**PREDICTING HIE: CONVENTIONAL RISK FACTORS COMPARED TO AGNOSTIC MACHINE LEARNING ALGORITHMS**

**INTRODUCTION**

While mothers report that the well-being of their unborn infant is the single biggest priority for them1 there is little evidence to guide them or the professionals looking after their baby2. The prediction of which infants will become compromised around birth is poorly understood3, and has been identified as a priority for the RCOG4 and the UK Department of Health5. We have presented some work that shows that modelling of risk is feasible6 and we know that simple interventions can improve neonatal and maternal7,8 outcomes.

One significant cause of perinatal brain injury is perinatal asphyxia, leading to hypoxic-ischaemic encephalopathy (HIE). HIE if often devastating, with life-long impacts for the infant13 and their family, as well as costing society millions of pounds in medical compensation, lost earnings and welfare support14. As well as the direct impact on infants and families, obstetric practice represents the biggest proportion of legal claims against the NHS15 and even small improvements in outcomes would yield substantial health benefits for individuals and economic benefits for health care services. Indeed, perinatal asphyxia is the 12th biggest cause of disability life years worldwide16 (i.e. a bigger impact than diabetes mellitus or tuberculosis), and even those infants with mild asphyxia have worse measures in cognition, movement and social metrics when compared to their peers17–19 and the true impact of this and other post-term related pathologies, and the economic implications, are unclear20. However interventions, such as induction of labour or operative delivery, can be employed if the risks of continuing the pregnancy are higher than delivery: for either the mother or the infant21. This lack of clear data on the perinatal risks and long term outcomes of these infants likely contributes to the variation in management of mothers with post-term babies11 and current NICE guidelines recognise this, and suggest that a research priority is to “identify babies at particularly high risk of morbidity and mortality who will benefit from induction and therefore avoid induction for babies who do not need it”22.

Risk factors for perinatal asphyxia and encephalopathy have been derived by a number of papers; although one of the most cited remains the work by Badawi *et al*10,26. This work identifies 35 potential risk factors for encephalopathy in an Australian population (Table 1) and together have been cited over 700 times (Data extracted Web of Science 17/12/2019). Analysing such data with Machine Learning (ML) algorithms to predict health outcomes is currently of great interest. However often algorithms require cleaned data, careful feature selection/engineering and significant training and expertise to develop, before they are able to meet, or exceed, clinical prediction. Recently more automated approaches to ML have become available, collectively known as Automated Machine Learning (AutoML).

This work is based on the Collaborative Perinatal Project (CPP) 24. Collection of data was from 14 units across the United States and showed little evidence of selection bias25. The dataset includes data on approximately 60,000 pregnancies, and 58,000 live born infants born between 1959 and 1965. Data was collected throughout the prenatal period, labour and delivery, postpartum and as the child grew.

**AIMS**

1. To evaluate if AutoML approaches applied to a large dataset with minimal human input and domain knowledge can predict poor outcomes as well, or better than conventional approaches with established risk factors.
2. To test if measures of infant growth can improve the prediction of hypoxic-ischaemic encephalopathy.
3. Estimate the amount of hypoxic brain injury potentially preventable using the above models.

**METHODS**

**Feature selection & engineering**

* + **Badawi**
  + **AutoML**
    - Drop high missingness
    - unordered categorical variables were recoded as dummy variables
    - Data minority (disease) class was resampled
    - Drop low variance
    - Predictors with high Pearson correlation (>80%) were eliminated by randomly selecting a single predictor from each pair
    - Feature selection using RF feature importance
    - Continuous predictors were standardised to Z-score in both training and testing data by using the mean and standard deviation of the training data estimated prior to pre-processing
  + **Lasso & Elastic net**

**Comparison of ML models using both sets of features**

* **RF**
* **NN**
* **NB**
* **Ensembl**

regularized trees or regularized random forest, CFS, RFE). Also Lasso/Elastic net for the standard regression.

**Outcomes – HIE & PD**

**Training = first half; test = second half**

**Best model = probability calibration**

**Outcomes**

The primary outcome was hypoxic-ischaemic encephalopathy (HIE) defined as having definite seizures, hypertonia, jitteriness, hypotonia, abnormal reflexes, or abnormal cry; after having a low 5 minute Apgar score (<7)27. Analyses were repeated for perinatal death (the need for resuscitation after birth and the presence of a low Apgar score [<7 at 5 minutes]).

