods from children with a range of developmental outcomes; we also included scans of poor quality. The intra and inter-rater agreement (Kendall’s tau) for TIS was 0.82 and 0.77 respectively (p<0.001).

2.7. Neurodevelopmental and functional outcomes at 18–24 months

Children were assessed at 18m using the Bayley Scales of Infant and Toddler Development-III (Bayley-III, [34,35]) which generates distinct Cognitive, Language and Motor Composite scores, with a normative mean (SD) of 100(15). Bayley-III was administered by one assessor (SF) unaware of the MRI scores. Inter-rater agreement for scoring the Bayley-III examinations in 10 children from video recordings was 97% [35]. As the lowest thresholds for Bayley-III Cognitive, Language and Motor composite scores are 55, 47 and 46 respectively, scores for children below these thresholds were allocated based on Bayley-III raw scores and clinical records by two assessors unaware of MRI findings. Children were also reviewed at 24m when the presence or absence of CP was confirmed and its severity graded using the Gross Motor Function Classification System (GMFCS) [36]. Independent ambulation (defined as 4–5 unaided steps), epilepsy defined according to the International League Against Epilepsy [37], presence of gastrostomy at this examination, and severe hearing or visual impairment were also recorded at 24 months.

2.8. Definition of binary favourable or adverse outcome

We defined a composite adverse outcome as death or moderate or severe disability (Bayley-III Cognitive/Language score (CLC)<85, CP GMFCS levels III-V, or severe hearing or severe visual impairment) as defined in previous RCTs [1,2]. We chose a Bayley-III CLC-score cut-off < 85 for binary analysis as we have published that this score is comparable to the Bayley-II Mental Developmental Index score < 70 evaluated contemporaneously in hypothermia-treated children [35]. We defined favourable outcome as all of survival with Bayley-III average CLC score ≥ 85, no or mild CP (GMFCS levels I-II), no severe hearing loss or visual impairment[2]. Follow-up medical letters from around 24m for children not attending Bayley-III assessment, allowed us to classify them as having adverse or favourable outcomes including CP.

2.9. Statistics

Statistical analyses were performed using SPSS–24 (SPSS, Chicago, IL, USA). Demographic and clinical data are summarised as median (95% CI or full range) and n(%). Analysis was undertaken in four patient groups: (1) all 168 scanned infants; (2) excluding infants with additional diagnoses (n = 158, 2 died and not scanned); (3) excluding infants with PNC (n = 156); (4) only infants fulfilling strict CoolCap/TOBY trial criteria excluding additional diagnoses and PNC (n = 146).

We first explored outcome prediction using stepwise binary logistic regression with regional MRI scores as explanatory variables only. The scores for WM, BGT, CX, PLIC and TIS were entered as potential
explanatory variables. Interaction between explanatory MRI variables was explored by allowing product terms in the list of explanatory variables.

The use of stepwise binary logistic regression with Bayesian Information Criterion (BIC) and significance probability as model selection criteria and how regression coefficients may be used to make outcome predictions for individual infants are explained in Supplementary Document 1 and Supplementary Tables 3 and 4.

In a second series of stepwise logistic regression analysis we added a large number of standard biochemical and clinical data obtained before and during TH as potential explanatory variables; most never entered any of the final regression equations. The regression coefficients from the logistic regressions (Supplementary Document 1) were used to make equations for binary prediction of adverse outcome for individual infants. Two of the six MR variables, the product of WM and BGT (WMxBGT) or the TIS, were always found to be the two most significant MR variables from the logistic regression. The effect of different cut-off values of these two MRI variables on outcome-prediction in the bivariate cross-tabulations was also explored and the best presented in Table 1, right panel.

The ability of logistic regression of (1) MR-variables only, (2) MR-variables combined with clinical/biochemical variables or (3) clinical/biochemical variables alone to predict adverse outcome in the four groups listed above, was explored. Negative (NPV) and positive (PPV) predictive values for adverse outcome, specificity and sensitivity and predictive accuracy (PA) were calculated. Binary logistic regression was used to predict severe CP and all CP. In a separate analysis we tested the predictive ability of the six entry-criteria for TH available within 6 h after birth as used in the CoolCap and TOBY trials.

