The Development and Validation of a Model for Predicting Neurological Disability following Neonatal Intensive Care.

Thesis submitted for the degree of

Doctor of Medicine

at the University of Leicester

by

Jon Stewart Dorling MBChB (Dundee)

Department of Health Sciences
University of Leicester
2007

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

The Development and Validation of a Model for Predicting Neurological Disability following Neonatal Intensive Care.

Thesis Abstract **Jon Dorling 24/9/07**

Aims

This thesis set out to test whether it is possible to predict neurodevelopmental outcome at two years of age using data collected in the first 12 hours of life or during the entire admission of a premature newborn infant. Outcomes were tested separately (severe disability in survivors against other survivors) or in combination (death or severe disability against survival without severe disability)

Methods

The hypotheses were tested in three cohorts; the East Anglian Very Low Birthweight Database, the Trent Neonatal Survey 'ABC study' and the United Kingdom Trial of Oscillation (UKOS).

After exploration of the cohort data quality, each was used in turn to develop a model for predicting outcome using data in the first 12 hours of life. The UKOS dataset was also used to test prediction using data available from the entire admission.

Univariate analysis was used to determine which variables were associated with the outcome of interest. Logistic regression was then used to develop the models and ROC curve analysis performed to test the predictive ability of the models.

Results

Each of the three models for predicting the combined outcome of death or severe disability appeared to predict reasonably well with areas under the curve of 0.808, 0.793 and 0.798. Further testing however showed that these models were good at predicting death but that they were very poor at predicting disability.

Data from the entire admission were successfully used to develop a model that predicted disability adequately in survivors (Az= 0.842) suggesting the importance of this time-period.

Conclusions

Further study is needed to determine whether additional data from before 12 hours of age would enable outcome prediction. Prediction of disability amongst survivors appears possible using data from the whole admission, and after testing in other cohorts, offers a number of future epidemiological uses.

To Carole, Anna, Matthew, and Fiona

Acknowledgements

I should like to thank my family for their encouragement, support and love during the completion of this thesis.

I am especially grateful to my supervisor David Field, whose patience, enthusiasm and encouragement made this thesis possible.

I would also like to thank Brad Manktelow and Liz Draper and all the members of the Neonatal Survey group within the University of Leicester.

In addition I am extremely grateful for the help of Neil Marlow, Janet Peacock and all the staff involved in the data collection for each of the study cohorts.

Contents

Tables

Figures

Abbreviations

Az	Area under the curve
BSID	Bayley Scales of Infant Development
CMR	Crude Mortality Rate
Cranial US	Cranial Ultrasound
CRIB	Clinical Risk Index for Babies
DNA	Deoxyribonucleic acid
EAVLBW	East Anglian Very Low Birthweight Cohort
FiO_2	Fraction of inspired oxygen
g	Gram
HIE	Hypoxic Ischaemic Encephalopathy
HRQOL	Health Related Quality of Life
ICIDH	International Classification of Impairments, Disabilities, and Handicaps
IMR	Infant Mortality Rate
Kg	Kilogram
mEq/L	Milliequivalents of solute per litre of solvent
mmHg	Millimetres of mercury
MRI	Magnetic Resonance Imaging
NBRS	Nursery Neurobiologic Risk Score
NHS	National Health Service
NICHHD	National Institute of Child Health and Human Development
NICU	Neonatal Intensive Care Unit
NMPI	Neonatal Mortality Prognosis Index
NMPI	Neonatal Mortality Prognosis Index
NTISS	National Therapeutic Intervention Scoring System
OHSQ	Oxford Health Status Questionnaire
paO_2	Partial pressure of oxygen dissolved in the plasma
PEDI	Pediatric Evaluation of Disability Inventory
ROC	Receiver Operating Characteristic
SGA	Small for Gestational Age
SMR	Standardised Mortality Rate
SNAP	Score for Neonatal Acute Physiology
SNAP-PE	Score for Neonatal Acute Physiology - Perinatal Extension
SPSS	Statistical Package for the Social Sciences
TISS	Therapeutic Intervention Scoring System
TNS	Trent Neonatal Survey
UK	United Kingdom
UKOS	United Kingdom Oscillation Study
US	United States
VLBW	Very Low Birthweight
WHO	World Health Organisation

CHAPTER ONE: INTRODUCTION & REVIEW OF THE LITERATURE

1.1 INTRODUCTION TO THESIS

This thesis documents the development of models to predict neurological outcome in premature infants. It goes on to discuss attempts at validating these models in other datasets and describes some of the difficulties with different types of data. The models were produced to play a role in future studies identifying the causes of variation in these outcomes and to allow risk correction in observational studies. Results from observational studies into these factors are difficult to interpret due to selection bias and differences in case mix. By enabling risk correction, models could allow identification of factors that are not easily tested by randomised controlled trials. An organisational factor such as whether extremely premature infants should be treated locally in a smaller unit or be transferred to a large tertiary unit is an example of the potential use of this methodology.

This thesis begins with an introduction to neonatal care, a literature review of evaluation methods, a review of neonatal prediction scoring systems in general and then specifically for neurological outcome. This is followed by a discussion of how to define outcome and common problems for research in this area and an explanation and listing of the hypotheses analysed in this work, followed by the study of each of these hypotheses.

1.2 NEONATAL CARE FOR PREMATURE INFANTS: A BRIEF HISTORY

In order to inform the reader, this thesis begins with a review of neonatal care and its historical background. In the twenty first century medical science has progressed to a point where infants born as much as 17 weeks prematurely and weighing less than 500 grams are treated with intensive care, as and when needed, in order to save their lives. Exceptions to this are those born before the gestational age limit of what is technically possible. This limit of what is known as 'viability' is dependent on the resources available to care for the infant and the local attitudes or beliefs about what is best for that baby or the population. There is a particularly notable difference between the approaches in Holland and the United States[1, 2]. Rhoden described these approaches[3] as a "statistical approach" for the Dutch policy of withholding intensive care from infants considered to be at too much risk of disability and the "wait until certainty" strategy for the American inclination to treat all infants until the prognosis is clearly futile. A third method, the "individualised prognostic strategy" was also described and is probably the one employed most often in the United Kingdom[2]. Due to financial constraints and higher mortality rates, there is also a different expectation of what should be done in developing countries. Treatment in these countries may depend on the ability of parents to provide the finance to support the treatment of their infants.

In a similar way, attitudes towards the weakest infants have also varied through history. The practice of the Spartans for leaving infants on Mount Taygetos is well known and an often quoted example of infanticide[4]. Seneca also justified this practice in Roman times saying "We drown the weakling and the monstrosity. It is not passion, but reason, to separate the useless from the fit"[5].This practice of

abandoning or killing children shortly after birth was eventually outlawed by Christianity marking a switch away from Eugenics[6]. The spread of Christianity across the world led to the provision of care for the poor, sick, widows and strangers as advised by the first council of Nicaea in 325AD[7]. In the middle ages, hospitals developed from almshouses and hostelrys attached to monasteries, but these probably had a bigger interest in saving patient's souls rather than lives[8]. During the 18th and 19th centuries, hospitals were established in the majority of Industrial Towns and Cities of the United Kingdom via prosperous sponsors[9].

Traditionally, delivering infants and managing the final stages of pregnancy was the province of midwives treating mothers at home. As Dunn describes[10], there was a gradual involvement of physicians in the care of both mothers and their infants; 'the man-midwives, as they came to be called - men such as William Smellie (1697-1763) and William Hunter (1718-1783) achieved professional recognition in the 18th and 19th centuries as physician-accoucheurs'. These surgeons were involved when instrumental interventions were required for abnormal labours[11]. This coincided with the development of 'lying-in' hospitals with the first maternity hospital in the British Isles (London's General Lying-In Hospital) being opened in London in 1739. However due to the development of gynaecological operations and caesarean sections, these all-round specialists were displaced by surgeons who had such ample obstetric and gynaecological surgical workloads that the care of the infants was 'consigned to midwives and nursery nurses'[10].

The first hospital dedicated to treating children opened on April 24th, 1769. George Armstrong's Dispensary for the Infant Poor, located at 7 Red Lion Square, London, treated the infants of the poor for free, including nearly 35,000 infants during it's 12 years of existence[12]. Other children's hospitals were also successfully founded by

public subscription, including Great Ormond Street in 1852[13]. By the 1920s, most cities in the United Kingdom had children's hospitals. These were staffed by medical physicians with a special interest in Paediatrics with the first full time Paediatrician (Sir Frederic Still) being the president of the British Paediatric Association at it's founding in 1928[14]. Unfortunately the majority of newborn infants were still cared for in the maternity wards in general hospitals but Paediatricians became more involved by visiting, often in an honorary capacity.

The first unit for treating premature infants was set up by Victoria Mary Crosse in Birmingham in 1931 who also produced the first British textbook on the subject of Prematurity [15]. The establishment of the NHS[16] in 1948 by Aneurin Bevan opened up the role of Paediatricians to work within maternity hospitals. Junior Paediatricians were also employed at this time to perform the time consuming but life saving new procedure of umbilical exchange transfusion for rhesus incompatibility[17]. Unfortunately the early days of paediatric care of newborn infants were associated with a number of adverse outcomes, as the specialty established itself and learnt what was appropriate for babies as well as what was possible. Retrolental fibroplasia[18], now known as retinopathy of prematurity[19] was caused by injudicious use of oxygen, whilst starving was also employed to prevent aspiration, and hypothermia was not uncommon[10].

Since 1950, the intervening decades to the present time in the early 21st Century, have seen rapid and impressive strides in neonatal care. Important advances during this time are documented in Table 1[10, 20]. As time has passed there has been increasing interest in the quality of this care, with methods being developed for monitoring and comparing outcomes. The next section describes these approaches.

Table 1: Progress in Neonatal Intensive Care 1950-2000

1952	Apgar Score for assessing newborns
1952	Necrotizing entercolitis described
1952	Initial trial showing link between excessive oxygen and retinopathy of prematurity
1953	Structure of DNA published
1953	Respiratory distress syndrome described
1953	First neonatal surgical unit opened at Alder Hey Children's Hospital, Liverpool
1955	Trial showing that withholding fluid immediately after birth is not beneficial
1956	Chance observation of effect of sunlight on serum unconjugated bilirubin
1956	Correct number of 46 human chromosomes published
1958	Trial of hypothermia demonstrates link to decreased survival
1958	Description of the benefits of light for hyperbilirubinaemia
1959	Surfactant deficiency identified as the cause of respiratory distress syndrome
1959	Trisomy 21 identified in Down's Syndrome
1961	Thalidomide linked to birth defects
1963	First report of intrauterine fetal transfusion
1963	Newborn screening test for phenylketonuria
1963	First published description of fetal intrauterine transfusion
1964	Group B streptococcus as common cause of neonatal sepsis
1966	Prevention of Rhesus Haemolytic Disease of the Newborn by use of Anti-D
1967	First heart transplant in newborn
1968	Total parenteral nutrition for newborns described
1969	Trial of phototherapy treatment of hyperbilirubinemia
1972	Intermittent mandatory ventilation for respiratory distress syndrome
1972	Trial of antenatal glucocorticoids for prevention of respiratory distress syndrome
1974	Indomethacin constriction of the ductus arteriosus observed
1975	Patency of ductus arteriosus by prostaglandin E depicted
1975	Extra-corporeal membrane oxygenation first used in infants
1980	Surfactant treatment for respiratory distress syndrome
1980	Human immunodeficiency virus and acquired immunodeficiency syndrome first described
1985	Trial of cryotherapy for retinopathy of prematurity
1987	Oxygen Saturation monitoring for newborns
1990	Surfactant treatment for respiratory distress syndrome
1992	Supine sleeping position for infants reduces sudden infant deaths
1994	Maternal zidovudine reduces perinatal transmission of HIV
1997	Nitric oxide for neonatal pulmonary hypertension
1997	Trial of nitric oxide for pulmonary hypertension in the newborn
2000	Initial mapping of human genome completed

1.3 EVALUATING NEONATAL CARE

As is only natural, once a system is established to provide healthcare the issue of how well the system performs becomes pertinent. In the United Kingdom the outcomes of pregnancy have been of interest to governments and public health staff for many years. Statistics on Births and Deaths were first consistently collated in the 19th Century following a law passed in 1812 that made those who failed to register births and deaths liable to a fine[21]. Further analyses of stillbirths, infant and perinatal mortality were first performed in 1905[22] and have been continually used since then[23]. It is not surprising that this information is of interest to those planning and monitoring health care services in the maternity sector. As well as comparison with past performance, the collection of good quality, accurate data allows the comparison of populations of patients, such as those from geographical areas, or of groups treated in the same hospital.

Intensive care services are particularly costly and due to limited financial resources are therefore subject to widespread scrutiny of both outcomes and costs[24-27]. In addition, there has been an acceptance that inevitably mistakes do occur in hospitals and intensive care is the most important location for these errors[28, 29]. Following media and public interest there has consequently been increased acceptance of the need for the measurement and comparison of quality of care. These measurements can then be used to stimulate the development and improvement of services. These assessments have gained additional importance in the United Kingdom with the focus of both the media and government on outcomes following cardiac surgery in Bristol[30]. Neonatal services are not exempt from these considerations. Important neonatal outcomes that can be measured and compared

include death rates, lengths of stay and common and important complications such as chronic lung disease, retinopathy of prematurity, or neurodevelopment.

Mortality Rates

Crude Mortality Rates (CMR) were the first technique used to measure the quality of health of populations and were a progression from what were known as 'bills of mortality'. These 'bills' were essentially a list of the number of people dying in a certain area of different conditions. These bills were first produced in the 16th century principally to identify the early signs of plague epidemics[31]. Crude mortality rates are simple to calculate, being the total number of deaths per 1000 people. They are not however very useful as they can give a misleading impression as an older population may have a higher death rate when compared to a younger population[32]. This is an important issue, as western populations age with an increasing crude mortality being seen over time[33].

Standardised Mortality Measurements

In order to correct for this problem Standardised Mortality Rates (SMRs) were developed. SMRs are derived by correcting for the age of the population being studied typically to a standard American population[32]. This explains why they are also known as 'age-adjusted rates'. To calculate a SMR, one takes the number of observed deaths in a group of individuals of a certain age and divides it by the number of deaths that would be expected from the reference population[34]. Whilst being widely used in studies of adult populations, this methodology has been little used in studies of newborn infants. SMRs were used in the 1960s and 1970s to compare geographical areas[35-37] but were superseded by risk correction methods due to concerns about variations in birthweight and gestational ages[38].

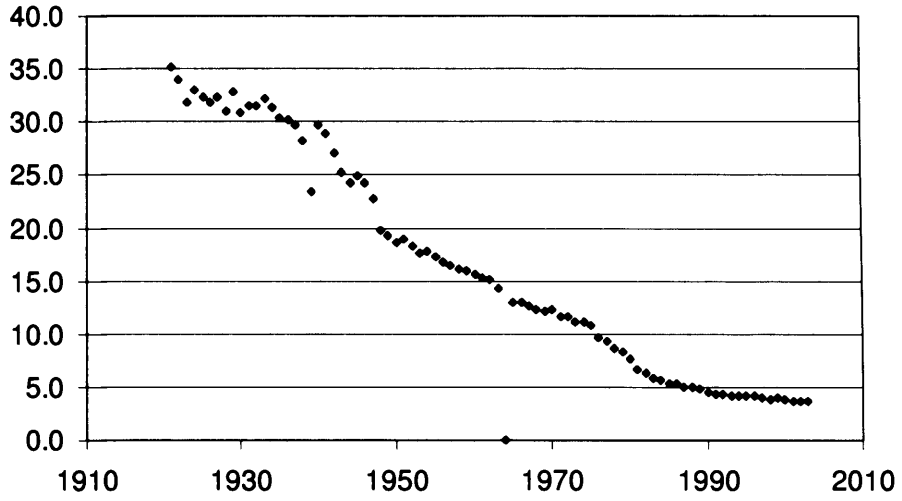
Infant Mortality Measurements

Throughout history, the period of infancy from birth to the first birthday has had the highest risk of death even into the twentieth century. Writing about Dublin in the 1780s, Moore[39] stated that '16.5 per cent of infants born died before they were two weeks old' . By the simple expedients of 'opening the windows and attending to ventilation and cleanliness this was reduced to 4 per cent within a few years mainly by the elimination of tetanus of the newborn, to which previously more than nine-tenths of the deaths were due'. Similar figures were also seen in England, with Percival writing in 1789[40], that 'in Manchester half the children born die before reaching the fifth year'.

From these concerns about infant death a widely used method of comparing mortality was developed. This infant mortality rate (IMR) consists of the number of deaths before the first birthday per 1000 live births. Unlike CMRs in adults, IMRs only cover the first year of life and are therefore not affected by the age of the population. The IMR in Sweden, Scotland, England and France during the early 1860s were documented in the *Journal of the Statistical Society* of London, in 1866[41]. In 1860-61, the infant mortality per 1,000 registered births was 141 in Sweden, 149 in Scotland, 170 in England, and 223 in France. There has been a continual improvement in this marker of healthcare quality until recent years when a plateau may have been reached (see Figure 1).

Figure 1: Neonatal Mortality Rate per 1,000 live births, England and Wales 1921 - 2003.

Data from Office for National Statistics, Mortality Statistics, Series DH3.



Comparing the performance of neonatal units

Comparing the rates of important outcomes can give useful information regarding the care of infants in a particular area. They have a number of flaws however and it was because of these that alternative techniques such as risk correction were developed. The biggest problem is inherent variations in the level of risk (e.g. ill health) within the populations that are compared, with for example, mortality rates being higher in poorer communities[42]. These risk factors are often a more significant cause of variations in outcomes than the medical care the infants receive. For this reason, predictive scores that enable correction for levels of risk (risk-adjustment) were developed to compare mortality in different countries, geographical areas or neonatal units[43]. These will be discussed in detail in the next section. An example of a risk score in action is Tarnow-Mordi's comparison of two neonatal units, showing a difference in survival after risk correction where there was no apparent difference in crude mortality[44]. The UK neonatal staffing study[45] also used risk correction for comparing different factors related to outcome in 54 randomly selected NICUs in the

United Kingdom. This study demonstrated that risk-adjusted mortality was unrelated to patient volume or staffing provision and suggested that only transferring the most premature infants for tertiary care offered any potential to improve outcomes.

1.4 USES AND BENEFITS OF OUTCOME PREDICTION

Predicting the outcome of an admission to hospital or an operation is important for a number of reasons: some of these are listed in Table 2. Individual patients, their relatives or those caring for them are frequently interested in prognosis, especially the chance of survival. Prediction may also have a role in allocating patients to suitable treatments or for ensuring a similar mix of patients in trials[46]. Whilst adequate randomisation should mean that this is unnecessary, the small size of trials in neonatal care may make this a useful process. Scoring systems such as the clinical risk index for babies (CRIB, described on page 25) are often used for comparing treatment and control groups of infants[47, 48]. In this latter situation they are used to measure risk: an ability that can enhance the analysis of outcomes in groups of patients by allowing comparison of 2 groups for similarity of risk (e.g. the population of two intensive care units), for looking at patterns of care and outcomes in groups of patients or over differing timescales.

In neonatal intensive care two main uses of scoring systems predominate. First, and foremost they are used to correct for risk, enabling comparison of the outcome of different populations of babies. Examples of such 'group predictions' are given in the next section and in Table 2. Second, an 'individual prediction' of the risk for a particular infant can be helpful for counselling parents about the chances of their infant surviving to discharge. Individual predictions may also be useful for identifying patients likely to benefit from certain treatments or to stratify infants in trial groups to ensure similarity of risk. A third use: using an individual's outcome in decision making about appropriateness of care is highly controversial and not widely practised. Potentially a risk score could be used to ration intensive care, but there is

considerable disquiet over this approach as scores do not give a 100% chance of death and do not take into account family wishes[49-51]. A slightly less controversial approach has also been suggested where the score is used to confirm that continued intensive care is futile or is becoming so[52, 53]. It has however been noted that physicians identify futility early and little extra resource use would be saved by identifying these situations earlier[54].

Table 2: Uses of Outcome Prediction

INDIVIDUAL PREDICTIONS

1. Giving prognostic information
2. Stratifying infants in trials (to ensure similarity of risk)
3. Determining individual treatment
4. Use as surrogate markers for later outcome

GROUP PREDICTIONS

1. Comparing study groups for similarity of risk
2. Auditing the severity of illness in different units
3. Comparing the performance of different units
4. Determining trends in results over time
5. Reviewing if infants are treated appropriately for risk (e.g. number of septic screens or ventilation days)
6. Comparing rates of complications; are some preventable?

GROUP PREDICTIONS

Risk correction is the most frequent research technique utilising scoring systems. Also known as risk adjustment this approach attempts to tease out whether it is the quality of care or the quality of the patients that leads to variation in outcomes[55]. Risk correction was developed for premature infants to attempt to explain the wide variations in crude survival rates between different neonatal intensive care units[56]. It was unclear whether these differences were primarily due to the compared populations comprising infants of differing risk, or due to the quality of care being significantly different. Units treating few sick infants usually see much lower crude mortality rates than tertiary referral units who treat the smallest, sickest infants[44, 56].

In risk correction, a summary score is derived from measurements of each individual's illness severity, in order to determine the risk of an outcome before treatment is received. If suitably sized populations are taken, comparison can then be made between the expected and observed outcome in a population. Using this method it is therefore possible to analyse the admissions to a neonatal unit over a suitable time period to determine the unit's performance. This might be needed as sicker infants might be offered treatment or obstetric practices might improve the condition of admitted infants. Kaaresen and colleagues used the CRIB score in this way to show that risk corrected survival improved over 16 years[57]. It has been suggested that this methodology may be used increasingly for comparing outcomes over time and between units since the Kennedy report into Paediatric Cardiac Surgery[30]. Indeed a national neonatal audit data collection system including CRIB II risk of mortality was launched in April 2007[58].

When using these scoring systems for risk adjustment it is vital that the score reflects the condition of the infant at birth or as soon afterwards as possible. Any delay in measurement allows the quality of treatment received by the infant to affect the score and could introduce bias. CRIB, for example is measured at 12 hours of age and a unit performing poorly at stabilising infants would have infants with higher CRIB scores due to needing a higher concentration of oxygen or having a lower temperature as a consequence of poor respiratory or thermoregulatory care in the first 12 hours of life. Any increase in the scores would therefore reflect the quality of care rather than the innate risk of the infants at birth. It is also possible to innocently or intentionally manipulate scores by over-treating an infant and increasing the score for some infants and therefore the expected number of deaths in a population. In both of these situations the observed number of deaths may then be less than that predicted and the unit would appear falsely good. Scoring at 12 hours of age also allows the quality of treatment to impact on the score. CRIB is however a relative improvement on other scores that measure data after 12 hours[59].

Disease severity scores have been used to compare mortality rates (for example in Scotland and Australia)[60]. The authors examined 1628 admissions in six Australian NICUs, 775 in five Scottish tertiary NICUs, and 148 in three Scottish non-tertiary NICUs. They demonstrated risk adjusted hospital mortality that was approximately 50% higher in Scottish neonatal units as opposed to the Australian units. They have also been used to investigate other outcomes as variables in multiple regression analysis. An example of this is Kahn's study which showed a very wide variation in the use of narcotic pain relief which was also related to illness severity[61], Vyas and co-workers looked at actual and expected rates of death in 5 neonatal units, along with rates of retinopathy of prematurity in the same units[62]. The unit with the lowest

mortality also had the highest rate of retinopathy of prematurity suggesting an association.

PREDICTING INDIVIDUAL OUTCOMES

Decisions on whether to resuscitate infants are commonly based on information available at the time of birth; a birthweight of less than 500g or a gestational age of less than 23 weeks are often used as a reason for not resuscitating infants. The use of more complex prognostic scoring systems in other circumstances is controversial; raising both legal and ethical concerns. From a practical point of view there are major difficulties; using different risk scores may give similar group predictions but individual estimates can differ significantly, lessening the usefulness of a score in a clinical situation[63]. Most clinicians prefer not to use percentage chances when counselling but to alert the parents to the risk by using terms such as likely, more than likely or very concerned to express the level of risk. For predictive scores to be useful in clinical practice, they must be easy to use, predict outcomes reproducibly, and be applicable to the required group of neonates to be studied[64].

When considering the future of an individual infant, clinicians may be able to prognosticate as accurately as any scoring system, since they can take account of the full clinical picture in that child[2, 65]. Indeed it has been suggested that combining clinical assessment with a scoring system would improve the accuracy of risk assessment[66]. While this could be important in clinical practice for individuals, using clinicians' views for group predictions and research would introduce an unacceptable level of subjectivity and potential bias.

COMMONLY USED SCORES FOR PREDICTING MORTALITY

Risk prediction for newborn infants has to a great extent focussed on mortality [67]. In the United Kingdom the most commonly used risk correction tool is the CRIB score. This 'Clinical Risk Index for Babies' was created to predict mortality for very low birthweight (VLBW) infants born at less than 31 weeks gestation at birth and was derived from a dataset of 812 infants admitted to four UK tertiary neonatal units between 1988 and 1990[56]. A quarter of the infants died. Six variables were identified by logistic regression. These are birthweight, gestational age, presence of a congenital malformation, along with the physiological markers of maximum base deficit in the first 12 hours of life and minimum and maximum appropriate inspired oxygen concentration in the first 12 hours.

The final score is based on a weighted sum of these six factors. In the original study, the score had good discriminatory ability (area under the ROC curve: Az = 0.90), considerably better than birth weight alone (Az = 0.78)[56, 68]. Rautenon produced an area under the ROC curve using CRIB of 0.87 in another cohort [69]. The ease of data collection is a major advantage of CRIB, as calculation takes five minutes per infant, compared with 20– 30 minutes for some of the more complex scores which will be described below[70]. CRIB is also assessed over the first 12 hours of life, making it somewhat less susceptible to treatment effects biasing the value[71, 72].

CRIB was recently updated to CRIB II[73]. It uses a mortality risk grid known as the 'Draper grid'[74]. This gives a percentage chance of survival to discharge according to gestational age and birth weight. CRIB II combines this and admission temperature and base excess to predict mortality. It was intended to improve predictions for smaller, very premature infants and to exclude variables that could be influenced by

care given to the infant, although the inclusion of admission temperature complicates this attempt.

In the United States the most widely accepted and utilised score is the SNAP score. This 'Score for Neonatal Acute Physiology' was developed in 1990 using data drawn from three units in Boston, USA covering 1643 infants, 154 of whom weighed less than 1.5kg at birth[75]. Unlike CRIB, SNAP can be applied to any infant admitted to a neonatal unit, but perhaps because of the small number of VLBW infants in the population from which it was derived, it has reduced sensitivity to differences between the most premature infants[76]. SNAP scores are based on 28 items collected over the first 24 hours of life from a variety of sources including each body system and selected blood test results. Unlike the CRIB score, where parameters are weighted according to their statistical relation to death, the variables were devised and weighted by experts. The original cohort was also used to extend SNAP to form the SNAP-PE score (Score for Neonatal Acute Physiology—Perinatal Extension) by adding birth weight, small for gestational age (weight under 5th centile for gestation), and low Apgar score at
five minutes[76]. In Richardson's comparison, SNAP predicted death better than birth weight alone (Az 0.87 v 0.77), and SNAP-PE was even better (Az 0.93).

In a similar way that CRIB was updated, SNAP and SNAP-PE have also been updated to SNAP-II AND SNAPPE-II[77]. This update was primarily designed to make data collection easier but also improved validity by using impressively large derivation and verification cohorts of 10,819 and 14,610 infants respectively. Changes included shortening the period of data collection to 12 hours and reducing the number of variables to six (mean blood pressure, lowest temperature, PaO_2 : FiO_2 ratio, serum pH, multiple seizures, and urine output). These factors were

assessed as having the strongest statistical association with mortality. As with the original SNAP score, SNAP II was also extended to produce the SNAPPE-II by adding the perinatal extension factors. Richardson demonstrated good discrimination (Az 0.91) and calibration (Hosmer-Lemeshow 0.90) for SNAPPEII in predicting mortality[77].

Four other mortality scores have also been developed, these are NTISS, NICHHD, and the Berlin Score. The National Therapeutic Intervention Scoring System (NTISS)[78] was derived by an expert panel as a modification of the adult intensive care score known as the therapeutic intervention scoring system (TISS). NTISS is atypical as it is based on the treatments received by an infant. As therapy is determined to a large extent by policy and practice in units, it cannot be used to compare quality of care between units.

The National Institute of Child Health and Human Development (NICHHD) score[79] was created using factors recorded on admission to one of seven US units for 1823 infants born between 1987 and 1989 and weighing 501– 1500g. Logistic regression was used to select the variables, with validation using another 1780 infants. It has been little used since development however. The Berlin Score was developed using logistic regression methods for only 396 VLBW development infants and 176 VLBW validation infants from 1988 to 1991[68]. It suffers from the inclusion of a number of subjective factors which limit its role as a means of objective comparison between units. An additional Mexican score exists: Neonatal Mortality Prognosis Index (NMPI)[80] but this too is little used.

As mortality has improved over time, there has been increasing interest in predicting morbidity, particularly as improved survival may mean infants surviving but with limited quality of life[1,81]. Section 1.6 therefore describes previous attempts at predicting neurodevelopmental outcome. Wishing to give the reader a greater understanding of the detailed discussion of these endeavours at predicting neurodevelopmental outcome, section 1.5 first introduces the methods used to develop and test such models.

1.5 DERIVATION AND VALIDATION OF SCORING SYSTEMS

In order to explain the methodology and avoid using large methods sections in each of the study chapters, a few words now follow on how models are derived and validated. Two steps must be carried out when producing a new scoring system for use by other researchers, these are to develop a score from one population and then to test it in another population. These steps are known as derivation and validation.

Derivation

Derivation describes the identification of predictive variables, measuring and weighting these variables in a 'derivation cohort' which can then be validated in further data samples.

There are two methods used to develop models for risk correction. Clinical knowledge and experience from expert panels can be used to select and give relative weights to variables to be included in a score. Alternatively, 'statistical' predictive models can be developed by analysis of data for variables that have a strong association with the outcome of interest and their relative weights. Multivariable methods are then used to combine and select the different variables for predicting the outcome of interest and to select the appropriate variable[43]. Generally, these statistical scores outperform medical scores and today most scores are 'statistical' as relevant data is usually available for model development. Clinical knowledge should however be used to determine which variables should be analysed, with the intention of improving a model's performance in other groups of infants and making it more convincing to those considering using it in the future.

When choosing the variables to derive a score, researchers must balance the need for easy data collection against potentially improved predictive ability. Highly predictive models can be developed using many variables but these often take a long time to collect, thereby limiting their use in the clinical setting. The optimum score will predict the outcome adequately enough for the investigator's needs whilst being easy to collect.

As described by Richardson[43], variables must be 'predictive', meaning that they are biologically related to outcome and severity. They must also be 'available, measurable, frequent, accurately recorded and reliable'. Reliability is important and indicates that another observer will produce the same score for the same observation. Simplicity, objectivity and the use of minimum or maximum values all help to improve reliability.

Validation

Following derivation, the next stage is to test the model in the original cohort and then to validate it in another similar dataset. Testing a scoring system in another cohort of data is employed to measure external validity (i.e. the extent to which the results of a study are generalisable or transferable). Successful validation confirms that it predicts future events, with an adequate accuracy (calibration). To be clinically useful the predicted event rates from the score and those actually observed should closely match[82]. The Hosmer & Lemeshow Goodness-of-Fit test is usually used to assess this characteristic[83]. For this test the observations are categorised into groups according to their predicted risk. The number of predicted and observed outcomes within each of these groups are then compared, with a well-calibrated

score showing no evidence for statistically significant differences (i.e. p-value > 0.05). Table 3 provides an example of this process.

Table 3: Observed and Estimated Frequencies Within Each Decile of Risk of Mortality Using a Logistic Regression Model.

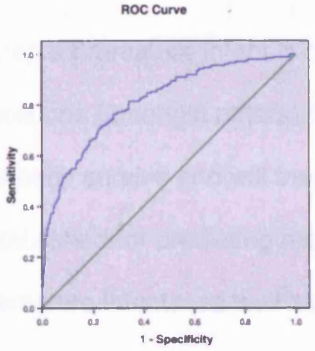
In this example, Chi-squared = 14.743, degrees of freedom = 7, p = 0.039.

Decile of Risk	1	2	3	4	5	6	7	8	9	10
Died										
Observed	0	3	0	0	1	5	1	3	14	0
Expected	0.1	1.7	1.2	0.7	0.9	1.3	1.9	4.9	14.3	0
Observed - Expected	-0.1	1.3	-1.2	-0.7	0.1	3.7	-0.9	-1.9	-0.3	0.0
Survived										
Observed	3	79	51	29	29	25	28	30	10	0
Expected	2.9	80.3	49.8	28.3	29.1	28.7	27.1	28.1	9.7	0
Observed - Expected	0.1	-1.3	1.2	0.7	-0.1	-3.7	0.9	1.9	0.3	0.00

Discrimination describes the ability of a score to differentiate between infants with different outcomes: most commonly whether infants die or not. Discrimination is assessed by the area under the Receiver Operating Characteristic (ROC) curve[84]. This figure is obtained by plotting the true positive rate against the false positive rate (or 1 – specificity) for the full range of values that can be taken as a cut-off. The area under this curve (abbreviated to Az) indicates the overall discriminatory ability of a scoring system. A perfectly discriminating test would have an area under the curve of 1.0 with no false positives nor any false negatives, whereas a score no better than chance alone has the value 0.5. As shown in Figure 2, the line at 45 degrees has an area under it of 0.5 indicating discrimination no better than chance. A value above 0.8 is usually taken to indicate that a score may be useful in practice.

Figure 2: A Receiver Operating Characteristic Curve.

(Az=0.816, 95% C.I. 0.780 – 0.852)



An older method of describing model discriminatory ability is the use of a classification table as presented in Table 4 below. This presents the ability of the model using a single cut-off point. Useful values are given, such as sensitivity, specificity, negative and positive predictive values. This is equivalent to presenting a single point on the ROC curve. The ROC curve is much better than this, as it allows a full range of cut-offs to be presented.

Table 4: Components of the Classification Table

	Predicted to survive	Predicted to die
Survivors	a	b
Non-Survivors	c	d

Sensitivity = true prediction of death / total deaths or d / (c+d).
Specificity = true prediction of survival / total survival or a /(a+b).
Predictive value of living = predicted survivors / total survivors or (a+c)/(a+b)
Predictive value of dying = predicted deaths / total deaths or (b+d)/(c+d)
Misclassification rate = total incorrect prediction/all predictions or (b+c)/(a+b+c+d).

1.6 PREDICTING NEURODEVELOPMENTAL OUTCOMES

When a premature infant is being cared for in neonatal intensive care, there are two questions (amongst others) that parents and clinicians would like the answers to: will the baby survive and will they be normal if they survive[85]. Methods are well established for predicting mortality and have demonstrated the improving survival of premature infants, as well as enabling the approaches listed in the previous section. Considerable concern exists that the increasing survival of extremely premature infants may be leading to larger numbers of infants surviving with disability [1,81]. This may be further exacerbated if survival data are scrutinised by external bodies or the public, as clinicians may be pressurised into carrying out a survival at all costs policy. In the United Kingdom the best interests of each individual infant are taken as paramount, with Judges recently confirming this by upholding the wishes of Doctors to treat infants conservatively, despite contrary parental wishes[86, 87].

In Trent and Yorkshire, mortality rates in high risk infants are monitored and compared each year with extra resources or service reviews being employed in situations where outcomes are worse than expected[88, 89]. Government and / or public scrutiny of these death rates might lead to the pursuit of survival at all costs. This might lead to the survival of some infants with poor quality of life. It would therefore be useful to monitor both death rates and later outcomes to obtain a fuller picture of unit performance. A predictive model would also allow us to determine if more babies are surviving with disability compared to the past or whether improvements in the quality of obstetric and neonatal care that have resulted in significant improvements in survival have also enhanced long term outcomes. It might also be possible to use a risk-correction model to identify processes or procedures that could lead to improvements in neurological outcome. This approach

has already been employed for mortality[45, 90] and involves large populations of infants and testing factors, such as staffing numbers or experience against risk corrected outcomes.

Three risk adjustment scores have already been assessed for use in predicting later neurodisability after neonatal intensive care. Two of these (CRIB and SNAP) were developed for predicting death, whilst the Nursery Neurobiologic Risk Score (NBRS) was developed for predicting disability and is described below on page 36.

CRIB Score and Neurological Morbidity

Four publications have examined the use of the CRIB score for predicting neurodevelopmental outcome [91, 92, 93, 94]. Table 5 summarises the results from these studies. Data on the outcome of 695 infants from the derivation cohort suggested that CRIB could predict a combined outcome of death or impairment[94]. However, in a further study containing infants from the original study, a close relation between CRIB at 12 hours and severe disability at 24 months of age was not demonstrated[93].

Two studies not containing infants from the original cohort showed that CRIB discriminated poorly in the role of predicting outcome at 12 months (Az = 0.70)[91], and 18 months (0.77)[92]. Lago also found that birth weight alone was similar (Az = 0.70), and gestational age alone was better (Az = 0.83) than CRIB[92]. Although these findings probably indicate that the CRIB score cannot reliably be used to predict later outcome, it is possible that the results reflect that neurodevelopmental testing before two years is unreliable and misclassifies some infants.

Fowlie[95] combined CRIB with cranial ultrasonography in 297 infants from the original cohort surviving beyond 72 hours. CRIB scoring was performed at 72 hours, with

ultrasound appearances from "around" 72 hours. 99 infants had missing CRIB, ultrasound, or follow up data. A CRIB score greater than 4 with a grade 3 or 4 intraventricular haemorrhage was predictive of severe disability, but there were only five infants in this group. In comparison with birth weight (Az = 0.70) and gestational age (Az = 0.74), CRIB and ultrasonography improved the model's discrimination (Az = 0.89). To implement this simple approach would require an alteration to current practice for collecting CRIB scores and, probably, ultrasound data. In addition, interpretations of cranial ultrasound findings have been shown to vary between clinicians[96].

Table 5: Ability of the Clinical Risk Index for Babies (CRIB) Score, With and Without Ultrasound (US) to Predict Neurodisability

	Age at assessment (months)	Number of infants assessed	Method of developmental assessment	Outcome	Predictive value (Az)
CRIB[91]	12	351	Griffith's test	Major impairment	0.703
CRIB[94]	18	695	Questionnaires from doctors, health visitors, and community nurses	Death or impairment	0.83
CRIB[92]	18	81	Amiel-Tison method and Bayley development scales	Major disability	0.77
CRIB[93]	24	398	Health visitor: standardised questionnaire	Severe disability	0.71
CRIB & cranial Ultrasound at 72 hours of age[95]	18	240	Health visitor completed questionnaire	Severe disability	0.89

SNAP and Neurological Morbidity

A retrospective case note review of 173 inborn infants from Minnesota examined the ability of the SNAP score to predict neurological outcome in premature infants born in 1993 and 1994 before 30 weeks gestation[97]. A score was collected for every day of each admission to produce a "cumulative SNAP score". This was then examined in relation to assessments at around 1 year of life and during the 3rd year of life. Although the authors did not use ROC curve analysis, they did show that the quartile of infants with the worst cumulative SNAP score had significantly lower motor development indices at 1 year, as well as lower psychomotor development indices at both assessments.

Nursery Neurobiologic Risk Score (NBRS)

One score, the NBRS was developed specifically for predicting neurological outcome in VLBW infants[98]. Brazy chose and weighted 13 factors, correlating these with outcome in 57 infants at 24 months of age from 1986 to 1988. A "revised NBRS" was developed from the seven factors accounting for almost all of the differences in outcome (length of ventilation, serum pH, seizures, intraventricular haemorrhage, periventricular leukomalacia, infection and hypoglycaemia). Scored at 14 days of age, taking five minutes per infant, it was highly repeatable, with all infants scoring over 5 having abnormal development at 24 months corrected age. Table 6 summarises the use of the NBRS in predicting neurodisability.