**Feature selection**

Two sets of prediction models were developed. The first used established risk factors for HIE, based on *a-priori* proposed risk factors (Table 1)10,26(‘Conventional analysis’). The second were set were identified by AutoML approaches using all data available in the cohort (‘AutoML analysis’).

**Conventional analysis**

Data were cleaned and harmonised where possible with the measures previously proposed 10,26. A logistic regression model was developed on the first half of the data (infants born 1959-1962) and then tested on the second half (infants born 1963 to 1965). The primary model included only variables measurable at, or before 37 weeks of gestation (Table 1). The initial model was then repeated with antenatal measures and the identified measure of fetal growth (birthweight centile (>90th, 10-90th, 3rd-9th, <3rd). A third model was developed using the antenatal variables and additional variables available to clinicians at the onset, or duration labour, but before the birth of the infant. The model was independently repeated for the other outcome measures of interest.

For all models, a prediction of the outcome of interest was derived and infants allocated to one of 10 risk deciles, ROC curves were derived alongside AUC measures. We calculated the number of infants with HIE in the highest decile to estimate the possible number of infants where HIE may be avoidable by targeted interventions prior to birth.

Comparisons between the three models were performed to test if the addition of growth, or intrapartum measures, improved the prediction of the model.

**AutoML** **analysis**

An agnostic ML model was developed using the same testing-validating cohorts from the previous component. 28 out of 518 variables were discarded as they contained >5% of missing data values; leaving a potential 490 exposure data fields for the prediction models.

All variables were identified as either antenatal (measurable before 37 weeks gestation), growth (birth measures of growth), and intrapartum (measures only available at or after 37 weeks, up to the point to delivery), and classified as either categorical or numerical.

All analyses were performed using a dockerised implementation of Jupyter notebook (jupyter/tensorflow-notebook:7a0c7325e470) with additional packages installed as specified. Each variable set-indication were individually processed prior to training as follows. First, unordered categorical variables were recoded as dummy variables. Second, the training data minority (disease) class was resampled using SMOTE (Synthetic Minority Over-sampling Technique [imbalanced-learn v0.5.0]) to improve class balance between affected/unaffected observations. Third, predictors with high Pearson correlation (>80%) in the training dataset were eliminated by randomly selecting a single predictor from each pair. Fourth, a random forest was applied to the training data to select features with a conservative feature importance threshold (>0.01). Finally, continuous predictors were standardised to Z-score in both training and testing data by using the mean and standard deviation of the training data estimated prior to pre-processing. Following data processing, ML models were trained using predictors identified by the random forest and evaluated using the testing dataset with the following algorithms using default parameters unless otherwise indicated.

A) Logistic regression (sci-kit learn). The maximum number of iterations taken for convergence was increased to 100,000.

B) Random Forest (sci-kit learn)

C) Neural Network (TensorFlow). One hidden layer with number of nodes equal to number of predictors. Rectified linear unit activation function.

D) Adanet ([v0.8.0]). An Ensembl/AutoML binary classifier was prepared using the following models: linear model, neural network (hidden layers/nodes 20, 50, 100, 20 & 10, 50 & 20, 100 & 40, 100 & 40 & 20), gradient boosted trees.

For all methods, the three models will be derived as above, and then repeated for the other outcomes of interest. Infants were allocated a risk decile as in the conventional model; and comparisons made between models as above, and the top 50 measures identified from the ML model as most predictive reported/defined. Finally, we will then test if the corresponding ML derived model differs from the conventional model’s predictive value.

**RESULTS**

**Population demographics**

The dataset was based on the full CPP variable file dataset; containing data on 58,760 infants. A total of 12,005 infants were born preterm (<37 weeks of completed gestation), 5476 were born after 42 weeks, and 964 were born to a mother of less than 16 years age; leaving a total of 40,315 for the analyses. 19,487 infants were born between 1959 and 1962 (and were placed in the first cohort), while 20,828 were born between 1963 and 1966 (and were placed in the second).

Table 2 shows the demographics of the population, split by HIE status. Overall 209 (0.5%) had evidence of HIE, 549 (1.4%) died in perinatal period, 1228 (3.1%) had a low Apgar score at 5 minutes and 2013 (5.1%) required resuscitation after birth. With regard to antenatal factors, infants with HIE were more likely to have older but primiparous mothers, without private health insurance. Mothers were also more likely to have placenta previa and infants more likely to be male and from multiple births; but otherwise antenatal risk factors did not appear to difference substantially. With regard to growth measures, infants with HIE were more likely to be poorly grown. Infants with, and without HIE, differed for most of the intrapartum factors except the recording of a nuchal cord.