2.10. Role of the funding source

The funders did not have a role in the study design and the analysis or interpretation of the data.

3. Results

Neonatal and maternal demographic data are presented in Supplemental Table 1. 178 infants met criteria for cooling; 17 infants died in the neonatal period, 145/161(90%) of survivors had Bayley-III assessments at 18–24m and outcome was classified from medical records in 16; 168/178(94%) infants had MRI scans and the regional distribution of injury is presented in (Fig. 1). Passive cooling started at a median age of 0.7h (IQR 0.5–1.0 h), active cooling at 3.6h(IQR 3.3–4.9h) and target temperature was reached at 4.2h(IQR 3.9–5.9h). Pre-cooling, 14% of infants had a normal voltage aEEG + seizures, 70% a moderately depressed aEEG and 16% a severely depressed aEEG [38]. The severity distribution of the background aEEG pattern at start of cooling was stable over the study (Supplementary Figure 2).

3.1. Developmental outcomes in survivors at 18–24 m

No patient was lost to follow up, however 16 of the 133 survivors did not undergo Bayley-III but other examinations. The majority of survivors 133/161(83%) had Bayley-III scores ≥85 and were without severe CP, epilepsy, hearing or visual impairment or need for gastrostomy feeding. Twenty-two (14%) of the surviving children were diagnosed with CP (GMFCS Level-I: 13, Level-III: 1, Level-IV: 1, Level-V: 7). Seven of the 13 children with Level-I CP had Bayley-III CLC scores ≥85. Epilepsy was diagnosed in eight (5%) children and nine (6%) had a gastrosomy, eight of whom had CP (GMFCS Level-V: 7, Level-I: 1). Seven (4%) children had severe hearing loss and six (3.5%) severe visual impairment. Seventeen (11%) children were unable to walk independently at 18m, 11 of whom had CP (GMFCS Levels-III-V: 9, Level-I: 2) and four of the remaining six children had additional chromosomal or metabolic diagnoses (Supplementary Table 1).

3.2. MRI regional scores

Fig. 1 shows the frequency distribution of the regional MRI injury scores [4]. WM signal abnormality was most common, present in 72% and fairly evenly distributed across the severity range. For the other regions the commonest finding was no injury. Cortical injury was seen in 42%. Only 31% of infants had BGT injury and 24% abnormal PLIC signal.

3.3. MRI total injury score (range 0–11) and WMxBGT product score versus individual Bayley-III scores

The relationships between MRI TIS and Bayley-III CLC scores are shown in Fig. 2. Bayley-III CLC scores decrease with increasing TIS but there is a wide range of Bayley scores for any one TIS score. However, all children with a TIS of 0 or 1 had a Bayley-III CLC score >85 and only five children without PNC or additional diagnoses had a TIS between 2 and 5 and a Bayley-III CLC score <85. A TIS of 7 or more was almost always associated with CLC scores <85. More detail of the infants within each TIS score from 0 to 11 is given in the figure legend.

The relationship between the WMxBGT product and Bayley-III Motor score is shown in Fig. 3. The presence and severity of CP is coded according to GMFCS levels. The table inset shows all possible combinations of WM and BGT scores and their corresponding binary outcome. Note that WMxBGT will be 0 when either WM or BGT score is 0. Of the 109 infants with WMxBGT score of 0, 100 had a favourable outcome. In contrast, the higher the WM × BGT product, the severer the CP.

3.4. Combining MRI and clinical variables

Supplementary Table 2 describes all variables explored in the regression analysis. Significant variables in individual regressions are presented in Tables 1 and 2.

In the left upper block of Table 1 are results from the total scanned cohort (n = 168). The second, third and bottom blocks show the three clinical subgroups within the cohort. Results of regressions similar to Supplementary Table 4, are given in a condensed form to the left of Table 1. For clarity, the ‘framed’ 2 × 2 table at the top is extracted from Supplementary Table 4. Specificity, sensitivity, NPV, PPV and PA are given.