Using this score, Nunes[99] studied 77 infants at 12 months of age. Of those infants with a score of 8 or more, 80% developed a major handicap. Lefebvre[100] retrospectively collected the NBRS and outcome at 18 months in 121 infants, obtaining remarkably different results from Brazy[98]. Lefebvre's ROC curve value of

0.79 is similar to that of CRIB[100]. Contractor analysed outcomes at three years of age in 56 extremely premature infants, showing that a high NBRS at discharge was associated with four times the risk of an abnormal outcome. After modifying the score (to comprise acidosis, hypoxaemia, hypotension, intraventricular haemorrhage, infection, and hypoglycaemia), they also showed very good sensitivity and specificity[101].

Although it is a reasonable predictor of neurological outcome, the NBRS cannot be used for risk adjustment because of the delayed timing of data collection and the consequent effect of care. A new score with reliable accuracy using data from early in an infant's life is therefore needed to enable risk correction. This thesis documents attempts to produce such a model.

Table 6: Ability of the NBRS to Predict Neurodisability

Score	Age at Assessment	Score value	Sensitivity	Specificity	Positive Predictive value (PPV)	Negative Predictive value (NPV)
NBRS[98]	24 months	5 or more	52%	100%	100%	
NBRS[100]	18 months	*5 or more*	81%	54%	49%	84%
		8 or more	56%	87%	71%	78%
Modified NBRS[101]	3 years	'high' NBRS	100%	98%	92%	100%
			Any handicap	**Major handicap**		
NBRS[99]	12 months	<5	20%	5%		
		5 to 7	41%	23%		
		8 or more	95%	80%		

1.7 DEFINING OUTCOMES: ASSESSING AND CLASSIFYING DEVELOPMENTAL OUTCOME

Many life-ending decisions are taken due to concern about long-term outcomes: it is therefore possible that a unit which has high mortality may have very good long term outcomes. Similarly a unit that 'saves' many infants may have poor outcomes in infancy and childhood. In a new era of closely auditing and comparing outcomes, monitoring mortality rates may cause units to strive to save as many babies as possible. Measuring developmental outcomes in a standardised way to monitor the performance of many units would hopefully prevent this situation occurring.

Unfortunately there are many outcomes that can be assessed and an equally large number of ways to classify these. Because of different definitions and outcomes it has been difficult to compare outcomes from published studies and audit data from geographical populations. To be widely applicable, a scoring system for this population would predict many of the varied neurological outcome measures that exist. This would help to ensure the widespread usability of the model by as many clinicians as possible, who could apply it to existing data and not need to change assessment methods.

FUNCTIONAL CLASSIFICATION

Two systems have been developed to establish a universal method of measuring outcome in infancy according to the functional abilities of a child. These are the World Health Organization's 'International Classification of Impairments, Disabilities, and Handicaps' (ICIDH) and the Oxford Health Status Questionnaire (OHSQ).

WHO published the ICIDH in 1980 for classifying the consequences of disease in adults; the methodology has since been applied to VLBW children[02, 103]. Impairment was defined as any loss or abnormality of physiologic or anatomic structure; disability as any restriction or loss of ability (attributable to an impairment) in performing an activity in a manner and range considered normal for a human being. A handicap was defined as a disadvantage for a given individual, resulting from a disability or impairment, that limits or prevents the fulfilment of a role that is normal for that individual (depending on age, sex and social, and cultural factors). WeeFIM, a tool for assessing and characterising children's functional independence was developed from this classification and can be used in children aged 6 months to 7 years[104]. WeeFim analyses the parent's view of the child's need for assistance to complete 18 routine tasks of daily life (see Appendix 1). The level of independence is scored for each task in 7 grades from completely dependent to full independence. The tasks can be divided to give scores for three domains: self-care, mobility, and cognition. The Pediatric Evaluation of Disability Inventory (PEDI) is a similar tool[105]. The PEDI is composed of 73 self-care items, 59 mobility items, and 65 social function items: it is therefore much harder to administer.

The WHO classification was updated in 2001 as the International Classification of Functioning, Disability and Health (ICF)[106]. The main purpose of reclassification was to move to a more sociological model reflecting society's inability to adapt to help the individual with difficulties rather than the inability of the individual to cope with society. As it was developed in adults, it may not recognise the developmental nature of children's abilities and much work is needed to convert it into a version for use with children including recalibration and validation[107].

The OHSQ was developed by an expert working party and has been substantially used in the United Kingdom[108], [109]. Information collected at two years of age includes health and functional status, severity of functional loss within various domains as well as documentation of the underlying disease or impairment. Continued documentation until school age has also been recommended to identify lesser learning difficulties. This system can be used to classify children according to functional ability into severe, moderate, mild and no disability [see Appendix 2]. Focussing on the different grades of disability has been criticised, as it may fail to analyse the high incidence of milder problems[110], although these are harder to define and classify in infancy, becoming easier to detect with age.

CEREBRAL PALSY

Papers describing neonatal outcome following neonatal intensive care often compare the incidence of cerebral palsy. Cerebral palsy has been defined as a group of conditions that are 'non-progressive, but often changing, motor impairment syndromes secondary to lesions or anomalies of the brain arising in the early stages of its development'[111]. Cerebral palsy is seen in between 8-10% of infants in most VLBW cohorts. It is more common in extremely premature infants and in boys, often taking the form of spastic diplegia which is due to white matter damage in the area of the internal capsule[112]. As Nelson highlights, there are concerns about comparing rates of cerebral palsy as diagnosis can be difficult and cerebral palsy has been seen to progress or even resolve in the first few years of life[113]. This is also a 'medical model' classification which labels children and can lead to stigmatisation. Many children with cerebral palsy have a good functional outcome, so it is preferable to use the functional outcome measures that are described above. These are likely be of more interest to parents and surviving children[114].

SENSORY OUTCOMES

Sensory outcomes are also important, with survivors of prematurity at high risk of blindness and deafness. It has been recommended that these should form part of the classification of severe disability. The UK working party guideline suggested classifying as severely disabled those children with blindness or the ability to see light only or children with hearing impairment uncorrectable by aids. In a similar manner, speech and language ability assessed at two years of corrected age can be classified as severe disability if an infant is unable to comprehend a word or sign in a cued situation or is unable to produce more than five recognisable sounds[108].

QUALITY OF LIFE

Health Related Quality of Life (HRQOL) assessments (including quality-adjusted life-years) reflect what is important from the perspective of the child and family. They can be defined as the 'physical, psychological, and social domains of health, which can be influenced by an individual's experiences and perception'[115]. Infants are not able to give an appropriate opinion on these and evidence suggests parents or clinicians views are a poor proxy[116], [117,118]. They are therefore not useful in infancy although they have been used in research, particularly in older children and adults[119, 120,121,122]. Other factors such as maternal depression, chronic lung disease of prematurity, poverty, social background and cognitive development may all affect a child's interpretation of his or her quality of life[123]. Having to wait until a child reaches five years or more reduces the value of this information for neonatal care providers; 18 months to two years of age corrected for prematurity is usually used instead for this purpose, negating the use of personally reported HRQL in assessing neonatal care. Interestingly, objective measures of HRQL demonstrate poorer outcomes in high risk infants, but subjective measures of HRQL suggest that these are not necessarily perceived as such by the individual[124].

DEVELOPMENTAL OR INTELLIGENCE QUOTIENTS

Developmental or intelligence quotients are useful outcome measures. They summarise overall performance and ability allowing comparisons with reference standards or populations. They can, however, mask subtle and isolated areas of abnormal function[125]. As quotients are calculated by averaging functions in a number of different domains, it is important to look at the underlying function in each area. A correction for prematurity is usually used until the age of two years from the due date[126]. This is a simple process with the expected date of delivery being used in place of the actual date of birth to calculate the age.

The Bayley Scales of Infant Development (BSID)[127, 128, 129] are widely accepted as the best measures with the updated BSID-II being the current 'gold-standard' model for infancy. A third edition, BSID-III has also recently been published[129] The original BSID was published in 1969 and can be used for infants of 2 to 30 months of age[127]. It gives a score for a Mental Developmental Index (MDI) score and a Psychomotor Developmental Index (PDI) score. It was updated in 1993[128] for infants of 1 to 42 months of age and also gives a Behaviour Rating Scale. A MDI score of less than 70 indicates significant cognitive developmental delay. Other similar scores are the Stanford-Binet Intelligence Scale (Stanford-Binet IQ) and Wechsler Intelligence Scale for Children-Revised (WISC-R). An important issue is how one should classify infants who have such severe disability that they cannot be scored. Perhaps the best approach is to score these infants as very low (for example assigning a quotient of 40) and therefore avoiding potential bias by excluding them[130].

1.8 PROBLEMS WITH OUTCOME PREDICTION

Predicting neurodevelopmental outcomes is not without difficulties. As discussed above, problems in this area include which measurement or classification of outcome to use, inaccuracies of scores and balancing ease of use with improvements in accuracy that can be obtained from measuring more variables. Whilst prediction scores are subject to inaccuracy, outcome assessments can also be inaccurate, for example, the Denver-II developmental screening test, although widely used in the United States, suffers from modest sensitivity and poor specificity with a high false positive rate[131].

AGE

Age at testing is another difficult issue requiring a balance of accuracy and expediency. Tools for assessing developmental acquisition are less well validated and harder to perform in younger infants due to poor co-operation, whilst cognitive function is also difficult to measure due to lack of communication skills and understanding. As children age, classification of their functional ability becomes more stable with some infants even 'growing out' of cerebral palsy[113]. This must however be balanced with the need to determine outcome sooner to instigate potentially beneficial treatments, to inform neonatal care practices and provision, or to confirm the safety of randomised controlled trials.

Earlier assessments, before two years of age are better for informing the management and planning of neonatal care provision. Any time lag between the admission and obtaining later outcomes means neonatal care processes and practices may have already changed and the data may no longer be relevant. Early assessment is also mandatory when aiming to identify infants who might benefit

from an early intervention program[132] [133]. Later assessments suffer from a greater loss of patients due to migration and disengagement from the neonatal care team potentially biasing the results, as children receiving care and therapy may be less likely to participate in follow up[134]. Later assessments do however provide more detailed information and the severity of disability becomes easier to classify. Recent publications have also demonstrated that abnormal outcomes persist into adolescence and early adulthood[135, 136].

In view of these difficulties, infants are usually seen at various ages, initially as part of clinical care, with outcome being assessed at 18 to 24 months corrected age for the purposes of research. It must also be remembered that when comparing a child to age-related standards, a correction for prematurity must be used until the corrected age reaches two years of age[137].

COST CONSIDERATIONS

Neonatal trials typically involve many centres and therefore data collection can be costly, both financially and in terms of organisational resources. Trials are often therefore performed without formal later follow up assessments. This is a significant concern as a number of interventions have recently been shown to have harmful effects on the brain[138],[139]. In observational studies, large cohorts are needed to determine important statistically significant findings. For these reasons, good quality follow up assessment is costly. Dobrez estimated that in 2001, medical staff could administer a half hour screening test (Denver-II) for approximately $55 - $60 per child[140]. Parent questionnaires in comparison cost $10- $15, mainly for consultation time to explain the results.

Parent questionnaires have been used in both the clinical situation and in research studies and can also be used as a screening tool to identify infants who should be seen for a closer assessment[141],[142]. These are an efficient, cheap method of obtaining data that do not require highly trained staff[143]. Johnson and colleagues recently demonstrated good correlation of parental reports with Bayley scale assessments for determining a motor development index of less than 70[144]. Parents who are unable to fill out the questionnaire can be assisted by translators or health visitors. Data can also be collected over the telephone by clerical staff[145]. The major concern with questionnaires is parental bias. Parents may be keen to please and overrate their child or they might be overly critical of their child following a difficult experience of the perinatal period or in the presence of poor emotional attachment. The parents of infants with severely disabled children may find it distressing to fill out a questionnaire, so it is possible fewer of these infants may be reported using this method. Varying patterns of ethnic diversity, culture or parental language skills may also introduce bias: Janson having demonstrated differences between Norwegian and American children[146].

In 1999 in the UK a trial of two methods for collecting outcome data at two years of corrected age indicated that the cost of obtaining follow up data from Community Child Health Service clerks, using routine records was approximately £61 per patient[147]. Similar data obtained directly from the parents by questionnaire cost £37 per patient. These costs are minimal when one considers the overall cost of neonatal intensive care and of special educational provision[148]. The earlier identification of problems may increase assessment and treatment costs, but if early intervention can prevent or reduce disability, this may be offset by a reduction in the need for services when the child is older[149].

DATA COLLECTION

As mentioned above the source of the data has an important impact on costs. Data are however already collected by community paediatric services, general practitioners, health visitors and schools. Obtaining information from such routine records could avoid the need for expensive follow -up programmes[147]. Although attempts have been made to establish widespread follow-up programmes, these have largely floundered. This approach would involve extracting information from hospital notes, clinic letters, GP records or Community records. After extraction, these data must be categorised for them to be usable. The Oxford Health Status Questionnaire, whilst designed as a prospective tool, is based on functional outcomes that can be readily assessed in clinic, by health visitors or on reviewing notes. Further research is needed before this source can be utilised, in particular testing the reproducibility of individual and multiple data collectors.

Any assessment measure can potentially be biased by sociodemographic factors. For example, financially poorer parents are less likely to attend assessments and less likely to complete questionnaires. This should be taken into account by utilising health visitors or translators to ensure complete follow up and assistance with questionnaires. Even if measures are taken to ensure good rates of attendance, neurodevelopmental test results are themselves known to be affected by socioeconomic status and maternal education and this should be born in mind when comparing different populations[150, 151].

SELECTION OF INFANT GROUPS

Many infants are at high risk of developmental problems following neonatal intensive care. Many term infants, apparently at low risk, subsequently develop problems and a screening system should ideally identify all such infants early in life so that

intervention might be undertaken to improve later outcome. Due to cost and resource implications, rather than employing a universal approach, screening is focused on high risk groups. These are typically cohorts of infants such as those with a birthweight under 1kg, under 1.5kg, or defined by gestational age: the potential for the introduction of bias from birthweight criteria is discussed in chapter two of this thesis. Likewise the issue of study setting is important: many papers report data from tertiary units, however these are subject to selection bias as the infants must be sick enough to require referral but also well enough to survive the transfer. Variations in referral pattern and ambulance and transport services can therefore affect the study results and should be avoided by the use of geographical based cohorts[152].

ETHICAL ISSUES

Measuring and comparing outcomes following premature birth inevitably raises difficult ethical questions regarding the treatment and resuscitation of very high risk infants. Interested parties are divided, with some suggesting that the treatment of very high risk infants is inappropriate, Likewise, supporters of intensive care point to the survival of some of these children with good quality of life[153]. It has also been noted that there is poor correlation of medical functional measures and quality of life measures[154], leading to questions about the use of medical models in making life or death decisions in high-risk infants[110]. Interestingly two-thirds of children with cerebral palsy weigh more than 1500 grams at birth[155] and only 10% of cases of mental retardation have been estimated to be due to very low birthweight[156].

1.9 AIMS AND HYPOTHESES

Based on the evidence presented in the review above, this Thesis assesses two hypotheses:

1. ***Using data from the first day of life, a scoring system can be developed to accurately predict death or severe disability at two years of age in VLBW or preterm infants born before 32 weeks of gestational age.***

2. ***A screening test based on diagnostic criteria can be used near to the time of discharge to accurately identify individuals who will develop severe disability at two years of age.***

In order to test these hypotheses, the first step in this thesis was to look in detail at the cohorts available for study. Three different types of cohort were available for study: a birthweight defined geographical cohort, a gestational age defined geographical cohort and a group of children recruited to a randomised trial. This examination is described in chapter two, the aims of which were to identify the best cohort for study and by defining the inclusion criteria to explore methods for reducing the potential for bias from the nature of the cohorts.

CHAPTER TWO: DETERMINING THE BEST COHORT FOR THE STUDY

2.1 INTRODUCTION

Because of the concerns described in chapter one, the first stage of the work presented in this thesis was to examine the cohorts that were available for study. The characteristics of an ideal cohort include: a wide selection of patients; collection of data from a complete cohort or an adequate selection of cases and controls; well defined exposures and outcomes; and minimal loss of patients during the study and through into follow-up[157, 158]. Attempts to identify or measure potential confounders and their distribution between groups are also important[159, 160].

In order for a predictive model to be widely applicable to many other infants in the future, it should be developed from as broad a population as possible. For newborn infants, this implies using a geographical population which does not exclude any infants. Ideally this would include infants of all gestational ages, however this approach would require an extremely large dataset and the causes of poor neurodevelopmental outcome are different for very preterm and term infants[161]. With funding and time it would be possible to collect a dataset specifically for

developing a model for predicting disability in premature infants. In the interests of expediency and cost, for this thesis three existing datasets, available for secondary analysis were assessed. These were the East Anglian Very Low Birthweight Database, The Trent Neonatal Survey 'ABC study' (TNS) and the United Kingdom Trial of Oscillation (UKOS). Each of these cohorts contained a large number of babies, with outcome data being available from assessment at two years of age. The first step in this thesis was to identify any weaknesses of the datasets and to determine the best dataset to use for further modelling. A number of checklists exist for use in assessing the quality of cohort studies[157, 162, 163]. Four particularly important points were identified for assessing in detail the quality of the data in the cohorts: these were selection bias, information bias, confounding and external validity.

Selection bias involves the systematic inclusion into a study of a higher proportion of patients with an increased or reduced risk of disease than that of the general population[164, 162, 165]. An example of this type of bias would be the inclusion of a higher than expected proportion of babies born by vaginal delivery rather than caesarean section in a trial of an early neuroprotective strategy. This could arise if obtaining informed consent from mothers delivered by emergency caesarean section was hindered by the anaesthetic drugs. The cohort would therefore be biased (containing more vaginally delivered babies) and not necessarily representative of the general population in whom the results would be applied. Selection bias is a significant issue that can markedly affect the results of outcome studies following neonatal intensive care[166]. Another common mistake is to compare death rates between centres, whilst failing to include infants that are stillborn or die before admission to a neonatal unit[164]. This may be important when the groups are treated in a different way: one hospital may have a practice that involves resuscitating all

infants, regardless of condition at birth, whereas another hospital might treat only those that are seen as having a good chance of survival. The former hospital would admit more infants and have more deaths on the neonatal unit: they would therefore have proportionally poorer outcomes amongst it's group of 'admissions to neonatal intensive care'[166].

Information bias is related to how the data on individuals are collected[157]. For infants in trials, attempts are usually made to ensure that data collection is carried out in a similar way between individuals. Blinding is used to try and prevent the researchers inadvertently biasing data collection. This might happen if they (perhaps subconsciously) prefer one treatment to another: they might then look harder for complications or score the outcome lower in the group treated with the less favoured treatment. For cohorts information bias is an important consideration and relates to whether the same methods are used for all the patients. The need to collect data from patients or parents by telephone or postal questionnaire is an example of how such a bias could come about. Parents are unlikely to score their child in the same objective way that a blinded medical assessment would. This is further compounded by the fact that those individuals who are difficult to follow-up, appear to be more likely to suffer with significant disability[134].

Confounding has been described as the most likely cause of a spurious association in an epidemiological study[163]. In essence, confounding describes the situation where a factor is not itself a cause of the outcome of interest, but is associated with a range of other factors that do increase the risk of the outcome. The differing rates of cigarette smoking amongst women of different socioeconomic status is an example of this[163, 167]. Women of lower income are more likely to smoke but this is just one of a host of factors related to their social situation. If socioeconomic status

is not adequately measured or inadequate statistical corrections are applied, then researchers may conclude that smoking is associated with many complications. Presumed pregnancy complications from smoking include placenta previa, abruptio placentae, risk of miscarriage, intrauterine growth retardation, preterm birth, and reduced infant birth weight[168, 169, 170]. Smoking may be to blame for some of these outcomes, but social factors confound the studies exaggerating the effect of smoking[167].

External validity describes the usability of the study results in other populations: a process known as extrapolation[171]. In particular trials are at risk of limited external validity as, by necessity, patients are selected by inclusion criteria [172165]. Trial cohorts often contain a mix of patients who are not as unwell as the general population as those who are very unwell may not be approached or cannot give consent to take part in a trial. In the UKOS study this issue may have been compounded by the exclusion of very well infants who did not need ventilating despite being born before 29 weeks gestation. As Sackett puts it when considering treating a patient according to the results of a trial, the question can be articulated as: "is my patient so different from those in the trial that its results cannot help me make my treatment decision?"[173].

2.2 METHODS

Data for the analysis was obtained from three datasets. In order to determine whether selection bias was an issue, the patterns of gestational age, birthweight and sex were examined for each of the cohorts. To compare rates of intra-uterine growth restriction, growth centiles of the 1990 British standards were used to determine if infants were below the 10th Centile of Birthweight for gestational age[174]. These were also used to determine the proportion of infants that would have been excluded by the birthweight cut off of 1500g from an imaginary cohort. In order to compare like for like, the cohorts were compared using all infants, a 1.5kg upper weight limit and an upper gestational age limit of 28+6 weeks. In order to study in further detail the pattern of gestational age, a large population-based cohort of births from Canada[175] was used as a reference population, as no similar dataset was available in the UK. United Kingdom population data obtained from birth certificates is published regularly by the Office of National Statistics[176, 177], but unfortunately, gestational age is not collected by this system.

Statistical analyses were carried out using SPSS software. The Chi-squared test was used for comparing categorical variables, the students t-test for normally distributed numerical data and the Mann-Whitney U test for non-parametric numerical data.

Information bias, confounding and external validity are difficult to assess and deal with, but these issues are discussed in the investigation chapters three and four. Information bias was assessed by presenting the proportions of infants who were lost to follow up or had underlying risk factors and comparing them for any statistically significant differences. Confounding was dealt with by the use of logistic

regression analyses in the model development, whereby the effect of one variable on the outcome of interest is assessed whilst controlling for the effect of other important variables[157]. Issues of external validity are also discussed in the individual chapters.

United Kingdom Trial of Oscillation (UKOS)

The **UKOS trial**[178] recruited infants from 25 U.K. and international centres during the period of August 1998 to January 2001. Infants were eligible for the study if their gestational age was between 23 weeks and 28 weeks plus 6 days; if they were born in a participating centre; if they required endotracheal intubation from birth; and if they required ongoing intensive care. Infants were excluded if they had to be transferred to another hospital for intensive care shortly after birth or if they had a major congenital malformation. Outcomes were determined by medical examination, developmental tests or parental questionnaire and classified according to the Oxford Health Status Questionnaire (OHSQ)[108]. A predefined age window of 22 - 28 months of age (after correction for prematurity) was used to exclude infants who were not assessed inside this timeframe. This approach was taken, as the severity of disability is difficult to categorise before 24 months of age and because a later assessment can detect more problems[179]. Further information on this issue and the neurological outcomes has been published elsewhere[180].

Trent Neonatal Survey ABC Trial (TNS)

Data were obtained from a 1997 cluster randomised comparison of two methods of outcome data collection[147]. This study collected data prospectively on infants born to mothers resident in Nottingham or Leicestershire. The original trial also used data from infants in Wessex but for this thesis, data from the Trent babies alone was utilised as additional information was extracted from the Trent Neonatal Survey.

TNS has collected information about infants admitted to the neonatal units of the hospitals in the former Trent Region since 1993. Babies were eligible for inclusion in the study if they were born before 33 completed weeks of gestation (that is, up to and including 32 weeks and 6 days). Gestational age was recorded according to a hierarchical dating algorithm on the basis of last menstrual period, ultrasound scanning, and other clinical information. Outcome was obtained either by a postal questionnaire to parents or from routine community health records according to a randomised allocation[147]. This was then classified according to the OHSQ classification.

East Anglian Very Low Birthweight (EAVLBW)

The East Anglian Very Low Birthweight Database prospectively collects data on infants born under 1.5kg. Each of the eight neonatal intensive care units in the East Anglian region participates. One is located in a teaching hospital, three in large district general hospitals (>3500 deliveries per year) and four in small district general hospitals (<3500 deliveries per year). In total, the units deal with approximately 25,000 births each year including around 250 VLBW babies per year. A common dataset comprising routine antenatal, neonatal and follow-up data was collected on infants who were:

- liveborn with a birthweight of less than 1500 g
- born between 1st January 1993 and 31st December 2000
- born to mothers ordinarily resident in the eight district health authorities within the region.
- babies born outside the region if their mothers were ordinarily resident in the region.

Data were collected for transferred infants from each of the hospitals that admitted the infant. Twin or triplet infants with a birthweight above 1500 g were excluded even if the co-twin weighed less than 1500g. Clinicians prospectively collected data from the medical notes of the infants and their mothers. Developmental outcome at two years of age was determined by medical examination, developmental tests or a questionnaire completed by the general practitioner or health visitor and followed the OHSQ classification. Further details of the data collection methods and outcomes have been published elsewhere[181, 182].

2.3 RESULTS

A total of 797 infants were enrolled in the UKOS study, 311 in the TNS and 1938 in EAVLBW database (Table 7). Infants in the UKOS trial were considerably lighter and of earlier gestation than the other two cohorts, but this was also true of the EAVLBW cohort in comparison with TNS ($p= 9.88E*10^{-9}$ for gestational age and $p=4.95E*10^{-23}$ for birthweight). Maternal ages were similar in each of the cohorts. Although there were less male infants in the EAVLBW cohort, this difference was not statistically significant when compared to UKOS (p=0.15), or when compared to TNS (p=0.07). A high proportion (31.9%) of infants in the EAVLBW weighed less than the 10^{th} Centile at birth, and this was statistically significantly more than the UKOS cohort (17.3%, $p=9.6*10^{-15}$) and the TNS cohort (14.8%, $p=2.6*10^{-19}$): there was no statistical difference between the UKOS and TNS cohorts (p=0.31). The proportion of infants receiving antenatal steroids was considerably higher in the UKOS trial in comparison to the EAVLBW cohort (91.6% v 59.7%, $p<7.5*10^{-60}$). A significant difference was also seen in the proportion of infants receiving antenatal steroids in the TNS in comparison to the EAVLBW cohort (80.7% v 59.7%, $p< 3.6*10^{-6}$).

From the patterns of birthweight and gestation shown in Figures 3, 4, & 5, it is clear that the cohorts contain a very different mix of infants. The TNS pattern of gestational age is closest to that seen in the full cohort of all births from Canada, both showing increasing proportions as gestational age increases. The UKOS and EAVLBW cohorts had markedly different patterns providing strong evidence of selection bias. The UKOS trial had a smaller proportion of infants born at 28 completed weeks when looking at infants under 29 weeks alone (19.7%) compared to TNS (31.3%, p= 0.0086) and EAVLBW (27.2%, p= 0.00027). The histogram of gestational age of the whole cohort (Figure 3) appears to show a normal distribution for the EAVLBW gestations however this has been falsely produced by removing heavier infants. Figure 4 clarifies this by showing the normal distributions of birthweight that can be anticipated with gestational age criteria used to define the cohort. The right hand side of the EAVLBW dataset has been cut off at 1.5kg removing its normal distribution (see Figure 8). Figure 6 documents the estimated proportion of infants excluded from this dataset due to weighing more that 1.5kg at birth. This is a large proportion of infants above 29 weeks and also tends to exclude more males than females. By selecting only infants below 29 weeks, very few infants are excluded by the 1.5 kg cut-off and the normal distribution of infants by birthweight returns to the EAVLBW cohort (see Figure 7). The high proportion of infants weighing less than the 10th Centile at birth was also removed by applying this gestational age criterion as documented in Table 7c.

Table 7: Details of Infants Enrolled in the Three Cohorts

a. ALL INFANTS	UKOS	TNS	EAVLBW
Total infants	797	311	1938
Gestational age, median	26	30	29
Gestational age, inter-quartile range	25 – 27	28 - 32	27-31
Birthweight, median	840g	1335g	1125g
Birthweight, inter-quartile range	690 – 1000	1002 - 1660	886 - 1330
Maternal age, mean (SD)	28.8 (6.1)	28.4 (5.8)	27.9 (5.8)*
Male sex, % (n)	53.7 (428)	56.9 (177)	50.7 (982)
Small for Gestational Age: <10th Cent,% (n)	17.3 (138)	14.8 (46)	31.9 (618)
Antenatal Steroids Received % (n)	91.6 (727)	80.7 (251)	59.7 (1157)
Data on Antenatal Steroids unavailable % (n)	0.4 (3)	1.3 (4)	0.1 (2)
* n = 1424			

b. VLBW INFANTS	UKOS	TNS	EAVLBW
VLBW infants	796	190	1938
Gestational age, median	26	28	29
Gestational age, inter-quartile range	25 – 27	27-31	27-31
Birthweight, median	839g	1090	1125g
Birthweight, inter-quartile range	690 – 1000	900 - 1262.5	886 - 1330
Maternal age, mean (SD)	28.8 (6.1)	28.7 (5.5)	27.9 (5.8)*
Male sex, % (n)	53.7 (428)	52.1 (99)	50.7 (982)
Small for Gestational Age: <10th Cent,% (n)	17.3 (138)	24.2 (46)	31.9 (618)
Antenatal Steroids Received % (n)	91.2 (726)	82.6 (157)	59.7 (1157)
Data on Antenatal Steroids unavailable % (n)	0.4 (3)	1.0 (2)	0.1 (2)
* n = 1424			

c. Extremely Premature infants (<29 weeks)	UKOS	TNS	EAVLBW
Extremely Premature infants (<29 weeks)	797	96	896
Gestational age, median	26	27	27
Gestational age, inter-quartile range	25 – 27	26 - 28	25 - 28
Birthweight, median	840g	940g	894.5g
Birthweight, inter-quartile range	690 - 1000	781 - 1061	723 - 1070
Maternal age, mean (SD)	28.8 (6.1)	28.4 (5.1)	27.9 (5.9)*
Male sex, % (n)	53.7 (428)	57.3 (55)	55.6 (498)
Small for Gestational Age: <10th Cent,% (n)	17.3 (138)	6.3 (6)	14.3 (128)
Antenatal Steroids Received % (n)	91.6 (727)	80.2 (77)	59.6 (534)
Data on Antenatal Steroids unavailable % (n)	0.4 (3)	0	0
Proportion of Infants by Gestational Age			
22 weeks % (n)	0.0% (0)	2.1% (2)	1.2% (11)
23 weeks % (n)	5.0% (40)	6.3% (6)	5.7% (51)
24 weeks % (n)	12.9% (103)	5.2% (5)	10.3% (92)
25 weeks % (n)	17.7% (141)	10.4% (10)	13.2% (118)
26 weeks % (n)	19.6% (156)	14.6% (14)	17.6% (158)
27 weeks % (n)	25.1% (200)	30.2% (29)	24.8% (222)
28 weeks % (n)	19.7% (157)	31.3% (30)	27.2% (244)
* n = 632			

Figure 3: Histograms of Gestational Age by Cohort Source

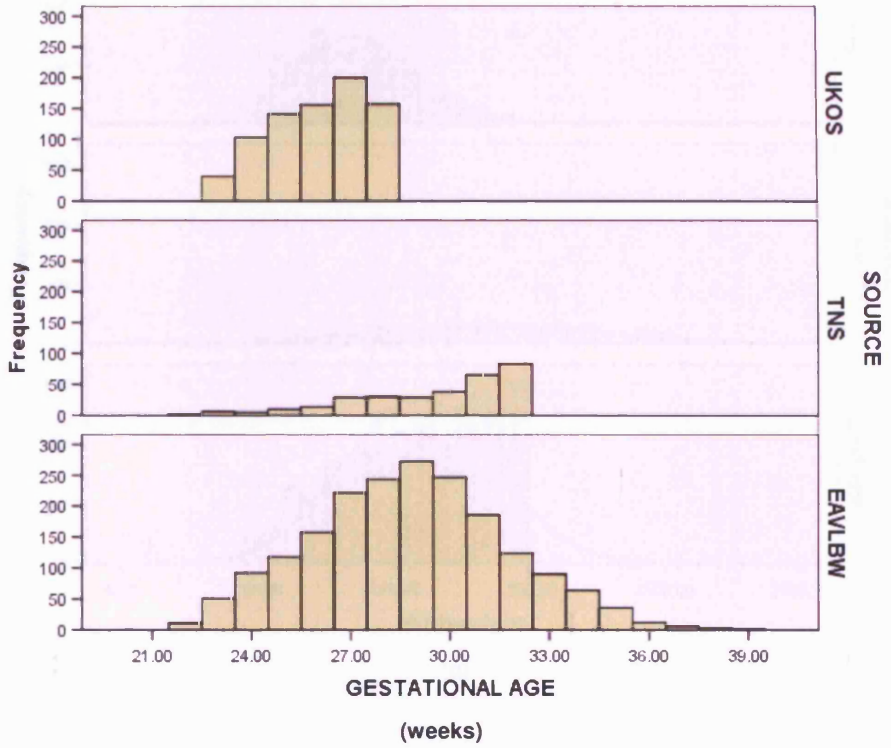


Figure 4: Histogram of Birthweight by Cohort Source

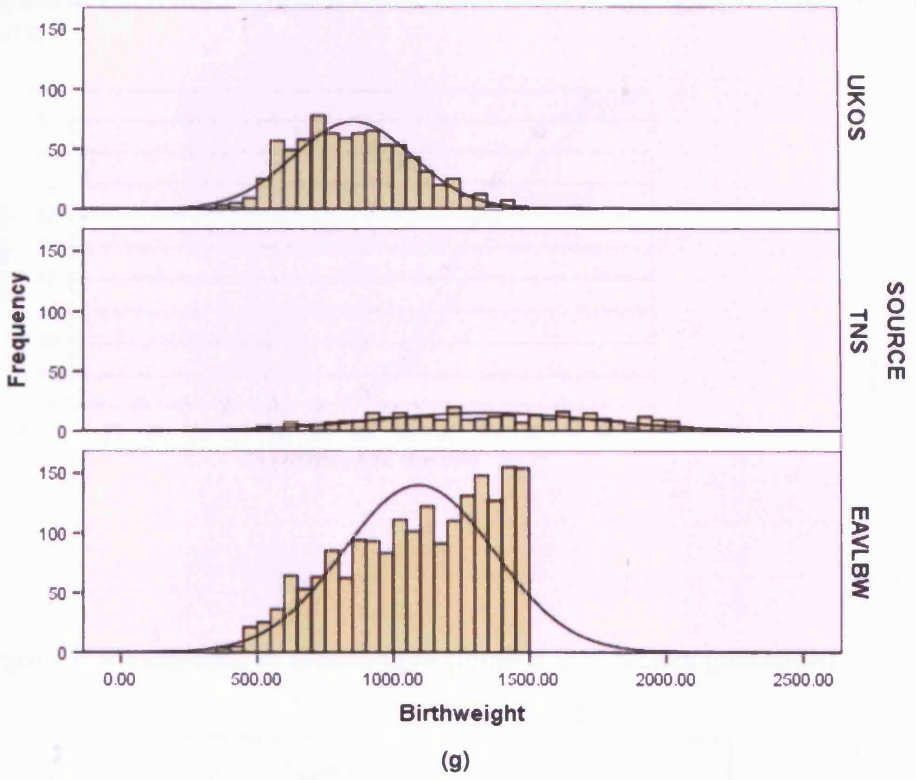


Figure 5: Proportions of Infants at Each Gestational Age By Study

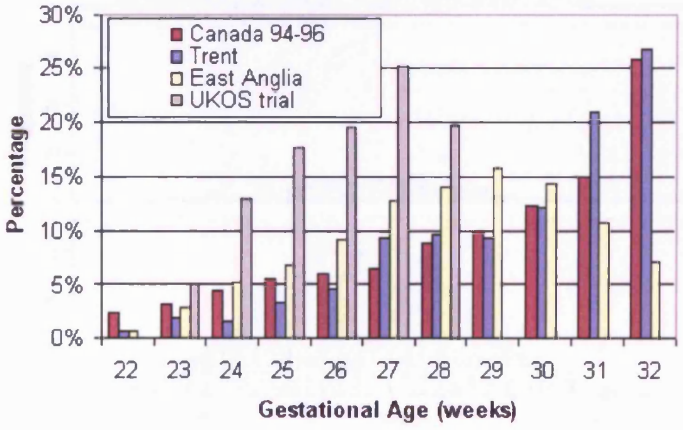


Figure 6: Estimated Proportions of Each Gestational Age Excluded by a 1.5kg Cut-off

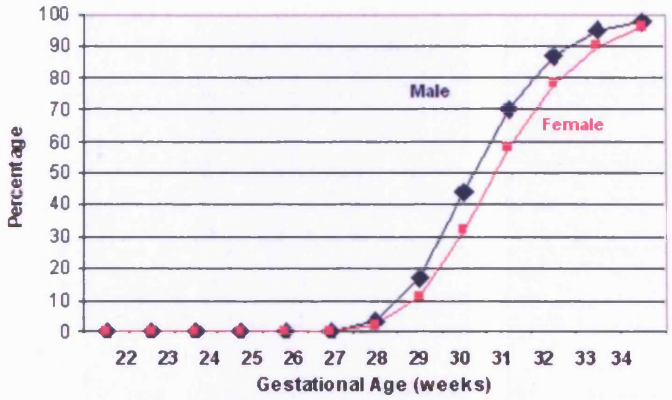


Figure 7: Histograms of Birthweight (infants < 29 weeks gestation)

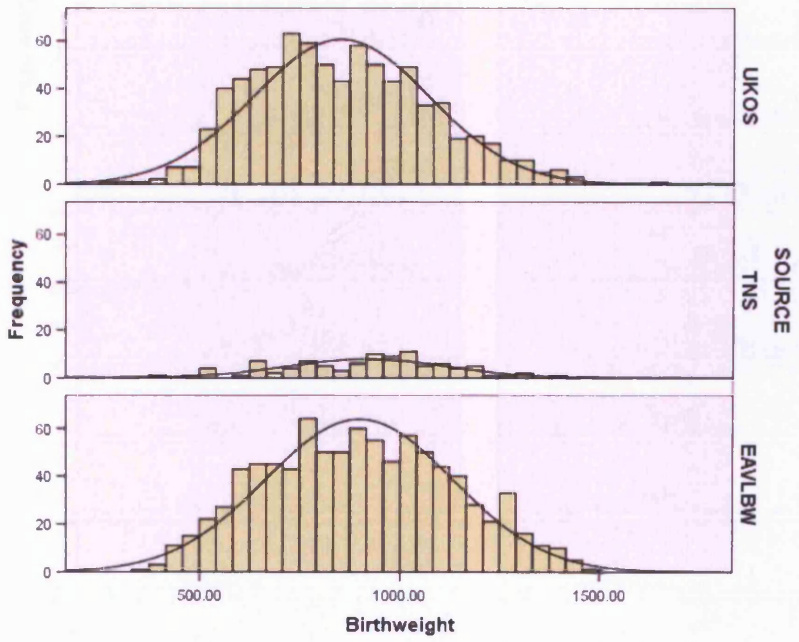
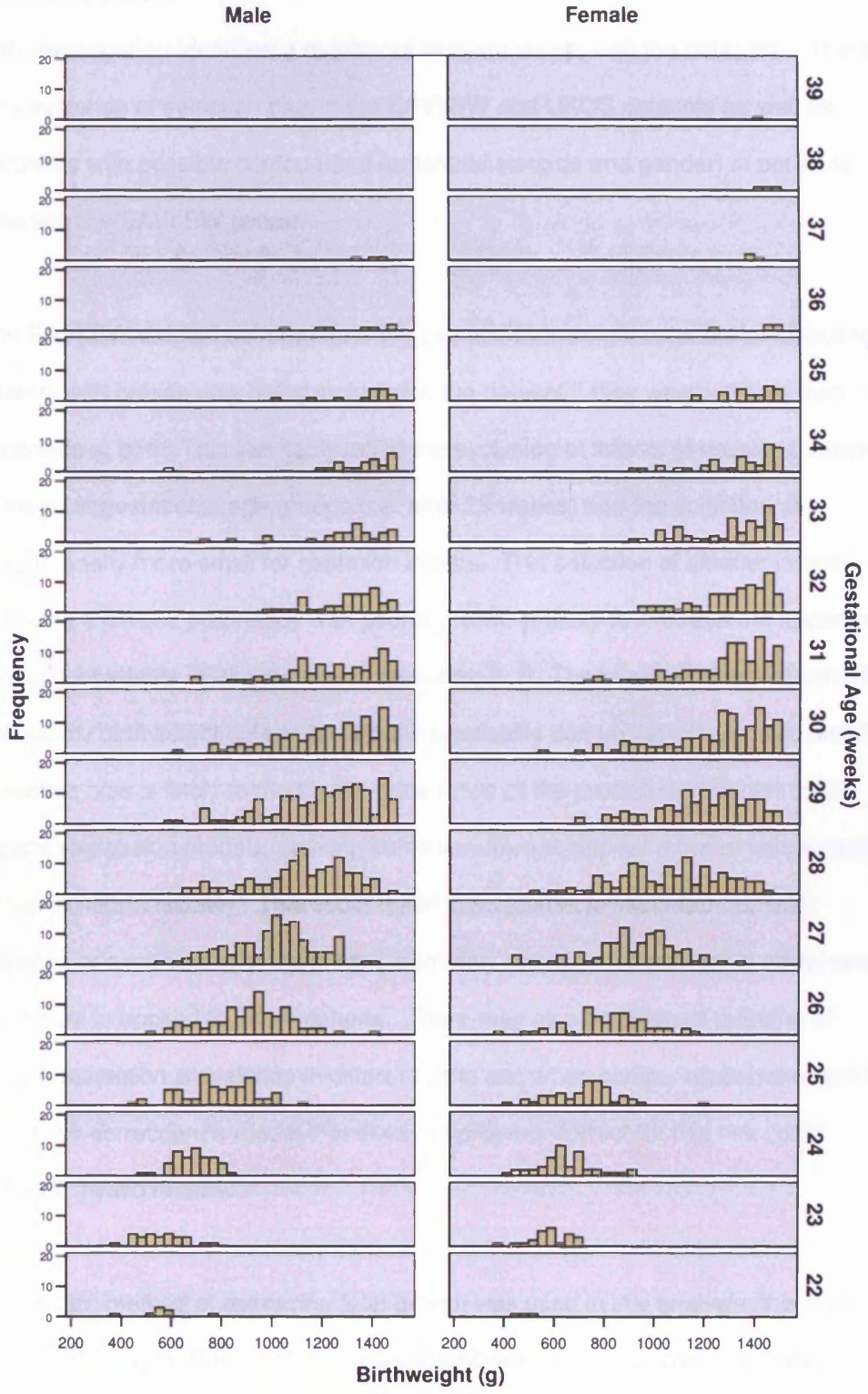


Figure 8: EAVLBW Cohort Histograms of Birthweight by Gestational Age



2.4 DISCUSSION

This investigation identified a number of relevant issues with the datasets. There was evidence of selection bias in the EAVLBW and UKOS datasets as well as problems with possible confounders (antenatal steroids and gender) in particular affecting the EAVLBW cohort.