**Model performance**

Figure XX shows the area under the curve (AUC) for the prediction models, when applied to the testing (later) pregnancies. The antenatal model reported an AUC of 0.71 (0.64-0.77), which improved to 0.73 (0.67-0.79) (p=0.09) with the addition of infant birth weight to the model, but not when intrapartum measures were included (p=0.68). Addition of growth factors increased the number of infants in the highest risk decile with HIE from 17 (28.8%) to 22 (37.3%).

The AutoML model produced similar AUC measures to the conventional analysis for antenatal (p=0.91) and the antenatal and growth measures (p=0.56). However, the antenatal and intrapartum ML model appeared to predict better than the corresponding conventional model (0.82 (0.78-0.87) vs 0.70 (0.64-0.77), p<0.001); and was able to identify a further 13 infants (44 (49.0%) vs 31 (34.4%)) than the antenatal along model. Factors identified from the ML HIE model as most predictive are shown in table 3. ROC curves for the 6 HIE models are shown in Figure 2.

Other ML results shown in Table 2….

**DISCUSSION**

**Conclusion**

In this work, on a historical cohort, a machine learning model with minimum data preparation was able to match and in some examples exceed the prediction of conventional analysis in predicting which infants would develop HIE after birth. The prediction was substantially improved when measures of growth were included; supporting the role for routine antenatal measures of growth during pregnancies using modern imaging techniques. Routine growth measures, and automated ML models on other routinely collected health data may provide an additional tool to obstetric services to help identify infants at high risk of brain injury around birth, and help target additional observation or interventions.

**Data and code availability**

R code are available from <https://ieugit-scmv-d0.epi.bris.ac.uk/ml18692/hie-ml>. NCPP data are freely available from XXX.

**REFERENCES**

1. Kingdon, C. *et al.* Choice and birth method: mixed-method study of caesarean delivery for maternal request. *BJOG* **116**, 886–895 (2009).

2. Berger, B., Schwarz, C. & Heusser, P. Watchful waiting or induction of labour--a matter of informed choice: identification, analysis and critical appraisal of decision aids and patient information regarding care options for women with uncomplicated singleton late and post term pregnancies: a . *BMC Complement. Altern. Med.* **15**, 143 (2015).

3. Odd, D. E. *et al.* Risk of low Apgar score and socioeconomic position: a study of Swedish male births. *Acta Paediatr.* **97**, 1275–1280 (2008).

4. Gynaecologists, R. C. of O. and. Easy Baby Counts. https://www.rcog.org.uk/eachbabycounts.

5. Health, D. of. New ambition to halve rate of stillbirths and infant deaths. https://www.gov.uk/government/news/new-ambition-to-halve-rate-of-stillbirths-and-infant-deaths.

6. Odd, D., Heep, A., Luyt, K. & Draycott, T. Hypoxic-Ischaemic Brain Injury: Delivery Before Intrapartum Events. in *Joined European Neonatal Societies Congress* (2015).

7. Draycott, T. *et al.* Does training in obstetric emergencies improve neonatal outcome? *Bjog* **113**, 177–182 (2006).

8. Chiossim, G. Timing of Delivery and Adverse Outcomes in Term Singleton Repeat Cesarean Deliveries. *Obs Gynecol* **121**, (2013).

9. Martinez-Biarge, M., Madero, R., González, A., Quero, A. & García-Alix, A. Perinatal morbidity and risk of hypoxic-ischemic encephalopathy associated with intrapartum sentinel events. *Am J Obs. Gynecol2* **206**, 148.e1–7 (2012).

10. Badawi, N. *et al.* Intrapartum risk factors for newborn encephalopathy: the Western Australian case-control study. *Bmj* **317**, 1554–1558 (1998).

11. Gülmezoglu, A., Crowther, C., Middleton, P. & Peatley, E. Induction of labour for improving birth outcomes for women at or beyond term. *Cochrane Database Syst Rev.* (2012) doi:10.1002/14651858.CD004945.pub3.

12. Campbell, M. K., Ostbye, T. & Irgens, L. M. Post-term birth: risk factors and outcomes in a 10-year cohort of Norwegian births. *Obstet. Gynecol.* **89**, 543–548 (1997).