For all four groups (blocks) in Table 1, the first significant step in the regression is WMxBGT. The second and third steps vary depending on which subgroup is analysed. The lower part within each of the four groups (blocks) shows the same regression after removing WMxBGT from the list of possible explanatory variables. This allows us to compare logistic regression with TIS as the MR variable with TIS in the cross-tables. These two MR variables are strongly correlated, and by removing WMxBGT, this allowed us to examine the strength of TIS. Now the first significant step in all groups is TIS. The 2nd or 3rd steps vary between LDH[28], LDH[pea], lactate[burst=5mmol]31, the number of inotropes used and the number of adrenalin doses given during resuscitation, pH, 10-minute Apgar score, and HIE grade at entry were never significant. Milder types of injury usually do not have BGT injury and the WMxBGT product will always be 0. When comparing the 13 infants with mild CP with the 13 infants who have poor outcome but no CP, the children with CP have higher cognitive scores but also higher TIS. The last subgroup with strict trial entry criteria (not PNC or additional diagnosis) has fewest patients (n = 146) and the highest PPV predictions: 95% when WMxBGT was allowed, and 90% for TIS. NPV is ≥90% and similar in all 4 groups and analyses. The median PA was 91%.

When examining the uncertainty of results in the upper part of Table 1, left, the 2 × 2 table, B for WMxBGT = 0.693 and SE =0.137 (not stated in the Table 1). The 95% confidence interval(CI) for SE will be 0.419–0.967. A bootstrap analysis gives a 95% CI 0.501- 1.022
Table 1

Left part shows four blocks of regression results, each with the regression applied to one of 4 sub-group of infants: 1: \( n = 168 \) total scanned cohort, 2: \( n = 158 \) excluding 10 scanned infants with coexisting diagnosis. 12 PNC are included, 3: \( n = 156 \) excluding 12 infants with PNC including 10 with coexisting diagnosis, 4: \( n = 146 \) excluding 22 scanned infants with PNC or other diagnosis. In each block, the upper part gives the results from regressions corresponding to the one performed in Supplementary document 1. The upper part of the upper block repeats the results obtained in the second regression in Supplementary document 1. In the lower half of each block, the product WMxBGT is not allowed in the regression, but all other MRI and biochemical and clinical variables are allowed. For each regression, the significant factors in the regression equation are listed with the corresponding P-values. The resulting 2 \( \times \) 2 tables are shown with the Positive Predictive Value (PPV) for adverse outcome, the Negative Predictive Value (NPV) for adverse outcome, Specificity (Sp), Sensitivity (Se) and Predictive Accuracy (PA).

| Variables allowed | Steps in binary logistic regression | B value (B0, B1) | Outcome using binary logistic regression | Cut off variable used | Cut-off for binary outcome prediction | Outcome using cross-tables |
|-------------------|------------------------------------|-----------------|----------------------------------------|----------------------|--------------------------------------|---------------------------|
|                   |                                    |                 | Favourable | Adverse | Total | NPV, PPV & Predictive Accuracy | Favourable | Adverse | Total | NPV, PPV & Predictive Accuracy |
| All MRI, clinical and biochemical variables | 0: constant | -3.370 | 129 | 14 | 143 | 90%<sup>NPV</sup> | WMxBGT | ≤2 | 126 | 13 | 139 | 91%<sup>NPV</sup> |
|                   | 1: WMxBGT | 0.693 | 3 | 22 | 25 | 88%<sup>PPV</sup> | WMxBGT | ≥2 | 6 | 23 | 29 | 79%<sup>PPV</sup> |
|                   | 2: LDH72h | 0.036 | 132 | 36 | 168 | 90%<sup>PA</sup> | WMxBGT | above | 122 | 9 | 131 | 91%<sup>PA</sup> |
|                   | 3: no inotrope | -1.113 | 98%<sup>NPV</sup> | 92%<sup>PPV</sup> | 98%<sup>PA</sup> | WMxBGT | ≤2 | 122 | 9 | 131 | 91%<sup>PA</sup> |
|                   | 4: no adren bolus | -3.405 | 97%<sup>NPV</sup> | 89%<sup>PA</sup> | WMxBGT | ≤2 | 122 | 9 | 131 | 91%<sup>PA</sup> |
| As above except WMxBGT | 0: constant | -4.425 | 127 | 14 | 141 | 90%<sup>NPV</sup> | WMxBGT | ≤2 | 122 | 9 | 131 | 91%<sup>PA</sup> |
|                   | 1: TIS | 0.587 | 5 | 22 | 27 | 81%<sup>PPV</sup> | WMxBGT | ≥2 | 5 | 21 | 26 | 81%<sup>PPV</sup> |
|                   | 2: LDH72h | 0.028 | 132 | 36 | 168 | 90%<sup>PA</sup> | WMxBGT | ≤2 | 122 | 9 | 131 | 91%<sup>PA</sup> |
|                   | 3: no inotrope | -1.113 | 126 | 13 | 139 | 90%<sup>PA</sup> | WMxBGT | ≤2 | 122 | 9 | 131 | 91%<sup>PA</sup> |
|                   | 4: no adren bolus | -3.405 | 97%<sup>NPV</sup> | 89%<sup>PA</sup> | WMxBGT | ≤2 | 122 | 9 | 131 | 91%<sup>PA</sup> |