The EAVLBW dataset suffered from the use of a birthweight criterion for selecting infants, with babies only being included in the dataset if they weighed less than 1.5 kilograms at birth. This inevitably led to the exclusion of infants of heavier birthweight in the later gestational age groups (i.e. after 29 weeks) and the selection of proportionally more small for gestation infants. This selection of smaller infants following a difficult pregnancy with poorer growth is likely to increase the apparent mortality of infants in 29 – 34 weeks gestation[74, 183]. These weaknesses of cohorts defined by birthweight criteria have been previously demonstrated and discussed[184]. Selection bias is likely to modify the odds ratios of the predictive variables in the logistic regression models, causing some variables to appear more or less strongly linked to neurodisability. This could result in a variable being inappropriately included or excluded from the model, ultimately reducing the predictive ability when the model is applied to future cohorts. There may also be different patterns of growth restriction prevalence in different units and when comparing between units using risk correction, a model that does not properly correct for this risk could produce invalid results.

A standard method of assessing fetal growth was used in this analysis: the 1990 British Birthweight Standards[174]. There are however some problems with this approach, although it is unlikely that these would have affected the results

significantly, as the same standards were used for each of the cohorts. The chief concern with growth restricted infants is the much higher risk of death that results from a failing intrauterine environment[185-187]. The use of birthweight centiles as a method for determining intrauterine growth restriction has been criticised, particularly as a significant number of healthy infants who are genetically small are picked up by this method when they have actually achieved their growth potential and not suffered intrauterine growth restriction[185-187]. The alternative is to use a computerised growth standard for each individual, taking into account gestational age, birthweight, gender, maternal height, maternal weight at first visit, ethnic group, and parity[188 189]. This was not possible in this study due to the data that was not collected in our cohorts, namely maternal weight at first visit, ethnic group and parity.

Selection bias was anticipated in the UKOS trial due to the requirement of needing ventilation for enrolment. Additional selection is likely to have occurred due to the nature of the units that took part in the trial. These were all tertiary neonatal units in the United Kingdom, with a large number (280 [40%[) of the trial infants being transferred to a study hospital before delivery[178]. This may have imparted selection bias, as the infant's mothers had to be fit for travel but needing tertiary level antenatal, postnatal or neonatal care[166]. Whilst selection bias does not affect the conclusions of the trial which was developed to determine the better ventilation strategy, it does have significant implications for the use of the data for secondary analysis. In particular there is a lack of infants at 28 weeks who did not need ventilating or were not considered sick enough to enrol in the trial. This could cause bias in the analysis of variables associated with the outcome at two years of age. Having less heterogeneity of risk in the cohort might weaken the calibration or discriminatory ability of any model developed from the cohort. Discrimination describes the ability of a predictive model to distinguish which patients will have

which outcome, by assessing its sensitivity and specifity across all possible values of the test whilst calibration evaluates how close the predicted chance of the outcome is to that seen in reality[190]. For instance, if a model for predicting outcome at two years of age was developed using trial infants with moderate and severe hypoxic ischemic encephalopathy (HIE) it would not be surprising if the model did not perform well at predicting the outcome of a big population cohort containing lots of infants with mild HIE.

In addition to gestation and birthweight, other confounders were examined. There was a trend towards fewer male infants in the EAVLBW dataset. As female babies weigh less than their male counterparts at the same gestation, this is likely to be due to more males being excluded than females. This could be important in developing a model to predict outcome at two years, as female infants generally have better neurodevelopmental outcome following prematurity in comparison to males[191-193]. Antenatal steroids have a well recognised benefit to later outcomes[194-197] and could therefore be an important confounder that could affect the cohorts. Notably, the UKOS trial dataset had a very good rate of antenatal steroid administration. This may reflect a further selection bias, perhaps due to being born in a tertiary unit and / or with a larger proportion of infants being delivered electively after transfer, giving time for antenatal steroids to be administered. Alternatively it could reflect the benefits of being in a trial and the consequent improvement in the quality of care that this is associated with[198, 199]. A much larger proportion of the EAVLBW cohort did not receive antenatal steroids: this may be due to the data coming from as far back as 1993, and rates have improved in more recent times[181]. Existing practice in the Trent region is measured on a yearly basis and indicates that between 71 and 100 percent of infants in each of the hospitals born before 33 weeks receive at least one dose of an antenatal steroid[89]. This issue raises the question of whether a model

developed from this data would be applicable to infants born at the present time, with higher prevailing usage of antenatal steroids.

As mentioned above in the introduction, in order for the model to be widely applicable to many other infants in the future, it should be developed from as broad a population as possible. Without including all infants in a new dataset, the best compromise appears to be to use a reasonably high gestational age cut-off such as the one used by the Trent Neonatal Survey: up to and including 32 weeks and 6 days. Developing a model based around the EAVLBW dataset would mean that it could only be applied to infants under 1.5 kilograms in the future and in view of the selection bias identified above, would also be best limited to infants below 29 weeks. The UKOS dataset has a similar problem with gestational age which is further compounded by the need for infants to be ventilated.

2.5 CONCLUSIONS

The best dataset for developing a model appears to be the TNS cohort followed by the EAVLBW using a gestational age cut-off of up to and including 28 weeks and 6 days. The UKOS trial dataset might also be used for analysis if due consideration of the problems of selection and limited future applicability is undertaken. Ideally one study cohort would be used for development and the others for testing and validating the model[190]. The option of combining the models is not feasible due to different collection of variables with some factors not being available to all the cohorts.

CHAPTER THREE: MODELS FOR PREDICTING NEURODEVELOPMENTAL OUTCOME FROM EARLY DATA

3.1 PREDICTING NEURODEVELOPMENTAL OUTCOME: TRENT NEONATAL SURVEY DATA

3.1.1 Introduction

As discussed in the previous chapter, for a number of reasons, the best dataset available for study into predicting outcome at two years of age was the Trent Neonatal Survey data from the ABC study. As a result the first attempt at derivation of a model was carried out using this dataset.

3.1.2 Methods

Data were obtained from the Trent Neonatal Survey for 311 infants born to mothers resident in Nottingham or Leicestershire in 1997. Babies were eligible for inclusion in the study if they were born at ≤32 completed weeks of gestation (i.e. up to and including 32 weeks and 6 days) and if their mothers lived in the health authorities covering Nottingham and Leicestershire. Gestational age was determined hierarchically according to the date of last menstrual period, ultrasound scanning, or other clinical information[147].

Outcome at two years of age was established from parental questionnaire or extracted from routinely available community data[147]. These data were collected during a pragmatic cluster randomised controlled trial to compare two approaches for collecting and collating outcome information on preterm infants who had required neonatal intensive care and had survived to two years of age. Although data were obtained from both Trent and Wessex in the original paper, for this study the information on infants in Trent alone was analysed. Excluding Wessex infants was necessary as data from the early course of infants who died was only available for

Trent infants from the Trent Neonatal Survey database. Outcome classification was performed according to the Oxford Health Status Questionnaire (OHSQ)[108].

Statistical analysis was performed using SPSS v11. The outcome of individual infants was classified as either survival without severe disability (S) or death or severe disability (D). These are illustrated in Figure 9. Death was defined as death by age 24 months of age after correction for prematurity, severe disability was defined as detailed in appendix 2. In terms of the predictive model only variables collectable in the first 12 hours of life were analysed to minimise the potential for the quality of treatment received during the infant's early course to influence the score[79]. Associations with outcome at two years of age corrected for prematurity were identified by the Chi-squared test, Fishers exact test or logistic regression analysis as appropriate. The variables were then assessed in combination using a stepwise multivariable logistic regression analysis. Variables were only included in the models if they had a significance level of less than 0.2 for an association with the combined outcome of death and severe disability. Predictive variables were added to the model in order of significance, a receiver operator characteristic (ROC) curve for the prediction of severe disability was then determined after each variable was added. The variable was then either left in the model or removed if the area under the ROC curve was less than before that variable was added. The recommendation of Wasson and colleagues[200] was used to limit the number of variables to a maximum of one variable for every ten cases of the outcome of interest.

Figure 9: Graphical Depiction of the Different Outcome Groups

← **SPECTRUM OF POSSIBLE OUTCOMES** →

Normal	Mildly Disabled	Moderately Disabled	Severely Disabled	Died
Group A			**Group B**	**Group C**

Survival without severe disability = Group A
Death or severe disability = Group B & C combined

As there was a combined outcome of either survival without severe disability or death or severe disability, to check that the model was predicting disability as well as death, the predictive ability of the probability score for the combined outcome from each individual was also assessed against the separated outcomes of death and severe disability. To test the ability of the model to discriminate between death and other outcomes, all cases were left in the cohort. The test therefore analysed the model's ability to discriminate a Group A outcome from either a Group B or C outcome. To test the ability of the model to discriminate between severe disability and milder disability and normality, deaths were first removed from the cohort. The test therefore analysed the ability of the model to discriminate Group B outcome from Group A outcome (see Figure 9).

3.1.3 Results

311 infants born in Leicestershire and Nottingham in 1997 were identified from the Trent Neonatal Survey and the ABC study. Outcomes and demographic details are documented in Table 8. Of the 311 infants, 27 died and 24 were assessed as being severely disabled at two years of age. The outcome was unknown at two years of age for 46 infants who were lost to follow up. Datapoints were missing for one or more variables for a further 17 infants. Information on the maximum base excess in the first 12 hours of life was missing for 63 infants. In order to avoid losing these infants from the analysis, missing values were replaced with the mean value of the other infants (-6.1 mEq/L). The association of predictive variables with death or severe disability at two years of age is documented in Table 9. The 'Draper grid' is a published tool that indicates the percentage chance of survival following birth according to gestational age, gender and birthweight in 250 gram blocks[74].

The final model for predicting death or severe disability at two years of age is documented in Table 10. The derivation of this model by stepwise forward regression and the predictive ability (discrimination) of the model at each step is detailed in Figure 10. The discrimination of the final model for predicting the combined outcome (death or severe disability), and the separated outcomes of death, and disability are shown with the ROC curves in Figure 11.

Table 8: Comparison of Included and Excluded Infants (Exclusion due to Unknown Outcome at two Years of Age)

	All infants		**Included**		**Excluded**		**Included v Excluded**
	n = 311	% or SD	n = 265	% or SD	n = 46	% or SD	
Gestational Age (completed weeks) Median	30		30		31		**p=0.004**
Birthweight (g) Median	1335		1280		1590		**p=0.002**
SGA Number	46	(14.8%)	43	(16.2%)	3	(6.5%)	p=0.09
Apgar Score at 1 Minute Mean	6.55	2.5	6.4	2.5	7.4	2.0	**p=0.003**
Apgar Score at 5 Minutes Mean	8.65	1.6	8.6	1.7	9.1	0.96	**p=0.002**
Sex							
Female	134	(43.1%)	112	(42.3%)	20	(43.5%)	p=0.95
Male	177	(56.9%)	151	(57.0%)	26	(56.5%)	
Congenital Malformation Present	18	(5.8%)	16	(6.0%)	2	(4.3%)	p=0.65
Multiple Pregnancy	65	(20.9%)	54	(20.4%)	11	(23.9%)	p=0.59
Surfactant Treatment	150	(48.2%)	133	(50.2%)	17	(37.0%)	P=0.089
Antenatal Steroid Treatment	251	(80.7%)	218	(82.3%)	33	(75.0%)	p=0.21
Maternal Age (Years)	28.4	5.8	28.4	5.7	28.0	6.1	p=0.59
Maximum Base Excess			-6.1	4.9	-5.5	2.7	p=0.32
Died	27	(8.7%)	27	(10.2%)			
Severe Disability at 2 years of age	24	(7.7%)	24	(9.0%)			
Survived without severe disability	214	(68.8%)	214	(80.8%)			
Outcome unknown at 2 years	46	(14.8%)	0		46	(100%)	

Table 9: Associations of Predictive Variables with Death or Severe Disability at two Years of Age

Variable	Coefficient	Standard Error	Wald	Number	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
							Lower	Upper
Probability of Survival (from Draper grid)	-0.047	0.008	36.7	265	0.0000000014	0.954	0.940	0.969
Gestational Age (weeks)	-0.367	0.066	31.0	265	0.000000026	0.692	0.608	0.788
Birthweight (g)	-0.002	0.000	21.3	265	0.0000038	0.998	0.997	0.999
Maximum base excess V2[1]	-0.156	0.038	17.0	265	0.000037	0.856	0.795	0.921
Maximum base excess	-0.157	0.038	16.8	215	0.000041	0.855	0.793	0.921
Apgar Score at 5 minutes	-0.329	0.089	13.5	260	0.00024	0.720	0.604	0.858
Apgar Score at 1 minute	-0.186	0.061	9.3	251	0.0023	0.831	0.737	0.936
Congenital malformation[2]	Chi-Square test used				0.010	Chi-Square test used		
Congenital malformation[2]	*1.288*	*0.531*	*5.9*	*265*	*0.015*	*3.624*	*1.281*	*10.253*
Antenatal steroids[2]	Chi-Square test used				0.023	Chi-Square test used		
Antenatal steroids[2]	-0.826	0.370	5.0	263	0.026	0.438	0.212	0.905
Gender	Chi-Square test used			265	0.22	Chi-Square test used		
Small for Gestational Age[3]	Chi-Square test used			265	0.34	Chi-Square test used		
Maternal Age (years)	-0.021	0.028	0.6	263	0.45	0.979	0.928	1.034
Number of Fetuses	0.039	0.280	0.0	265	0.89	1.040	0.600	1.802

Notes

[1] Maximum Base Excess Version 2: mean value of -6.1 substituted for missing values

[2] Chi square test used to assess association, *logistic regression* used to assess odds ratio

[3] Defined as birthweight below 10th centile for gestational age

Table 10: Model for Predicting Death or Severe Disability at Two Years of Age. Logistic Regression Coefficients and Odds Ratios

Variables	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Probability of Survival	-0.051	0.000000036	1.0	0.93	0.97
Maximum Base Excess V2	-0.15	0.00056	0.86	0.79	0.94
Apgar at 1 minute	0.13	0.14	1.14	0.96	1.35
Congenital Malformation	1.45	0.016	4.26	1.31	13.85
Constant	-0.78	0.47	0.46		

Figure 10: Process of Stepwise regression modeling for predicting death or severe disability at two years of age

STEP 1 Probability of Survival entered n = 265

Logistic Regression Model					
Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Probability of Survival*	-0.047	0.0000000014	0.95	0.94	0.97
Constant	2.33	0.00023	10.249		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.731	0.044	0.00000030	0.64	-	0.82

STEP 2 Gestational Age added n = 265

Logistic Regression Model					
Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Probability of Survival*	-0.060	0.00094	0.94	0.91	0.98
Gestational Age	0.12	0.43	1.13	0.83	1.54
Constant	-0.22	0.95	0.80		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.715	0.049	0.0000018	0.62	-	0.81

STEP 3 Gestational Age removed n = 265

See Result of Step 1

Continued overleaf

STEP 4 Birthweight added n = 265

Logistic Regression Model					
Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Probability of Survival*	-0.05	0.000024	0.95	0.92	0.97
Birthweight	0.00045	0.50	1.00	1.00	1.00
Constant	2.30	0.00036	9.98		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.743	0.042	0.000000072	0.66	-	0.83

STEP 5 Maximum Base Excess Version 2 added n = 265

Logistic Regression Model					
Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Probability of Survival*	-0.049	0.00018	0.95	0.93	0.98
Birthweight	0.00031	0.67	1.00	1.00	1.00
Max Base Excess V2	-0.13	0.00066	0.88	0.81	0.95
Constant	1.24	0.083	3.47		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.797	0.039	0.000000000045	0.72	-	0.87

Continued overleaf

STEP 6 Birthweight removed n = 265

Logistic Regression Model					
Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Probability of Survival*	-0.045	0.000000022	0.96	0.94	0.97
Max Base Excess V2	-0.13	0.00056	0.88	0.81	0.94
Constant	1.26	0.077	3.50998		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.798	0.039	0.000000000035	0.72	-	0.87

STEP 7 Apgar at 5 minutes added n = 251

Logistic Regression Model					
Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Probability of Survival*	-0.046	0.0000012	0.95	0.94	0.97
Max Base Excess V2	-0.14	0.0024	0.87	0.80	0.95
Apgar at 5 minutes	0.065	0.60	1.07	0.84	1.36
Constant	0.75	0.51	2.12		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.780	0.043	0.0000000040	0.70	-	0.86

STEP 8 Apgar at 5 minutes removed n = 265

See Result of Step 6

Continued overleaf

STEP 9 Apgar at 1 minute added **n = 260**

Logistic Regression Model					
Variable(s)	**Coefficient**	**Significance**	**Odds Ratio**	**95.0% C.I. for Odds Ratio**	
				Lower	**Upper**
Probability of Survival*	-0.049	0.000000078	0.95	0.93	0.97
Max Base Excess V2	-0.15	0.00048	0.86	0.80	0.94
Apgar at 1 minute	0.090	0.29	1.09	0.93	1.29
Constant	0.94	0.24	2.56		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	**Standard Error**	**Asymptotic Significance**	**Asymptotic 95% Confidence Interval**		
			Lower	-	**Upper**
0.802	0.038	0.00000000004 5	0.73	-	0.88

STEP 10 Congenital Malformation added **n = 260**

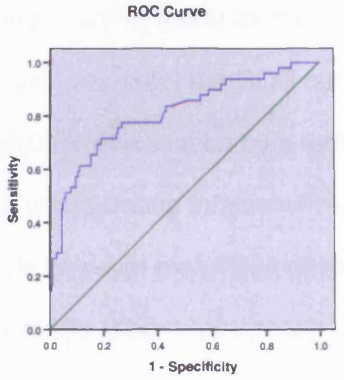
Logistic Regression Model					
Variable(s)	**Coefficient**	**Significance**	**Odds Ratio**	**95.0% C.I. for Odds Ratio**	
				Lower	**Upper**
Probability of Survival*	-0.051	0.000000036	1.0	0.93	0.97
Max Base Excess V2	-0.15	0.00056	0.86	0.79	0.94
Apgar at 1 minute	0.13	0.14	1.14	0.96	1.35
Congenital Malformation	1.45	0.016	4.26	1.31	13.85
Constant	-0.78	0.47	0.46		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	**Standard Error**	**Asymptotic Significance**	**Asymptotic 95% Confidence Interval**		
			Lower	-	**Upper**
0.808	0.038	0.00000000001 8	0.73	-	0.88

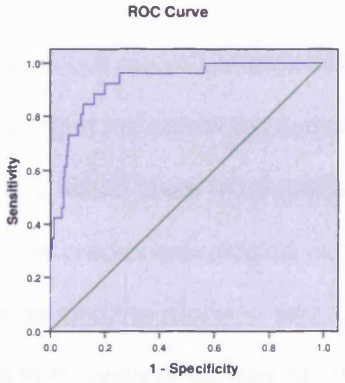
* Probability of survival obtained from 'Draper Grid'

Figure 11: ROC Curves of the Predictive Ability for the Different Outcomes (TNS)

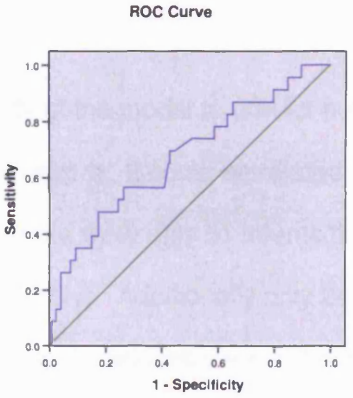
A. Severe disability or Death v Other **(Az = 0.808, 95% C.I. 0.735 – 0.882)**



B. Death v Other **(Az = 0.923, 95% C.I. 0.875 – 0.971)**



C. Severe disability v Other or No disability **(Az = 0.673, 95% C.I. 0.551 – 0.795)**



3.1.4 Discussion

This chapter describes the development of a model that at first glance appears to have good predictive ability for outcome at two years of age from data obtained at birth with an area under the ROC curve above 0.8. This cutoff is usually taken to indicate good discrimination by a test[201]. Further analysis shows that the model is very good at separating infants who will die from those who will survive to two years of age. It is however much less effective at determining which infants will develop severe disability at two years of age with an area under the ROC curve of only 0.673.

This raises serious questions about the use of a predictive tool for risk correction studies looking at the combined outcome of death and disability at two years of age. In particular it would seem most unwise to use this model for this purpose, as it does not accurately predict neurological outcome. Whilst it does predict death well, it offers little beyond the scope of well validated and often used scoring systems developed from much larger groups of infants. These scores for predicting death include CRIB, CRIB II and SNAP and are described in chapter one of this thesis.

The inability of the model to predict neurodevelopmental outcome may be due to a number of factors. It could be related to the relatively small number of infants in the cohort as there were only 51 infants that had the outcome of interest (death or severe disability). Additionally only 24 were assessed as being severely disabled. This significantly reduces the number of variables that can be included in the model, as it has been suggested that a minimum of ten such patients have the outcome of interest for each variable in the model[200]. Wasson's guidance on power in predictive modeling also indicates that the TNS dataset is not suitable for developing a model for predicting death or disability as separated outcomes. The results of a larger

study with more cases would allow the inclusion of more variables in the model, and hopefully improve predictive ability. In addition it would probably be seen as more reliable and plausible to clinicians and therefore more generalisable to other cohorts.

Another possible reason for the model failing to adequately predict neurodevelopment is that due to being a secondary analysis of a prior dataset, some variables were not available for analysis. The early development of the human brain can affected by many factors[161, 202] and it may be that in this study some very important variables were omitted. Although there is likely to be some overlap, it is not likely that the factors that influence development are the same as those that predict death. It might therefore be possible to repeat the study prospectively and collect additional data to improve the prognostic precision of the model. Examples of data that might improve the model include socioeconomic status and antenatal infection. Numerous studies have shown that social status is a vitally important factor related to the development of young children following prematurity [203, 204-207]. Recent studies have also suggested an important role for antenatal infection in the development of brain damage[208-210].

Another concern about the model comes from the number of infants on whom the outcome was unknown at two years of age. The original study compared collecting data from families completing a questionnaire and data from local community child health records. The families of 46 infants either did not participate in the trial or did not return data. Although this was less than 20% of the cohort, there were statistically significant differences between the included and missing groups for gestational age, birthweight, and Apgar scores at 1 and 5 minutes. The missing infants were more mature, weighed more at birth and had higher Apgar scores. This suggests that infants were lost to follow up due to there being less concern about

their outcome from clinicians or parents. Other reasons for being unable to complete the form are language or writing difficulties, potentially introducing bias from educational ability, race, language or culture. Tin and colleagues have also shown that in the United Kingdom, infants lost to follow up are more likely to suffer from severe disability[134]. A number of reasons have been suggested for this finding, including lower socioeconomic status as well as already being diagnosed and / or attending other health care services.

A further weakness of the study was that a significant proportion of the data from the community service was out of date, as clerks recorded details from the last clinic visit. For the original study this amounted to 28% of the community data coming from before 18 months of age. As the data was out of date, some younger children's milestones were inappropriately compared with the two year developmental stages. This may have resulted in an overestimation of the number of children with severe disability. The reported outcome of 10% of the original trial cohort was checked at home by a trained assessor who was not aware of the reported outcome. All parental questionnaire assignments agreed with the clinical assessment, whilst five out of 24 from the Community data were assessed as being severely disabled, when in fact this was due to the data being out of date. It is very likely that inaccurate outcome categorization would subsequently reduce the discrimination of the model.

The TNS dataset also has considerable strengths some of which are discussed in chapter two. Of the three study populations used in this thesis, it had the widest breadth of gestations. Double data entry was used to maximize the quality of the data and data from the neonatal admission was obtained in a reliable fashion by experienced, trained nurses who regularly visited the neonatal units. In view of these strengths, it was hoped that a model developed from the study could be used

in a wider population of future infants. The other two available cohorts are limited by lower gestational age and birthweight criteria and so models developed from them would be applicable to a smaller group of infants in the future.

There have been a number of other previous attempts to develop a model predictive of disability and these are detailed in Chapter one of this thesis. Previous studies concur with the finding of this report showing that it appears possible to predict a combined outcome of death and disability but not disability alone. It is likely that neurodevelopmental outcome after prematurity simply cannot be predicted using data available in the first few hours of life. It may be that the environment (and quality of care) that a premature infant experiences in the womb and in the first few weeks of life are so important for brain development that they outweigh inherent factors such as gestational age, birthweight, and Apgar scores. Recently Ambalavanan and colleagues have attempted to develop a model in a similar way to the studies described in this thesis, albeit using classification tree analysis instead of logistic regression[211]. They concluded that their model lacked the predictive accuracy needed for clinical use. MacKendrick, commenting on Ambalavanan's study in an editorial, suggested that it might not be possible to predict later brain development as it could be subject to chaos theory rather than the simple cause and effect paradigm of linear dynamics[212]. He describes chaos theory as having *'certain characteristic properties: they are exquisitely sensitive to the initial conditions imposed on the system; cause and effect are not proportional; and they appear to exhibit random, disorganized behavior that is not truly random but is governed by complex, nonlinear relationships'.*

Further study is needed to test the reasons for the models inability to predict neurodisability accurately before precise conclusions can be drawn about the use of

predictive models for disability following premature birth. The following two sections of this thesis look at predicting the combined outcome in two other cohorts. As there may be a significant difference between the aetiological factors for death and disability, chapter four documents an attempt to develop the model after first removing the infants who died.

3.1.5 Comments and implications

This investigation suggested that developing a model may be possible but in a larger dataset. This would allow the use of more predictive variables and alleviate concerns arising from the size of the cohort regarding the reliability of the results. As mentioned in chapter two, two further cohorts were available for further study: these were the EAVLBW dataset and the UKOS trial data. After applying a gestational age cut-off of up to and including 28 weeks and 6 days, the EAVLBW offered fewer issues regarding selection than the UKOS trial.

Alternatively, and perhaps the best option, would be to develop a new cohort and collect other data that could potentially predict developmental outcome. Due to the constraints of both time and cost this was not an option for this thesis but remains a suggestion for future work.

3.2 PREDICTING NEURODEVELOPMENTAL OUTCOME: EAST ANGLIAN VERY LOW BIRTHWEIGHT DATA

3.2.1 Introduction

As discussed in the previous section of this chapter, although the best dataset available for study into predicting outcome at two years of age appeared to be the Trent Neonatal Survey data from the ABC study, further analysis is needed to confirm these findings. This section of the chapter therefore details the second attempt at deriving a model using the EAVLBW dataset.

As mentioned in Chapter two the exclusion of infants of heavier birthweight amongst the more mature infants introduces selection bias in the EAVLBW dataset with a high proportion of the more mature infants being affected by intrauterine growth restriction. In order to avoid this, a gestational age cut-off was applied to the dataset: infants born after 28 weeks and 6 days being excluded from the analysis in this chapter.

3.2.2 Methods

Data were obtained from the East Anglian Very Low Birthweight Database infants born between between 1st January 1993 and 31st December 2000. Babies were eligible for inclusion in this study if they were born before 29 completed weeks of gestation (i.e. up to and including 28 weeks and 6 days). Their mothers were ordinarily resident in one of the eight district health authorities within the region. Data were collected for babies born inside or outside the region if their mothers were ordinarily resident in the region. For infants involved in transfers, information was collected from every involved hospital.

Gestational age was determined according to the date of last menstrual period, ultrasound scanning, or other clinical information. Neurodevelopmental follow-up at 24 months of age was performed by named local Paediatricians, experienced in developmental follow-up, using the Oxford Health Status Questionnaire to record findings[108]. If children were not able to attend this appointment, the child's Health Visitor or General Practitioner was asked to complete the forms based on routine surveillance information. For those who had moved or failed to respond, as much information as possible was obtained by letter either from a health professional or directly from a parent.

Statistical analysis was performed using SPSS v11. The outcome of individual infants was classified as no disability, mild disability or moderate disability: group A, severe disability: group B or death: Group C (see Figure 12). Only variables collectable in the first 12 hours of life were analysed to minimise the potential for the quality of treatment received during the infant's early course to influence the score[79]. Associations with the combined outcome of death or severe disability at two years of age corrected for prematurity were identified by the Chi-squared test, Fishers exact test or logistic regression analysis as appropriate. The variables were then assessed in combination using a stepwise multivariable logistic regression analysis.

Variables were only included in the models if they had a significance level of less than 0.2 for an association with the combined outcome of death or severe disability (Group B&C). Predictive variables were added to the model in order of significance, a receiver operator characteristic (ROC) curve for the prediction of severe disability was then determined after each variable was added. The variable was then either left in the model or removed if the area under the ROC curve was less than before that variable was added. The recommendation of Wasson and colleagues[200] was

used to limit the number of variables to a maximum of one variable for every 10 cases of the outcome of interest. Where variable data points were missing for just one or two infants, analysis was performed with the remaining infants. Where a significant number of numerical data points were missing, in order to avoid losing these infants from the analysis, missing values were replaced with the mean value of the remaining infants. This is one simple way of dealing with this potential loss of data in logistic regression models[213]. This substitution weakens rather than strengthens the statistical association between the variables and the outcome.

As there was a combined outcome, to check that the model was predicting disability as well death, the predictive ability of the probability score for the combined outcome from each individual was also assessed against the separated outcomes of death and severe disability. To test the ability of the model to discriminate between death and other outcomes all cases were left in the cohort. The test therefore analysed the model's ability to discriminate a Group A outcome from a Group B or C outcome (see Figure 12). Furthermore to test the ability of the model to discriminate between severe disability and milder disability and normality, deaths were first removed from the cohort. The test therefore analysed the ability of the model to discriminate Group B from Group A outcomes.

Figure 12: Graphical Depiction of the Different Outcome Groups

← **SPECTRUM OF POSSIBLE OUTCOMES** →

Normal	Mildly Disabled	Moderately Disabled	Severely disabled	Died
Group A			**Group B**	**Group C**

Survival without severe disability = Group A
Death or severe disability = Group B & C combined

3.2.3 Results

896 infants meeting the criteria and born in East Anglia between 1st January 1993 and 31st December 2000 were identified from the East Anglia Database[181]. Outcomes and demographic details are documented in Table 11. Of the 896 infants, 323 died and 81 were assessed as being severely disabled at two years of age. The outcome was unknown at two years of age for 12 infants who were lost to follow up and their data was excluded from the analysis. The association of predictive variables with death or severe disability at two years of age is documented in Table 12. As in Chapter 3.1, the Draper Grid prediction of survival was used[74].

A full set of datapoints for the variables included in the final model were only available for 139 infants. For the minimum appropriate percentage of inspired oxygen in the first 12 hours missing data amounted to 459 infants for whom the mean value of 39.94% was substituted. Similarly, the maximum appropriate percentage of inspired oxygen in the first 12 hours involved 455 infants and a mean value of 63.68%. Corresponding values for the maximum base deficit in first 12 hours, were 443 infants and - 5.38mEq/L, for the CRIB Score, 410 infants and 6.16 points, for head circumference at birth, 376 infants and 25.07cm, for the Apgar score at 1 minute of age, 81 infants and 4.96 points, and for the Apgar Score at 5 minutes, 77 and 7.54 points.

The final model for predicting death or severe disability at two years of age is documented in Table 13. The derivation of this model by stepwise forward regression and the predictive ability (discrimination) of the model at each step is detailed in Figure 13. The discrimination of the final model for predicting the combined outcome (death or severe disability), and the separated outcomes of death, and disability are shown with the ROC curves in Figure 14.

Table 11: Comparison of Included and Excluded Infants (Exclusion Due to Unknown Outcome at Two Years of Age)

	All infants		Included		Excluded		Included v Excluded
	n = 896	(%) or SD	n = 884	(%) or SD	n = 12	% or SD	
Gestational Age (completed weeks) Median	27.0	1.6	27.0	1.6	27.5	0.97	p=0.022
Birthweight (g) Median	894.5	233.4	893	232.9	1088.5	217.7	**p=0.018**
SGA	128	(14.3%)	127	(14.4%)	1	(8.3%)	p=0.55
Mean Apgar Score at 1 Minute	4.85	2.4	4.83	2.4	6.09	2.3	p=0.089
Mean Apgar Score at 5 Minutes	7.3	2.4	7.3	2.4	8.27	2.0	p=0.18
Sex							
Female	398	(44.4%)	391	(44.2%)	7	(58.3%)	p=0.33
Male	498	(55.6%)	493	(55.8%)	5	(41.7%)	
Congenital Malformation Present	44	(7.7%)	43	(7.6%)	1	(11.1%)	p=0.70
Multiple Pregnancy	237	(26.5%)	233	(26.4%)	4	(33.3%)	p=0.59
Surfactant Treatment	673	(75.2%)	665	(75.3%)	8	(66.7%)	p=0.49
Antenatal Steroid Treatment	534	(59.6%)	524	(59.3%)	10	(83.3%)	p=0.092
Maternal Age (Years)	27.9	5.9	27.9	5.9	29.5	6.7	p=0.50
Maximum Base Excess	-5.38	9.2	-5.35	9.3	-7.5	2.8	p=0.51
Outcome							
Died	323	(36.0%)	323	(36.5%)	-		
Severe Disability at 2 years of age	81	(9.0%)	81	(9.2%)	-		
Survived without severe disability	480	(53.6%)	480	(54.2%)	-		
Outcome unknown at 2 years	12	(1.3%)	-	-	12		

Table 12: Associations of predictive variables with death or severe disability at two years

Variable	Coefficient	Standard Error	Wald	Number	Significance	Odds Ratio	95.0% C.I. for Odds Ratio Lower	Upper
Probability of Survival	-0.04	0.00	143.3	884	5.06E-33	0.96	0.95	0.97
Gestational Age (weeks)	-0.58	0.05	126.8	884	2.10E-29	0.56	0.51	0.62
Birthweight (g)	0.00	0.00	83.3	884	7.01E-20	1.00	1.00	1.00
Apgar Score at 5 minutes	-0.34	0.04	82.8	807	8.97E-20	0.71	0.66	0.77
CRIB Score	0.22	0.02	81.5	474	1.73E-19	1.25	1.19	1.31
Apgar Score at 5 minutes[1]	-0.34	0.04	80.9	884	2.32E-19	0.71	0.66	0.77
CRIB Score[1]	0.22	0.02	79.7	884	4.29E-19	1.24	1.19	1.31
Apgar Score at 1 minute	-0.27	0.03	70.3	803	5.01E-17	0.76	0.72	0.81
Apgar Score at 1 minute[1]	-0.27	0.03	69.4	884	7.96E-17	0.76	0.72	0.81
Maximum Percentage O2 in first 12 hours	0.03	0.00	52.1	429	5.24E-13	1.03	1.02	1.04
Maximum Percentage O2 in first 12 hours[1]	0.03	0.00	51.5	884	7.02E-13	1.03	1.02	1.04
Minimum FiO2 in first 12 hours	0.04	0.01	49.4	425	2.10E-12	1.04	1.03	1.05
Minimum FiO2 in first 12 hours[1]	0.04	0.01	49.0	884	2.58E-12	1.04	1.03	1.05
Antenatal Steroids[2]				884	*4.13E-10*			
Antenatal Steroids[2]	-0.87	0.14	38.4	884	5.80E-10	0.42	0.32	0.55
Neonatal Transfer[2]				884	2.70E-06			
Neonatal Transfer[2]	-0.71	0.15	21.6	884	3.28E-06	0.49	0.36	0.66
Caesarean section[2]				882	*3.20E-06*			
Caesarean section[2]	0.63	0.14	21.5	882	3.53E-06	1.89	1.44	2.47
Gender[2]				884	*9.60E-05*			
Gender[2]	-0.54	0.14	15.1	884	0.00010	0.59	0.45	0.77
Head Circumference	-0.16	0.04	13.0	508	0.00031	0.86	0.79	0.93
Head Circumference[1]	-0.14	0.04	12.0	884	0.00053	0.87	0.80	0.94
Spontaneous Vaginal Delivery[2]				882	*0.0010*			
Spontaneous Vaginal Delivery[2]	0.45	0.14	10.8	882	0.0010	1.57	1.20	2.06
Preeclampsia[2]				882	0.0011			
Preeclampsia[2]	-0.65	0.20	10.3	882	0.0013	0.52	0.35	0.78
Other vaginal delivery[2]				883	0.0022			
Other vaginal delivery[2]	-1.71	0.63	7.5	883	0.0063	0.18	0.05	0.62
Breech[2]				841	*0.0092*			
Breech[2]	0.52	0.20	6.7	841	0.0098	1.68	1.13	2.50
Congenital Abnormality[2]				564	*0.014*			
Congenital Abnormality[2]	0.78	0.32	5.8	564	0.016	2.18	1.15	4.11
Year of Birth	-0.07	0.03	5.6	884	0.018	0.93	0.88	0.99
*In utero transfer**				884	0.038			
In utero transfer*	-0.36	0.17	4.3	884	0.038	0.70	0.50	0.98
*Multiple infant**				884	0.038			
Multiple infant*	0.32	0.15	4.3	884	0.039	1.37	1.02	1.85
*Any Transfer**				884	0.16			
Any Transfer*	0.20	0.14	1.9	884	0.16	1.22	0.92	1.61
Maximum Base Deficit in first 12 hours	0.01	0.01	1.8	441	0.18	1.01	0.99	1.04