13. Azzopardi, D. V *et al.* Moderate hypothermia to treat perinatal asphyxial encephalopathy. *N Engl J Med.* **361**, 1349–1358 (2009).

14. Odd, D. E., Gunnell, D., Lewis, G. & Rasmussen, F. Long-term Impact of Poor Birth Condition on Social and Economic Outcomes in Early Adulthood. *Pediatrics* **May 9**; **eFi**, e1498-504 (2011).

15. *Ten Years of Maternity Claims: An Analysis of NHS Litigation Authority Data*. (2012) doi:978-0-9565019-2-9.

16. Murray, C. J. L. *et al.* Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet (London, England)* **380**, 2197–2223 (2012).

17. Odd, D. E., Lewis, G., Whitelaw, A. & Gunnell, D. Resuscitation at birth and cognition at 8 years of age: a cohort study. *Lancet* **373**, 1615–22 (2009).

18. Odd, D. E., Gunnell, D., Whitelaw, A. & Lewis, G. The association between birth condition and neuropsychological functioning and educational attainment at school age. A cohort study. *Arch Dis Child* (2010).

19. Odd, D. E., Rasmussen, F., Gunnell, D. J., Lewis, G. & Whitelaw, A. A Cohort Study of Low Apgar Scores and Cognitive Outcomes. *Arch Dis Child Fetal Neonatal Ed* **93**, F115-20 (2008).

20. Goeree, R., Hannah, M. & Hewson, S. Cost-effectiveness of induction of labour versus serial antenatal monitoring in the Canadian Multicentre Postterm Pregnancy Trial. *Can Med Assoc J* **9.**, 1445–50. (1995).

21. Molina, G. *et al.* Relationship Between Cesarean Delivery Rate and Maternal and Neonatal Mortality. *JAMA* **314**, 2263–2270 (2015).

22. Induction of labour. *NICE Clin. Guidel. (July 2008)*.

23. National Institute for Health and Clinical Excellence (NICE). Inducing labour (CG70). (2008).

24. Hardy, J. B. The Johns Hopkins Collaborative Perinatal Project. Descriptive background. *Johns Hopkins Med. J.* **128**, 238–243 (1971).

25. Niswander, K. & Gordon, M. T. *The Women and Their Pregnancies: The Collaborative Perinatal Study of the NINDS*. (USGov. Printing Press, 1972).

26. Badawi, N. *et al.* Antepartum risk factors for newborn encephalopathy: the Western Australian case-control study. *Bmj* **317**, 1549–1553 (1998).

27. Odd, D., Lewis, G., Whitelaw, A. & Gunnell, D. J. Resuscitation at birth and cognition at 8 years of age: a cohort study. *Lancet* **9**, 1615–1622 (2009).

**Table 1. Established risk factors**

|  |  |  |
| --- | --- | --- |
| Antenatal Factors | Growth Measures | Intrapartum Factors |
| Maternal age (<20, 20-24, 25-29, 30-34, >35)  Parity 0, 1,>1  Maternal Employment  Health Insurance  Maternal race  FHx of seizures (recurrent non-febrile seizures)  FHx of neurological disorder (excludes seizures)  Infertility Treatment  Maternal Hypertension  Maternal height (<160, 160-164, >164)  Maternal Thyroid Disease  Pre-eclampsia  Antenatal bleeding (mod or severe)  Viral Illness  Alcohol (some, none, unknown)  Birthweight centile (>90th, 10-90th, 3rd-9th, <3rd)  Sex  Abnormal placenta  Late or no antenatal care  Multiple births | Birth weight centile (>90th, 10-90th, 3rd-9th, <3rd) | Gestation (37-42)  OP presentation  Maternal Pyrexia  Maternal Intrapartum Event (Haemorrhage, convulsions, uterine rupture, snapped cord, out of hospital birth)  Membrane rupture >12 hours  Blood Pressure abnormalities – Captured above  Nuchal cord  Cord prolapse  Onset of labour (spontaneous, induced, none)  Mode of delivery (Spontaneous, induced vaginal, elective CS, emergency CS, breech manoeuvre)  Shoulder dystocia  Epidural Anaesthetic  Breech Presentation  ROM>12 hours |