M. Thoresen et al. / EClinicalMedicine 36 (2021) 100885
which is somewhat skewed, but not very different from the 95% CI based on the standard error. This shows that the model is robust. Using bootstrap analysis, the “next to enter” variable is still not significant. This degree of uncertainty and skewness is typical for the whole table.

Table 1 (right), shows corresponding cross-table results when prediction is made using either WMxBGT or TIS with two different cut-off values, for each of the four patient subgroups. The PPVs from cross-tables were 12% lower than PPV based on binary logistic regression when WMxBGT or TIS was the single MRI variable.

Fig. 4 shows a graphic summary of all PPV, NPV and PAs from the 4 groups. For each analysis, regression gave significantly better PPV prediction (median 88%) than cross-tabulation (median 76%) as shown in the upper panel. There were no differences for NPV or PA values between any analysis method or subgroup.

Table 2 shows the prediction of outcome of two defined clinical outcome groups; severe CP and all CP. We present prediction first allowing all MRI variables and all clinical and biochemical variables. Only one variable, WMxBGT is significant for predicting all CP and severe CP. In the second analysis, we use all MRI variables except WMxBGT, allowing also all clinical variables. Now TIS is the first significant variable for both outcome groups and time to starting cooling also comes in as a second significant variable for predicting severe CP. For all CP, only TIS is significant. Thirdly, we explored clinical variables only as early prediction is important and MRI is not always available. The significant variables in both CP groups without MR
variables are the number of anticonvulsants given during TH and the aEEG pattern before 6h of age.

Three of the 13 infants with CLC < 85 also had other diagnoses. All the 125 infants with CLC ≥ 85 and without CP were correctly predicted to have good outcome (NPV = 100%).

In a separate stepwise regression (not shown), we tested whether the six CoolCap/TOBY cooling entry-criteria (10-minute Apgar score, base-excess, pH at <1h, need for ventilation at 10 min, worst HIE and aEEG grade < 6h), were predictive of outcome. We entered either all six or only five variables excluding aEEG. Without aEEG, only four of the 45 infants with adverse outcome were correctly predicted. When including aEEG this improved to 15 of 45.

4. Discussion

We present novel analyses of MRI findings using two new scores (WMxBGT and TIS), developed from the regional scores of Rutherford [4] in a 7-year post-RCT population cohort of hypothermia-treated infants with neonatal encephalopathy. The results show improved prediction of adverse binary outcome and particularly good prediction of very severe outcomes compared to earlier data [39]. This predictive value improved further with the addition of early readily available clinical data. The PPV was 88% for adverse binary outcome for the whole cohort and 95% in the group (n = 146) excluding PNC and infants with additional diagnoses, and most comparable to the
original TH trial recruitment. We have shown how the inclusion of infants who were either not accepted in the trials or excluded from analysis, affect overall outcome prediction; this is an important aspect of our study and distinguishes it from those of others and has not been done before.