Continued overleaf

Maximum Base Deficit in first 12 hours[1]	0.01	0.01	1.8	884	0.18	1.01	0.99	1.04
Small for Gestational Age < 10th Centile[2,3]				884	*0.18*			
Small for Gestational Age < 10th Centile[2,3]	-0.26	0.19	1.8	884	0.18	0.77	0.53	1.13
Number of infants in Pregnancy	0.17	0.13	1.7	884	0.19	1.19	0.92	1.53
PROM 7 days[2]				881	0.26			
Birth order	0.18	0.18	1.0	884	0.32	1.19	0.84	1.69
Diabetes[2]				883	0.55			
Intubation[2]				883	0.55			
Maternal Hypertension[2]				882	0.63			
Antepartum Haemorrhage[2]				883	0.70			
Maternal Age	0.00	0.01	0.0	626	0.88	1.00	0.97	1.02
Eclampsia[2]				883	0.94			

Notes

[1] Mean value substituted for missing values

[2] Chi square test used to assess association, *logistic regression* used to assess odds ratio if $p < 0.2$

[3] Defined as birthweight below 10th centile for gestational age

Table 13: Two Models for predicting death or severe disability at two years of age. Logistic regression coefficients and odds ratios

A. Model using all factors

Variables	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Probability of Survival	-0.029	9.93E-11	0.97	0.96	0.98
Apgar Score at 5 minutes with mean (7.54) substituted for missing values	-0.225	2.40E-07	0.80	0.73	0.87
CRIB Score with mean (6.16) substituted for missing values	0.095	0.0050	1.10	1.03	1.17
Maximum Percentage O2 in first 12 hours with mean (63.68) substituted for missing values	0.007	0.26	1.0069	0.99	1.02
Minimum FiO2 in first 12 hours with mean (39.94) substituted for missing values	0.017	0.034	1.017	1.0012	1.03
Antenatal Steroids	-0.58	0.00067	0.56	0.40	0.78
Neonatal Transfer	-0.68	0.00063	0.51	0.34	0.75
Gender	-0.46	0.0075	0.63	0.45	0.89
Head Circumference with mean (25.07) substituted for missing values	0.079	0.12	1.08	0.98	1.20
Spontaneous Vaginal Delivery	0.22	0.20	1.24	0.89	1.74
Other vaginal delivery	-1.95	0.0067	0.14	0.035	0.58
In utero transfer	0.31	0.17	1.37	0.87	2.15
Multiple infant	1.13	0.059	3.11	0.96	10.08
Maximum Base Deficit in first 12 hours with mean value (5.38) substituted for missing values	-0.013	0.33	0.99	0.96	1.01
Number of infants in Pregnancy	-0.76	0.14	0.47	0.17	1.27
Constant	1.46	0.32	4.31		

B. Parsimonious Model after removal of some factors

Variables	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Probability of Survival	-0.027	0.0039	0.97	0.97	0.98
Apgar Score at 5 minutes with mean (7.54) substituted for missing values	-0.23	0.042	0.80	0.73	0.86
CRIB Score with mean (6.16) substituted for missing values	0.14	0.028	1.15	1.09	1.21
Antenatal Steroids	-0.59	0.16	0.55	0.40	0.76
Gender	-0.57	0.16	0.57	0.41	0.78
Constant	3.42	0.49	30.71		

Figure 13 Process of Stepwise regression modelling for predicting death or severe disability at 2 years of age

STEP 1 Probability of Survival[a] n = 884

Logistic Regression Model					
Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Probability of Survival[a]	-0.04	5.1E-33	0.96	0.95	0.97
Constant	2.30	3.8E-25	10.01		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.736	0.017	9.2E-34	0.70	-	0.77

STEP 2 Gestational age added n = 884

Logistic Regression Model					
Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Probability of Survival[a]	-0.043	5.4E-06	0.96	0.94	0.98
Gestational Age	0.035	0.81	1.04	0.78	1.37
Constant	1.52	0.64	4.59		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.736	0.017	8.4E-34	0.70	-	0.77

STEP 3 Gestational age removed n = 884

See Result of Step 1

STEP 4 Birthweight added n = 884

Logistic Regression Model					
Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Probability of Survival[a]	-0.05	1.9E-18	0.95	0.94	0.96
Birthweight	0.00	0.020	1.001	1.0002	1.002
Constant	1.77	2.6E-08	5.86		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.743	0.017	1.6E-35	0.71	-	0.78

Continued overleaf

STEP 5 Apgar Score at 5 minutes[b] added — n = 884

Logistic Regression Model					
Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Probability of Survival[a]	-0.047	2.5E-14	0.95	0.94	0.97
Birthweight	0.0013	0.024	1.0013	1.0002	1.003
Apgar Score at 5 mins[b]	-0.243	2.2E-09	0.78	0.72	0.85
Constant	3.250	1.3E-14	25.80		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.769	0.016	3.3E-43	0.74	-	0.80

STEP 6 CRIB Score[b] added — n = 884

Logistic Regression Model					
Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Probability of Survival[a]	-0.043	5.01E-12	0.96	0.95	0.97
Birthweight	0.0019	0.0019	1.0019	1.0007	1.003
Apgar Score at 5 mins[b]	-0.24	3.8E-09	0.79	0.72	0.85
CRIB Score[b]	0.15	1.4E-07	1.16	1.10	1.22
Constant	1.61	0.0015	5.01		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.786	0.015	8.6E-49	0.76	-	0.82

STEP 7 Apgar Score at 1 minute[b] added — n = 884

Logistic Regression Model					
Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Probability of Survival[a]	-0.042	1.07E-11	0.96	0.95	0.97
Birthweight	0.0019	0.0022	1.0019	1.0007	1.003
Apgar Score at 5 mins[b]	-0.22	2.5E-05	0.80	0.73	0.89
CRIB Score[b]	0.14	2.04E-07	1.16	1.094	1.22
Apgar Score at 1 min[b]	-0.034	0.46	0.97	0.88	1.058
Constant	1.60	0.0016	4.93		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.787	0.015	5.47E-49	0.76	-	0.82

STEP 8 Apgar Score at 1 minute[b] removed — n = 884

See Result of Step 6

Continued overleaf

STEP 9 Maximum O_2 in first 12 hours[b] added — n = 884

Logistic Regression Model					
Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Probability of Survival[a]	-0.042	9.5E-12	0.96	0.95	0.97
Birthweight	0.0017	0.0050	1.00	1.00052	1.0029
Apgar Score at 5 mins[b]	-0.24	9.1E-09	0.79	0.73	0.86
CRIB Score[b]	0.11	0.00056	1.11	1.048	1.19
Maximum O_2 in first 12 hours[b]	0.013	0.012	1.01	1.0028	1.023
Constant	1.10	0.044	3.01		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.790	0.015	4.4E-50	0.76	-	0.82

STEP 10 Minimum O_2 in first 12 hours[b] added — n = 884

Logistic Regression Model					
Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Probability of Survival[a]	-0.043	9.7E-12	0.96	0.95	0.97
Birthweight	0.0016	0.010	1.0016	1.00037	1.0028
Apgar Score at 5 mins[b]	-0.23	1.6E-08	0.79	0.73	0.86
CRIB Score[b]	0.086	0.0089	1.089	1.02	1.16
Maximum O_2 in first 12 hours[b]	0.0059	0.31	1.0060	0.99	1.018
Minimum O_2 in first 12 hours[b]	0.018	0.017	1.019	1.0033	1.034
Constant	1.07	0.054	2.90		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.791	0.015	2.1E-50	0.76	-	0.82

STEP 11 Antenatal steroids added — n = 884

Logistic Regression Model					
Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Probability of Survival[a]	-0.040	2.0E-10	0.96	0.95	0.97
Birthweight	0.0013	0.032	1.0013	1.00012	1.0026
Apgar Score at 5 mins[b]	-0.22	1.1E-07	0.80	0.74	0.87
CRIB Score[b]	0.083	0.012	1.087	1.019	1.16
Maximum O_2 in first 12 hours[b]	0.0062	0.29	1.0062	0.995	1.018
Minimum O_2 in first 12 hours[b]	0.016	0.033	1.017	1.0013	1.032
Antenatal Steroids	-0.50	0.0023	0.607	0.44	0.84
Constant	1.43	0.012	4.17		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.793	0.015	4.2E-51	0.76	-	0.82

Continued overleaf

STEP 12 Neonatal transfer added — n = 884

Logistic Regression Model					
Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Probability of Survival[a]	-0.040	4.5E-10	0.96	0.95	0.97
Birthweight	0.0014	0.029	1.0014	1.00014	1.0026
Apgar Score at 5 mins[b]	-0.21	4.2E-07	0.81	0.74	0.88
CRIB Score[b]	0.086	0.010	1.090	1.021	1.16
Maximum O_2 in first 12 hours[b]	0.0068	0.25	1.0068	0.995	1.02
Minimum O_2 in first 12 hours[b]	0.016	0.039	1.016	1.00078	1.03
Antenatal Steroids	-0.51	0.0019	0.60	0.43	0.83
Neonatal Transfer	-0.53	0.0027	0.59	0.42	0.83
Constant	1.44	0.012	4.2		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.800	0.015	2.8E-53	0.77	-	0.83

STEP 13 Caesarean section added — n = 884

Logistic Regression Model					
Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Probability of Survival[a]	-0.038	2.01E-06	0.96	0.95	0.98
Birthweight	0.0012	0.056	1.0012	0.99997	1.0025
Apgar Score at 5 mins[b]	-0.21	5.21E-07	0.81	0.75	0.88
CRIB Score[b]	0.087	0.0068	1.091	1.022	1.16
Maximum O_2 in first 12 hours[b]	0.0070	0.24	1.0070	0.995	1.02
Minimum O_2 in first 12 hours[b]	0.016	0.039	1.016	1.001	1.03
Antenatal Steroids	-0.50	0.0024	0.61	0.44	0.84
Neonatal Transfer	-0.53	0.0031	0.59	0.42	0.84
Caesarean section	0.13	0.43	1.14	0.82	1.60
Constant	1.20	0.064	3.32		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.799	0.015	5.4E-53	0.77	-	0.83

STEP 14 Caesarean section removed — n = 884

See Result of Step 12

Continued overleaf

STEP 15 Gender added — n = 884

Logistic Regression Model					
Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Probability of Survival[a]	-0.034	5.2E-07	0.97	0.95	0.98
Birthweight	0.00075	0.27	1.00075	0.999	1.0021
Apgar Score at 5 mins[b]	-0.21	4.8E-07	0.81	0.74	0.88
CRIB Score[b]	0.093	0.0053	1.098	1.028	1.17
Maximum O_2 in first 12 hours[b]	0.0062	0.30	1.0062	0.99	1.02
Minimum O_2 in first 12 hours[b]	0.015	0.049	1.016	1.00009	1.03
Antenatal Steroids	-0.53	0.0014	0.59	0.43	0.82
Neonatal Transfer	-0.54	0.0025	0.58	0.41	0.83
Gender	-0.45	0.011	0.64	0.45	0.90
Constant	2.31	0.00061	10.07		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.804	0.015	1.1E-54	0.77	-	0.83

STEP 16 Head Circumference at birth[b] added — n = 884

Logistic Regression Model					
Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Probability of Survival[a]	-0.035	3.57E-07	0.97	0.95	0.98
Birthweight	0.00054	0.44	1.00054	0.999	1.0019
Apgar Score at 5 mins[b]	-0.21	7.54E-07	0.81	0.75	0.88
CRIB Score[b]	0.096	0.0044	1.10088	1.030	1.18
Maximum O_2 in first 12 hours[b]	0.0059	0.33	1.0059	0.99	1.018
Minimum O_2 in first 12 hours[b]	0.016	0.042	1.016	1.0006	1.032
Antenatal Steroids	-0.52	0.0017	0.59	0.43	0.82
Neonatal Transfer	-0.55	0.0022	0.58	0.41	0.82
Gender	-0.43	0.016	0.65	0.46	0.92
Head Circumference[b]	0.072	0.16	1.074	0.97	1.19
Constant	0.67	0.61	1.96		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.805	0.015	4.8E-55	0.78	-	0.83

Continued overleaf

STEP 17 Spontaneous vaginal delivery added — n = 882

Logistic Regression Model					
Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Probability of Survival[a]	-0.032	6.9E-06	0.97	0.96	0.98
Birthweight	0.00032	0.65	1.00032	0.999	1.0017
Apgar Score at 5 mins[b]	-0.21	8.2E-07	0.81	0.746	0.88
CRIB Score[b]	0.10	0.0043	1.10	1.031	1.18
Maximum O_2 in first 12 hours[b]	0.006	0.28	1.0065	0.995	1.018
Minimum O_2 in first 12 hours[b]	0.016	0.041	1.016	1.00067	1.032
Antenatal Steroids	-0.52	0.0018	0.60	0.43	0.82
Neonatal Transfer	-0.53	0.0029	0.59	0.41	0.83
Gender	-0.45	0.011	0.64	0.45	0.90
Head Circumference[b]	0.069	0.17	1.07	0.97	1.18
Spontaneous Vaginal Delivery	0.24	0.17	1.27	0.90	1.78
Constant	0.64	0.63	1.91		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.806	0.015	3.2E-55	0.78	-	0.83

STEP 18 Birthweight removed — n = 882

Logistic Regression Model					
Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Probability of Survival[a]	-0.029	4.4E-11	0.97	0.96	0.98
Apgar Score at 5 mins[b]	-0.21	8.5E-07	0.81	0.75	0.88
CRIB Score[b]	0.095	0.0048	1.10	1.029	1.17
Maximum O_2 in first 12 hours[b]	0.0066	0.28	1.0066	0.99	1.019
Minimum O_2 in first 12 hours[b]	0.016	0.038	1.017	1.00094	1.032
Antenatal Steroids	-0.53	0.0014	0.59	0.43	0.82
Neonatal Transfer	-0.53	0.0030	0.59	0.41	0.83
Gender	-0.48	0.0042	0.62	0.45	0.86
Head Circumference[b]	0.074	0.14	1.08	0.98	1.19
Spontaneous Vaginal Delivery	0.26	0.13	1.29	0.93	1.80
Constant	0.69	0.60	2.00		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.805	0.015	3.5E-55	0.78	-	0.83

Continued overleaf

STEP 19 Preeclampsia added — n = 880

Logistic Regression Model					
Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Probability of Survival[a]	-0.029	5.5E-11	0.97	0.96	0.98
Apgar Score at 5 mins[b]	-0.21	1.50E-06	0.81	0.75	0.89
CRIB Score[b]	0.10	0.0044	1.10	1.030	1.18
Maximum O_2 in first 12 hours[b]	0.006	0.29	1.0064	0.99	1.018
Minimum O_2 in first 12 hours[b]	0.016	0.040	1.016	1.00071	1.032
Antenatal Steroids	-0.52	0.0016	0.59	0.43	0.82
Neonatal Transfer	-0.53	0.0030	0.59	0.41	0.83
Gender	-0.46	0.0064	0.63	0.45	0.88
Head Circumference[b]	0.073	0.15	1.075	0.97	1.19
Spontaneous Vaginal Delivery	0.21	0.23	1.24	0.88	1.75
Preeclampsia	-0.21	0.40	0.81	0.50	1.31
Constant	0.72	0.59	2.05		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.805	0.015	6.8E-55	0.78	-	0.83

STEP 20 Preeclampsia removed — n = 882

See Result of Step 18

STEP 21 Other vaginal delivery added — n = 882

Logistic Regression Model					
Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Probability of Survival[a]	-0.029	8.1E-11	0.97	0.96	0.98
Apgar Score at 5 mins[b]	-0.22	4.5E-07	0.81	0.74	0.88
CRIB Score[b]	0.095	0.0048	1.10	1.029	1.17
Maximum O_2 in first 12 hours[b]	0.0070	0.25	1.007	0.995	1.02
Minimum O_2 in first 12 hours[b]	0.016	0.048	1.02	1.0001	1.03
Antenatal Steroids	-0.55	0.00090	0.58	0.42	0.80
Neonatal Transfer	-0.55	0.0021	0.57	0.40	0.82
Gender	-0.50	0.0033	0.61	0.44	0.85
Head Circumference[b]	0.079	0.12	1.08	0.98	1.19
Spontaneous Vaginal Delivery	0.20	0.23	1.23	0.88	1.71
Other Vaginal Delivery	-1.91	0.0074	0.15	0.037	0.60
Constant	0.72	0.59	2.06		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.810	0.015	7.1E-57	0.78	-	0.84

Continued overleaf

STEP 22 Year of Birth added — n = 882

Logistic Regression Model					
Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Probability of Survival[a]	-0.029	8.5E-11	0.97	0.96	0.98
Apgar Score at 5 mins[b]	-0.22	4.6E-07	0.80	0.74	0.86
CRIB Score[b]	0.094	0.0053	1.10	1.03	1.17
Maximum O_2 in first 12 hours[b]	0.0070	0.25	1.007	0.995	1.02
Minimum O_2 in first 12 hours[b]	0.016	0.047	1.02	1.0002	1.03
Antenatal Steroids	-0.57	0.0014	0.57	0.40	0.80
Neonatal Transfer	-0.56	0.0021	0.57	0.40	0.82
Gender	-0.49	0.0034	0.61	0.44	0.85
Head Circumference[b]	0.079	0.12	1.082	0.98	1.19
Spontaneous Vaginal Delivery	0.20	0.24	1.22	0.88	1.71
Other Vaginal Delivery	-1.91	0.0074	0.15	0.04	0.60
Year of Birth	0.0079	0.84	1.0079	0.93	1.09
Constant	-15.0	0.85	3.0E-07		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.810	0.015	7.7E-57	0.78	-	0.84

STEP 23 Year of Birth removed — n = 882

See Result of Step 21

STEP 24 In Utero Transfer added — n = 882

Logistic Regression Model					
Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Probability of Survival[a]	-0.029	6.5E-11	0.97	0.96	0.979623212
Apgar Score at 5 mins[b]	-0.22	3.4E-07	0.80	0.74	0.87
CRIB Score[b]	0.094	0.005	1.10	1.029	1.17
Maximum O_2 in first 12 hours[b]	0.0069	0.26	1.0069	0.995	1.019
Minimum O_2 in first 12 hours[b]	0.016	0.040	1.016	1.00073	1.03
Antenatal Steroids	-0.59	0.00049	0.56	0.40	0.77
Neonatal Transfer	-0.68	0.00056	0.51	0.34	0.74
Gender	-0.50	0.0033	0.61	0.44	0.85
Head Circumference[b]	0.081	0.109	1.084	0.98	1.20
Spontaneous Vaginal Delivery	0.21	0.23	1.23	0.88	1.72
Other Vaginal Delivery	-1.94	0.0076	0.14	0.035	0.60
In Utero Transfer	0.37	0.11	1.44	0.92	2.3
Constant	0.69	0.61	2.00		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.811	0.015	3.5E-57	0.78	-	0.84

Continued overleaf

STEP 25 Multiple Pregnancy added — n = 882

Logistic Regression Model					
Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Probability of Survival[a]	-0.029	1.3E-10	0.97	0.96	0.980038617
Apgar Score at 5 mins[b]	-0.23	2.0E-07	0.80	0.73	0.87
CRIB Score[b]	0.094	0.0053	1.099	1.028	1.17
Maximum O_2 in first 12 hours[b]	0.0068	0.26	1.0068	0.99	1.02
Minimum O_2 in first 12 hours[b]	0.016	0.040	1.016	1.0007	1.03
Antenatal Steroids	-0.56	0.00083	0.57	0.41	0.79
Neonatal Transfer	-0.68	0.00055	0.50	0.34	0.74
Gender	-0.48	0.0050	0.62	0.45	0.87
Head Circumference[b]	0.077	0.13	1.080	0.98	1.19
Spontaneous Vaginal Delivery	0.21	0.22	1.23	0.88	1.73
Other Vaginal Delivery	-1.92	0.0079	0.15	0.036	0.61
In Utero Transfer	0.35	0.13	1.42	0.90	2.22
Multiple Pregnancy	0.28	0.13	1.33	0.92	1.92
Constant	0.72	0.59	2.05		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.812	0.015	1.2E-57	0.78	-	0.84

STEP 26 Maximum Base Deficit in first 12 hours[b] added — n = 882

Logistic Regression Model					
Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Probability of Survival[a]	-0.029	1.1E-10	0.97	0.96	0.979867357
Apgar Score at 5 mins[b]	-0.23	2.1E-07	0.80	0.73	0.87
CRIB Score[b]	0.095	0.0048	1.10	1.029	1.17
Maximum O_2 in first 12 hours[b]	0.0070	0.25	1.0071	0.995	1.019
Minimum O_2 in first 12 hours[b]	0.017	0.035	1.017	1.0012	1.033
Antenatal Steroids	-0.56	0.00088	0.57	0.41	0.79
Neonatal Transfer	-0.69	0.00054	0.50	0.34	0.74
Gender	-0.47	0.0056	0.62	0.45	0.87
Head Circumference[b]	0.079	0.12	1.082	0.98	1.20
Spontaneous Vaginal Delivery	0.22	0.20	1.25	0.89	1.74
Other Vaginal Delivery	-1.94	0.0074	0.14	0.035	0.60
In Utero Transfer	0.33	0.15	1.39	0.89	2.19
Multiple Pregnancy	0.28	0.13	1.33	0.92	1.91
Max Base Deficit in first 12hr[b]	-0.012	0.36	0.99	0.96	1.014
Constant	0.67	0.62	1.96		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.813	0.015	7.5E-58	0.78	-	0.84

Continued overleaf

STEP 27 Small for Gestation (<10th Centile) added — n = 882

Logistic Regression Model					
Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Probability of Survival[a]	-0.029	1.1E-10	0.97	0.96	0.98
Apgar Score at 5 mins[b]	-0.23	2.1E-07	0.80	0.73	0.87
CRIB Score[b]	0.097	0.0051	1.1016	1.030	1.18
Maximum O_2 in first 12 hours[b]	0.007	0.25	1.007	0.995	1.019
Minimum O_2 in first 12 hours[b]	0.017	0.038	1.017	1.00090	1.03
Antenatal Steroids	-0.56	0.00094	0.57	0.41	0.80
Neonatal Transfer	-0.69	0.00054	0.50	0.34	0.74
Gender	-0.47	0.0054	0.62	0.45	0.87
Head Circumference[b]	0.078	0.12	1.081	0.98	1.19
Spontaneous Vaginal Delivery	0.21	0.24	1.23	0.87	1.75
Other Vaginal Delivery	-1.95	0.0073	0.14	0.034	0.59
In Utero Transfer	0.33	0.15	1.39	0.89	2.19
Multiple Pregnancy	0.28	0.14	1.32	0.91	1.91
Max Base Deficit in first 12hr[b]	-0.012	0.36	0.99	0.96	1.014
Small for Gestation (<10th Centile)	-0.057	0.82	0.94	0.58	1.54
Constant	0.73	0.59	2.07		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.813	0.015	1.02E-57	0.78	-	0.84

STEP 28 Small for Gestation (<10th Centile) removed — n = 882

See Result of Step 26

STEP 29 Number of infants in pregnancy added — n = 882

= ***Model using all factors***

Logistic Regression Model					
Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Probability of Survival[a]	-0.029	9.9E-11	0.97	0.96	0.98
Apgar Score at 5 mins[b]	-0.23	2.4E-07	0.80	0.73	0.87
CRIB Score[b]	0.095	0.0050	1.10	1.03	1.17
Maximum O_2 in first 12 hours[b]	0.0068	0.26	1.007	0.99	1.019
Minimum O_2 in first 12 hours[b]	0.017	0.034	1.02	1.0012	1.03
Antenatal Steroids	-0.58	0.00067	0.56	0.40	0.78
Neonatal Transfer	-0.68	0.00063	0.51	0.34	0.75
Gender	-0.46	0.0075	0.63	0.45	0.89
Head Circumference[b]	0.079	0.12	1.08	0.98	1.20
Spontaneous Vaginal Delivery	0.22	0.20	1.24	0.89	1.74
Other Vaginal Delivery	-2.0	0.0067	0.14	0.035	0.58
In Utero Transfer	0.31	0.17	1.37	0.87	2.15
Multiple Pregnancy	1.1	0.059	3.11	0.96	10.1
Max Base Deficit in first 12hr[b]	-0.013	0.33	0.99	0.96	1.01
Number of Infants in Pregnancy	-0.76	0.14	0.47	0.17	1.27
Constant	1.5	0.32	4.31		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.814	0.015	2.9E-58	0.79	-	0.84

 Continued overleaf

STEP 30 Maximum Base Deficit in first 12 hours[b] removed — n = 882

Logistic Regression Model					
Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Probability of Survival[a]	-0.029	1.2E-10	0.97	0.96	0.979937604
Apgar Score at 5 mins[b]	-0.23	2.3E-07	0.80	0.73	0.87
CRIB Score[b]	0.094	0.0054	1.10	1.028	1.17
Maximum O_2 in first 12 hours[b]	0.0066	0.28	1.007	0.995	1.019
Minimum O_2 in first 12 hours[b]	0.016	0.040	1.016	1.0007	1.032
Antenatal Steroids	-0.58	0.00063	0.56	0.40	0.78
Neonatal Transfer	-0.68	0.00064	0.51	0.34	0.749201213
Gender	-0.46	0.0067	0.63	0.45	0.879848324
Head Circumference[b]	0.076	0.13	1.079	0.98	1.19
Spontaneous Vaginal Delivery	0.21	0.22	1.23	0.88	1.73
Other Vaginal Delivery	-1.93	0.0072	0.14	0.035	0.59
In Utero Transfer	0.33	0.15	1.39	0.89	2.19
Multiple Pregnancy	1.12	0.063	3.05	0.94	9.91
Number of Infants in Pregnancy	-0.75	0.15	0.47	0.17	1.30
Constant	1.49	0.30	4.44		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.813	0.015	4.9E-58	0.78	-	0.84

STEP 31 Maximum FiO2 in first 12 hours removed — n = 882

Logistic Regression Model					
Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Probability of Survival[a]	-0.029	1.0E-10	0.97	0.96	0.979857077
Apgar Score at 5 mins[b]	-0.23	1.9E-07	0.80	0.73	0.87
CRIB Score[b]	0.10	0.0016	1.11	1.040	1.18
Minimum O_2 in first 12 hours[b]	0.020	0.0031	1.021	1.0069	1.03
Antenatal Steroids	-0.58	0.00065	0.56	0.40	0.78
Neonatal Transfer	-0.67	0.00071	0.51	0.35	0.75
Gender	-0.47	0.0056	0.62	0.45	0.87
Head Circumference[b]	0.079	0.12	1.08	0.98	1.19
Spontaneous Vaginal Delivery	0.20	0.25	1.22	0.87	1.70
Other Vaginal Delivery	-1.92	0.0079	0.15	0.036	0.60
In Utero Transfer	0.33	0.15	1.40	0.89	2.19
Multiple Pregnancy	1.13	0.059	3.10	0.96	10.01
Number of Infants in Pregnancy	-0.76	0.14	0.47	0.17	1.28
Constant	1.66	0.25	5.3		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.812	0.015	2.3E-57	0.78	-	0.84

Continued overleaf

STEP 32 Spontaneous Vaginal Delivery removed — n = 883

Logistic Regression Model					
Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Probability of Survival[a]	-0.030	1.3E-11	0.97	0.96	0.98
Apgar Score at 5 mins[b]	-0.23	1.8E-07	0.80	0.73	0.87
CRIB Score[b]	0.10	0.0020	1.11	1.038	1.18
Minimum O_2 in first 12 hours[b]	0.020	0.0033	1.02	1.0067	1.034
Antenatal Steroids	-0.59	0.00053	0.56	0.40	0.78
Neonatal Transfer	-0.68	0.00059	0.51	0.34	0.75
Gender	-0.47	0.0058	0.63	0.45	0.87
Head Circumference[b]	0.084	0.096	1.09	0.99	1.20
Other Vaginal Delivery	-1.99	0.0057	0.14	0.033	0.560280005
In Utero Transfer	0.33	0.15	1.40	0.89	2.19
Multiple Pregnancy	1.13	0.056	3.10	0.97	9.88
Number of Infants in Pregnancy	-0.76	0.13	0.47	0.17	1.25
Constant	1.71	0.24	5.53		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.812	0.015	2.1E-57	0.78	-	0.84

STEP 33 In utero transfer removed — n = 883

Logistic Regression Model					
Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Probability of Survival[a]	-0.030	1.6E-11	0.97	0.96	0.98
Apgar Score at 5 mins[b]	-0.22	2.2E-07	0.80	0.73	0.869595009
CRIB Score[b]	0.10	0.0018	1.11	1.038	1.179508969
Minimum O_2 in first 12 hours[b]	0.020	0.0041	1.02	1.006	1.033500132
Antenatal Steroids	-0.56	0.00092	0.57	0.41	0.80
Neonatal Transfer	-0.56	0.0018	0.57	0.40	0.810429624
Gender	-0.47	0.0059	0.63	0.45	0.87
Head Circumference[b]	0.082	0.10	1.09	0.98	1.20
Other Vaginal Delivery	-1.97	0.0055	0.14	0.035	0.56018733
Multiple Pregnancy	1.18	0.044	3.27	1.031	10.4
Number of Infants in Pregnancy	-0.80	0.11	0.45	0.17	1.20
Constant	1.77	0.22	5.87		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.810	0.015	5.8E-57	0.78	-	0.84

Continued overleaf

STEP 34 Number of Infants in Pregnancy removed — n = 883

Logistic Regression Model					
Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Probability of Survival[a]	-0.030	1.9E-11	0.97	0.96	0.98
Apgar Score at 5 mins[b]	-0.22	1.9E-07	0.80	0.73	0.87
CRIB Score[b]	0.10	0.0016	1.11	1.039	1.180464784
Minimum O_2 in first 12 hours[b]	0.020	0.0040	1.02	1.0063	1.03
Antenatal Steroids	-0.54	0.0013	0.58	0.42	0.81
Neonatal Transfer	-0.57	0.0016	0.57	0.40	0.81
Gender	-0.48	0.0043	0.62	0.44	0.86
Head Circumference[b]	0.082	0.098	1.09	0.98	1.20
Other Vaginal Delivery	-1.95	0.0061	0.14	0.035	0.57
Multiple Pregnancy	0.29	0.11	1.34	0.93	1.93
Constant	0.95	0.47	2.59		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.809	0.015	1.2E-56	0.78	-	0.84

STEP 35 Multiple Pregnancy removed — n = 883

Logistic Regression Model					
Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Probability of Survival[a]	-0.03	9.5E-12	0.97	0.96	0.98
Apgar Score at 5 mins[b]	-0.22	3.3E-07	0.80	0.74	0.87
CRIB Score[b]	0.10	0.0016	1.11	1.040	1.18
Minimum O_2 in first 12 hours[b]	0.020	0.0039	1.020	1.0063	1.03
Antenatal Steroids	-0.56	0.00077	0.57	0.41	0.79
Neonatal Transfer	-0.56	0.0019	0.57	0.40	0.81
Gender	-0.50	0.0029	0.61	0.44	0.84
Head Circumference[b]	0.087	0.083	1.09	0.99	1.20
Other Vaginal Delivery	-1.97	0.0059	0.14	0.034	0.57
Constant	0.94	0.48	2.55		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.808	0.015	4.6E-56	0.78	-	0.84

Continued overleaf

STEP 36 Head Circumference removed n = 884

Logistic Regression Model					
Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Probability of Survival[a]	-0.027	1.9E-11	0.97	0.97	0.98
Apgar Score at 5 mins[b]	-0.22	2.0E-07	0.80	0.74	0.87
CRIB Score[b]	0.096	0.0023	1.1024	1.035	1.17
Minimum O_2 in first 12 hours[b]	0.019	0.0042	1.020	1.0062	1.03
Antenatal Steroids	-0.58	0.00049	0.56	0.41	0.78
Neonatal Transfer	-0.55	0.0023	0.58	0.41	0.82
Gender	-0.55	0.00090	0.58	0.42	0.80
Other Vaginal Delivery	-1.96	0.0066	0.14	0.034	0.58
Constant	3.07	5.9E-09	21.5		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.806	0.015	1.4E-55	0.78	-	0.84

STEP 37 Other vaginal delivery removed n = 884

Logistic Regression Model					
Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Probability of Survival[a]	-0.028	4.6E-12	0.97	0.96	0.98
Apgar Score at 5 mins[b]	-0.21	4.0E-07	0.81	0.74	0.88
CRIB Score[b]	0.097	0.0024	1.10	1.035	1.17
Minimum O_2 in first 12 hours[b]	0.020	0.0033	1.02	1.0067	1.034
Antenatal Steroids	-0.55	0.00080	0.58	0.42	0.80
Neonatal Transfer	-0.53	0.0028	0.59	0.41	0.83
Gender	-0.53	0.0013	0.59	0.43	0.81
Constant	2.94	1.8E-08	18.9		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.801	0.015	6.5E-54	0.77	-	0.83

Continued overleaf

STEP 38 Minimum FIO2 in first 12 hours[b] removed — n = 884

Logistic Regression Model					
Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Probability of Survival[a]	-0.026	5.1E-11	0.97	0.97	0.98
Apgar Score at 5 mins[b]	-0.22	1.3E-07	0.80	0.74	0.87
CRIB Score[b]	0.14	3.2E-07	1.15	1.09	1.22
Antenatal Steroids	-0.60	0.00022	0.55	0.40	0.75
Neonatal Transfer	-0.53	0.0026	0.59	0.41	0.83
Gender	-0.58	0.00041	0.56	0.41	0.77
Constant	3.47	1.5E-12	32.2		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.799	0.015	3.9E-53	0.77	-	0.83

STEP 39 Neonatal Transfer removed — n = 884
= PARSIMONIOUS MODEL

Logistic Regression Model					
Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Probability of Survival[a]	-0.027	6.9E-12	0.97	0.97	0.98
Apgar Score at 5 mins[b]	-0.23	3.7E-08	0.80	0.73	0.86
CRIB Score[b]	0.14	4.8E-07	1.15	1.089	1.21
Antenatal Steroids	-0.59	0.00025	0.55	0.40	0.76
Gender	-0.57	0.00047	0.57	0.41	0.78
Constant	3.42	1.7E-12	30.7		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.793	0.015	4.7E-51	0.76	-	0.82

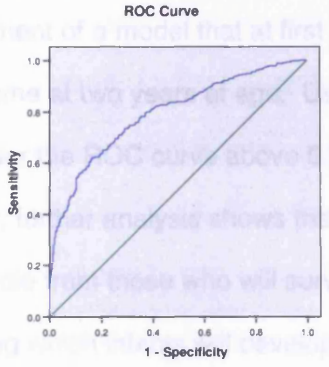
Notes

a: Probability of survival obtained from 'Draper Grid'

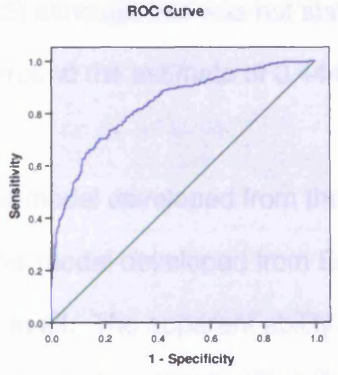
b: Mean value substituted for missing values

Figure 14: ROC Curves of Predictive Ability for the Different Outcomes (EAVLBW) - PARSIMONIOUS MODEL

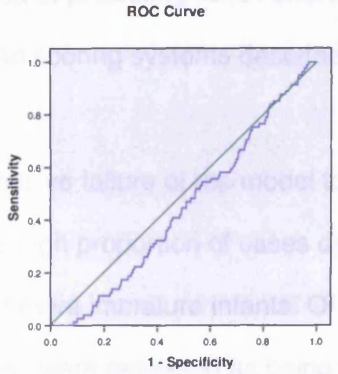
A. Severe disability or Death v Other (Az = 0.793, 95% C.I. 0.763 – 0.823)



B. Death v Other (Az = 0.835, 95% C.I. 0.807 – 0.863)



C. Severe disability v Other or No disability (Az = 0.444, 95% C.I. 0.385 – 0.504)



3.2.4 Discussion

As in the previous section of this chapter, this section also describes the development of a model that at first glance appears to have good predictive ability for outcome at two years of age. Using data obtained in the first 12 hours of life the area under the ROC curve above 0.8 implies good discrimination[201]. Once again however, further analysis shows that the model is very good at separating infants who will die from those who will survive to two years of age, but it is very poor at identifying which infants will develop severe disability at two years of age. Indeed the area under the ROC curve of just 0.444 is worse than could be expected by chance alone (0.5) although this was not statistically significant with a 95% confidence interval around the estimate of 0.444 of 0.385 – 0.504.

Whilst the model developed from the Trent cohort appeared to predict disability poorly, this model developed from East Anglian infants does not appear to predict disability at all. The apparent ability to predict the combined outcome is almost certainly due to the model's good ability for predicting death. Once again however this method of predicting death offers little beyond the scope of other well validated, established scoring systems described in chapter one of this thesis.

The impressive failure of the model to predict neurodevelopmental outcome may be due to the high proportion of cases dying in this cohort as a result of the greater number of more immature infants. Of the 896 infants in the EAVLBW cohort, 323 died and 81 were assessed as being severely disabled at two years of age. In contrast, the Trent cohort detailed in section 3.1 contained a balanced mix of outcomes with 27 infants who died, and 24 infants being severely disabled. It could therefore be anticipated that the model derived from the East Anglian cohort would be better at predicting death than the later outcome.

Strengths of the EAVLBW cohort included the size of the cohort, it's geographical population basis and an impressive follow-up rate. The number of infants involved would indicate that a model could have as many as 40 variables, potentially enhancing the models predictive ability[200]. Although this might be beneficial for prediction it makes the model cumbersome and harder to use. Successful models that are widely used by other investigators contain a small number of readily available datapoints[77]. For this reason, in this study a large model (Step 29) was developed and then reduced down to a parsimonious model (Step 39) for which the ROC curve analysis was produced. As can be seen from the ROC curve values in Figure 14, some predictive ability was lost, but considerably less data was needed for the parsimonious model.

Collecting data from a geographical population removes a common cause of selection bias and the questionable generalisability that tertiary cohorts suffer from[166]. These issues come from the different mix of risk in tertiary cohorts as many of the infants have both been referred for expert tertiary care indicating high risk, but have also survived transfer either before or after birth indicating less severe illness than those who do not survive so far.

The follow up rate was impressive[158] with outcome being determined for 98.7% of the infants born before 29 completed weeks of gestation. These are recognized as the highest risk infants[179] and they were therefore kept under clinical review by local clinicians who determined the outcome around 24 months of age corrected for prematurity. Whilst this system enabled the impressive follow up rate, it may also have introduced bias from non-blinded assessments[157].

As discussed in chapter two, the complete dataset of all infants weighing under 1.5kg at birth was subject to selection bias due to the use of a birthweight criterion. This problem was dealt with by applying an upper gestation limit but this limits the future use of the model because of a lack of generalisability and resulted in a higher proportion of infants dying due to prematurity. This may have resulted in the model predicting death in preference to severe disability.

A major weakness of the study was missing data: this affected all but 139 infants. In order to include all infants and variables in the analysis, mean values were used in substitution for missing continuous values[213]. This weakened the association of the variables with the combined outcome (Table 12) and potentially reduced the discriminatory ability of the model. Limiting the model to just infants on whom data was available would have introduced an unacceptable risk of bias[157]. The final model contained six such variables utilising the mean value of the other infants, while the parsimonious model contained 2 of these variables. If the model was being used in other cohorts this would lead to significant concerns about it's generalisability meriting substantial validation in at least one similar cohort with full data ascertainment[190, 214]. It is clear however that this model does not predict the combined outcome and it will not be used in other studies: substituted values have therefore been left in the model.

Another potential weakness of the study was that different assessment procedures were employed by different clinicians to determine the outcome. For example, to determine the developmental progress of the infants some clinicians used the Denver score, others the Bayley score. Whilst this may have been a concern, the use of a standardized classification; the OHSQ[108] should have alleviated this issue

and also replicates the clinical situation that a prediction score might be employed in the future.

As in section 3.1 the inclusion of other predictive variables might have improved the model. It might however be better to use models that predict death and disability separately as it is unlikely that the factors which influence development are the same as those which predict death. In a similar way, it might be possible to predict more accurately the individual disabilities that make up severe disability[194, 215] such as severe hearing loss, blindness or inability to feed. As mentioned in section 3.2.1, the use of other predictive variables such as socioeconomic status and antenatal infection also warrant further investigation.