**Table 2. Demographics of study population (split by HIE)**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Characteristic** | | **Non-HIE infants** | **HIE infants** | **P** |
| **Ante-natal Measures** | | | | |
| Late Booking\*\*\* |  | 11,405 (29.1%) | 60 (28.7%) | 0.905 |
| Thyroid Disease |  | 1,028 (2.6%) | 5 (2.4%) | 0.837 |
| Maternal Age | < 20 years | 11,057 (28.2%) | 57 (27.3%) | <0.001 |
|  | 20-24 | 11,690 (29.8%) | 49 (23.4%) |  |
|  | 25-29 | 8809 (22.5%) | 36 (17.2%) |  |
|  | 30-34 | 4,644 (11.8%) | 39 (18.7%) |  |
|  | 35 or more | 3,022 (7.7%) | 28 (13.4%) |  |
| Parity | 0 | 10,434 (26.7%) | 78 (37.3%) | 0.001 |
|  | 1 | 8579 (22.0%) | 29 (13.9%) |  |
|  | 2 or more | 20,049 (51.3%) | 102 (48.8%) |  |
| Employed |  | 5,989 (15.5%) | 23 (11.2%) | 0.084 |
| Private Insurance |  | 2546 (7.0%) | 5 (2.7%) | 0.022 |
| Race | White | 19,560 (49.9%) | 63 (30.1%) | <0.001 |
|  | Black | 16,898 (43.1%) | 123 (58.9%) |  |
|  | Other | 2,764 (7.1%) | 23 (11.0%) |  |
| FHx of Seizures |  | 2560 (6.7%) | 16 (7.9%) | 0.500 |
| FHx Neurology\* |  | 1479 (3.9%) | 12 (5.9%) | 0.133 |
| Fertility Ix |  | 1008 (2.6%) | 6 (2.9%) | 0.797 |
| Hypertension |  | 167 (0.4%) | 1 (0.5%) | 0.911 |
| Preeclampsia |  | 1284 (3.3%) | 19 (9.1%) | <0.001 |
| Maternal Height | <160cm | 13,221 (36.4%) | 81 (41.3%) | 0.354 |
|  | 160-164cm | 10,961 (30.2%) | 54 (27.6%) |  |
|  | >164cm | 12,172 (33.5%) | 61 (31.1%) |  |
| Pre-labour bleeding |  | 10,792 (28.1%) | 69 (33.8%) | 0.071 |
| A/N Viral Illness |  | 2,688 (6.9%) | 15 (7.2%) | 0.846 |
| Alcoholism |  | 44 (0.11%) | 0 (0.0%) | 0.628 |
| Fever |  | 5,068 (13.0%) | 26 (12.4%) | 0.817 |
| Male |  | 19,842 (50.6%) | 134 (62.1%) | <0.001 |
| Placental Previa |  | 160 (0.41%) | 3 (1.5%) | 0.020 |
| Multiple Birth |  | 290 (0.74) | 5 (2.4%) | 0.006 |
| **Growth Measures** | | | | |
| Birth weight centile | Less than 3rd | 1208 (3.1%) | 29 (14.1%) | <0.001 |
|  | 3rd to 10th | 2898 (7.4%) | 28 (13.6%) |  |
|  | 10th to 90th | 31,265 (79.8%) | 125 (60.7%) |  |
|  | Above 90th | 3824 (9.8%) | 23 (11.7%) |  |
| **Intra-partum measures** | | | | |
| OP presentation |  | 2512 (6.6%) | 34 (16.8%) | <0.001 |
| Breech Presentation |  | 1023 (2.7%) | 31 (15.3%) | <0.001 |
| ROM>12 hours |  | 5706 (16.5%) | 50 (30.5%) | <0.001 |
| Caesarean Section |  | 2076 (5.3%) | 38 (18.2%) | <0.001 |
| MIE\*\* |  | 3000 (7.7%) | 42 (20.1%) | <0.001 |
| Nuchal cord |  | 10,225 (26.3%) | 52 (24.9%) | 0.636 |
| Prolapsed cord |  | 311 (0.8%) | 12 (5.7%) | <0.001 |
| Onset | No Labour | 1,146 (3.0%) | 12 (5.8%) | 0.019 |
|  | Spontaneous | 35,124 (90.3%) | 177 (85.1%) |  |
|  | Induced | 2636 (6.8%) | 19 (9.1%) | 0.019 |
| Shoulder Dystocia |  | 230 (0.6%) | 9 (4.3%) | <0.001 |
| Epidural |  | 617 (1.6%) | 10 (4.9%) | <0.001 |

\* Motor, sensory or developmental disorder in siblings

\*\* APH, eclampsia, uterine rupture or ruptured cord

\*\*\* >26 weeks of gestational age