We also show that whilst predictions about adverse outcome are easier to make from cross-tables (Table 1, right part) than from the regression results (Table 1, left part), this comes at a price. As shown in Fig 4, the PPV of adverse outcome is about 10/C0\% lower using the cross-tables rather than the regression data.

MRI alone was excellent for predicting severe outcomes like death and severe CP. We found a strong interaction between WM and BGT injury scores. A WM score of 2 or 3 was only associated with adverse outcome if there was concurrent BGT injury. WM injury, even with score 3, without BGT injury, usually resulted in a favourable 2-year outcome. In infants with severe CP, eight of nine had a WMxBGT score $\geq 6$. It is important to note that we had a low incidence of CP, only 22 of 161 infants, 9 of whom were severe and 13 mild. The PLIC signal was the best single predictor of mild CP in the logistic regression analysis.

Important strengths of the study are that our cohort was recruited over a long period to an experienced large cooling-centre with a stable severity of encephalopathy through the study. We used the same entry criteria and outcome definitions as the randomised CoolCap- and TOBY-trials, to which we also recruited; importantly we included aEEG, in contrast to most other published cohort studies [19,40]. Passive and active cooling were started very early and care was delivered following a strict clinical management protocol. There was a high rate (94%) of neonatal MRI and neurodevelopmental assessments (90%) at 18–24 months.

Information on maternal educational level was not collected in our study but we collected the UK English postcode-based deprivation score, reflecting economic, domestic, educational and social information [32]. We present the use of deprivation scores in
Table 2 shows the results from regression analyses of two different subgroups: 1: Predicting severe CP, 2: Predicting all CP. Each block the upper part shows the results when all MRI variables and all clinical variables are allowed. The middle part, the results when all MRI variables except WMxBGT and clinical variables are allowed, and the lower part, the results when only clinical variables are allowed in the regression.

| Patient cohorts | Dependent variable | Steps in logistic regression | Logistic regression & BIC | Actual outcome |
|-----------------|--------------------|-----------------------------|---------------------------|---------------|
|                 |                    | B value | Significance per step | −2ln likely hood | BIC | N | Favourable | Adverse | Total | %NPV, %PPV, %PA |
| 1: Predicting severe CP n = 160 excluding 17 dead and 1 survivor not scanned - | All MRI, clinical and biochemical variables | 0: constant | 0 |  | 7.710 | 1.334 | 0.002 | 14.615 | 29.841 | 160 | 148 | 1 | 149 | 99% NPV, 73% PPV, 98% PA |
|                 | 1: WMxBGT | | | | | | | | | | 151 | 98% SP, 89% Se, 98% PA |
|                 | All MRI, except WMxBGT, and clinical and biochemical variables | 0: constant | 1: TIS | 1.457 | 0.002 | 23.147 | 38.371 | 160 | 149 | 3 | 152 | 98% NPV, 75% PPV, 97% PA |
|                 | 2: start EXT cool | | | | | | 18.690 | 38.990 | 151 | 98% SP, 67% Se, 97% PA |
|                 | Clinical and biochemical variables only | 0: constant | 1: # anticonv | 1.138 | 0.000 | 53.123 | 68.348 | 160 | 149 | 4 | 158 | 94% NPV, 71% PPV, 96% PA |
|                 | 2: aEEG | | | | | | 31.916 | 52.217 | 151 | 99% SP, 56% Se, 96% PA |
| 2: Predicting all CP n = 160 excluding 17 dead and 1 survivor not scanned | All MRI, clinical and biochemical variables | 0: constant | | | 3.300 | 0.915 | 0.000 | 70.255 | 85.480 | 160 | 135 | 3 | 10 | 145 | 93% NPV, 80% PPV, 92% PA |
|                 | 1: WMxBGT | | | | | | | | | | 138 | 98% SP, 55% Se, 91% PA |
|                 | All MRI, except WMxBGT, and clinical and biochemical variables | 0: constant | 1: TIS | 5.755 | 0.913 | 0.000 | 60.288 | 75.514 | 160 | 134 | 4 | 10 | 144 | 93% NPV, 75% PPV, 91% |
|                 | | | | | | | | | | | 138 | 97% SP, 55% Se, 91% |
|                 | Clinical and biochemical variables only | 0: constant | 1: # anticonv | 6.439 | 0.789 | 0.000 | 103.287 | 118.513 | 160 | 135 | 3 | 14 | 149 | 91% NPV, 73% PPV, 89% PA |
|                 | 2: aEEG | | | | | | 87.437 | 107.738 | 138 | 98% SP, 36% Se, 89% PA |
Supplementary Document 1. There was no relationship between deprivation scores and the occurrence or severity of HIE at birth in our region.