3.2.5 Comments and implications

This investigation provides further evidence that predicting a combined outcome of death and disability is difficult. It remains unclear whether this is due to lack of variables in the model or whether later physical insults, complications, or environmental factors are more important for long term outcome than inherent risk at birth. As mentioned in the previous section, to determine whether these factors are more important predictors it would seem best to develop a new cohort and collect other data that could potentially predict developmental outcome. In addition it may also be advisable to look at individual disabilities separately.

Whilst it is also hampered by possible selection bias, the UKOS trial dataset contains considerable information on later complications and was therefore analysed to determine the relative importance of early and late factors in the next two sections.

3.3 PREDICTING NEURODEVELOPMENTAL OUTCOME: UNITED KINGDOM OSCILLATION STUDY

3.3.1 Introduction

As discussed in the previous two sections of this chapter, although the best datasets available for study into predicting outcome at two years of age appear to be the Trent Neonatal Survey data from the ABC study and the East Anglian Very Low Birthweight Database, further analysis is needed to confirm these findings. This section of the chapter therefore details the third attempt at deriving a model using the UKOS dataset.

3.3.2 Methods

Data collected prospectively were obtained from the United Kingdom randomised trial of high frequency oscillation (UKOS)[178]. This was a randomized, controlled comparison between conventional and oscillatory ventilation commenced in the first hour of life in babies born between August 1998 and January 2001. Babies were eligible for inclusion in the trial and this secondary analysis of the data if they were born between 23 weeks and 28 weeks plus 6 days of gestational age; if they were born in a participating centre; if they required endotracheal intubation from birth; and if they required ongoing intensive care. Infants were excluded if they had to be transferred to another hospital for intensive care shortly after birth or if they had a major congenital malformation. Randomization was stratified according to centre and gestational age (23 to 25 weeks or 26 to 28 weeks).

The UKOS study showed no difference in the composite primary outcome of death or chronic lung disease, diagnosed at 36 weeks of postmenstrual age. This outcome occurred in 66 percent of the infants assigned to receive high-frequency

oscillatory ventilation and 68 percent of those in the conventional-ventilation group (RR risk in oscillated group: 0.98; 95 percent confidence interval, 0.89 to 1.08). Neurodevelopmental outcome data was determined between 22 and 28 months of age corrected for prematurity by a questionnaire completed by the local clinician who followed the infant clinically to at least two years of age corrected for prematurity[180]. Questionnaires were mailed to the local paediatrician responsible for follow up when each infant reached 21 months post-term age, with a request that the child be evaluated as close to 24 months post-term age as possible and within a "window" of 22–28 months. Up to two reminders were sent to paediatricians when questionnaires had not been returned to the coordinating centre by 25 months post-term age. If questionnaires were still not returned, in the United Kingdom the child's local health visitor was telephoned and asked to complete the forms. Outcome was classified at two years of age as normal, mild disability, moderate disability, severe disability, or death, according to the Oxford Health Status Questionnaire (OHSQ)[108].

Statistical analysis was performed using SPSS v11. The outcome of individual infants was classified as no disability, mild disability or moderate disability: group A, severe disability: group B or death: Group C (see Figure 15). Only variables collectable in the first 12 hours of life were analysed to minimise the potential for the quality of treatment received during the infant's early course to influence the score[79]. Associations with outcome at two years of age corrected for prematurity were identified by the Chi-squared test, Fishers exact test or logistic regression analysis as appropriate. The variables were then assessed in combination, using a stepwise multivariable logistic regression analysis.

Variables were only included in the models if they had a significance level of less than 0.2 for an association with the combined outcome of death and severe

disability. Predictive variables were added to the model in order of significance, a receiver operator characteristic (ROC) curve for the prediction of severe disability was then determined after each variable was added. The variable was then either left in the model or removed if the area under the ROC curve was less than before that variable was added. The recommendation of Wasson and colleagues[200] was used to limit the number of variables to a maximum of one variable for every 10 cases of the outcome of interest. Where variable data points were missing for just one or two infants or for categorical variables, analysis was performed with the remaining infants only. Where a significant number of numerical data points were missing, in order to avoid losing these infants from the analysis, missing values were replaced with the mean value of the remaining infants[213].

As there was a combined outcome, to check that the model was predicting disability in addition to death, the predictive ability of the probability score for the combined outcome from each individual was also assessed against the separated outcomes of death and severe disability. To test the ability of the model to discriminate between death and other outcomes, all cases were left in the cohort. The test therefore analysed the model's ability to discriminate a Group A outcome from a combined Group B or C outcome (see Figure 15). To test the ability of the model to discriminate between severe disability and milder disability and normality, deaths were first removed from the cohort. The test therefore analysed the ability of the model to discriminate Group B from Group A outcomes.

Figure 15: Graphical Depiction of the Different Outcome Groups

← SPECTRUM OF POSSIBLE OUTCOMES →

Normal	Mildly Disabled	Moderately Disabled	Severely disabled	Died
Group A			**Group B**	**Group C**

Survival without severe disability = Group A
Death or severe disability = Group B & C combined

3.3.3 Results

797 infants were enrolled in the study between August 1998 and January 2001. Outcomes and demographic details are documented in Table 14. Of the 797 infants, 212 died and 32 were assessed as being severely disabled at two years of age. The outcome was not measured within the pre-specified age window for 203 infants and their data was excluded from the analysis. The association of predictive variables with death or severe disability at two years of age is documented in Table 15. As in the other sections of Chapter three, the Draper Grid prediction of survival was used[74].

A full set of datapoints for the variables included in the final model were only available for 340 infants. Where variable data points were missing for just one or two infants or for categorical variables, analysis was performed with the remaining infants only. Where a significant number of numerical data points were missing, in order to avoid losing these infants from the analysis, missing values were replaced with the mean value of the remaining infants.

For the Apgar score at 10 minutes this amounted to 278 infants for whom the mean value of 8.37 points was substituted. Similarly, the head circumference involved 236 infants and a mean value of 24.4cm was used regardless of gestational age. Corresponding values for the mean blood pressure at 2 hours, were 45 infants and 31.5 mmHg, base excess at 12 hours, 42 infants and -4.66 mEq/L, inspired percentage of oxygen at 12hrs, 37 infants and 37.7%, mean blood pressure at 12 hours, 27 infants and 31.7 mmHg, base excess at 2 hours, 21 infants and -4.66 mEq/L, Apgar score at 5 minutes, 19 infants and 7.78 points, serum pH at 12 hours, 19 infants and 7.346. Apgar score at 1 minute, 11 infants and 5.23 points, Apgar

score for heart rate at five minutes, eight infants and 2 points, inspired percentage of oxygen at two hours, 5 infants and 46.0%, time to intubation, three infants and 3.74 minutes, and serum pH at two hours, two infants and 7.333.

For categorical variables missing data that resulted in infants being removed from the analysis involved nine infants for cardiac massage & adrenaline during resuscitation, five infants for volume expander received during resuscitation, two infants for inotropes used by 12 hours and one infant for volume expander received by 12 hours. Due to multiple variables being missing for the same infant, this involved a total of 12 infants.

The final model for predicting death or severe disability at two years of age is documented in Table 16. The derivation of this model by stepwise forward regression and the predictive ability (discrimination) of the model at each step is detailed in Figure 16. The discrimination of the final model for predicting the combined outcome (death or severe disability), and the separated outcomes of death, and disability are shown with the ROC curves in Figure 17.

Table 14: Comparison of Included and Excluded infants (exclusion due to unknown outcome at two years of age)

	All infants		Included		Excluded		Included v Excluded
	n =797	% or SD	n =594	% or SD	n =203	% or SD	
Gestational Age (completed Weeks) Median	25	-	26	-	27	-	p=0.06
Birthweight (g) Median	840	-	820g	-	885g	-	**p=0.001**
SGA	138	(18.9%)	110	(18.5%)	28	(20.3%)	p=0.125
Mean Apgar Score at 1 Minute	5.34	2.3	5.23	2.4	5.66	2.1	**p=0.024**
Mean Apgar Score at 5 Minutes	7.89	1.8	7.71	1.9	8.21	1.4	**p=0.003**
Sex							
Female	369	(46.3%)	263	(44.3%)	106	(52.2%)	p=0.05
Male	428	(53.7%)	331	(55.7%)	97	(47.8%)	
Multiple Pregnancy	190	(23.8%)	148	(24.9%)	42	(20.7%)	p=0.22
Surfactant Treatment	769	(96.5%)	577	(97.1%)	192	(94.6%)	p=0.088
Antenatal Steroid Treatment	727	(91.2%)	543	(91.4%)	203	(90.6%)	p=0.31
Maternal Age (Years)	28.8	6.1	29.0	6.1	28.4	6.2	p=0.26
Base Excess at 2 hours	-4.48	4.1	-4.66	3.4	-3.97	4.0	**p=0.04**
Base Excess at 12 hours	-4.55	3.3	-4.66	3.3	-4.24	3.3	p=0.13
Outcome							
Died	212	(26.6%)	212	(35.7%)	0	0	NA
Severe Disability at 2 years of age	47	(5.9%)	32	(5.4%)	15	(7.4%)	**p=0.30**
Survived without severe disability	538	(32.5%)	350	(58.9%)	188	(92.6%)	NA
Outcome unknown at 2 years	203	(25.5%)	0	-	203	-	NA

Table 15: Associations of predictive variables with death or severe disability at two years

Variable	Coefficient	Standard Error	Wald	Number	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
							Lower	Upper
Probability of Survival	-0.039	0.0042	84.6	594	3.60E-20	0.96	0.95	0.97
Birthweight (g)	-0.0039	0.00047	68.2	594	1.47E-16	0.996	0.995	0.997
Gestational Age (weeks)	-0.50	0.062	64.3	594	1.06E-15	0.61	0.54	0.69
Inspired Oxygen at 12hrs (%)	2.69	0.45	35.9	557	2.09E-09	14.7	6.1	35.3
Inspired Oxygen at 12hrs[1]	2.69	0.45	35.7	594	2.31E-09	14.7	6.1	35.5
Inspired Oxygen at 2hrs (%)	2.12	0.38	31.6	589	1.93E-08	8.3	4.0	17.4
Inspired Oxygen at 2hrs[1]	2.12	0.38	31.6	594	1.89E-08	8.3	4.0	17.4
Head Circumference	-0.32	0.062	26.7	358	2.41E-07	0.73	0.64	0.82
Head Circumference[1]	-0.31	0.060	26.0	594	3.37E-07	0.74	0.65	0.83
Base Excess at 2 hours	-0.11	0.02	23.6	573	1.17E-06	0.89	0.85	0.93
Base Excess at 2 hours[1]	-0.11	0.02	23.6	594	1.21E-06	0.89	0.85	0.94
Apgar Score at 5 minutes	-0.24	0.049	23.2	575	1.45E-06	0.79	0.72	0.87
Apgar Score at 5 minutes[1]	-0.24	0.049	23.2	594	1.46E-06	0.79	0.72	0.87
Inotropes used by 12 hours	-0.80	0.18	20.1	592	7.48E-06	0.45	0.32	0.64
Apgar Score at 1 minute	-0.16	0.037	18.7	583	1.53E-05	0.85	0.79	0.92
Apgar Score at 1 minute[1]	-0.16	0.037	18.7	594	1.54E-05	0.85	0.79	0.92
Volume expander received by 12 hours	-1.025	0.25	16.3	593	5.30E-05	0.36	0.22	0.59
Base Excess at 12 hours	-0.10	0.03	13.8	552	0.00021	0.90	0.86	0.95
Base Excess at 12 hours[1]	-0.10	0.03	13.5	594	0.00024	0.90	0.86	0.95
Serum pH at 2 hours	-2.96	0.84	12.48	592	0.00041	0.052	0.01	0.27
Serum pH at 2 hours[1]	-2.97	0.84	12.50	594	0.00041	0.051	0.01	0.27
Apgar score at 10 minutes	-0.23	0.068	11.1	316	0.00085	0.80	0.70	0.91
Apgar score at 10 minutes[1]	-0.23	0.068	11.2	594	0.00082	0.80	0.70	0.91
Apgar Score for Heart Rate at 5 minute	-0.61	0.20	8.8	586	0.0030	0.54	0.36	0.81
Apgar Score for Heart Rate[1]	-0.59	0.20	8.4	594	0.0038	0.55	0.37	0.83
Cardiac Massage & Adrenaline during resuscitation	0.84	0.30	8.0	585	0.0048	2.31	1.29	4.12
Mean Blood Pressure at 2 hours	-0.031	0.012	6.39	549	0.012	0.97	0.95	0.99
Mean Blood Pressure at 2 hours[1]	-0.031	0.012	6.38	594	0.012	0.97	0.95	0.99
Mean Blood Pressure at 12 hours	-0.031	0.013	5.47	567	0.019	0.97	0.94	0.99
Mean Blood Pressure at 12 hours[1]	-0.031	0.013	5.36	594	0.021	0.97	0.95	1.00

Continued overleaf

Serum pH at 12 hours	-2.04	0.90	5.09	575	0.024	0.13	0.022	0.77
Serum pH at 12 hours[1]	-2.01	0.90	5.01	594	0.025	0.13	0.023	0.78
Time to Intubation	-0.046	0.021	4.70	591	0.030	0.95	0.92	1.00
Time to Intubation[1]	-0.046	0.021	4.70	594	0.030	0.95	0.92	1.00
PaCO2 at 12 hours	-0.073	0.03	4.55	593	0.033	0.93	0.87	0.99
Small for Gestation[2]	0.45	0.21	4.40	594	0.036	1.56	1.03	2.37
Male Gender	0.34	0.17	4.07	594	0.044	1.41	1.01	1.96
Volume expander received during resuscitation	1.09	0.55	3.86	589	0.050	2.97	1.00	8.80
Multiple pregnancy	0.34	0.19	3.14	594	0.077	1.40	0.96	2.04
Normal Vaginal Delivery	0.28	0.17	2.83	594	0.092	1.33	0.95	1.85
Delivery for Intra-Uterine Growth Retardation	-0.40	0.27	2.25	594	0.13	0.67	0.39	1.13
Number of infants in pregnancy	0.17	0.15	1.41	594	0.23	1.19	0.89	1.59
Maternal Labour	0.20	0.18	1.27	594	0.26	1.22	0.86	1.72
Antenatal steroids	-0.30	0.30	1.1	593	0.30	0.74	0.41	1.32
Delivery by / for Spontaneous preterm labour	0.17	0.17	1.04	594	0.31	1.19	0.85	1.66
Family number	-0.23	0.24	0.91	594	0.34	0.80	0.50	1.27
Delivery for other reasons	-0.17	0.20	0.78	594	0.38	0.84	0.57	1.23
Pregnancy induced diabetes	0.66	0.77	0.73	577	0.39	1.93	0.43	8.70
Pre-existing IDDM	0.77	0.92	0.70	592	0.40	2.15	0.36	12.99
Birth Order	0.18	0.21	0.68	593	0.41	1.19	0.79	1.81
Delivery for Hypertension or Pre-Eclampsia	-0.18	0.22	0.67	594	0.41	0.83	0.54	1.29
Pregnancy induced hypertension	0.17	0.22	0.64	588	0.42	1.19	0.78	1.82
Ruptured membranes > 24 hours	-0.14	0.19	0.54	585	0.46	0.87	0.60	1.26
Maternal Age	-0.010	0.014	0.48	592	0.49	0.99	0.96	1.02
Cephalic Presentation	-0.12	0.17	0.46	569	0.50	0.89	0.63	1.25
PaO2 at 2 hours	-0.014	0.02	0.38	590	0.54	0.99	0.94	1.03
Delivery for Chorioamnionitis	-0.18	0.29	0.37	594	0.54	0.84	0.47	1.49
Breech Presentation	0.10	0.18	0.32	569	0.57	1.11	0.78	1.58
Ruptured membranes > 22 hours	0.21	0.44	0.23	583	0.63	1.24	0.52	2.91
PaO2 at 12 hours	-0.013	0.03	0.21	575	0.64	0.99	0.93	1.04
PaCO2 at 2 hours	-0.021	0.05	0.19	591	0.67	0.98	0.89	1.08
Smoking in Pregnancy	-0.08	0.20	0.17	551	0.68	0.92	0.63	1.36
Antenatal transfer	-0.068	0.17	0.2	583	0.69	0.93	0.67	1.31
Antepartum haemorrhage	-0.071	0.31	0.05	588	0.817	0.93	0.51	1.70
Chorioamnionitis	0.039	0.27	0.02	572	0.89	1.04	0.61	1.78
Race - white or other	-0.022	0.23	0.009	592	0.92	0.98	0.63	1.53
Delivery for Antepartum Haemorrhage	-0.026	0.31	0.01	594	0.93	0.97	0.54	1.77
Pre-existing hypertension	-0.0063	0.44	0.0002	587	0.99	0.99	0.42	2.36

Notes

[1] Mean value substituted for missing values

[2] Defined as birthweight below 10th centile for gestational age

Table 16: Model for predicting death or severe disability at two years of age, Logistic regression coefficients and odds ratios for factors in the model

Variables	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Probability of Survival	-0.024	0.0051	0.98	0.96	0.99
Birthweight	-0.002	0.019	0.998	0.996	1.00
FiO2 at 12 hours with mean substituted for missing values	2.16	0.00016	8.6	2.8	26.4
FiO2 at 2 hours with mean substituted for missing values	1.11	0.022	3.03	1.17	7.86
Base Excess at 2 hours with mean substituted for missing values	-0.024	0.45	0.98	0.92	1.04
Apgar Score at 1 minute with mean substituted for missing values	-0.046	0.31	0.96	0.88	1.04
Volume expander received by 12 hours	-0.37	0.22	0.69	0.38	1.25
Base Excess at 12 hours with mean substituted for missing values	-0.070	0.056	0.93	0.87	1.002
pH at 2 hours with mean substituted for missing values	-1.16	0.34	0.31	0.029	3.43
Mean Blood Pressure at 12 hours with mean substituted for missing values	0.046	0.0057	1.05	1.014	1.08
pH at 12 hours with mean substituted for missing values	1.26	0.32	3.52	0.29	42.0
Time to intubation with mean substituted for missing values	-0.039	0.13	0.96	0.92	1.01
Male Gender	0.39	0.075	1.48	0.96	2.29
Multiple infant	-0.63	0.0074	0.53	0.34	0.85
Delivery for Intrauterine Growth Retardation	-0.35	0.32	0.71	0.36	1.40
Constant	0.084	0.99	1.09		

Figure 16: Process of Stepwise regression modeling for predicting death or severe disability at 2 years of age

STEP 1 Probability of Survival[a] n = 594

Logistic Regression Model

Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Probability of Survival[a]	-0.039	3.60E-20	0.962	0.954	0.970
Constant	1.73	6.09E-13	5.65		

Discrimination of Model Prediction (ROC Curve Analysis)

Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.729	0.021	1.70E-21	0.688	-	0.771

STEP 2 Birthweight added n = 594

Logistic Regression Model

Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Probability of Survival[a]	-0.028	8.761E-06	0.972	0.960	0.984
Birthweight	-0.0015	0.023	0.998	0.997	1.000
Constant	2.41	5.73E-10	11.1		

Discrimination of Model Prediction (ROC Curve Analysis)

Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.738	0.021	4.62E-23	0.698	-	0.779

STEP 3 Gestational Age (days) added n = 594

Logistic Regression Model

Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Probability of Survival[a]	-0.034	0.019	0.967	0.940	0.994
Birthweight	-0.0014	0.082	0.999	0.997	1.0002
Gestational Age (days)	0.010	0.66	1.01	0.965	1.058
Constant	0.65	0.87	1.9		

Discrimination of Model Prediction (ROC Curve Analysis)

Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.739	0.021	4.24E-23	0.698	-	0.779

STEP 4 Gestational Age (days) removed n = 594

See Result of Step 2

STEP 5 FIO2 at 12 hours[b] added n = 594

Logistic Regression Model

Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Probability of Survival[a]	-0.024	0.00022	0.976	0.964	0.989
Birthweight	-0.0019	0.0067	0.998	0.997	0.999
FIO2 at 12 hours[b]	2.51	3.78E-07	12.3	4.7	32.5
Constant	1.53	0.00031	4.6		

Discrimination of Model Prediction (ROC Curve Analysis)

Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.762	0.02	1.55E-27	0.723	-	0.801

Continued overleaf

STEP 6 Inotrope received by 12 hours added — n = 592

Logistic Regression Model					
Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Probability of Survival[a]	-0.024	0.00024	0.976	0.96399918	0.989
Birthweight	-0.0019	0.0069	0.998	0.997	0.999
FiO2 at 12 hours[b]	2.48	1.18E-06	11.9	4.4	32.3
Inotrope received by 12 hours	-0.064	0.76	0.94	0.62	1.41
Constant	1.65	0.0018	5.19		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.764	0.020	6.73E-26	0.726	-	0.803

STEP 7 FiO2 at 2 hours[b] added — n = 592

Logistic Regression Model					
Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Probability of Survival[a]	-0.023	0.00050	0.977	0.965	0.990
Birthweight	-0.0020	0.0056	1.00	0.997	0.999
FiO2 at 12 hours[b]	2.00	0.00015	7.4	2.6	20.8
Inotrope received by 12 hours	-0.031	0.88	0.97	0.64	1.46
FiO2 at 2 hours[b]	1.39	0.0014	4.03	1.72	9.44
Constant	1.11	0.045	3.04		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.773	0.019	1.37E-29	0.735	-	0.811

STEP 8 Inotrope received by 12 hours removed — n = 594

Logistic Regression Model					
Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Probability of Survival[a]	-0.023	0.00049	0.977	0.965	0.990
Birthweight	-0.0020	0.0056	0.9980462	0.997	0.999
FiO2 at 12 hours[b]	2.00	9.97E-05	7.4	2.7	20.1
FiO2 at 2 hours[b]	1.46	0.00068	4.3	1.9	10.1
Constant	1.025	0.022	2.79		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.772	0.019	1.66E-29	0.734	-	0.810

STEP 9 Head Circumference[b] added — n = 594

Logistic Regression Model					
Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Probability of Survival[a]	-0.023	0.00056	0.977	0.965	0.990
Birthweight	-0.0020	0.0072	0.998	0.997	0.999
FiO2 at 12 hours[b]	2.00	9.96E-05	7.4	2.7	20.1
FiO2 at 2 hours[b]	1.46	0.00074	4.3	1.8	10.1
Head Circumference[b]	0.0087	0.91	1.01	0.87	1.17
Constant	0.84	0.61	2.32		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.772	0.019	1.42E-29	0.7340428	-	0.810

Continued overleaf

STEP 10 Head Circumference[b] removed — n = 594

See Result of Step 8

STEP 11 Base Excess at 2 hours[b] added — n = 594

Logistic Regression Model					
Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Probability of Survival[a]	-0.023	0.00047	0.977	0.965	0.990
Birthweight	-0.0018	0.013	0.998	0.997	1.000
FiO2 at 12 hours[b]	1.99	0.00013	7.3	2.6	20.3
FiO2 at 2 hours[b]	1.32	0.0027	3.7	1.6	8.8
Base Excess at 2 hours[b]	-0.07	0.0050	0.93	0.89	0.98
Constant	0.61	0.19	1.85		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.778	0.019	6.99E-31	0.741	-	0.815

STEP 12 Apgar Score at 5 mins[b] added — n = 594

Logistic Regression Model					
Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Probability of Survival[a]	-0.021	0.0015	0.979	0.966	0.992
Birthweight	-0.0019	0.0097	0.998	0.997	1.000
FiO2 at 12 hours[b]	1.97	0.00016	7.1	2.6	19.8
FiO2 at 2 hours[b]	1.26	0.0044	3.5	1.5	8.4
Base Excess at 2 hours[b]	-0.062	0.018	0.94	0.89	0.99
Apgar Score at 5 mins[b]	-0.065	0.27	0.94	0.84	1.05
Constant	1.18	0.091	3.25		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.779	0.019	5.48E-31	0.741	-	0.817

STEP 13 Apgar Score at 1 min[b] added — n = 594

Logistic Regression Model					
Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Probability of Survival[a]	-0.021	0.0019	0.979	0.966	0.992
Birthweight	-0.0019	0.0091	0.998	0.997	1.000
FiO2 at 12 hours[b]	1.98	0.00015	7.2	2.6	20.1
FiO2 at 2 hours[b]	1.26	0.0045	3.5	1.5	8.4
Base Excess at 2 hours[b]	-0.062	0.019	0.94	0.89	0.99
Apgar Score at 5 mins[b]	-0.033	0.66	0.97	0.83	1.12
Apgar Score at 1 min[b]	-0.036	0.51	0.96	0.87	1.06
Constant	1.12	0.11	3.06		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.779	0.019	4.49E-31	0.742	-	0.817

Continued overleaf

STEP 14 Volume expander received by 12 hours added — n = 594

Logistic Regression Model					
Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Probability of Survival[a]	-0.021	0.0020	0.979	0.966	0.992
Birthweight	-0.0018	0.012	0.998	0.997	1.000
FiO2 at 12 hours[b]	1.94	0.00020	7.0	2.5	19.3
FiO2 at 2 hours[b]	1.16	0.0084	3.2	1.3	7.7
Base Excess at 2 hours[b]	-0.058	0.029	0.94	0.90	0.99
Apgar Score at 5 mins[b]	-0.032	0.67	0.97	0.83	1.12
Apgar Score at 1 min[b]	-0.035	0.53	0.97	0.87	1.08
Volume expander by 12 hours	-0.31	0.29	0.73	0.41	1.30
Constant	1.48	0.059	4.41		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.781	0.019	2.20E-31	0.744	-	0.819

STEP 15 Base Excess at 12 hours[b] added — n = 593

Logistic Regression Model					
Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Probability of Survival[a]	-0.021	0.0019	0.979	0.966	0.992
Birthweight	-0.0018	0.012	0.998	0.997	1.000
FiO2 at 12 hours[b]	1.84	0.00050	6.3	2.2	17.7
FiO2 at 2 hours[b]	1.20	0.0074	3.3	1.4	8.0
Base Excess at 2 hours[b]	-0.042	0.14	0.96	0.91	1.01
Apgar Score at 5 mins[b]	-0.035	0.64	0.97	0.83	1.12
Apgar Score at 1 min[b]	-0.040	0.47	0.96	0.86	1.07
Volume expander by 12 hours	-0.28	0.34	0.76	0.42	1.35
Base Excess at 12 hours[b]	-0.050	0.13	0.95	0.89	1.01
Constant	1.38	0.060	3.96		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
				-	

STEP 16 Serum pH at 2 hours[b] added — n = 593

Logistic Regression Model					
Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Probability of Survival[a]	-0.021	0.0025	0.979	0.966	0.993
Birthweight	-0.0019	0.011	0.998	0.997	1.000
FiO2 at 12 hours[b]	1.84	0.00050	6.3	2.2	17.7
FiO2 at 2 hours[b]	1.13	0.017	3.1	1.2	7.9
Base Excess at 2 hours[b]	-0.036	0.24	0.96	0.91	1.03
Apgar Score at 5 mins[b]	-0.034	0.66	0.97	0.83	1.12
Apgar Score at 1 min[b]	-0.039	0.49	0.96	0.86	1.07
Volume expander by 12 hours	-0.28	0.35	0.76	0.43	1.35
Base Excess at 12 hours[b]	-0.050	0.13	0.95	0.89	1.01
Serum pH at 2 hours[b]	-0.51	0.67	0.60	0.059	6.13
Constant	5.20	0.56	181.7		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.783	0.019	1.08E-31	0.745	-	0.820

Continued overleaf

STEP 17 Apgar Score at 10 mins[b] added — n = 593

Logistic Regression Model					
Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Probability of Survival[a]	-0.021	0.0026	0.980	0.967	0.993
Birthweight	-0.0019	0.0097	0.998	0.997	1.000
FiO2 at 12 hours[b]	1.83	0.00052	6.3	2.2	17.6
FiO2 at 2 hours[b]	1.13	0.017	3.1	1.2	7.8
Base Excess at 2 hours[b]	-0.037	0.24	0.96	0.91	1.02
Apgar Score at 5 mins[b]	-0.0025	0.98	1.00	0.84	1.18
Apgar Score at 1 min[b]	-0.042	0.46	0.96	0.86	1.07
Volume expander by 12 hours	-0.29	0.33	0.75	0.42	1.34
Base Excess at 12 hours[b]	-0.050	0.13	0.95	0.89	1.014
Serum pH at 2 hours[b]	-0.46	0.70	0.63	0.061	6.5
Apgar Score at 10 mins[b]	-0.070	0.46	0.93	0.78	1.12
Constant	5.19	0.56	179.5		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.783	0.019	1.04E-31	0.745	-	0.820

STEP 18 Apgar Score at 5 mins[b] removed — n = 593

Logistic Regression Model					
Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Probability of Survival[a]	-0.021	0.0026	0.980	0.967	0.993
Birthweight	-0.0019	0.0097	0.998	0.997	1.000
FiO2 at 12 hours[b]	1.83	0.00051	6.3	2.2	17.6
FiO2 at 2 hours[b]	1.13	0.017	3.1	1.2	7.8
Base Excess at 2 hours[b]	-0.037	0.24	0.96	0.91	1.02
Apgar Score at 1 min[b]	-0.043	0.35	0.96	0.88	1.05
Volume expander by 12 hours	-0.28	0.33	0.75	0.42	1.34
Base Excess at 12 hours[b]	-0.050	0.13	0.95	0.89	1.01
Serum pH at 2 hours[b]	-0.46	0.70	0.63	0.06	6.5
Apgar Score at 10 mins[b]	-0.071	0.39	0.93	0.79	1.09
Constant	5.19	0.56	179.5		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.783	0.019	1.02E-31	0.745	-	0.820

STEP 19 Heart Rate Score at 5 mins[b] added — n = 593

Logistic Regression Model					
Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Probability of Survival[a]	-0.021	0.0026	0.980	0.967	0.993
Birthweight	-0.0019	0.0097	0.998	0.997	1.000
FiO2 at 12 hours[b]	1.84	0.00057	6.3	2.2	17.8
FiO2 at 2 hours[b]	1.13	0.017	3.1	1.2	7.8
Base Excess at 2 hours[b]	-0.037	0.24	0.96	0.91	1.02
Apgar Score at 1 min[b]	-0.043	0.36	0.96	0.87	1.05
Volume expander by 12 hours	-0.28	0.34	0.75	0.42	1.34
Base Excess at 12 hours[b]	-0.050	0.13	0.95	0.89	1.01
Serum pH at 2 hours[b]	-0.46	0.70	0.63	0.06	6.6
Apgar Score at 10 mins[b]	-0.071	0.39	0.93	0.79	1.10
Heart Rate Score at 5 mins[b]	0.0041	0.99	1.004	0.59	1.70
Constant	5.17	0.57	175.9		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.783	0.019	9.75E-32	0.745	-	0.820

Continued overleaf

STEP 20 Heart Rate Score at 5 mins[b] removed — n = 593

See Result of Step 18

STEP 21 Cardiac Massage or Adrenaline Resuscitation added n = 584

Logistic Regression Model					
Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Probability of Survival[a]	-0.020	0.0038	0.980	0.967	0.994
Birthweight	-0.0019	0.010	0.998	0.997	1.000
FiO2 at 12 hours[b]	1.79	0.00070	6.0	2.1	16.8
FiO2 at 2 hours[b]	1.035	0.029	2.8	1.1	7.2
Base Excess at 2 hours[b]	-0.034	0.27	0.97	0.91	1.03
Apgar Score at 1 min[b]	-0.040	0.40	0.96	0.88	1.05
Volume expander by 12 hours	-0.33	0.26	0.72	0.40	1.30
Base Excess at 12 hours[b]	-0.046	0.16	0.96	0.90	1.02
Serum pH at 2 hours[b]	-0.49	0.58	0.61	0.06	6.34
Apgar Score at 10 mins[b]	-0.075	0.37	0.93	0.79	1.09
Cardiac Massage or Adrenaline	0.11	0.76	1.12	0.54	2.30
Constant	5.5	0.54	256.3		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.779	0.019	2.11E-30	0.741	-	0.817

STEP 22 Cardiac Massage or Adrenaline Resuscitation removed n = 593

See Result of Step 18

STEP 23 Mean Blood Pressure at 2 hours[b] added — n = 593

Logistic Regression Model					
Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Probability of Survival[a]	-0.021	0.0024	0.979	0.966	0.993
Birthweight	-0.0019	0.0098	0.998	0.997	1.000
FiO2 at 12 hours[b]	1.86	0.00047	6.4	2.3	18.2
FiO2 at 2 hours[b]	1.13	0.017	3.1060135	1.2278731	7.85683558
Base Excess at 2 hours[b]	-0.037	0.24	0.964	0.907	1.024
Apgar Score at 1 min[b]	-0.044	0.33	0.957	0.875	1.047
Volume expander by 12 hours	-0.28	0.33	0.75	0.42	1.34
Base Excess at 12 hours[b]	-0.050	0.12	0.95	0.89	1.01
Serum pH at 2 hours[b]	-0.47	0.70	0.63	0.061	6.46
Apgar Score at 10 mins[b]	-0.057	0.42	0.94	0.80	1.10
Mean Blood Pressure at 2 hours[b]	0.0056	0.69	1.01	0.98	1.03
Constant	5.1	0.57	157.4		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.784	0.019	5.57E-32	0.747	-	0.821

Continued overleaf

STEP 24 Small for Gestational Age (<10thc) added — n = 593

Logistic Regression Model					
Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Probability of Survival[a]	-0.021	0.015	0.979	0.963	0.996
Birthweight	-0.0020	0.059	0.998	0.996	1.00
FiO2 at 12 hours[b]	1.86	0.00048	6.4	2.3	18.2
FiO2 at 2 hours[b]	1.13	0.017	3.1	1.2	7.9
Base Excess at 2 hours[b]	-0.037	0.24	0.96	0.91	1.02
Apgar Score at 1 min[b]	-0.044	0.33	0.96	0.87	1.05
Volume expander by 12 hours	-0.28	0.33	0.75	0.42	1.34
Base Excess at 12 hours[b]	-0.050	0.12	0.95	0.89	1.01
Serum pH at 2 hours[b]	-0.46	0.70	0.6300364	0.061	6.54
Apgar Score at 10 mins[b]	-0.067	0.42	0.94	0.79	1.10
Mean Blood Pressure at 2 hours[b]	0.0056	0.69	1.0056	0.979	1.033
Small for Gestational Age (<10thc)	-0.013	0.97	0.99	0.51	1.93
Constant	5.0	0.57	155.3		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.784	0.019	5.7E-32	0.747	-	0.821

STEP 25 Small for Gestational Age (<10thc) removed — n = 593

See Result of Step 18

STEP 26 Mean Blood Pressure at 12 hours[b] added — n = 593

Logistic Regression Model					
Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Probability of Survival[a]	-0.025	0.00056	0.976	0.962	0.989
Birthweight	-0.0019	0.011	0.998	0.997	1.000
FiO2 at 12 hours[b]	1.93	0.00032	6.9	2.4	19.9
FiO2 at 2 hours[b]	1.19	0.012	3.3	1.3	8.4
Base Excess at 2 hours[b]	-0.032	0.30	0.968	0.910	1.029
Apgar Score at 1 min[b]	-0.041	0.38	0.96	0.88	1.05
Volume expander by 12 hours	-0.40	0.19	0.67	0.37	1.21
Base Excess at 12 hours[b]	-0.053	0.11	0.95	0.89	1.01
Serum pH at 2 hours[b]	-0.43	0.72	0.65	0.063	6.8
Apgar Score at 10 mins[b]	-0.058	0.49	0.94	0.80	1.11
Mean Blood Pressure at 2 hours[b]	-0.0047	0.75	0.995	0.967	1.024
Mean Blood Pressure at 12 hours[b]	0.040	0.020	1.04	1.01	1.08
Constant	4.019	0.55	55.7		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.787	0.019	1.06E-32	0.750	-	0.824

Continued overleaf

STEP 27 Serum pH at 12 hours[b] added — n = 593

Logistic Regression Model					
Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Probability of Survival[a]	-0.025	0.00048	0.975	0.961	0.988
Birthweight	-0.0019	0.013	0.998	0.997	1.000
FiO2 at 12 hours[b]	2.03	0.00031	7.6	2.5	22.9
FiO2 at 2 hours[b]	1.16	0.015	3.2	1.3	8.2
Base Excess at 2 hours[b]	-0.033	0.29	0.97	0.91	1.03
Apgar Score at 1 min[b]	-0.039	0.40	0.96	0.88	1.05
Volume expander by 12 hours	-0.40	0.19	0.67	0.37	1.21
Base Excess at 12 hours[b]	-0.061	0.089	0.94	0.88	1.01
Serum pH at 2 hours[b]	-0.46	0.70	0.63	0.061	6.53
Apgar Score at 10 mins[b]	-0.055	0.51	0.95	0.80	1.11
Mean Blood Pressure at 2 hours[b]	-0.0046	0.75	1.00	0.97	1.02
Mean Blood Pressure at 12 hours[b]	0.040	0.020	1.04	1.01	1.08
Serum pH at 12 hours[b]	0.71	0.57	2.04	0.18	23.5
Constant	-1.075	0.93	0.34		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.787	0.019	1.01E-32	0.750	-	0.824

STEP 28 Time to Intubation[b] added — n = 593

Logistic Regression Model					
Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Probability of Survival[a]	-0.024	0.0010	0.976	0.962	0.990
Birthweight	-0.0020	0.0097	0.998	0.997	1.000
FiO2 at 12 hours[b]	2.06	0.00026	7.9	2.6	23.7
FiO2 at 2 hours[b]	1.18	0.014	3.2	1.3	8.3
Base Excess at 2 hours[b]	-0.032	0.31	0.97	0.91	1.03
Apgar Score at 1 min[b]	-0.028	0.54	0.97	0.89	1.06
Volume expander by 12 hours	-0.42	0.16	0.66	0.36	1.18
Base Excess at 12 hours[b]	-0.060	0.093	0.94	0.88	1.01
Serum pH at 2 hours[b]	-0.51	0.67	0.60	0.06	6.3
Apgar Score at 10 mins[b]	-0.062	0.45	0.94	0.80	1.10
Mean Blood Pressure at 2 hours[b]	-0.0050	0.73	0.99	0.97	1.02
Mean Blood Pressure at 12 hours[b]	0.041	0.018	1.04	1.007	1.078
Serum pH at 12 hours[b]	0.83	0.51	2.3	0.2	26.5
Time to Intubation[b]	-0.038	0.14	0.96	0.92	1.01
Constant	-1.41	0.91	0.24		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.790	0.019	2.66E-33	0.753	-	0.827

Continued overleaf

STEP 29 Male Gender added — n = 593

Logistic Regression Model					
Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Probability of Survival[a]	-0.018	0.019	0.982	0.967	0.997
Birthweight	-0.0026	0.0018	0.997	0.996	0.999
FiO2 at 12 hours[b]	2.077	0.00024	8.0	2.6	24.2
FiO2 at 2 hours[b]	1.18	0.014	3.2	1.3	8.3
Base Excess at 2 hours[b]	-0.025	0.43	0.98	0.92	1.04
Apgar Score at 1 min[b]	-0.031	0.51	0.97	0.89	1.06
Volume expander by 12 hours	-0.40	0.18	0.67	0.37	1.21
Base Excess at 12 hours[b]	-0.059	0.10	0.94	0.88	1.01
Serum pH at 2 hours[b]	-0.79	0.51	0.45	0.04	4.8
Apgar Score at 10 mins[b]	-0.065	0.43	0.94	0.80	1.10
Mean Blood Pressure at 2 hours[b]	-0.0025	0.86	1.00	0.97	1.03
Mean Blood Pressure at 12 hours[b]	0.041	0.017	1.04	1.01	1.08
Serum pH at 12 hours[b]	0.82	0.51	2.3	0.2	26.1
Time to Intubation[b]	-0.042	0.10	0.96	0.91	1.01
Male Gender	0.42	0.053	1.53	1.00	2.34
Constant	0.75	0.96	2.13		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.791	0.019	1.37E-33	0.755	-	0.828