As we have shown previously, biomarkers available earlier than the MRI, such as aEEG,[24] LDH, [26] lactate<sub>hr</sub> < 5mmol [21] and the need for inotropic or anticonvulsant drugs have good predictive abilities [27]. These markers improve outcome predictability when added to MRI and may be helpful when early prediction is needed or when MRI is not available.

Using the six TH entry-criteria, only 15 of 45 infants with adverse outcome were correctly predicted, and if early aEEG was not included (many European and US centres do not use aEEG as an entry criterion), the prediction was only correct in four. Thus entry criteria for recruitment cannot be used as outcome predictors, particularly if

---

**Fig. 4.** Fig 4. compares graphically three methods of outcome prediction analysis in infants having MRI scans. (data from Table 1). The whole cohort (n = 168) has the darkest colour grade fading towards the smallest cohort n = 146 where infants with postnatal collapse and/or additional diagnosis to HIE were excluded.

The 146 cohort would fulfill the original cooling entry-criteria in the CoolCap and TOBY trials. The upper panel shows the positive predictive value (PPV) for adverse outcome. The first 4 shaded bars show results from binary logistic regression from the best model allowing all six MRI and all clinical and biochemical variables. WMxBGT is the strongest MRI variable. For the n = 146-group (palest colour), the best PPV from logistic regression is 95%. In the second vertical set of bars, WMxBGT is removed from the allowed variables and total injury score (TIS) is now the most significant. Again the 146 group has the best prediction, now 90%. The next two vertical sets of bars use cross-tabulation analysis with the best cut-off for a single MRI variable, either WMxBGT or TIS. The sequence of results show that logistic regression is better than cross-tabulation and that WMxBGT is better than TIS for outcome prediction. The middle horizontal panel shows that the negative predictive value (NPV) for poor outcome is good, 90–93% between all groups and methods. The lowest horizontal panel shows the predictive accuracy (PA, the sum of all correct predictions, both adverse and favourable) compared to the whole group. Again, there is little difference between methods. In a dataset with 75% favourable outcome, it is the PPV for adverse outcome that is the most important predictor.
aEEG is not included. Twenty-five years ago, a 10-minute Apgar score of 0 predicted 100% adverse outcome, but in the cooled arm of the 2005 NICHD trial, 20% with Apgar scores of 0–2 at 10 mins survived with normal 2-year outcome [41] and in our (current) study, 55% had a favourable outcome and 45% died.

When comparing outcomes between different TH cohorts, it is important to know which infants are included in the analysis. Our cohort differs from other post-trial clinical studies in that it is population-based and does not include mild HIE but does include 12 asphyxiated infants later identified with an additional diagnosis to HIE and 12 with PNC. We therefore present the binary outcomes for the total cohort (n = 178) as well as 3 sub-cohorts where either ‘additional diagnosis’ or PNC (or both) were removed from the analysis. Our data shows that outcome prediction was best in the strict trial entry criteria group using combined MRI and clinical data. Infants with cerebral bleeds, congenital anomalies, cardiac disease, chromosomal or metabolic diagnoses are usually removed post-hoc from outcome analyses [19] even though most are diagnosed well after TH. Among the 12 infants with additional diagnoses, two died early, and five had adverse outcome but only one with severe CP, reflecting their underlying conditions. Of the 12 infants with PNC, nine had developmental scores in the normal range. Thus our PNC group had a more favourable outcome compared to the pre-cooling era [42] and we speculate that this improved outcome may relate to the early initiation of TH after effective resuscitation.

Compared to the first TH RCT data [1,2,43] from 1999 to 2006, mortality in our cohort was low (9.6% v 32%) and survival without adverse outcome much higher (75% v 51%). Good outcomes including low mortality in the standard TH group that were presented in a recent 4-group RCT were similar to ours [3]. Cohort studies are often not comparable if they include mild HIE [19] or have different clinical practice regarding redirection of care [40].

Recently, magnetic resonance spectroscopic measurements of thalamic N-acetyl-aspartate concentrations have been proposed as an excellent binary predictor of adverse neurodevelopmental outcome [19]. However, the statistics presented for that predictor are only valid for death and/or severe disability [44]. When including moderate disability as an adverse outcome, the specificity of adverse outcome remains high (98%), but the sensitivity is only 44%.

A limitation of studies of perinatal asphyxia recruiting to TH shortly after birth, including ours, is the lack of validity of a pre-cooling clinical neurological examination. Before TH, a post-insult neurological examination correlated well with MRI and outcome [45]. However, we found that Sarnat scoring after rewarming in the CoolCap trial was a better predictor in normothermic than cooled infants [46].

In summary, we present a population-based moderate-severe neonatal encephalopathy cohort where passive and active cooling was initiated early and 75% of infants have good outcome. We have developed two novel MRI algorithms, easily determined from conventional MRI; the WMxBGT product and the TIS, that are highly predictive of adverse outcome and CP (PPV 95%), though slightly lower than the regression equation in the supplementary documents and the table. Our data indicates two novel MRI algorithms, easily determined from conventional MRI; the WMxBGT product and the TIS, that are highly predictive of adverse outcome and CP (PPV 95%), though slightly lower than the regression equation in the supplementary documents and the table.

Funding
The staff maintaining the database were supported by the UK children’s charity SPARKS and external charitable donations. Equipment used to cool the patients and the aEEG machines to monitor the brain were supported by the Moulton foundation, the Lerdal Foundation for Acute Medicine and charitable donations. Use of the MRI scanners were supported by grants, the National Health Service and the University of Bristol. Clinical and research staff were supported by the National Health Service, the University of Bristol and the University of Oslo through their salaries.

Author contributions
All authors fulfill the four ICMJE criteria with 1: contribution to the conception, design and analysis of the work (MT, FC, SJ, and LW) and acquisition and interpretation of data (SJ, MT, LW, MMB, MK, EC and FC), 2: Drafting of the work or revision it critically for important intellectual content (SJ, MT, LW, MMB, EC and FC), 3: Final approval of the version to be submitted (SJ, MT, LW, MMB, MK, EC and FC) and 4: Agreement to be accountable for all aspects of the work (SJ, MT, LW, MMB, MK, EC and FC).

Data sharing statement
The datasets analyzed in the current study are not publicly available due to restricted access until Apr 1st 2023, when further information about the dataset is available from the corresponding author on reasonable request.

Declaration of Competing Interest
M Karlsson declares patents Method of Determining Hypoxia and Testing System for Determining Hypoxia Induced Cellular Damage. All other authors have nothing to disclose.

Acknowledgements
We thank our funders, clinical collaborators and parents of patients. Their support has been unaltering in allowing us to carry out treatment and to collect prospective data of infants undergoing Therapeutic Hypothermia which has been the standard of care at St Michael’s Hospital in Bristol since December 2006. We thank consultant neonatologists J Tooley, A Jain, K Luyt, P Cairns, D Harding, D Evans, S Jones, & A Whitelaw as well as other medical staff E Scull-Brown, CS Bond, S Okano, X Liu, B Robbins and J Stone who have been key collaborators. We also thank JK Gundersen for the figures.

Supplementary materials
Supplementary material associated with this article can be found in the online version at doi:10.1016/j.eclinm.2021.100885.