STEP 30 Volume expander resuscitation added — n = 593

Logistic Regression Model					
Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Probability of Survival[a]	-0.018	0.021	0.9819275	0.96686981	0.99721977
Birthweight	-0.0027	0.0017	0.9973319	0.99567023	0.99899641
FiO2 at 12 hours[b]	2.11	0.00020	8.2572998	2.71745805	25.090728
FiO2 at 2 hours[b]	1.06	0.028	2.8905275	1.12068192	7.45541549
Base Excess at 2 hours[b]	-0.024	0.45	0.9766137	0.91820747	1.03873501
Apgar Score at 1 min[b]	-0.029	0.54	0.9715943	0.88588091	1.06560067
Volume expander by 12 hours	-0.44	0.16	0.646391	0.35381563	1.1809011
Base Excess at 12 hours[b]	-0.055	0.12	0.9460777	0.88182095	1.01501675
Serum pH at 2 hours[b]	-0.74	0.54	0.4776224	0.04462719	5.11175327
Apgar Score at 10 mins[b]	-0.07	0.43	0.935744	0.79394154	1.10287327
Mean Blood Pressure at 2 hours[b]	-0.0015	0.92	0.9984641	0.97048649	1.02724828
Mean Blood Pressure at 12 hours[b]	0.042	0.015	1.0425792	1.00805043	1.07829072
Serum pH at 12 hours[b]	0.79	0.53	2.2043207	0.18795956	25.8514631
Time to Intubation[b]	-0.039	0.12	0.9617093	0.9153555	1.01041052
Male Gender	0.41	0.060	1.5125014	0.98332729	2.32644865
Volume expander resuscitation	0.54	0.42	1.7140904	0.45996973	6.38760736
Constant	0.56	0.96	1.7562556		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.790	0.019	4.39E-33	0.754	-	0.827

Continued overleaf

STEP 31 Volume expander resuscitation removed — n = 593

See Result of Step 29

STEP 32 Multiple Pregnancy added — n = 593

Logistic Regression Model					
Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Probability of Survival[a]	-0.020	0.011	0.980	0.965	0.995
Birthweight	-0.0026	0.0020	0.997	0.996	0.999
FiO2 at 12 hours[b]	2.13	0.00020	8.4	2.7	26.0
FiO2 at 2 hours[b]	1.11	0.022	3.0	1.2	7.9
Base Excess at 2 hours[b]	-0.024	0.45	0.98	0.92	1.04
Apgar Score at 1 min[b]	-0.036	0.44	0.96	0.88	1.06
Volume expander by 12 hours	-0.36	0.22	0.69	0.38	1.25
Base Excess at 12 hours[b]	-0.068	0.062	0.93	0.87	1.00
Serum pH at 2 hours[b]	-1.04	0.39	0.35	0.032	3.85
Apgar Score at 10 mins[b]	-0.046	0.58	0.96	0.81	1.12
Mean Blood Pressure at 2 hours[b]	-0.0022	0.88	1.00	0.97	1.03
Mean Blood Pressure at 12 hours[b]	0.047	0.0069	1.05	1.01	1.09
Serum pH at 12 hours[b]	1.14	0.37	3.1	0.3	37.0
Time to Intubation[b]	-0.039	0.13	0.96	0.91	1.01
Male Gender	0.43	0.048	1.54	1.00	2.37
Multiple Infant	-0.61	0.0099	0.55	0.34	0.86
Constant	0.33	0.96	1.39		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.798	0.018	5.43E-35	0.762	-	0.834

STEP 33 Normal Delivery added — n = 593

Logistic Regression Model					
Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Probability of Survival[a]	-0.019	0.045	0.982	0.964	1.000
Birthweight	-0.0028	0.0034	0.997	0.995	0.999
FiO2 at 12 hours[b]	2.13	0.00021	8.4	2.7	25.9
FiO2 at 2 hours[b]	1.12	0.022	3.1	1.2	7.9
Base Excess at 2 hours[b]	-0.023	0.46	0.98	0.92	1.04
Apgar Score at 1 min[b]	-0.037	0.43	0.96	0.88	1.06
Volume expander by 12 hours	-0.39	0.21	0.68	0.37	1.24
Base Excess at 12 hours[b]	-0.069	0.060	0.93	0.87	1.00
Serum pH at 2 hours[b]	-1.04	0.39	0.35	0.03	3.9
Apgar Score at 10 mins[b]	-0.048	0.57	0.95	0.81	1.12
Mean Blood Pressure at 2 hours[b]	-0.0022	0.88	1.00	0.97	1.03
Mean Blood Pressure at 12 hours[b]	0.048	0.0066	1.05	1.01	1.09
Serum pH at 12 hours[b]	1.11	0.38	3.0	0.3	36.3
Time to Intubation[b]	-0.039	0.13	0.96	0.91	1.01
Male Gender	0.44	0.046	1.56	1.01	2.41
Multiple Infant	-0.61	0.0094	0.54	0.34	0.86
Normal Delivery	0.073	0.77	1.08	0.66	1.75
Constant	0.52	0.97	1.68		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.798	0.018	5.40E-35	0.762	-	0.834

Continued overleaf

STEP 34 Delivery for Intrauterine Growth Retardation added — n = 593

Logistic Regression Model					
Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Probability of Survival[a]	-0.021	0.028	0.979	0.960	0.998
Birthweight	-0.0024	0.018	0.998	0.996	1.000
FiO2 at 12 hours[b]	2.12	0.00022	8.4	2.7	25.7
FiO2 at 2 hours[b]	1.11	0.023	3.0	1.2	7.8
Base Excess at 2 hours[b]	-0.022	0.49	0.98	0.92	1.04
Apgar Score at 1 min[b]	-0.039	0.41	0.96	0.88	1.05
Volume expander by 12 hours	-0.40	0.20	0.67	0.36	1.23
Base Excess at 12 hours[b]	-0.070	0.056	0.93	0.87	1.00
Serum pH at 2 hours[b]	-1.12	0.36	0.33	0.03	3.6
Apgar Score at 10 mins[b]	-0.046	0.58	0.96	0.81	1.12
Mean Blood Pressure at 2 hours[b]	-0.0028	0.85	1.00	0.97	1.03
Mean Blood Pressure at 12 hours[b]	0.047	0.0078	1.05	1.01	1.09
Serum pH at 12 hours[b]	1.19	0.35	3.28	0.27	39.5
Time to Intubation[b]	-0.040	0.12	0.96	0.91	1.01
Male Gender	0.41	0.071	1.50	0.97	2.34
Multiple Infant	-0.63	0.0061	0.53	0.34	0.85
Normal Delivery	0.097	0.69	1.10	0.68	1.79
Delivery for Intrauterine Growth Retardation	-0.36	0.31	0.70	0.35	1.39
Constant	0.76	0.96	2.13		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.796	0.018	4.64E-35	0.762	-	0.834

STEP 35 Mean Blood Pressure at 2 hours[b] removed — n = 593

Logistic Regression Model					
Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Probability of Survival[a]	-0.021	0.927	0.979	0.960	0.998
Birthweight	-0.0024	0.018	0.998	0.996	1.000
FiO2 at 12 hours[b]	2.13	0.00019	8.4	2.8	25.9
FiO2 at 2 hours[b]	1.11	0.023	3.0	1.2	7.8
Base Excess at 2 hours[b]	-0.022	0.49	0.98	0.92	1.04
Apgar Score at 1 min[b]	-0.040	0.40	0.96	0.88	1.05
Volume expander by 12 hours	-0.40	0.20	0.67	0.36	1.23
Base Excess at 12 hours[b]	-0.070	0.055	0.93	0.87	1.00
Serum pH at 2 hours[b]	-1.12	0.36	0.33	0.030	3.6
Apgar Score at 10 mins[b]	-0.044	0.59	0.96	0.81	1.13
Mean Blood Pressure at 12 hours[b]	0.046	0.0065	1.05	1.01	1.08
Serum pH at 12 hours[b]	1.19	0.35	3.3	0.3	39.6
Time to Intubation[b]	-0.040	0.12	0.96	0.91	1.01
Male Gender	0.41	0.067	1.51	0.97	2.34
Multiple Infant	-0.63	0.0061	0.53	0.34	0.85
Normal Delivery	0.097	0.70	1.10	0.68	1.79
Delivery for Intrauterine Growth Retardation	-0.35	0.31	0.70	0.35	1.40
Constant	0.70	0.96	2.01		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.796	0.018	4.45E-35	0.7619729	-	0.834

Continued overleaf

STEP 36 Normal Delivery removed — n = 593

Logistic Regression Model					
Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Probability of Survival[a]	-0.023	0.0066	0.977	0.961	0.994
Birthweight	-0.0022	0.017	0.998	0.996	1.000
FIO2 at 12 hours[b]	2.14	0.00018	8.5	2.8	26.0
FIO2 at 2 hours[b]	1.10	0.024	3.0	1.2	7.8
Base Excess at 2 hours[b]	-0.023	0.47	0.98	0.92	1.04
Apgar Score at 1 min[b]	-0.039	0.41	0.96	0.88	1.05
Volume expander by 12 hours	-0.38	0.22	0.69	0.38	1.25
Base Excess at 12 hours[b]	-0.069	0.058	0.93	0.87	1.00
Serum pH at 2 hours[b]	-1.12	0.36	0.33	0.03	3.59
Apgar Score at 10 mins[b]	-0.042	0.61	0.96	0.82	1.13
Mean Blood Pressure at 12 hours[b]	0.045	0.0070	1.05	1.01	1.08
Serum pH at 12 hours[b]	1.22	0.34	3.4	0.3	40.6
Time to Intubation[b]	-0.039	0.12	0.96	0.91	1.01
Male Gender	0.40	0.073	1.49	0.96	2.30
Multiple Infant	-0.62	0.0067	0.54	0.34	0.86
Delivery for Intrauterine Growth Retardation	-0.34	0.33	0.71	0.36	1.41
Constant	0.44	0.97	1.55		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.798	0.018	4.06E-35	0.762	-	0.834

STEP 37 Apgar Score at 10 mins[b] removed — n = 593

Logistic Regression Model					
Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Probability of Survival[a]	-0.024	0.0051	0.976	0.960	0.993
Birthweight	-0.0022	0.019	0.998	0.996	1.000
FIO2 at 12 hours[b]	2.2	0.00016	8.6	2.8	26.4
FIO2 at 2 hours[b]	1.1	0.022	3.0	1.2	7.9
Base Excess at 2 hours[b]	-0.024	0.45	0.98	0.92	1.04
Apgar Score at 1 min[b]	-0.046	0.31	0.96	0.88	1.04
Volume expander by 12 hours	-0.37	0.22	0.69	0.38	1.25
Base Excess at 12 hours[b]	-0.070	0.056	0.93	0.87	1.00
Serum pH at 2 hours[b]	-1.16	0.34	0.3	0.03	3.4
Mean Blood Pressure at 12 hours[b]	0.046	0.0057	1.05	1.01	1.08
Serum pH at 12 hours[b]	1.26	0.32	3.5	0.3	42.0
Time to Intubation[b]	-0.039	0.13	0.96	0.92	1.01
Male Gender	0.39	0.075	1.48	0.96	2.29
Multiple Infant	-0.63	0.0074	0.53	0.34	0.85
Delivery for Intrauterine Growth Retardation	-0.35	0.32	0.71	0.36	1.40
Constant	0.084	0.99	1.09		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.798	0.018	3.96E-35	0.762	-	0.835

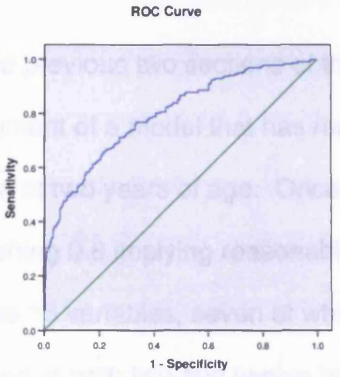
Notes

a: Probability of survival obtained from 'Draper Grid'

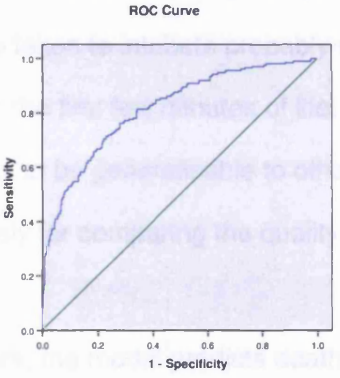
b: Mean value substituted for missing values

Figure 17: ROC Curves of Predictive Ability for the Different Outcomes (UKOS)

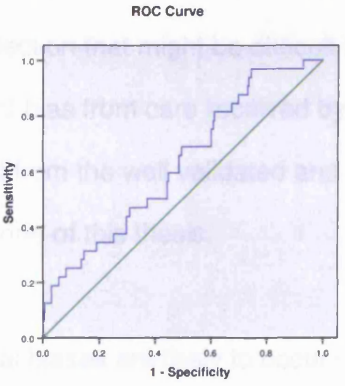
A. Severe disability or Death v Other **(Az = 0.798, 95% C.I. 0.762 – 0.835)**



B. Death v Other **(Az = 0.816, 95% C.I. 0.780 – 0.852)**



C. Severe disability v Other or No disability **(Az = 0.627, 95% C.I. 0.529 – 0.724)**



3.3.4 Discussion

As in the previous two sections of this chapter, this section also depicts the development of a model that has reasonable predictive ability for the combined outcome at two years of age. Once again, the area under the ROC curve is approaching 0.8 implying reasonable discrimination[201]. The model does however comprise 15 variables, seven of which are physiological including three variables measured at both two and twelve hours of age. These physiological data points are very likely to be affected by the quality of care received by the infants. In addition, the time taken to intubate probably represents the quality of care provided to the infant in the first few minutes of life. These considerations indicate that the model is not likely to be generalisable to other populations and could only be used very cautiously for comparing the quality of care in the trial centres.

As before, the model predicts death much better than it predicts disability, although with an area under the curve of 0.816 this is less dramatic than the investigations detailed earlier in this chapter. This model is however large and requires significant data collection that might be difficult to achieve and is liable to be affected by treatment bias from care received by the infants. It therefore offers little beyond that obtained from the well validated and frequently used scoring systems described in chapter one of this thesis.

Additional biases are likely to occur in a trial scenario and chapter two documents concerns about this trial cohort in comparison to the population based cohorts. Of particular concern is selection bias which can be introduced by the use of inclusion and exclusion criteria as well as researchers' attitudes to trials. The units involved in the trial were all tertiary referral units who were capable of providing the intervention,

namely high frequency oscillatory ventilation. Selection bias is therefore a problem, as some infants were referred from other hospitals for tertiary care. This transfer of care (if the infant was not delivered before the in-utero transfer was completed), and the selection criteria of the trial, will have caused selection bias. Although there was a suggestion that the 28 week infants were not representative of a normal group of infants (see chapter two), these infants were left in the model in order to maintain the numbers and the potential future use of the model. Additionally, some units or clinicians may have been more enthusiastic than others about the trial and recruited a wider variety of patients. They might also have included infants that were less sick than other centres. Therefore comparing centres using a model derived in a trial setting could be misleading. These factors that introduce selection bias weaken the possible generalisability of the model[166].

Once again the inability of the model to predict neurodevelopmental outcome may be partly due to the proportion of cases dying in the cohort. In this trial, of the 594 infants included in the predictive models, 212 died and 32 were assessed as being severely disabled at two years of age. Of the three cohorts studied in this thesis, mortality was highest in the UKOS trial due to the extreme prematurity of the infants and the requirement of the infants to be ventilated. As discussed earlier, developing the model from a cohort comprised of a wider gestational mix might improve this prediction, as well as extending the future applicability of the model.

Due to concerns about inappropriate classification of outcome, a predefined time 'window' was applied to when the infants should be assessed. Although data were available from outside of this time frame, such data were disregarded and the infants excluded from the analysis. This meant that the follow up rate was only 74.5% with the outcome of 203 infants being unknown between 22–28 months. In addition the

comparison of included and excluded infants (see Table 14) indicates that the excluded infants were of heavier birthweight, had a higher Apgar Score at 1 and 5 minutes, and a higher base excess at 2 hours of age. This is most likely due to the fact that the outcome was known for all infants who died and they were all included in the model, whereas the 213 infants with unknown outcome were all survivors.

Once again, even for the infants that were included, a major weakness of the study was missing data: this affected all but 340 infants. In order to include all infants and variables in the analysis, mean values were used in substitution for missing continuous values. This weakened the association of the variables with the combined outcome (see Table 15) and is likely to have reduced the discriminatory ability of the model. Limiting the model to just infants on whom full data was available, would have introduced an unacceptable risk of bias[157]. It remains a major concern that the final model contained 9 variables using mean values for missing data. If the model was being used in other cohorts this would lead to significant concerns about it's generalisability and merit substantial validation in at least one similar cohort with full data ascertainment[190, 214]. As with the model derived in section 2 of this chapter, it is clear however that this model does not predict the combined outcome and it will not be used in other studies: substituted values have therefore been left in the model.

As in sections 3.1 and 3.2 the inclusion of other predictive variables might have improved the model. As mentioned in section 3.2.1, variables such as socioeconomic status and antenatal infection also warrant further investigation. Alternatively it might be better to use models that predict death and disability separately, as it is unlikely that the factors which influence development are the same as those which predict death. In a similar way, it might be possible to predict

more accurately the individual disabilities that make up severe disability[194, 215] such as severe hearing loss, blindness or inability to feed.

3.3.5 Comments and implications

This investigation provides further evidence that predicting a combined outcome of death and disability is difficult. It remains unclear whether this is due to lack of variables in the model or whether later insults, complications, or environmental factors are more important for long term outcome than inherent risk at birth. As mentioned in the previous sections, to determine whether these factors are more important predictors, it would seem best to develop a new cohort and collect other data that could potentially predict developmental outcome. In addition it may also be advisable to look at individual disabilities separately.

Whilst it is also hampered by possible selection bias, the UKOS trial dataset contains considerable information on later complications and was therefore analysed to determine the relative importance of early and late factors in the chapter that follows.

CHAPTER FOUR: PREDICTING NEURODEVELOPMENTAL OUTCOME USING DATA FROM THE WHOLE NEONATAL ADMISSION: UNITED KINGDOM OSCILLATION STUDY.

4.1 INTRODUCTION

As discussed in the previous chapter, two remaining questions are left following the attempts to predict outcome at two years of age using data available in the first 12 hours of an infant's life.

The models described in chapter three all predict death much better than disability, suggesting that including deaths in the model development might be a flawed approach, preventing the model from predicting disability. This could be due either to there being many more deaths than disabled infants in the cohorts, or that the factors causing death are not the same as those causing disability. For these reasons this chapter looks at predicting disability after removing deaths from the modeling.

Secondly, as each of the three previous analyses failed to predict neurological outcome at two years of age, it is possible that this outcome simply cannot be predicted by data available in the first 12 hours of an infant's life. This is supported by the fact that no other studies have developed models that successfully predict

neurodevelopmental outcome unless later factors are included. Brazy and colleagues developed the nursery neurobiologic risk score from a population of very low birth weight infants[98]; In order to predict outcome at 24 months, they needed to use data from later in the admission, such as length of ventilation, serum pH, seizures, intraventricular haemorrhage, periventricular leukomalacia, infection and hypoglycaemia. Unfortunately they did not perform ROC curve analysis but only presented a classification table making it difficult to assess how well it discriminated between outcomes across a wide range of points. Lefebvre tested the score at 14 days of age, using ROC curve analysis to show that even at this stage the score was barely adequate at predicting outcome at 18 months (Az = 0.79) [100]. Factors from later in an infant's course such as bronchopulmonary dysplasia, intraventricular haemorrhage, length of stay or length of ventilation are also so very strongly associated with neurological outcome, that predicting outcome with data from very early in the course of an admission is unlikely to be possible[211, 216-218].

This chapter therefore uses the same group of infants to develop two models; one from data available at 12 hours of life and one from data available at discharge. The two hypotheses tested are that it is not possible to develop a model for predicting severe disability using early data and that in the same group of infants a satisfactory model can be developed using data available at discharge to home. The UKOS study dataset was chosen for this analysis, as it had the best quality and most complete data with the largest number of variables measuring care received and complications.

4.2 METHODS

Prospectively collected data were obtained from the United Kingdom randomised trial of high frequency oscillation (UKOS)[178]. This was a randomized, controlled comparison between conventional and oscillatory ventilation commenced in the first hour of life in babies born between August 1998 and January 2001. Babies were eligible for inclusion in the trial and this secondary analysis of the data if they were born between 23 weeks and 28 weeks plus 6 days of gestational age; if they were born in a participating centre; if they required endotracheal intubation from birth; and if they required ongoing intensive care. Infants were excluded if they had to be transferred to another hospital for intensive care shortly after birth, or if they had a major congenital malformation. Randomization was stratified according to centre and gestational age (23 to 25 weeks or 26 to 28 weeks).

The UKOS study showed no difference in the composite primary outcome of death or chronic lung disease, diagnosed at 36 weeks of postmenstrual age. This outcome occurred in 66 percent of the infants assigned to receive high-frequency oscillatory ventilation and 68 percent of those in the conventional-ventilation group (RR risk in oscillated group: 0.98; 95 percent confidence interval, 0.89 to 1.08). Neurodevelopmental outcome was determined between 22 and 28 months of age, corrected for prematurity by a questionnaire completed by the local clinician who followed the infant clinically to at least two years of age corrected for prematurity[180]. Questionnaires were mailed to the local paediatrician responsible for follow up when each infant reached 21 months post-term age, with a request that the child be evaluated as close to 24 months post-term age as possible and within a "window" of 22–28 months. Up to two reminders were sent to paediatricians when questionnaires had not been returned to the coordinating centre by 25 months post-term age. If questionnaires were still not returned, in the United Kingdom the child's local health

visitor was telephoned and asked to complete the forms. Outcome was classified at two years of age as normal, mild disability, moderate disability, severe disability, or death according to the Oxford Health Status Questionnaire (OHSQ)[108].

Statistical analysis was performed using SPSS v11. The outcome of individual infants was classified as no disability, mild disability or moderate disability: group A, or severe disability: group B. All infants who died before 24 months of age after correction for prematurity were excluded from the analysis. Any variable collected during the inpatient stay until discharge home was used in the analysis. Associations with outcome at two years of age corrected for prematurity were identified by the Chi-squared test, Fishers exact test or logistic regression analysis as appropriate. The variables were then assessed in combination using a stepwise, multivariable logistic regression analysis.

Variables were only included in the models if they had a significance level of less than 0.2 for an association with the outcome of severe disability. Predictive variables were added to the model in order of significance, a receiver operator characteristic (ROC) curve for the prediction of severe disability was then determined after each variable was added. The variable was then either left in the model or removed if the area under the ROC curve was less than before that variable was added. The recommendation of Wasson and colleagues[200] was used to limit the number of variables to a maximum of 1 variable for every 10 cases of the outcome of interest. Where numerical data points were missing, in order to avoid losing these infants from the analysis, missing values were replaced with the mean value of the remaining infants.

ROC curve analysis was used to test the ability of the model to discriminate between severe disability and other outcomes in the infants who survived to two years of age. The test therefore analysed the model's ability to discriminate a Group A outcome from a Group B outcome (see Figure 15).

4.3 RESULTS

797 infants were enrolled in the study between August 1998 and January 2001. Outcomes and demographic details are documented in Table 17. Of the 797 infants, 212 died and 32 were assessed as being severely disabled at two years of age. The outcome was not measured within the pre-specified age window for 203 infants and their data was excluded from the analysis. This left 382 surviving infants for the analyses described in this chapter. There were however no significant differences between included and excluded infants. The association of predictive variables with severe disability at two years of age is documented in Table 18.

As before a full set of datapoints for the variables included in the final model was not available for all the 382 infants. In order to avoid losing some of the infants from the analysis, for variables that were associated with severe disability ($p<0.2$) but had a number of data points were missing, missing values were replaced with the mean value of those infants surviving to two years of age on whom data was available. Variables that were not associated with severe disability were not analysed in this way, as substituting values may have inappropriately caused variables to be significantly associated and therefore eligible for inclusion in the modeling. For the Apgar score at 10 minutes there were 179 infants with missing data for whom the mean value of 8.62 points was substituted. Similarly, the days of level 2 care involved 123 infants and a mean value of 11.69 days. Corresponding values for the

age when out of oxygen (and days of oxygen), were 100 infants and 61.6 days, bilirubin, 39 infants and 176.7 micromol/l, inspired oxygen at 12 hours, 19 infants and 33.6%, mean blood pressure at 24 hours, 10 infants and 35.03 mmHg, age at extubation, 6 infants and 18.9 days, days of level 1 care, 6 infants and 36.87 days, ventilation days, 2 infants and 14.1 days, age at discharge, 1 infant and 96.5 days, serum pH at 12 hours, 1 infant and 7.351, and time to intubation 1 infant and 4.04 minutes.

For categorical variables missing data that resulted in infants being removed from the analysis involved 350 infants for ruptured membranes over 22 hours, 175 infants for hearing test result, 30 infants for major abnormality on late cranial ultrasound scan, 20 for smoking in pregnancy, 17 for chorioamnionitis, 16 for breech presentation, 16 for meningitis, nine for antenatal transfer, eight for severe retinopathy of prematurity, five for other air leak, four for cardiac massage and adrenaline, four for pulmonary haemorrhage, four for pulmonary interstitial emphysema, four for systemic steroids, two for major abnormality on early cranial ultrasound scan, two for necrotizing enterocolitis, two for oxygen dependency at discharge, two for septicaemia, two for volume expansion, one for birth order, one for black race, and one infant for seizures. For cranial ultrasound results, a major abnormality comprised either grade 3 or 4 intraventricular haemorrhage or cystic periventricular leukomalacia[219]. Ultrasound results were therefore analysed using the following categorical yes or no variables.

1. Major abnormality on an early scan (infants with no scan result excluded).
2. Major abnormality on a late scan (infants with no scan result excluded).
3. Major abnormality on either an early or on a late scan with the 30 infants with no scan result being excluded.

4. Major abnormality on either an early or on a late scan with the 30 infants with no scan result being classed as normal. 'Cranial USS major abnormality (0 if no scan)'.

Two models for predicting severe disability at two years of age are documented in Table 19. The first of these was derived from variables that are available within the first 12 hours of an infant's life, the second was developed using variables obtained from the entire admission. The derivation of these models by stepwise forward regression and the predictive ability (discrimination) of the model at each step is detailed in Figure 18. The discrimination of the models for predicting severe disability is shown with the corresponding ROC curve in Figure 19.

Table 17: Comparison of Included and Excluded infants

Exclusion due to death or unknown outcome at two years of age

	All infants		Included		Excluded		Included v Excluded
	n =797	(%) or SD	n = 382	(%) or SD	n =415	(%) or SD	
Gestational Age (completed Weeks) Median	26 wks	-	27 wks	-	26 wks	-	**p=0.00000030**
Birthweight (g) Median	853.4 g	218.6	903.6 g	208.3	807.3 g	217.9	**p=0.00000000032**
SGA	138	(17.3%)	60	(15.7%)	78	(18.8%)	p=0.25
Mean Apgar Score at 1 Minute	5.34	2.32	5.59	2.36	5.11	2.26	**p=0.04**
Mean Apgar Score at 5 Minutes	7.89	1.77	8.09	1.65	7.70	1.86	**p=0.02**
Sex							
Female	369	(46.3%)	176	(46.1%)	193	(46.5%)	p=0.903
Male	428	(53.7%)	206	(53.9%)	222	(53.5%)	
Multiple Pregnancy	190	(23.8%)	87	(22.8%)	103	(24.8%)	p=0.499
Surfactant Treatment	769	(96.5%)	369	(96.6%)	400	(96.4%)	p=0.871
Antenatal Steroid Treatment	376	(47.4%)	351	(92.1%)	376	(91.0%)	p=0.58
Maternal Age (Years)	28.8	6.1	29.0	5.8	28.7	6.4	p=0.495
Maximum Base Excess at 2 hours	-4.48	4.1	-4.00	3.8	-4.93	4.3	**p=0.001**
Maximum Base Excess at 12 hours	-4.55	3.3	-4.28	2.9	-4.81	3.7	**P=0.028**
Outcome							
Died	212	(26.6%)	0	0	212	(51.1%)	NA
Severe Disability at 2 years of age	47	(5.9%)	32	(8.4%)	15	(3.6%)	NA
Survived without severe disability	538	(67.5%)	350	(91.6%)	227	(54.7%)	NA
Outcome unknown at 2 years	203	(25.5%)	0	0	203	(48.9%)	NA

Table 18: Association of predictive variables with severe disability at two years

Variable	Coefficient	Standard Error	Wald	Number	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
							Lower	Upper
Major Abnormality on Late Cranial US[1]	-	-	-	352	3.28E-17	-	-	-
Major Abnormality on any Cranial US (0 if no scan)[1]	-	-	-	382	4.34E-13	-	-	-
Major Abnormality on Early Cranial US[1]	-	-	-	380	5.28E-13	-	-	-
Major Abnormality on Late Cranial US	2.89	0.42	46.48	352	9.24E-12	17.94	7.82	41.14
Seizures	-	-	-	381	1.25E-11	-	-	-
Major Abnormality on any Cranial US (0 if no scan)	2.45	0.40	37.8	382	7.96E-10	11.6	5.3	25.3
Major Abnormality on Early Cranial US	2.44	0.40	37.6	380	8.82E-10	11.5	5.27	25.1
Patent Ductus Arteriousus	-	-	-	382	0.00038	-	-	-
Ventilation Days	0.03	0.010	11.81	380	0.00059	1.03	1.01	1.05
Ventilation Days with mean value	0.03	0.010	11.61	382	0.00066	1.03	1.01	1.05
Severe retinopathy of prematurity[1]	-	-	-	374	0.00074	-	-	-
Patent Ductus Arteriousus	-1.29	0.38	11.37	382	0.00075	0.28	0.13	0.58
Patent Ductus Arteriousus*	-	-	-	382	0.00038	-	-	-
Age at extubation	0.019	0.0057	10.56	376	0.0012	1.02	1.01	1.03
Number of surfactant doses*	-	-	-	382	0.0015	-	-	-
Severe retinopathy of prematurity	1.62	0.52	9.50	374	0.0021	5.03	1.80	14.06
Age at discharge	0.012	0.0038	9.35	382	0.0022	1.01	1.00	1.02
Age at discharge	0.012	0.0038	9.33	381	0.0023	1.01	1.00	1.02
Pulmonary Haemorrhage[1]	-	-	-	378	0.0059	-	-	-
Days Level 1 Care with mean value	0.014	0.0051	7.17	382	0.0074	1.01	1.00	1.02
Days Level 1 Care	0.014	0.0051	7.09	376	0.0078	1.01	1.00	1.02
Pulmonary Haemorrhage	-1.32	0.51	6.72	378	0.0095	0.27	0.10	0.73
Days Level 2 Care with mean value	0.020	0.0079	6.45	382	0.011	1.02	1.00	1.04
Bilirubin	0.011	0.0044	5.82	343	0.016	1.01	1.00	1.02
Bilirubin with mean value	0.010	0.0043	5.75	382	0.016	1.01	1.00	1.02
Days Level 2 Care	0.018	0.0077	5.57	*259*	0.018	1.02	1.00	1.03
Percent Inspired Oxygen at 12hrs with mean value	2.29	0.97	5.57	382	0.018	9.84	1.47	65.65
Systemic steroids[1]	-	-	-	378	0.019	-	-	-
Percent Inspired Oxygen at 12hrs	2.26	0.97	5.47	363	0.019	9.59	1.44	63.69
Systemic steroids	-0.86	0.37	5.29	378	0.021	0.42	0.20	0.88
Oxygen Dependency at discharge[1]	-	-	-	380	0.025	-	-	-
Oxygen Dependency at discharge	-0.85	0.39	4.79	380	0.029	0.43	0.20	0.92
Gender[1]	-	-	-	382	0.033	-	-	-
Oxygen Dependency at 36 weeks	-	-	-	382	0.035	-	-	-

CONTINUED OVERLEAF

Days in Oxygen	0.012	0.0056	4.39	282	0.036	1.01	1.00	1.02
Age out of oxygen	0.012	0.0056	4.39	*282*	0.036	1.01	1.00	1.02
Gender	0.85	0.41	4.32	382	0.038	2.33	1.05	5.18
Oxygen Dependency at 36 weeks	-	-	-	382	0.035	-	-	-
Delivery reason Hypertension or Pre-Eclampsia	0.77	0.38	4.13	382	0.042	2.15	1.03	4.49
Antenatal Transfer[1]	-	-	-	373	0.048	-	-	-
Oxygen Dependency at 28 days[1]	-	-	-	382	0.050	-	-	-
Days in Oxygen with mean value	0.010	0.0051	3.66	382	0.056	1.01	1.00	1.02
Age out O2	0.0097	0.0051	3.66	382	0.056	1.01	1.00	1.02
Oxygen Dependency at 28 days	-1.36	0.74	3.34	382	0.067	0.26	0.06	1.10
pH At 12 hours with mean value	-3.63	2.00	3.28	382	0.070	0.03	0.00	1.35
pH At 12 hours	-3.62	2.00	3.27	381	0.070	0.03	0.00	1.35
Apgar 10 Min with mean value	-0.20	0.13	2.65	382	0.10	0.81	0.64	1.04
Pneumothorax[1]	-	-	-	382	0.10	-	-	-
Other Air Leak[1]	-	-	-	377	0.11	-	-	-
Pneumothorax	-0.84	0.53	2.51	382	0.11	0.43	0.15	1.22
Pulmonary Interstitial Emphysema[1]	-	-	-	378	0.12	-	-	-
Pulmonary Interstitial Emphysema	-0.88	0.58	2.30	378	0.13	0.41	0.13	1.29
Apgar 10 Min	-0.19	0.12	2.29	203	0.13	0.83	0.65	1.06
Maternal Age	-0.047	0.031	2.25	382	0.13	0.95	0.90	1.01
Delivery reason Chorioamnionitis[1]	-	-	-	382	0.14	-	-	-
Draper Grid Prediction	-0.013	0.0088	2.13	382	0.14	0.99	0.97	1.00
Delivery reason Chorioamnionitis	-0.76	0.53	2.07	382	0.15	0.47	0.17	1.31
Time to Intubation with mean value	-0.13	0.091	2.06	382	0.15	0.88	0.73	1.05
Time to Intubation	-0.13	0.091	2.04	381	0.15	0.88	0.73	1.05
Other Air Leak	-1.75	1.24	1.99	377	0.16	0.17	0.02	1.98
Cardiac massage & adrenaline[1]	-	-	-	378	0.16	-	-	-
Mean blood pressure at 24 hours with mean value	-0.046	0.033	1.96	382	0.16	0.96	0.90	1.02
Mean blood pressure at 24 hours	-0.044	0.032	1.91	372	0.17	0.96	0.90	1.02
Cardiac massage & adrenaline	0.79	0.58	1.87	378	0.17	2.21	0.71	6.89
Delivery reason Hypertension or Pre-Eclampsia[1]	-	-	-	382	0.19	-	-	-
Gestational Age (days)	-0.024	0.019	1.66	382	0.20	0.98	0.94	1.01
Necrotizing Enterocolitis[1]	-	-	-	380	0.21	-	-	-
Ruptured Membranes over 24 hours[1]	-	-	-	382	0.22	-	-	-
Septicaemia[1]	-	-	-	380	0.23	-	-	-
Percent Inspired Oxygen at 2hrs	0.98	0.82	1.43	378	0.23	2.67	0.54	13.28

CONTINUED OVERLEAF

Delivery Reason Small for Gestation[1]	-	-	-	382	0.26	-	-	-
Base Excess at 12 hours	-0.073	0.066	1.19	368	0.27	0.93	0.82	1.06
Days Level 3 Care	-0.010	0.0096	1.08	357	0.30	0.99	0.97	1.01
Mean blood pressure at 12 hours	-0.030	0.031	0.93	377	0.34	0.97	0.91	1.03
Breech Presentation[1]	-	-	-	366	0.35	-	-	-
Chorioamnionitis[1]	-	-	-	365	0.38	-	-	-
Meningitis[1]	-	-	-	366	0.38	-	-	-
Birth Weight	-0.00076	0.00091	0.70	382	0.40	0.999	0.997	1.001
Age at pneumothorax	0.059	0.076	0.59	351	0.44	1.06	0.91	1.23
Birth Order[1]	-	-	-	381	0.44	-	-	-
Family Number[1]	-	-	-	382	0.45	-	-	-
Multiple pregnancy[1]	-	-	-	382	0.45	-	-	-
Antenatal Steroids 48 hours or more before delivery[1]	-	-	-	382	0.48	-	-	-
Volume expansion[1]	-	-	-	380	0.49	-	-	-
Base Excess at 2 hours	-0.035	0.052	0.45	370	0.50	0.97	0.87	1.07
Base Excess at 24 hours	-0.039	0.066	0.36	363	0.55	0.96	0.84	1.09
Antenatal Steroids[1]	-	-	-	382	0.57	-	-	-
Jaundice[1]	-	-	-	382	0.57	-	-	-
Apgar 5 Min	-0.057	0.11	0.29	371	0.59	0.94	0.77	1.16
Apgar Score for heart rate at 1 min*	-	-	-	382	0.59	-	-	-
Delivery reason Spontaneous preterm labour*	-	-	-	382	0.62	-	-	-
pH At 24 hours	-0.97	2.05	0.23	380	0.63	0.38	0.01	20.91
Antepartum Haemorrhage[1]	-	-	-	382	0.67	-	-	-
Inotrope treatment[1]	-	-	-	382	0.71	-	-	-
Delivery reason APH[1]	-	-	-	382	0.72	-	-	-
Delivery Reason Other[1]	-	-	-	382	0.72	-	-	-
Maternal Labour[1]	-	-	-	382	0.75	-	-	-
Mean blood pressure 2 hours	-0.0068	0.028	0.061	351	0.81	0.99	0.94	1.05
pH at 2 hours	0.36	1.89	0.04	380	0.85	1.44	0.04	58.45
Apgar 1 Min	-0.011	0.078	0.020	376	0.89	0.99	0.85	1.15
Hearing test result	-	-	-	207	0.89	-	-	-
Ruptured Membranes over 22 hours	-	-	-	32	0.92	-	-	-
Head Circumference	0.012	0.12	0.0087	244	0.93	1.01	0.79	1.29
Race[1]	-	-	-	381	0.98	-	-	-
Smoking in Pregnancy[1]	-	-	-	362	0.99	-	-	-

Note
[1] Chi square test used to assess association, logistic regression used to assess odds ratio if $p<0.2$

Table 19: Models for predicting severe disability at two years of age in infants surviving to discharge. Logistic regression coefficients and odds ratios for factors in the model

(A) Best model for predicting severe disability using early information available before 12 hours of age.

Variables	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Percent Inspired Oxygen at 12hrs with mean value	2.84	0.0065	17.2	2.21	133.1
Gender (Male 1, Female 0)	1.032	0.019	2.81	1.19	6.64
Antenatal Transfer	-0.91	0.045	0.40	0.17	0.98
Constant	-3.82	1.41E-10	0.022		

(B) Best model for predicting severe disability using information available before discharge to home.