References
[1] Gluckman PD, Wyatt JS, Azzopardi D, Ballard R, Edwards AD, Ferriero DM, et al. Selective head cooling with mild systemic hypothermia after neonatal encephalopathy: multicentre randomised trial. Lancet 2003;363(9440):663–70 (02/12/ed. 2005 Feb 19).
[2] Azzopardi DV, Strohm B, Edwards AD, Dyer L, Halliday HL, Juszczak E, et al. Moderate hypothermia to treat perinatal asphyxial encephalopathy. N Engl J Med 2009;361(14):1149–58 (10/03 ed. 2009 Oct 1).
[3] Shankaran S, Laptokat AR, Pappas A, McDonald SA, Das A, Tyson JL, et al. Effect of depth and duration of cooling on deaths in the NICU among neonates with hypoxic ischemic encephalopathy: a randomized clinical trial. JAMA 2014;312(24):2629–39 (12/24 ed. 2014 Dec 24).
[4] Rutherford M, Ramenghi LA, Edwards AD, Brocklehurst P, Halliday H, Levene M, 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 2009;8(9):39–45 (11/10 ed. 2010 Jan).
[5] Azzopardi D, Strohm B, Linsell L, Hobson A, Juszczak E, Kurinczuk JJ, et al. Implementation and conduct of therapeutic hypothermia for perinatal asphyxial encephalopathy in the UK: analysis of national data. PLoS ONE 2012;7(6):e38504.
[6] Smitt E, Liu X, Jary S, Cowan F, Thoresen M. Cooling neonates who do not fulfill the standard cooling criteria – short- and long-term outcomes. Acta Paediatr (Oslo, Norway: 1992) 2014;104(2):138–45 (08/29 ed. 2015 Feb).
Thoresen M. Patient selection and prognostication with hypothermia treatment.

Karlsson M, Wiberg-Itzel E, Chakkarapani E, Blennow M, Rutherford MA, et al. Predicting motor outcome and death in term hypoxic-ischemicencephalopathy. J Perinatol 2017;37(12):1539-42. 

Martinez-Biarge M, De Wispelaere LA, Ouwehand S, Smidt LCA, Groenendaal F, et al. A novel magnetic resonance imaging score predicts neurodevelopmental outcome after perinatal asphyxia and therapeutic hypothermia. J Perinatol 2018;38(2):190-9. 

Martinez-Biarge M, Bregant T, Wusthoff CJ, Chew ATM, Diez-Sebastian J, Kapellou O, et al. Correlation of conventional magnetic resonance imaging as a predictor of death or disability at 7 years following hypoxic-ischaemic encephalopathy. Arch Dis Child Fetal Neonatal Ed 2012;97(1):F30–7. 

Martinez-Biarge M, Becher J-C, Bhushan SS, Lyon AJ. Unexpected collapse in apparently healthy newborns–a prospective national study of a missing cohort of neonatal deaths and near-death events. Arch Dis Child Fetal Neonatal Ed 2012;97(1):F30–4. 

Becker J-C, Bregant T, Wusthoff CJ, Chew ATM, Diez-Sebastian J, Rutherford M, et al. Feasibility of using neonatal brain NAA to predict infant development. Lancet 2003;361(9359):736–40. 

Palisano R, Rosenbaum P, Walter S, Russell D, Wood E, Galuppi B. Development and reliability of a system to classify gross motor function in children with cerebral palsy. Dev Med Child Neurol 1997;39(4):214–23. 

Weeke LC, Groenendaal F, Mudigonda K, Blennow M, Lequin MH, Meiners LC, et al. Near-term electroencephalography and brain magnetic resonance imaging in asphyxia comparing cooled and non-cooled infants. Eur J Paediatr Neurol 2019;23(1):181–7. 

 Warner ME, Azzopardi D, Rutherford M, Groenendaal F, Bydder GM, et al. Overview | therapeutic hypothermia with intracorporeal temperature monitoring for hypoxic perinatal brain injury | guidance | NICE | [Internet]. NICE; [cited 2020 Nov 28]. Available from: https://www.nice.org.uk/guidance/ipg347.