Variables	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Major Abnormality on any Cranial US (assumed normal if no scan)	2.00	3.35E-06	7.4	3.2	17.2
Seizures*	-1.84	0.0037	0.16	0.046	0.55
Ventilation Days with mean value replacing missing values	0.024	0.062	1.025	0.999	1.051
Constant	0.10	0.94	1.10		

* Seizures classified as y =1, n =2

Figure 18: Process of Stepwise regression modeling for predicting severe disability at two years of age using (A) data available at 12 hours of age and (B) Data available at discharge

A: Model using data available by 12 hours of age

STEP 1 Inspired Oxygen at 12hrs[a] n = 382

Logistic Regression Model

Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Inspired Oxygen at 12hrs[a]	2.29	0.018	9.8	1.5	65.7
Constant	-3.22	1.6E-14	0.040		

Discrimination of Model Prediction (ROC Curve Analysis)

Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.645	0.048	0.0065	0.552	-	0.739

STEP 2 Gender added n = 382

Logistic Regression Model

Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Inspired Oxygen at 12hrs[a]	2.57	0.011	13.1	1.9	96.3
Gender	0.93	0.026	2.5	1.1	5.7
Constant	-3.91	2.9E-12	0.02		

Discrimination of Model Prediction (ROC Curve Analysis)

Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.683	0.046	0.00062	0.593	-	0.773

STEP 3 Delivery for hypertension or preeclampsia added n = 382

Logistic Regression Model

Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Inspired Oxygen at 12hrs[a]	2.58	0.012	13.2	1.8	99.3
Gender	0.95	0.023	2.59	1.14	5.87
Delivery for hypertension or preeclampsia	0.58	0.17	1.78	0.77	4.10
Constant	-4.06	2.6E-12	0.02		

Discrimination of Model Prediction (ROC Curve Analysis)

Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.686	0.050	0.00050	0.589	-	0.783

STEP 4 Antenatal transfer added n = 373

Logistic Regression Model

Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Inspired Oxygen at 12hrs[a]	2.82	0.0077	16.8	2.1	133.4
Gender	1.04	0.018	2.83	1.19	6.71
Delivery for hypertension or preeclampsia	0.83	0.15	1.88	0.80	4.40
Antenatal Transfer	-0.91	0.046	0.40	0.17	0.98
Constant	-3.96	1.08E-10	0.019		

Discrimination of Model Prediction (ROC Curve Analysis)

Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.723	0.0469	3.91E-05	0.631	-	0.815

Continued overleaf

STEP 5 Serum pH at 12 hours of age[b] added — n = 373

Logistic Regression Model					
Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Inspired Oxygen at 12hrs[a]	2.55	0.020	12.8	1.5	109.7
Gender	1.0016	0.023	2.72	1.15	6.47
Delivery for hypertension or preeclampsia	0.84	0.14	1.90	0.81	4.46
Antenatal Transfer	-0.90	0.047	0.41	0.17	0.99
Serum pH at 12 hours of age[b]	-2.19	0.31	0.11	0.0016	7.64
Constant	12.2	0.44	204469.4		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.732	0.046	1.8E-05	0.643	-	0.822

STEP 6 Apgar Score at 10 minutes[c] added — n = 373

Logistic Regression Model					
Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Inspired Oxygen at 12hrs[a]	2.36	0.033	10.6	1.2	92.7
Gender	0.99	0.025	2.69	1.13	6.40
Delivery for hypertension or preeclampsia	0.74	0.097	2.09	0.88	4.98
Antenatal Transfer	-0.87	0.055	0.42	0.17	1.02
Serum pH at 12 hours of age[b]	-2.19	0.31	0.11	0.0016	7.91
Apgar Score at 10 minutes[c]	-0.21	0.11	0.81	0.63	1.05
Constant	14.1	0.38	1271106.2		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.743	0.045	7.6E-06	0.654	-	0.831

STEP 7 Mother's age added — n = 373

Logistic Regression Model					
Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Inspired Oxygen at 12hrs[a]	2.36	0.032	10.6	1.2	90.8
Gender	0.99	0.025	2.70	1.13	6.46
Delivery for hypertension or preeclampsia	0.69	0.12	2.00	0.83	4.81
Antenatal Transfer	-0.84	0.067	0.43	0.18	1.06
Serum pH at 12 hours of age[b]	-2.59	0.24	0.075	0.0010	5.76
Apgar Score at 10 minutes[c]	-0.19	0.14	0.82	0.64	1.06
Mother's Age	-0.045	0.19	0.96	0.89	1.02
Constant	18.1	0.27	76174375		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.750	0.044	4.0E-06	0.663	-	0.837

Continued overleaf

STEP 8 Probability of Survival[b] added — n = 373

Logistic Regression Model					
Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Inspired Oxygen at 12hrs[a]	2.24	0.043	9.4	1.07	82.0
Gender	0.93	0.037	2.54	1.06	6.11
Delivery for hypertension or preeclampsia	0.72	0.11	2.06	0.85	4.99
Antenatal Transfer	-0.86	0.060	0.42	0.17	1.04
Serum pH at 12 hours of age[b]	-2.57	0.25	0.077	0.0010	6.02
Apgar Score at 10 minutes[a]	-0.18	0.18	0.836	0.644	1.086
Mother's Age	-0.048	0.17	0.953	0.891	1.020
Probability of Survival[b]	-0.011	0.25	0.989	0.971	1.008
Constant	18.7	0.26	128138638		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.760	0.044	1.6E-06	0.674	-	0.846

STEP 9 Delivery for Choriomamnionitis added — n = 357

Logistic Regression Model					
Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Inspired Oxygen at 12hrs[a]	2.12	0.058	8.3	0.93	74.9
Gender	0.94	0.038	2.55	1.05	6.18
Delivery for hypertension or preeclampsia	0.73	0.11	2.08	0.84	5.16
Antenatal Transfer	-0.86	0.062	0.42	0.17	1.05
Serum pH at 12 hours of age[b]	-2.59	0.25	0.07	0.0009	6.09
Apgar Score at 10 minutes[a]	-0.17	0.20	0.84	0.65	1.09
Mother's Age	-0.045	0.19	0.96	0.89	1.02
Probability of Survival[b]	-0.010	0.31	0.99	0.97	1.01
Delivery for Choriomamnionitis	0.31	0.63	1.37	0.39	4.80
Constant	18.7	0.27	127358360		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.758	0.042	2.1E-06	0.676	-	0.840

STEP 10 Delivery for Choriomamnionitis removed — n = 373

see results of step 8

STEP 11 Time to intubation[a] added — n = 373

Logistic Regression Model					
Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Inspired Oxygen at 12hrs[a]	2.62	0.022	13.7	1.5	128.4
Gender	1.01	0.026	2.7	1.1	6.7
Delivery for hypertension or preeclampsia	0.86	0.062	2.36	0.96	5.82
Antenatal Transfer	-0.86	0.058	0.42	0.17	1.03
Serum pH at 12 hours of age[b]	-2.56	0.25	0.077	0.0010	6.21
Apgar Score at 10 minutes[a]	-0.18	0.19	0.84	0.64	1.09
Mother's Age	-0.052	0.14	0.95	0.89	1.017
Probability of Survival[b]	-0.0088	0.36	0.99	0.97	1.010
Time to Intubation[a]	-0.16	0.11	0.85	0.70	1.037
Constant	18.9	0.26	159032396		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.774	0.043	4.5E-07	0.689	-	0.858

Continued overleaf

STEP 12 Cardiac Massage & Adrenaline added — n = 370

Logistic Regression Model

Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Inspired Oxygen at 12hrs[a]	2.54	0.026	12.6	1.4	117.9
Gender	0.98	0.031	2.67	1.09	6.51
Delivery for hypertension or preeclampsia	0.88	0.058	2.40	0.97	5.93
Antenatal Transfer	-0.89	0.056	0.41	0.17	1.02
Serum pH at 12 hours of age[b]	-2.86	0.21	0.058	0.00089	4.79
Apgar Score at 10 minutes[a]	-0.16	0.26	0.86	0.65	1.12
Mother's Age	-0.049	0.17	0.95	0.89	1.02
Probability of Survival[b]	-0.0084	0.38	0.99	0.97	1.01
Time to intubation[a]	-0.16	0.11	0.86	0.71	1.03
Cardiac Massage & Adrenaline	0.56	0.39	1.75	0.49	6.30
Constant	20.7	0.22	968796779		

Discrimination of Model Prediction (ROC Curve Analysis)

Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.770	0.044	6.5E-07	0.684	-	0.856

STEP 13 Gestational Age (Days) added — n = 370

Logistic Regression Model

Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Inspired Oxygen at 12hrs[a]	2.50	0.029	12.1	1.3	113.7
Gender	1.07	0.023	2.91	1.16	7.30
Delivery for hypertension or preeclampsia	1.03	0.043	2.80	1.03	7.58
Antenatal Transfer	-0.88	0.059	0.42	0.17	1.03
Serum pH at 12 hours of age[b]	-2.87	0.20	0.06	0.00067	4.73
Apgar Score at 10 minutes[a]	-0.16	0.26	0.86	0.65	1.12
Mother's Age	-0.049	0.17	0.95	0.89	1.02
Probability of Survival[b]	0.010	0.71	1.01	0.96	1.06
Time to intubation[a]	-0.15	0.11	0.86	0.71	1.03
Cardiac Massage & Adrenaline	0.82	0.34	1.86	0.51	6.77
Gestational Age (Days)	-0.043	0.46	0.96	0.86	1.07
Constant	27.6	0.15	9.984E+11		

Discrimination of Model Prediction (ROC Curve Analysis)

Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.767	0.044	8.3E-07	0.681	-	0.854

STEP 14 Probability of Survival[b] removed — n = 370

Logistic Regression Model

Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Inspired Oxygen at 12hrs[a]	2.50	0.028	12.2	1.3	114.4
Gender	1.02	0.024	2.8	1.1	6.7
Delivery for hypertension or preeclampsia	0.96	0.042	2.61	1.04	6.58
Antenatal Transfer	-0.88	0.057	0.41	0.17	1.03
Serum pH at 12 hours of age[b]	-2.85	0.21	0.058	0.0007	4.809
Apgar Score at 10 minutes[a]	-0.15	0.27	0.86	0.65	1.12
Mother's Age	-0.049	0.16	0.95	0.89	1.02
Time to intubation[a]	-0.15	0.11	0.86	0.71	1.04
Cardiac Massage & Adrenaline	0.59	0.37	1.80	0.50	6.47
Gestational Age (Days)	-0.023	0.27	0.98	0.94	1.02
Constant	24.4	0.16	3.89E+10		

Discrimination of Model Prediction (ROC Curve Analysis)

Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.767	0.044	8.3E-07	0.681	-	0.854

Continued overleaf

STEP 15 Cardiac Massage & Adrenaline removed — n = 373

Logistic Regression Model

Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Inspired Oxygen at 12hrs[a]	2.58	0.024	13.2	1.4	124.8
Gender	1.05	0.020	2.9	1.18	6.9
Delivery for hypertension or preeclampsia	0.94	0.045	2.5	1.02	6.45
Antenatal Transfer	-0.87	0.060	0.42	0.17	1.04
Serum pH at 12 hours of age[b]	-2.56	0.25	0.076	0.0010	6.308
Apgar Score at 10 minutes[c]	-0.18	0.19	0.84	0.64	1.09
Mother's Age	-0.053	0.14	0.95	0.89	1.02
Time to Intubation[d]	-0.16	0.11	0.86	0.71	1.04
Gestational Age (Days)	-0.023	0.27	0.98	0.94	1.02
Constant	22.5	0.19	6.096E+09		

Discrimination of Model Prediction (ROC Curve Analysis)

Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.773	0.043	4.9E-07	0.688	-	0.857

STEP 16 Gestational Age (Days) removed — n = 373

Logistic Regression Model

Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Inspired Oxygen at 12hrs[a]	2.71	0.017	15.1	1.6	139.3
Gender	1.06	0.018	2.9	1.2	7.0
Delivery for hypertension or preeclampsia	0.85	0.064	2.3	0.95	5.7
Antenatal Transfer	-0.87	0.061	0.42	0.17	1.04
Serum pH at 12 hours of age[b]	-2.58	0.25	0.08	0.0009	6.01
Apgar Score at 10 minutes[c]	-0.19	0.16	0.83	0.64	1.08
Mother's Age	-0.050	0.15	0.95	0.89	1.02
Time to Intubation[d]	-0.16	0.10	0.85	0.70	1.03
Constant	18.5	0.27	103160279		

Discrimination of Model Prediction (ROC Curve Analysis)

Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.767	0.044	8.8E-07	0.681	-	0.852

STEP 17 Serum pH at 12 hours of age[b] removed — n = 373

Logistic Regression Model

Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Inspired Oxygen at 12hrs[a]	2.962338551	0.007552848	19.343154	2.2010691	169.98903
Gender	1.098323874	0.014445809	2.9991349	1.24388269	7.2312366
Delivery for hypertension or preeclampsia	0.828343831	0.068971993	2.2895238	0.93764157	5.59053611
Antenatal Transfer	-0.880449936	0.056843433	0.4145963	0.16753973	1.0259663
Apgar Score at 10 minutes[c]	-0.183754218	0.169656967	0.8321403	0.64018018	1.08166031
Mother's Age	-0.043775907	0.200073076	0.9571684	0.89517481	1.02345529
Time to Intubation[d]	-0.160699734	0.103813067	0.8515477	0.70164404	1.03347778
Constant	-0.817223712	0.594676875	0.4416561		

Discrimination of Model Prediction (ROC Curve Analysis)

Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.753	0.046	2.9E-06	0.663	-	0.844

Continued overleaf

STEP 18 Mother's Age removed — n = 373

Logistic Regression Model					
Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Inspired Oxygen at 12hrs[a]	2.94	0.0081	18.9	2.1	167.0
Gender	1.08	0.016	2.94	1.23	7.07
Delivery for hypertension or preeclampsia	0.86	0.058	2.35	0.97	5.70
Antenatal Transfer	-0.89	0.052	0.41	0.17	1.01
Apgar Score at 10 minutes[a]	-0.19	0.15	0.82	0.63	1.07
Time to intubation[a]	-0.17	0.10	0.85	0.69	1.03
Constant	-1.94	0.13	0.14		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.748	0.047	4.6E-06	0.656	-	0.840

STEP 19 Apgar Score at 10 minutes[a] removed — n = 373

Logistic Regression Model					
Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Inspired Oxygen at 12hrs[a]	3.14	0.0044	23.1	2.7	199.7
Gender	1.08	0.016	2.93	1.23	7.02
Delivery for hypertension or preeclampsia	0.75	0.088	2.12	0.89	5.05
Antenatal Transfer	-0.93	0.042	0.39	0.16	0.97
Time to intubation[a]	-0.18	0.085	0.83	0.68	1.03
Constant	-3.58	3.6E-08	0.028		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.746	0.047	5.9E-06	0.653	-	0.838

STEP 20 Delivery for hypertension or preeclampsia removed — n = 373

Logistic Regression Model					
Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Inspired Oxygen at 12hrs[a]	3.19	0.0034	24.3	2.9	204.1
Gender	1.05	0.018	2.84	1.19	6.77
Antenatal Transfer	-0.93	0.042	0.40	0.16	0.97
Time to intubation[a]	-0.17	0.095	0.84	0.69	1.03
Constant	-3.43	5.3E-08	0.032		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.736	0.045	1.3E-05	0.649	-	0.824

STEP 21 Time to intubation[a] removed — n = 373

Logistic Regression Model					
Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Inspired Oxygen at 12hrs[a]	2.84	0.0065	17.2	2.2	133.1
Gender	1.03	0.019	2.81	1.19	6.64
Antenatal Transfer	-0.91	0.045	0.40	0.17	0.98
Constant	-3.82	1.4E-10	0.022		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.713	0.044	8.6E-05	0.626	-	0.799

Notes

a: Mean value substituted for missing values

b: Probability of survival obtained from 'Draper Grid'

B: Model from data available by discharge

STEP 1 Cranial USS major abn (0 if no scan) n = 382

Logistic Regression Model

Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Major Abnormality on any Cranial US (assumed normal if no scan)	2.448538006	7.95894E-10	11.571429	5.289808369	25.28467937
Constant	-3.113815309	4.22236E-30	0.0444444		

Discrimination of Model Prediction (ROC Curve Analysis)

Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.731	0.055	1.8E-05	0.624	-	0.838

STEP 2 Seizures added n = 381

Logistic Regression Model

Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Major Abnormality on any Cranial US (assumed normal if no scan)	2.12	5.7E-07	8.3	3.6	19.1
Seizures	-2.09	0.00052	0.12	0.038	0.40
Constant	0.96	0.41	2.6		

Discrimination of Model Prediction (ROC Curve Analysis)

Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.752	0.055	2.3E-06	0.645	-	0.860

STEP 3 Patent Ductus Arteriousus added n = 381

Logistic Regression Model

Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Major Abnormality on any Cranial US (assumed normal if no scan)	2.01	0.0000030	7.5	3.2	17.4
Seizures	-1.89	0.0022	0.15	0.045	0.51
Patent Ductus Arteriousus	-0.83	0.051	0.43	0.19	1.00
Constant	1.90	0.14	6.7		

Discrimination of Model Prediction (ROC Curve Analysis)

Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.786	0.050	0.00000009	0.687	-	0.885

STEP 4 Patent Ductus Arteriousus replaced by Ventilation Days[a] n = 381

Logistic Regression Model

Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Major Abnormality on any Cranial US (assumed normal if no scan)	2.00	0.0000034	7.4	3.2	17.2
Seizures	-1.84	0.0037	0.16	0.05	0.55
Ventilation Days[a]	0.02	0.082	1.02	1.00	1.05
Constant	0.10	0.94	1.10		

Discrimination of Model Prediction (ROC Curve Analysis)

Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.842	0.036	1.6E-10	0.771	-	0.913

FINAL MODEL

Continued overleaf

STEP 5 Ventilation Days[a] replaced by Severe ROP — n = 373

Logistic Regression Model					
Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Major Abnormality on any Cranial US (assumed normal if no scan)	1.99	4.9E-06	7.3	3.1	17.1
Seizures	-2.06	0.00079	0.13	0.04	0.43
Severe Retinopathy of Prematurity[a]	0.88	0.18	2.41	0.67	8.60
Constant	0.87	0.47	2.39		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.760	0.054	0.0000011	0.654	-	0.867

STEP 6 Severe ROP replaced by Age at extubation[a] — n = 381

Logistic Regression Model					
Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Major Abnormality on any Cranial US (assumed normal if no scan)	2.04	2.5E-06	7.7	3.3	17.8
Seizures	-1.92	0.0024	0.15	0.042	0.51
Age at extubation[a]	0.01	0.26	1.01	0.99	1.02
Constant	0.48	0.71	1.62		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.839	0.037	2.06E-10	0.767	-	0.911

STEP 7 Age at extubation[a] replaced by Number of surfactant doses — n = 381

Logistic Regression Model					
Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Major Abnormality on any Cranial US (assumed normal if no scan)	2.04	1.9E-06	7.7	3.33	17.9
Seizures	-1.93	0.0016	0.15	0.043	0.49
Number of Surfactant doses	0.53	0.056	1.69	0.98	2.91
Constant	-0.36	0.80	0.70		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.810	0.044	6.2E-09	0.723	-	0.897

STEP 8 Number of surfactant doses replaced by Age at discharge[a] — n = 381

Logistic Regression Model					
Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Major Abnormality on any Cranial US (assumed normal if no scan)	2.11	1.1E-06	8.2	3.5	19.2
Seizures	-1.92	0.0023	0.15	0.043	0.50
Age at discharge	0.011	0.014	1.012	1.0024	1.021
Constant	-0.55	0.69	0.56		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.827	0.043	8.4E-10	0.743	-	0.910

STEP 9 Age at discharge[a] replaced by Pulmonary Haemorrhage — n = 377

Logistic Regression Model					
Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Major Abnormality on any Cranial US (assumed normal if no scan)	2.072	2.4E-06	7.9	3.4	18.8
Seizures	-2.17	0.00043	0.11	0.034	0.38
Pulmonary Haemorrhage	-0.95	0.12	0.39	0.12	1.29
Constant	2.87	0.10	17.7		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.769	0.055	6.9E-07	0.662	-	0.876

Continued overleaf

STEP 10 Pulmonary Haemorrhage replaced by Days Level 1 Care[a] n = 377

Logistic Regression Model

Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Major Abnormality on any Cranial US (assumed normal if no scan)	2.10	1.1E-06	8.1	3.5	18.9
Seizures	-1.95	0.0018	0.14	0.042	0.48
Days Level 1 Care[a]	0.011	0.054	1.011	0.9998	1.023
Constant	0.24	0.85	1.3		

Discrimination of Model Prediction (ROC Curve Analysis)

Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.805	0.045	1.05E-06	0.718	-	0.893

STEP 11 Days Level 1 Care[a] replaced by Days Level 2 Care[a] n = 381

Logistic Regression Model

Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Major Abnormality on any Cranial US (assumed normal if no scan)	2.07	1.4E-06	7.8	3.4	18.4
Seizures	-2.09	0.00075	0.12	0.037	0.42
Days Level 2 Care[a]	0.017	0.053	1.017	0.9998	1.035
Constant	0.73	0.55	2.1		

Discrimination of Model Prediction (ROC Curve Analysis)

Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.746	0.057	3.9E-06	0.636	-	0.858

STEP 12 Days Level 2 Care[a] replaced by Bilirubin[a] n = 381

Logistic Regression Model

Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Major Abnormality on any Cranial US (assumed normal if no scan)	2.06	1.6E-06	7.8	3.4	17.9
Seizures	-2.05	0.00055	0.13	0.040	0.41
Bilirubin[a]	0.0070	0.15	1.007	0.998	1.017
Constant	-0.39	0.79	0.6770522		

Discrimination of Model Prediction (ROC Curve Analysis)

Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.788	0.047	6.6E-06	0.696	-	0.880

STEP 13 Bilirubin[a] replaced by Percent inspired oxygen at 12 hours of age[a] n = 381

Logistic Regression Model

Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Major Abnormality on any Cranial US (assumed normal if no scan)	2.12	6.8E-07	8.4	3.6	19.3
Seizures	-1.999	0.0011	0.14	0.041	0.45
Percent inspired oxygen at 12 hours of age[a]	1.99	0.11	6.6	0.66	66.1
Constant	0.11	0.93	1.11		

Discrimination of Model Prediction (ROC Curve Analysis)

Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.810	0.043	6.2E-09	0.726	-	0.894

STEP 14 Percent inspired oxygen at 12 hours of age[a] replaced by Systemic steroids n = 381

Logistic Regression Model

Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Major Abnormality on any Cranial US (assumed normal if no scan)	2.14	6.4E-07	8.5	3.7	19.7
Seizures	-2.04	0.00079	0.13	0.040	0.43
Systemic steroids	-0.83	0.050	0.43	0.19	1.00
Constant	2.23	0.10	9.3		

Discrimination of Model Prediction (ROC Curve Analysis)

Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.817	0.044	2.9E-09	0.731	-	0.903

Continued overleaf

STEP 15 Systemic steroids replaced by Oxygen Dependency at discharge n = 381

Logistic Regression Model					
Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Major Abnormality on any Cranial US (assumed normal if no scan)	2.13	8.2E-07	8.4	3.5	19.8
Seizures	-2.16	0.00051	0.12	0.034	0.39
Oxygen Dependency at discharge	-0.97	0.031	0.38	0.16	0.92
Constant	2.74	0.064	15.5		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.812	0.045	4.9E-09	0.723	-	0.901

STEP 16 Oxygen Dependency at discharge replaced by Gender n = 381

Logistic Regression Model					
Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Major Abnormality on any Cranial US (assumed normal if no scan)	2.00	3.19E-06	7.4	3.2	17.2
Seizures	-2.40	0.00020	0.091	0.026	0.32
Gender	1.01	0.038	2.7	1.06	7.1
Constant	0.94	0.44	2.6		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.828	0.037	8.5E-10	0.754	-	0.901

STEP 17 Gender replaced by Oxygen Dependency at 36 weeks n = 381

Logistic Regression Model					
Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Major Abnormality on any Cranial US (assumed normal if no scan)	2.20	3.9E-07	9.0	3.9	21.1
Seizures	-2.02	0.00096	0.13	0.040	0.44
Oxygen Dependency at 36 weeks GA	-0.96	0.041	0.38	0.15	0.96
Constant	2.10	0.12	8.1		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.795	0.047	3.2E-08	0.702	-	0.888

STEP 18 Oxygen Dependency at 36 weeks replaced by Delivery for hypertension or preeclampsia n = 381

Logistic Regression Model					
Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Major Abnormality on any Cranial US (assumed normal if no scan)	2.35	2.2E-07	10.5	4.3	25.7
Seizures	-2.20	0.00042	0.11	0.033	0.38
Delivery for hypertension or preeclampsia	1.21	0.014	3.37	1.27	8.90
Constant	0.78	0.52	2.2		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.813	0.0452	4.7E-09	0.724	-	0.901

STEP 19 Delivery for hypertension or preeclampsia replaced by Antenatal Transfer n = 372

Logistic Regression Model					
Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Major Abnormality on any Cranial US (assumed normal if no scan)	1.95	9.7E-06	7.0	3.0	16.6
Seizures	-2.19	0.00050	0.11	0.033	0.38
Antenatal Transfer	-0.53	0.27	0.59	0.23	1.51
Constant	1.38	0.27	4.0		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.794	0.045	6.1E-08	0.706	-	0.882

Continued overleaf

STEP 20 Antenatal Transfer replaced by Oxygen Dependency at 28 days n = 381

Logistic Regression Model					
Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Major Abnormality on any Cranial US (assumed normal if no scan)	2.04	1.9E-06	7.7	3.3	17.9
Seizures	-2.06	0.00074	0.13	0.037	0.42
Oxygen Dependency at 28 days	-0.97	0.20	0.38	0.085	1.70
Constant	2.05	0.16	7.8		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.764	0.053	7.4E-07	0.661	-	0.867

STEP 21 Oxygen Dependency at 28 days replaced by Days in Oxygen[a] n = 381

Logistic Regression Model					
Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Major Abnormality on any Cranial US (assumed normal if no scan)	2.12	6.5E-07	8.3	3.6	19.1
Seizures	-2.00	0.0011	0.14	0.041	0.45
Days in Oxygen[a]	0.0068	0.23	1.007	0.996	1.018
Constant	0.32	0.81	1.37		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.770	0.049	4.2E-07	0.674	-	0.866

STEP 22 Days in Oxygen[a] replaced by Age out of Oxygen[a] n = 381

Logistic Regression Model					
Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Major Abnormality on any Cranial US (assumed normal if no scan)	2.12	6.5E-07	8.3	3.6	19.1
Seizures	-2.00	0.0011	0.14	0.041	0.45
Age out of Oxygen[a]	0.0068	0.23	1.007	0.996	1.018
Constant	0.32	0.81	1.37		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.770	0.049	4.2E-07	0.674	-	0.866

STEP 23 Age out of Oxygen[a] replaced by Serum pH at 12 hours of age[a] n = 381

Logistic Regression Model					
Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Major Abnormality on any Cranial US (assumed normal if no scan)	2.09	9.4E-07	8.0	3.5	18.5
Seizures	-2.04	0.00067	0.13	0.040	0.42
Serum pH at 12 hours of age[b]	-1.98	0.39	0.14	0.0015	12.8
Constant	15.4	0.36	4951346.5		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.786	0.051	8.0E-08	0.687	-	0.886

STEP 24 Serum pH at 12 hours of age[a] replaced by Apgar score at 10 minutes[a] n = 381

Logistic Regression Model					
Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Major Abnormality on any Cranial US (assumed normal if no scan)	2.11	8.8E-07	8.2	3.5	19.0
Seizures	-2.07	0.00066	0.13	0.038	0.41
Apgar score at 10 minutes	-0.04	0.81	0.96	0.70	1.32
Constant	1.27	0.46	3.5		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.718	0.06	4.3E-05	0.601	-	0.836

Continued overleaf

STEP 25 Apgar score at 10 minutes[a] replaced by Pneumothorax n = 381

Logistic Regression Model					
Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Major Abnormality on any Cranial US (assumed normal if no scan)	2.09	9.6E-07	8.1	3.5	18.7
Seizures	-2.17	0.00040	0.11	0.034	0.36
Pneumothorax	-0.91	0.14	0.40	0.12	1.33
Constant	2.82	0.10	16.8		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.778	0.051	1.9E-07	0.678	-	0.879

STEP 26 Pneumothorax replaced by Other air leak n = 377

Logistic Regression Model					
Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Major Abnormality on any Cranial US (assumed normal if no scan)	2.00	4.0E-06	7.4	3.2	17.3
Seizures	-2.19	0.00027	0.11	0.035	0.36
Other Air Leak	-1.77	0.20	0.17	0.011	2.58
Constant	4.66	0.13	106.0		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.757	0.055	2.2E-06	0.650	-	0.864

STEP 27 Other air leak replaced by Pulmonary Interstitial Emphysema n = 377

Logistic Regression Model					
Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Major Abnormality on any Cranial US (assumed normal if no scan)	2.25	2.5E-07	9.5	4.0	22.3
Seizures	-1.76	0.0060	0.17	0.049	0.60
Pulmonary Interstitial Emphysema	-1.00	0.16	0.37	0.090	1.50
Constant	2.18	0.23	8.9		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.764	0.054	1.1E-06	0.657	-	0.871

STEP 28 Pulmonary Interstitial Emphysema replaced by Maternal Age n = 381

Logistic Regression Model					
Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Major Abnormality on any Cranial US (assumed normal if no scan)	2.13	6.0E-07	8.4	3.7	19.5
Seizures	-2.03	0.0010	0.13	0.039	0.44
Maternal Age	-0.039	0.27	0.96	0.90	1.03
Constant	1.96	0.19	7.1		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.800	0.044	1.9E-08	0.714	-	0.887

STEP 29 Maternal Age replaced by Probability of Survival[b] n = 381

Logistic Regression Model					
Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Major Abnormality on any Cranial US (assumed normal if no scan)	2.11	7.0E-07	8.3	3.6	19.0
Seizures	-2.05	0.00079	0.13	0.039	0.43
Probability of Survival[b]	-0.01	0.40	0.99	0.97	1.01
Constant	1.38	0.29	4.0		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.782	0.047	1.2E-07	0.690	-	0.875

Continued overleaf

STEP 30 Probability of Survival[a] replaced by Delivery for Chorioamnionitis n = 381

Logistic Regression Model					
Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Major Abnormality on any Cranial US (assumed normal if no scan)	2.15	4.3E-07	8.6	3.7	19.9
Seizures	-2.00	0.00085	0.14	0.0418	0.44
Delivery for Chorioamnionitis	-0.85	0.26	0.52	0.16	1.70
Constant	1.35	0.27	3.8		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.756820201	0.05518189	1.85448E-05	0.6478957	-	0.863974717

STEP 31 Delivery for Chorioamnionitis replaced by Time to Intubation[a] n = 381

Logistic Regression Model					
Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Major Abnormality on any Cranial US (assumed normal if no scan)	2.11	9.0E-07	8.2	3.5	19.1
Seizures	-2.09	0.00079	0.12	0.036	0.42
Time to Intubation[a]	-0.12	0.20	0.89	0.75	1.064
Constant	1.33	0.29	3.8		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.804	0.043	1.2E-08	0.721	-	0.887

STEP 32 Time to Intubation[a] replaced by Cardiac Massage & Adrenaline n = 381

Logistic Regression Model					
Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Major Abnormality on any Cranial US (assumed normal if no scan)	2.10	9.1E-07	8.2	3.5	18.9
Seizures	-2.11	0.00053	0.12	0.037	0.40
Cardiac Massage & Adrenaline	0.65	0.33	1.9	0.51	7.1
Constant	0.96	0.42	2.61		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.766	0.053	5.3E-07	0.665	-	0.871

STEP 33 Cardiac Massage & Adrenaline replaced by Mean blood pressure at 24 hours[a] n = 381

Logistic Regression Model					
Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Major Abnormality on any Cranial US (assumed normal if no scan)	2.09	1.23E-06	8.1	3.5	18.9
Seizures	-2.09	0.00053	0.12	0.038	0.40
Mean blood pressure at 24 hours[a]	-0.011	0.74	0.99	0.93	1.06
Constant	1.36	0.42	3.9		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.766	0.053	6.2E-07	0.662	-	0.870

STEP 34 Mean blood pressure at 24 hours[a] replaced by Gestational Age (days) n = 381

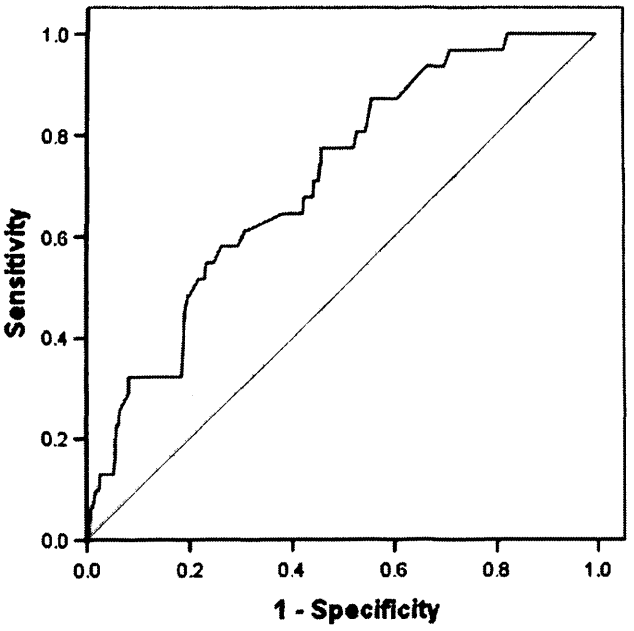
Logistic Regression Model					
Variable(s)	Coefficient	Significance	Odds Ratio	95.0% C.I. for Odds Ratio	
				Lower	Upper
Major Abnormality on any Cranial US (assumed normal if no scan)	2.11	7.5E-07	8.3	3.5	19.1
Seizures	-2.08	0.00061	0.12	0.038	0.41
Gestational Age (Days)	-0.0026	0.90	1.00	0.96	1.04
Constant	1.5	0.72	4.3		

Discrimination of Model Prediction (ROC Curve Analysis)					
Area under the curve (Az)	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval		
			Lower	-	Upper
0.758	0.051	1.4E-06	0.657	-	0.858

Figure 19: Ability of Models for Predicting Severe Disability at Two Years of Age (UKOS Cohort)

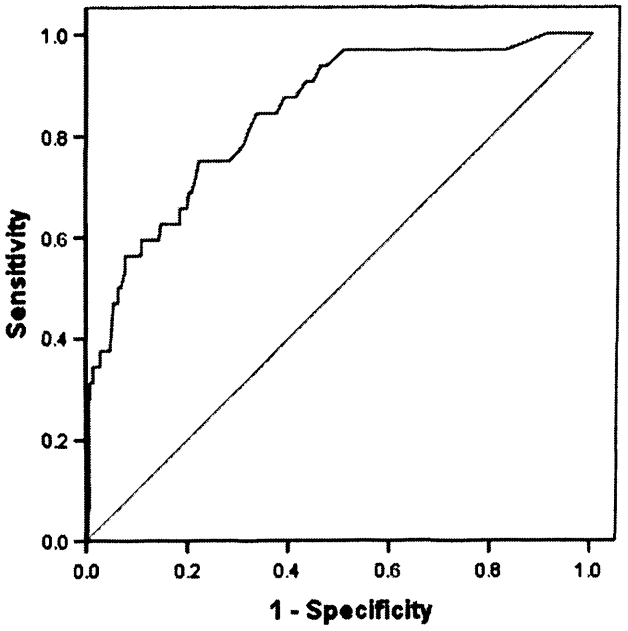
A. Model using information available before 12 hours of age.

(Az =0.713, 95% C.I.0.626– 0.799)



B. Model using information available before discharge.

(Az = 0.842, 95% C.I. 0.771 – 0.913)



4.4 DISCUSSION

This chapter describes the development of two models for predicting severe disability in survivors at two years of age after correcting for prematurity. Two models were developed in order to test two different hypotheses. These hypotheses were that it might be possible to predict severe disability in survivors at two years of age after correcting for prematurity using data either from the first 12 hours of life or from the entire admission respectively. This investigation suggests that it simply is not possible to do this using data from the first 12 hours alone. Indeed as can be seen in Table 18, many of the variables that were significantly associated with death and severe disability in Chapter 3.3 are either much more weakly associated, or not significantly associated, with severe disability amongst survivors at two years of age. An example of this is birthweight with a p value of 0.40 for severe disability in survivors in comparison to a p value of less than 0.00000000000000015 for death and disability (chapter 3.3). This finding suggests that the factors predicting death are not the same as those that predict disability. This picture is however confused by the fact that it is impossible to determine which of those infants who died would have developed severe disability.

The model for predicting severe disability using data from the entire admission does however appear to predict severe disability quite well, with an area under the ROC curve of 0.842. This model is small due to the application of Wasson's advice that the number of variables should be limited to a maximum of 1 variable for every 10 cases of the outcome of interest [200]. As there were only 32 cases of severe disability, three variables were therefore used in the model. It may be possible to

develop a better model for predicting severe disability by using a cohort containing more cases. This size does however mean that the model could be used easily in future studies of other cohorts of infants. The model is plausible, being made up of variables that have previously been shown to be markers of poor outcome[98, 211, 216-218]. Unfortunately the model cannot be used for comparing neonatal units for quality of care as the component variables, namely major abnormality on any cranial ultrasound scan, seizures and days of ventilation, are all likely to be affected by the quality of care[59]. Possible future uses of the model include as a surrogate marker for late outcome. In this situation the predicted risk of severe disability could be very useful as an early warning system for outcome at discharge, as waiting for two years to determine outcome is frustrating for parents, clinicians and researchers alike. In particular, the model might be very useful for identifying infants who would benefit from interventions that could be applied in infancy such as physiotherapy in an attempt to improve later outcome.

Missing data may have weakened the predictive value of the model. In order to perform the analysis without losing a large amount of data, missing values were replaced with the mean value of the infants for which the data existed. This approach could not be used for categorical variables. It is likely that this approach weakens the predictive association of the factor concerned. This was not a major problem for the factors selected in the final model, with 30 infants having a missing value for the presence of a major abnormality on any cranial ultrasound; these were assigned as having no abnormality as the most likely explanation was that a cranial ultrasound was not felt necessary. One infant had a missing value for seizures and was excluded from the analysis, two infants had a missing value for ventilation days and were assigned the mean value from the other infants. Complete data on some

of the other variables might however have altered the three variables that make up the model by altering the statistical selection process.

Loss of patients due to the inability to obtain information on outcome or perform the assessment between the prespecified age of 22 to 28 months was an important concern. Outcome in survivors was only available in this format for 65% of the infants. Studies have suggested that missing infants can have significantly different outcomes to included infants, although some studies have shown better outcome and some have shown a poorer outcome in those that are lost to follow up [134, 220-222]. There were no apparent differences between those assessed in the window and those excluded due to lack of information or an assessment outside the window. The possibility of bias must still however be born in mind when considering the results of this analysis. The potential for bias might have been improved by including the 55 infants who had assessments outside the prespecified age window, however it was felt that this approach was likely to reduce the accuracy of outcome assignment. This is because some children assessed early will be defined as having a poorer outcome than they would actually have in the window, whilst later assessments will give children longer to develop and therefore overscore their ability[147]. This approach was therefore not pursued in this study.

As mentioned elsewhere in the thesis, the trial cohort might also suffer from selection bias which can be introduced by the use of inclusion and exclusion criteria, as well as researcher's attitudes to recruiting patients into the trial. As mentioned previously in Chapter three, the inclusion of other predictive variables might have improved the model, variables such as socioeconomic status and antenatal infection in particular, warrant further investigation.

This model benefits from being derived from data collected from a large multicentre trial. This is likely to mean that the data is of good quality and drawn from infants of a wider social and ethnic background than a single unit or region might provide. The neurological assessments were performed by local paediatricians and therefore reflect the normal level of care and assessment although they might suffer from reporting bias. Further information may have been obtained by specific developmental testing[180], but the classification into severe disability was the outcome of interest in this thesis.

Previous attempts to predict neurodisabilty have also not been fully successful, although few publications have presented ROC curve analyses of model discrimination. An early model developed solely for predicting disability using data from the first day of life has not been published, apparently confirming the finding of this thesis that it is not possible to predict neurodevelopmental outcome using data from such early data. CRIB and SNAP, both scores for predicting mortality, have been tested in this way for predicting later outcome. These are detailed in section 1.5 of this thesis and elsewhere[67]. In summary, however the ability to predict disability separately from death using CRIB was inadequate (Az 0.70[91], 0.71[93], 0.77) except in the original cohort with the addition of cranial ultrasound results at 'around' 72 hours of age (Az = 0.89[95]). Unfortunately this timescale means that the score can no longer be used for risk correction to analyse quality of care. The late model presented in this chapter uses less data, making it easier to collect and is only a little less accurate. CRIB did appear to be a reasonable predictor for predicting the combined outcome of death or severe disability in the original cohort (Az = 0.83)[94]. This finding is however similar to that in chapter three of this thesis and is likely to represent the ability of the CRIB score to predict death as it was designed to do and not from its ability to predict disability.

SNAP, a similar but more time consuming risk score for predicting death has also been examined using retrospective data from 173 inborn infants[97]. This was not examined using ROC curve analysis and was modified by collecting information for every day of the infant's admission summed to produce a "cumulative SNAP score". The authors showed that the quartile of infants with the worst cumulative SNAP score had significantly lower motor development indices at 1 year, as well as lower psychomotor development indices at both 1 year and 3 years of age.

Brazy developed and tested a similar model for predicting neurodevelopmental outcome using data available at discharge, although it originally comprised 13 variables, these were shortened to just seven variables[98]. These were length of ventilation, serum pH, seizures, intraventricular haemorrhage, periventricular leukomalacia, infection and hypoglycaemia. Each of these was scored as a categorical variable recording a 0, 1, 2 or 4 points which were then summed to give a total score. As the derivation cohort only included 68 VLBW infants, the authors therefore disregarded Wasson suggestion for the maximum number of variables in a model according to the number of cases with the outcome of interest [200]. Expert opinion was also used instead of statistical methods to weight the categorical variables probably weakening the model's predictive ability. It is likely that the model developed in this thesis predicts disability better and is more reliable as it includes less variables and uses the continuous data for the length of ventilation. The NBRS model was used retrospectively by Lefebvre in 121 very premature infants born before 29 completed weeks of gestation who were assessed at 18 months[100]. Using score cut-offs defined by Brazy in the original NBRS paper, the model classified all the 26 infants with a severe neurosensory or developmental outcome as being at least at moderate risk (9 as infants at moderate risk, seventeen at high risk). An area

under the curve was only described for any disability (Az=0.79). This group included all the mild cases of cerebral palsy and infants with a Griffiths developmental quotient up to 90. The authors concluded that the model could still be used both for counselling parents and for identifying infants who should be offered close follow-up. The score presented in this thesis from the UKOS data is likely to be easier to collect as it relies on just 3 of the data points collected for NBRS. Ventilation days are also defined more precisely in the new model, giving more accuracy in the prediction.

4.5 COMMENTS AND IMPLICATIONS

This investigation suggests that it is possible to predict outcome at two years of age with reasonable accuracy using data available at discharge. Unfortunately it does not appear that data from earlier in the admission is adequate for this purpose. This means that risk correction cannot be used for comparing different units for the quality of their neurodevelopmental care. As the model was produced from a trial cohort of highly selected infants, it's generalisability to the general population needs to be tested. This could be achieved by validating it in another source of patients. Unfortunately the Trent Neonatal Survey data collection did not include ultrasound findings when outcome data at two years was collected for the ABC study analysed in chapter three of this thesis. Further neurodevelopmental outcome data is currently being collected for Trent infants born in 2001, 2002 and 2003 and the model could then be tested in this cohort.

Cranial ultrasound abnormalities, length of ventilation, and seizures are all likely to be affected significantly by quality of care. It may be worth measuring these

outcomes (perhaps using risk correction for these outcomes to assess quality of care) and also using the data to identify infants who would benefit from early intervention according to the late model. The model may also be useful as a guide in counselling parents about what the future may hold for their child and as a surrogate marker of later outcomes for randomised controlled trials or health economic assessments.

CHAPTER 5: COMMENTS AND CONCLUSIONS

5.1 THESIS AIMS

This thesis set out to develop a model for predicting severe disability or death at two years of age after correction for prematurity. After analysing the different cohorts in chapter two, this thesis presents investigations of predicting these outcomes in combination (death or severe disability against survival without severe disability) or separately (severe disability in survivors against other survivors). Chapter three documents the use of early data for this modeling, whereas chapter four uses data from the entire admission.

5.2 THESIS FINDINGS, IMPLICATIONS AND SUGGESTIONS FOR FUTURE RESEARCH

5.2.1 Predicting death or severe disability against survival without severe disability using early data

Three attempts were carried out to produce a suitable model for this task. On first glance the models produced in chapters 3.1, 3.2 and 3.3 all appeared to predict the combined outcome reasonably well. Indeed the area under the ROC curve values of 0.808, 0.793 and 0.798 are all considered good. However, a major finding of this thesis is that for all three models, further investigation demonstrated that these findings were due to the models being good at predicting death, but that they were very poor at predicting disability. There are four likely explanations for this consistent finding and these are discussed below in the order of importance in the view of the author of this thesis. Further research to answer these questions generated by this thesis would be extremely valuable.

The first possibility is that it simply is not possible to predict this combined outcome, as the factors related to survival are not the same as those that are related to severe disability. This is supported by the ability of the models to predict death accurately. A large number of variables were analysed in the UKOS and East Anglian cohorts making it less likely that these predictions could be substantially improved by collecting more variables. There were however some important variables overlooked by the cohorts that could have been collected at the time of birth. These include the presence of antenatal infection and socioeconomic status. To determine whether disability can be predicted using data from early in life, a large prospective cohort study collecting these and many other variables is required.

The second explanation which may well explain the inability to predict outcome using early data, is that events which occur later in an infant's stay are much more important than the baseline risk that is present at birth. These variables include the occurrence of things that either damage the brain, such as significant intraventricular haemorrhage and meningitis, or variables that highlight that there has been a problem such as seizures. The results of Chapter four in this thesis, where data from the entire admission were used to develop a model that predicted disability adequately (Az= 0.842), suggest that this is the case. This interpretation is supported by the results of Lefebvre who applied the NNBRS to data from the entire admission producing an area under the curve of 0.79 for outcome at 18 months of age[100]. If this scenario is true, comparisons of late outcomes between units or groups of individuals are likely to reflect the quality of care given to infants during their entire stay on the neonatal unit. The three variables combined in the model developed in chapter four could be relatively easily collected and therefore used to compare the quality of care in different units.

The third possibility is that this methodology for looking at early data and predicting a combined outcome of death and severe disability is flawed, as it is impossible to know what the neurodevelopmental outcome would have been for the infants who died. In order to test this, future research should use more complicated statistical methods whereby the likely neurological outcome is used for those infants who died. This situation is analogous to that of banking statistics and risk prediction. Mortgage companies attempt to predict who would default on a mortgage in order to determine risk. As they refuse credit to a significant number of individuals, they use credit scoring methodology[223] to determine the risk of defaulting for those people who are not given a mortgage.

The fourth possibility is that proportionally more infants died, than developed severe disability in each of the cohorts, thereby causing the models to be better at predicting deaths than disability. This could be investigated in more detail by including the moderately disabled infants in the disabled group so balancing the groups better. This is unlikely to be worthwhile however, as the relationship between the predictive variables and the outcome at two years of age will probably be weakened by including these infants due to the lower level of outcome severity. An alternative is to examine groups of infants at higher risk of death and disability. Such an approach would limit the models use to a smaller group of babies, such as those weighing less than 1 kilogram at birth, or being born before 28 weeks of gestation.

5.2.2 Predicting severe disability in survivors against survival without severe disability using late data

As described above in chapter four in this thesis, data from the entire admission were successfully used to develop a model that predicted disability adequately (Az= 0.842). This suggests that the experiences in the first few weeks of life may be the most important predictors of future outcome. The model offers many exciting potential uses both for research and in clinical use. These include giving prognostic information when counseling parents at discharge, identifying infants that might benefit from early developmental interventions, or as surrogate markers for later outcome. The model could also be used in cohorts for comparing study groups for similarity of risk, or for measuring trends in results over time. The use of the model as a marker of quality of care also warrants further investigation.

Before the model can be used in these ways the calibration of the model should be tested using a goodness of fit test such as that of Hosmer and Lemeshow[83]. It then needs validating in other cohorts to test its generalisability to the general population. In order to establish the model, reliability testing is also important as the model needs to be easy to use and not open to subjective assessments [224]. This is unlikely to be a significant problem but should be tested as ultrasound results and identifying whether an infant has had seizures are dependent on observer interpretation.

Future studies should also explore other modalities of imaging that could improve the accuracy of predictions. Magnetic resonance imaging is currently being examined in this way in many centres across the world. As this produces a higher resolution image of the brain, it might be possible to substantially improve predictions of outcome at two years of age using MRI instead of ultrasonography.

APPENDIX 1: SAMPLE WEEFIM RATING FORM. ADAPTED FROM THE FUNCTIONAL INDEPENDENCE MEASURE FOR CHILDREN (WEEFIM)

7 Complete independence (Timely, Safely)	No helper
6 Modified independence (Device)	
Modified dependence	
5 Supervision	Helper
4 Minimal assistance (Subject=75%+)	
3 Moderate assistance (Subject=50%+)	
Complete dependence	
2 Maximal assistance (Subject=25%+)	
1 Total assistance (Subject=0%+)	

Self-care or follow up

A. Eating ———
B. Grooming ———
C. Bathing ———
D. Dressing – upper body ———
E. Dressing – lower body ———
F. Toileting ———

Sphincter control

G. Bladder management ———
H. Bowel management ———

Transfers

I. Chair/wheelchair ———
J. Toilet ———
K. Tub/shower ———

Locomotion

L. Walk/wheelchair/crawl ———
M.Stairs ———

Communication

N. Comprehension ———
O. Expression ———

Social cognition

P. Social interaction ———
Q. Problem solving ———
R. Memory ———

Total WeeFIM ——

APPENDIX 2: OHSQ CLASSIFICATION OF DISABILITY

Definitions	Severe	Moderate	Mild
Respiration	O_2 >1 hour a day or mechanical ventilation	Limited exercise tolerance on drugs	Limited exercise tolerance - no drugs
Renal	Treated with dialysis	Treated with drugs or diet	Impairment but no treatment
Feeding	TPN or tube-feeding or stoma fed		Special diet
Fits	More than one seizure a month on treatment	Less than one seizure a month on treatment	No treatment required or no seizures on treatment
Walking	Unable to walk without assistance	Reduced mobility	Non fluent gait
Sitting	Unable to sit Sits only with support or sits unsupported but unstable		
Hand use	Unable to feed self	Some difficulty both hands	Some difficulty one hand
Head control	Unstable without support or no head control		
Vision	Blind or sees light only in two eyes	Blind in one eye only or not fully correctable	Normal vision with correction
Hearing	Impairment not corrected with aids (or hearing loss both ears worse than 70 dB)	Impairment corrected with aids or hearing in both ears between 40 and 70 dB	
Communication	No speech or vocabulary <6 words or unable to understand in cued situation	Vocabulary <10 words or unable to understand in unfamiliar situation	Single words only-vocabulary <10 words
Developmental age at 2 years	<12 months	12 to 18 months	>18 months <24 months

REFERENCES

1. Lorenz JM, Paneth N, Jetton JR, den Ouden L, Tyson JE. Comparison of management strategies for extreme prematurity in New Jersey and the Netherlands: outcomes and resource expenditure. *Pediatrics* 2001; Dec;108(6):1269-74.

2. Cuttini M, Nadai M, Kaminski M, Hansen G, de Leeuw R, Lenoir S, et al. End-of-life decisions in neonatal intensive care: physicians' self-reported practices in seven European countries. EURONIC Study Group. *Lancet* 2000; Jun 17;355(9221):2112-8.

3. Rhoden NK. Treating Baby Doe: the ethics of uncertainty. *Hastings Cent Rep* 1986; Aug;16(4):34-42.

4. Patterson C. "Not Worth the Rearing": The Causes of Infant Exposure in Ancient Greece. *Transactions of the American Philological Association 1985;* 1985;115:103-123.

5. SENECA LA. *De la Ira. Text i traduccio\0301 del Dr. Carles Cardo\0301.* Barcelona; 1924.

6. Jackson M. Infanticide. *Lancet* 2006; Mar 11;367(9513):809.

7. PERCIVAL HR. *The Seven Ecumenical Councils of the Undivided Church. Their canons and dogmatic decrees, together with the canons of all the local synods which have received ecumenical acceptance. Edited with notes gathered from the writings of the greatest scholars by H. R. Percival.* ; 1900.

8. Risse GB. Health care in hospitals: the past 1000 years. *Lancet* 1999; Dec;354 Suppl:SIV25.

9. McGrew RE. *Encyclopedia of medical history.* New York: McGraw-Hill; 1985.

10. Dunn PM. The birth of perinatal medicine in the United Kingdom. *Semin Fetal Neonatal Med* 2006; Dec;11(6):386-97.

11. Loudon I. Deaths in childbed from the eighteenth century to 1935. *Med Hist* 1986; Jan;30(1):1-41.

12. Dunn PM. George Armstrong MD (1719-1789) and his dispensary for the infant poor. *Arch Dis Child Fetal Neonatal Ed* 2002; Nov;87(3):F228-31.

13. Williams AN. "The joy to bless and to relieve mankind": child healthcare at Northampton General Infirmary 1744. *Arch Dis Child* 2005; Dec;90(12):1227-9.

14. British Paediatric Association (LONDON). *The British Paediatric Association, 1928-1952.* London; 1955.

15. Crosse VM. *The Premature Baby.* J. & A. Churchill Ltd; 1945.

16. Rivett G. *From cradle to grave : fifty years of the NHS.* London: King's Fund; 1998.

17. Diamond LK. Replacement Transfusion as a Treatment for Erythroblastosis Fetalis. *Pediatrics* 1948; November 1;2(5):520-4.

18. CLIFFORD SH. Postmaturity, with placental dysfunction; clinical syndrome and pathologic findings. *J Pediatr* 1954; Jan;44(1):1-13.

19. An international classification of retinopathy of prematurity. Prepared by an international committee. *Br J Ophthalmol* 1984; Oct;68(10):690-7.

20. Pearson HA, Anunziato D, Baker JP, Gartner LM, Howell DA, Strain JE, et al. Committee report: American Pediatrics: milestones at the millennium. *Pediatrics* 2001; Jun;107(6):1482-91.

21. Newman G. *Infant mortality : a social problem.* Methuen; 1906.

22. Thatham J. 68th report of the Registrar General of Births, Marriages and Deaths in England and Wales 1907 London: HMSO; 1905.

23. Office for National Statistics. Birth Statistics, London: ONS; 2005.

24. Rogowski J. Cost-effectiveness of care for very low birth weight infants. *Pediatrics* 1998; Jul;102(1 Pt 1):35-43.

25. Rogowski J. Measuring the cost of neonatal and perinatal care. *Pediatrics* 1999; Jan;103(1 Suppl E):329-35.

26. Petrou S, Mehta Z, Hockley C, Cook-Mozaffari P, Henderson J, Goldacre M. The impact of preterm birth on hospital inpatient admissions and costs during the first 5 years of life. *Pediatrics* 2003; Dec;112(6 Pt 1):1290-7.

27. Petrou S, Henderson J, Bracewell M, Hockley C, Wolke D, Marlow N, et al. Pushing the boundaries of viability: the economic impact of extreme preterm birth. *Early Hum Dev* 2006; Feb;82(2):77-84.

28. Simpson JH, Lynch R, Grant J, Alroomi L. Reducing medication errors in the neonatal intensive care unit. *Arch Dis Child Fetal Neonatal Ed* 2004; Nov 89(6):F480-2.

29. Snijders C, van Lingen RA, Molendijk H, Fetter WP. Incidents and errors in neonatal intensive care: A review of the literature. *Arch Dis Child Fetal Neonatal Ed* 2007;92: F391-F398.

30. *Learning from Bristol : the report of the public inquiry into children's heart surgery at the Bristol Royal Infirmary 1984-1995.* London. The Stationery Office; 2001.

31. Graunt J. Natural and Political Observations on the Bills of Mortality. London. Roycroft;1676.

32. Anderson RN, Rosenberg HM. Age standardization of death rates: implementation of the year 2000 standard. *Natl Vital Stat Rep* 1998; Oct 7;47(3):1,16, 20.

33. Treas J. *Older Americans in the 1990s and beyond.* Washington, D.C.: Population Reference Bureau; 1995.

34. Gould JB. Vital records for quality improvement. *Pediatrics* 1999; Jan;103(1 Suppl E):278-90.

35. Brimblecombe FS, Ashford JR. Significance of low birth weight in perinatal mortality. A study of variations within England and Wales. *Br J Prev Soc Med* 1968; Jan;22(1):27-35.

36. Ashford JR, Read KL, Riley VC. An analysis of variations in perinatal mortality amongst local authorities in England and Wales. *Int J Epidemiol* 1973; Spring;2(1):31-46.

37. Mallett R, Knox EG. Standardized perinatal mortality ratios: technique, utility and interpretation. *Community Med* 1979; Feb;1(1):6-13.

38. Kleinman JC. Indirect standardization of neonatal mortality for birth weight. *Int J Epidemiol* 1982; Jun;11(2):146-54.

39. Moore W. On some of the more prominent causes of excessive mortality in early life. National Association for the promotion of Social Science. Dublin; 1861.

40. Percival T. Observations on the state of population in Manchester; and other adjacent places, 1789.

41. Farr W. Mortality of Children in the Principal States of Europe. *Journal of the Statistical Society of London* 1866;29(1):1-12.

42. Dummer TJ, Parker L. Changing socioeconomic inequality in infant mortality in Cumbria. *Arch Dis Child* 2005; Feb;90(2):157-62.

43. Richardson D, Tarnow-Mordi WO, Lee SK. Risk adjustment for quality improvement. *Pediatrics* 1999; Jan;103(1 Suppl E):255-65.

44. Tarnow-Mordi W, Ogston S, Wilkinson AR, Reid E, Gregory J, Saeed M, et al. Predicting death from initial disease severity in very low birthweight infants: a method for comparing the performance of neonatal units. *BMJ* 1990; Jun 23;300(6740):1611-4.

45. The UK Neonatal Staffing Study,Group. Patient volume, staffing, and workload in relation to risk-adjusted outcomes in a random stratified sample of UK neonatal intensive care units: a prospective evaluation. *The Lancet* 2002/1/12;359(9301):99-107.

46. Herridge MS. Prognostication and intensive care unit outcome: the evolving role of scoring systems. *Clin Chest Med* 2003; Dec;24(4):751-62.

47. Thompson SW, McClure BG, Tubman TR. A randomized, controlled trial of parenteral glutamine in ill, very low birth-weight neonates. *J Pediatr Gastroenterol Nutr* 2003; Nov;37(5):550-3.

48. McClure RJ, Newell SJ. Randomised controlled study of clinical outcome following trophic feeding. *Arch Dis Child Fetal Neonatal Ed* 2000; Jan;82(1):F29-33.

49. Luce JM, Wachter RM. The ethical appropriateness of using prognostic scoring systems in clinical management. *Crit Care Clin* 1994; Jan;10(1):229-41.

50. Marcin JP, Pollack MM, Patel KM, Ruttimann UE. Decision support issues using a physiology based score. *Intensive Care Med* 1998; Dec;24(12):1299-304.

51. Zollo MB, Moskop JC, Kahn CE,Jr. Knowing the score: using predictive scoring systems in clinical practice. *Am J Crit Care* 1996; Mar;5(2):147-51.

52. Ridley SA. Uncertainty and scoring systems. *Anaesthesia* 2002; Aug;57(8):761-7.

53. Goh AY, Mok Q. Identifying futility in a paediatric critical care setting: a prospective observational study. *Arch Dis Child* 2001; Mar;84(3):265-8.

54. Sachdeva RC, Jefferson LS, Coss-Bu J, Brody BA. Resource consumption and the extent of futile care among patients in a pediatric intensive care unit setting. *J Pediatr* 1996; Jun;128(6):742-7.

55. Poloniecki J. Half of all doctors are below average. *BMJ* 1998; Jun 6;316(7146):1734-6.

56. Cockburn F, Cooke RWI, Gamsu HR, Greenough A, Hopkins A, McIntosh N, et al. The CRIB (clinical risk index for babies) score: A tool for assessing initial neonatal risk and comparing performance of neonatal intensive care units. *Lancet* 1993;342(8865):193-8.

57. Kaaresen PI, Dohlen G, Fundingsrud HP, Dahl LB. The use of CRIB (clinical risk index for babies) score in auditing the performance of one neonatal intensive care unit. *Acta Paediatr* 1998; Feb;87(2):195-200.

58. Lyon A. How should we report neonatal outcomes?. *Semin Fetal Neonatal Med* 2007; Jun 15;.

59. Pollack MM, Koch MA, Bartel DA, Rapoport I, Dhanireddy R, El-Mohandes AAE, et al. A comparison of neonatal mortality risk prediction models in very low birth weight infants. *Pediatrics* 2000;105(5):1051-7.

60. Risk adjusted and population based studies of the outcome for high risk infants in Scotland and Australia. International Neonatal Network, Scottish Neonatal Consultants, Nurses Collaborative Study Group. *Arch Dis Child Fetal Neonatal Ed* 2000; Mar;82(2):F118-23.

61. Kahn DJ, Richardson DK, Gray JE, Bednarek F, Rubin LP, Shah B, et al. Variation among neonatal intensive care units in narcotic administration. *Arch Pediatr Adolesc Med* 1998; Sep;152(9):844-51.

62. Vyas J, Field D, Draper ES, Woodruff G, Fielder AR, Thompson J, et al. Severe retinopathy of prematurity and its association with different rates of survival in infants of less than 1251 g birth weight. *Arch Dis Child Fetal Neonatal Ed* 2000; Mar;82(2):F145-9.

63. Iezzoni LI, Ash AS, Shwartz M, Daley J, Hughes JS, Mackiernan YD. Judging hospitals by severity-adjusted mortality rates: the influence of the severity-adjustment method. *Am J Public Health* 1996; Oct;86(10):1379-87.

64. Fleisher BE, Murthy L, Lee S, Constantinou JC, Benitz WE, Stevenson DK. Neonatal severity of illness scoring systems: a comparison. *Clin Pediatr (Phila)* 1997; Apr;36(4):223-7.

65. Noninitiation or withdrawal of intensive care for high-risk newborns. *Pediatrics* 2007; Feb;119(2):401-3.

66. Stevens SM, Richardson DK, Gray JE, Goldmann DA, McCormick MC. Estimating neonatal mortality risk: an analysis of clinicians' judgments. *Pediatrics* 1994; Jun;93(6 Pt 1):945-50.

67. Dorling JS, Field DJ, Manktelow B. Neonatal disease severity scoring systems. *Arch Dis Child Fetal Neonatal Ed* 2005; Jan;90(1):F11-6.

68. Maier RF, Caspar-Karweck UE, Grauel EL, Bassir C, Metze BC, Obladen M. A comparison of two mortality risk scores for very low birthweight infants: clinical risk index for babies and Berlin score. *Intensive Care Med* 2002; Sep;28(9):1332-5.

69. Rautonen J, Makela A, Boyd H, Apajasalo M, Pohjavuori M. CRIB and SNAP: assessing the risk of death for preterm neonates. *Lancet* 1994; May 21;343(8908):1272-3.

70. Bastos G, Gomes A, Oliveira P, da Silva AT. A comparison of 4 pregnancy assessment scales (CRIB, SNAP, SNAP-PE, NTISS) in premature newborns. Clinical Risk Index for Babies. Score for Neonatal Acute Physiology. Score for Neonatal Acute Physiology-Perinatal Extension. Neonatal Therapeutic Intervention Scoring System. *Acta Med Port* 1997; Feb-Mar;10(2-3):161-5.

71. Hughes-Davies TH. The CRIB score. *Lancet* 1993; Oct 9;342(8876):938.

72. Tarnow-Mordi W, Parry G. The CRIB score. *Lancet* 1993; Nov 27;342(8883):1365.

73. Parry G, Tucker J, Tarnow-Mordi W, UK Neonatal Staffing Study Collaborative Group. CRIB II: an update of the clinical risk index for babies score. *Lancet* 2003; May 24;361(9371):1789-91.

74. Draper ES, Manktelow B, Field DJ, James D. Prediction of survival for preterm births by weight and gestational age: retrospective population based study. *BMJ* 1999; Oct 23;319(7217):1093-7.

75. Richardson DK, Gray JE, McCormick MC, Workman K, Goldmann DA. Score for Neonatal Acute Physiology: A physiologic severity index for neonatal intensive care. *Pediatrics* 1993;91(3):617-23.

76. Richardson DK, Phibbs CS, Gray JE, McCormick MC, Workman-Daniels K, Goldmann DA. Birth weight and illness severity: Independent predictors of neonatal mortality. *Pediatrics* 1993;91(5 I):969-75.

77. Richardson DK, Corcoran JD, Escobar GJ, Lee SK. SNAP-II and SNAPPE-II: Simplified newborn illness severity and mortality risk scores. *J Pediatr* 2001; Jan;138(1):92-100.

78. Gray JE, Richardson DK, McCormick MC, Workman-Daniels K, Goldmann DA. Neonatal therapeutic intervention scoring system: a therapy-based severity-of-illness index. *Pediatrics* 1992; Oct;90(4):561-7.

79. Horbar JD, Onstad L, Wright E. Predicting mortality risk for infants weighing 501 to 1500 grams at birth: a National Institutes of Health Neonatal Research Network report. *Crit Care Med* 1993; Jan;21(1):12-8.

80. Garcia H, Villegas-Silva R, Villanueva-Garcia D, Gonzalez-Cabello H, Lopez-Padilla M, Fajardo-Gutierrez A, et al. Validation of a prognostic index in the critically ill newborn. *Rev Invest Clin* 2000; Jul-Aug;52(4):406-14.

81. Doyle LW, Victorian Infant Collaborative Study Group. Neonatal intensive care at borderline viability--is it worth it?. *Early Hum Dev* 2004; Nov;80(2):103-13.

82. Altman DG, Royston P. What do we mean by validating a prognostic model?. *Stat Med* 2000; Feb 29;19(4):453-73.

83. Hosmer DW. *Applied logistic regression.* 2nd ed. New York ; Chichester: Wiley; 2000.

84. van Erkel AR, Pattynama PM. Receiver operating characteristic (ROC) analysis: basic principles and applications in radiology. *Eur J Radiol* 1998; May;27(2):88-94.

85. Ward K. Perceived needs of parents of critically ill infants in a neonatal intensive care unit (NICU). *Pediatr Nurs* 2001; May-Jun;27(3):281-6.

86. Dyer O. Parents of disabled baby lose appeal against court order. *BMJ* 2005; Sep 3;331(7515):472.

87. Dyer C. Doctors need not ventilate baby to prolong his life. *BMJ* 2004; Oct 30;329(7473):995.

88. Parry G, Billingham K, Gibson A. Infant mortality rate in Sheffield. *Arch Dis Child Fetal Neonatal Ed* 2000; May;82(3):F257.

89. Field DJ, Draper ES, Manktelow BN. *Trent Neonatal Survey Report 2005—One of The Infant Mortality and Morbidity Studies.* 1st ed. Leicester: University of Leicester; 2006.

90. Stc Hamilton KE, Redshaw ME, Tarnow-Mordi W. Nurse staffing in relation to risk-adjusted mortality in neonatal care. *Arch Dis Child Fetal Neonatal Ed* 2007; Mar;92(2):F99-F103.

91. Buhrer C, Grimmer I, Metze B, Obladen M. The CRIB (Clinical Risk Index for Babies) score and neurodevelopmental impairment at one year corrected age in very low birth weight infants. *Intensive Care Med* 2000; Mar;26(3):325-9.

92. Lago P, Freato F, Bettiol T, Chiandetti L, Vianello A, Zaramella P. Is the CRIB score (clinical risk index for babies) a valid tool in predicting neurodevelopmental outcome inExtremely low birth weight infants?. *Biol Neonate* 1999; Oct;76(4):220-7.

93. Fowlie PW, Gould CR, Tarnow-Mordi WO, Strang D. Measurement properties of the Clinical Risk Index for Babies--reliabilty, validity beyond the first 12 hours, and responsiveness over 7 days. *Crit Care Med* 1998; Jan;26(1):163-8.

94. Pharoah PO. CRIB and impairment after neonatal intensive care. Clinical risk index for babies. *Lancet* 1995; Jul 1;346(8966):58-9.

95. Fowlie PW, Tarnow-Mordi WO, Gould CR, Strang D. Predicting outcome in very low birthweight infants using an objective measure of illness severity and cranial ultrasound scanning. *Arch Dis Child Fetal Neonatal Ed* 1998; May;78(3):F175-8.

96. Reynolds PR, Dale RC, Cowan FM. Neonatal cranial ultrasound interpretation: a clinical audit. *Arch Dis Child Fetal Neonatal Ed* 2001; Mar;84(2):F92-5.

97. Mattia FR, deRegnier RA. Chronic physiologic instability is associated with neurodevelopmental morbidity at one and two years in extremely premature infants. *Pediatrics* 1998; Sep;102(3):E35.

98. Brazy JE, Eckerman CO, Oehler JM, Goldstein RF, O'Rand AM. Nursery Neurobiologic Risk Score: important factor in predicting outcome in very low birth weight infants. *J Pediatr* 1991; May;118(5):783-92.

99. Nunes A, Melo F, Silva JE, Costa A, Bispo MA, Palminha JM. Importance of J. Brazy's neurobiological index. Prediction of the number and severity of complications in very low birth weight infants. *Acta Med Port* 1998; Jul;11(7):615-21.

100. Lefebvre F, Gregoire MC, Dubois J, Glorieux J. Nursery Neurobiologic Risk Score and outcome at 18 months. *Acta Paediatr* 1998; Jul;87(7):751-7.

101. Contractor CP, Leslie GI, Bowen JR, Arnold JD. The Neonatal Neurobiologic Risk Score: does it predict outcome in very premature infants?. *Indian Pediatr* 1996; Feb;33(2):95-101.

102. Schreuder AM, Veen S, Ens-Dokkum MH, Verloove-Vanhorick SP, Brand R, Ruys JH. Standardised method of follow-up assessment of preterm infants at the age of 5 years: use of the WHO classification of impairments, disabilities and handicaps. Report from the collaborative Project on Preterm and Small for gestational age infants (POPS) in The Netherlands, 1983. *Paediatr Perinat Epidemiol* 1992; Jul;6(3):363-80.

103. Veen S, Ens-Dokkum MH, Schreuder AM, Verloove-Vanhorick SP, Brand R, Ruys JH. Impairments, disabilities, and handicaps of very preterm and very-low-birthweight infants at five years of age. The Collaborative Project on Preterm and Small for Gestational Age Infants (POPS) in The Netherlands. *Lancet* 1991; Jul 6;338(8758):33-6.

104. Msall ME, DiGaudio K, Duffy LC, LaForest S, Braun S, Granger CV. WeeFIM. Normative sample of an instrument for tracking functional independence in children. *Clin Pediatr (Phila)* 1994; Jul;33(7):431-8.

105. Haley SM, New England Medical Center Hospital. PEDI Research Group. *Pediatric evaluation of disability inventory (PEDI) : development, standardization and administration manual.* Boston, MA: New England Medical Center Hospital, PEDI Research Group; 1992.

106. World Health Organization. International classification of functioning, disability, and health 2001;.

107. Jessen EC, Colver AF, Mackie PC, Jarvis SN. Development and validation of a tool to measure the impact of childhood disabilities on the lives of children and their families. *Child Care Health Dev* 2003; Jan;29(1):21-34.

108. National Perinatal Epidemiology Unit (Great Britain), Oxford Regional Health Authority. *Disability and perinatal care : measurement of health status at two years.* National Perinatal Epidemiology Unit; 1994.

109. Jones HP, Guildea ZE, Stewart JH, Cartlidge PH. The Health Status Questionnaire: achieving concordance with published disability criteria. *Arch Dis Child* 2002; Jan;86(1):15-20.

110. McCormick MC. The outcomes of very low birth weight infants: are we asking the right questions?. *Pediatrics* 1997; Jun;99(6):869-76.

111. Mutch L, Alberman E, Hagberg B, Kodama K, Perat MV. Cerebral palsy epidemiology: where are we now and where are we going?. *Dev Med Child Neurol* 1992; Jun;34(6):547-51.

112. Derrick M, Drobyshevsky A, Ji X, Tan S. A model of cerebral palsy from fetal hypoxia-ischemia. *Stroke* 2007; Feb;38(2 Suppl):731-5.

113. Nelson KB, Ellenberg JH. Children who "outgrew' cerebral palsy. *Pediatrics* 1982; May;69(5):529-36.

114. Marlow N, Wolke D, Bracewell MA, Samara M, EPICure Study Group. Neurologic and developmental disability at six years of age after extremely preterm birth. *N Engl J Med* 2005; Jan 6;352(1):9-19.

115. Testa MA, Simonson DC. Assesment of quality-of-life outcomes. *N Engl J Med* 1996; Mar 28;334(13):835-40.

116. Saigal S, Stoskopf BL, Feeny D, Furlong W, Burrows E, Rosenbaum PL, et al. Differences in preferences for neonatal outcomes among health care professionals, parents, and adolescents. *JAMA* 1999; Jun 2;281(21):1991-7.

117. Jenney ME, Campbell S. Measuring quality of life. *Arch Dis Child* 1997; Oct;77(4):347-50.

118. Theunissen NC, Vogels TG, Koopman HM, Verrips GH, Zwinderman KA, Verloove-Vanhorick SP, et al. The proxy problem: child report versus parent report in health-related quality of life research. *Qual Life Res* 1998; Jul;7(5):387-97.

119. Doyle LW, Victorian Infant Collaborative Study Group. Changing availability of neonatal intensive care for extremely low birthweight infants in Victoria over two decades. *Med J Aust* 2004; Aug 2;181(3):136-9.

120. Saigal S, Rosenbaum P, Stoskopf B, Hoult L, Furlong W, Feeny D, et al. Comprehensive assessment of the health status of extremely low birth weight

children at eight years of age: comparison with a reference group. *J Pediatr* 1994; Sep;125(3):411-7.

121. Saigal S, Feeny D, Furlong W, Rosenbaum P, Burrows E, Torrance G. Comparison of the health-related quality of life of extremely low birth weight children and a reference group of children at age eight years. *J Pediatr* 1994; Sep;125(3):418-25.

122. Saigal S, Feeny D, Rosenbaum P, Furlong W, Burrows E, Stoskopf B. Self-perceived health status and health-related quality of life of extremely low-birth-weight infants at adolescence. *JAMA* 1996; Aug 14;276(6):453-9.

123. Hack M. Consideration of the use of health status, functional outcome, and quality-of-life to monitor neonatal intensive care practice. *Pediatrics* 1999; Jan;103(1 Suppl E):319-28.

124. Donohue PK. Health-related quality of life of preterm children and their caregivers. *Ment Retard Dev Disabil Res Rev* 2002;8(4):293-7.

125. Aylward GP. Cognitive and neuropsychological outcomes: more than IQ scores. *Ment Retard Dev Disabil Res Rev* 2002;8(4):234-40.

126. Rickards AL, Kitchen WH, Doyle LW, Kelly EA. Correction of developmental and intelligence test scores for premature birth. *Aust Paediatr J* 1989; Jun;25(3):127-9.

127. Bayley N. *Manual for the Bayley Scales of Infant Development.* New York: Psychological Corporation; 1969.

128. Bayley N. *Bayley Scales of Infant Development* New York: Psychological Corporation; 1993.

129. Bayley N, Psychological Corporation. *Bayley scales of infant and toddler development.* 3rd ed. New York: Psychological Corporation; 2006.

130. Marlow N. Neurocognitive outcome after very preterm birth. *Arch Dis Child Fetal Neonatal Ed* 2004; May;89(3):F224-8.

131. Glascoe FP, Byrne KE, Ashford LG, Johnson KL, Chang B, Strickland B. Accuracy of the Denver-II in developmental screening. *Pediatrics* 1992; Jun;89(6 Pt 2):1221-5.

132. Shonkoff JP, Hauser-Cram P. Early intervention for disabled infants and their families: a quantitative analysis. *Pediatrics* 1987; Nov;80(5):650-8.

133. Kirby RS, Swanson ME, Kelleher KJ, Bradley RH, Casey PH. Identifying at-risk children for early intervention services: lessons from the Infant Health and Development Program. *J Pediatr* 1993; May;122(5 Pt 1):680-6.

134. Tin W, Fritz S, Wariyar U, Hey E. Outcome of very preterm birth: children reviewed with ease at 2 years differ from those followed up with difficulty. *Arch Dis Child Fetal Neonatal Ed* 1998; Sep;79(2):F83-7.

135. Anderson P, Doyle LW, Victorian Infant Collaborative Study Group. Neurobehavioral outcomes of school-age children born extremely low birth weight or very preterm in the 1990s. *JAMA* 2003; Jun 25;289(24):3264-72.

136. Doyle LW, Casalaz D, Victorian Infant Collaborative Study Group. Outcome at 14 years of extremely low birthweight infants: a regional study. *Arch Dis Child Fetal Neonatal Ed* 2001; Nov;85(3):F159-64.

137. Blasco PA. Preterm birth: to correct or not to correct. *Dev Med Child Neurol* 1989; Dec;31(6):816-21.

138. O'Shea TM, Kothadia JM, Klinepeter KL, Goldstein DJ, Jackson BG, Weaver RG,3rd, et al. Randomized placebo-controlled trial of a 42-day tapering course of dexamethasone to reduce the duration of ventilator dependency in very low birth weight infants: outcome of study participants at 1-year adjusted age. *Pediatrics* 1999; Jul;104(1 Pt 1):15-21.

139. Shinwell ES, Karplus M, Reich D, Weintraub Z, Blazer S, Bader D, et al. Early postnatal dexamethasone treatment and increased incidence of cerebral palsy. *Arch Dis Child Fetal Neonatal Ed* 2000; Nov;83(3):F177-81.

140. Dobrez D, Sasso AL, Holl J, Shalowitz M, Leon S, Budetti P. Estimating the cost of developmental and behavioral screening of preschool children in general pediatric practice. *Pediatrics* 2001; Oct;108(4):913-22.

141. Skellern CY, Rogers Y, O'Callaghan MJ. A parent-completed developmental questionnaire: follow up of ex-premature infants. *J Paediatr Child Health* 2001; Apr;37(2):125-9.

142. Pritchard MA, Colditz PB, Beller EM, Queensland Optimising Preterm Infant Outcomes Group. Parents' evaluation of developmental status in children born with a birthweight of 1250 g or less. *J Paediatr Child Health* 2005; Apr;41(4):191-6.

143. Bricker D, Squires J, Kaminski R, Mounts L. The validity, reliability, and cost of a parent-completed questionnaire system to evaluate at-risk infants. *J Pediatr Psychol* 1988; Mar;13(1):55-68.

144. Johnson S, Marlow N, Wolke D, Davidson L, Marston L, O'Hare A, et al. Validation of a parent report measure of cognitive development in very preterm infants. *Dev Med Child Neurol* 2004; Jun;46(6):389-97.

145. McCormick MC, Athreya BH, Bernbaum JC, Charney EB. Preliminary observations on maternal rating of health of children: data from three subspecialty clinics. *J Clin Epidemiol* 1988;41(4):323-9.

146. Janson H, Squires J. Parent-completed developmental screening in a Norwegian population sample: a comparison with US normative data. *Acta Paediatr* 2004; Nov;93(11):1525-9.

147. Field D, Draper ES, Gompels MJ, Green C, Johnson A, Shortland D, et al. Measuring later health status of high risk infants: randomised comparison of two simple methods of data collection. *BMJ* 2001; Dec 1;323(7324):1276-81.

148. Stevenson RC, Pharoah PO, Stevenson CJ, McCabe CJ, Cooke RW. Cost of care for a geographically determined population of low birthweight infants to age 8-9 years. II. Children with disability. *Arch Dis Child Fetal Neonatal Ed* 1996; Mar;74(2):F118-21.

149. Glascoe FP, Foster EM, Wolraich ML. An economic analysis of developmental detection methods. *Pediatrics* 1997; Jun;99(6):830-7.

150. Escalona SK. Babies at double hazard: early development of infants at biologic and social risk. *Pediatrics* 1982; Nov;70(5):670-6.

151. Der G, Batty GD, Deary IJ. Effect of breast feeding on intelligence in children: prospective study, sibling pairs analysis, and meta-analysis. *BMJ* 2006; Nov 4;333(7575):945.

152. Yu VY, Doyle LW. Regionalized long-term follow-up. *Semin Neonatol* 2004; Apr;9(2):135-44.

153. Tyson J. Evidence-based ethics and the care of premature infants. *Future Child* 1995; Spring;5(1):197-213.

154. Schneider JW, Gurucharri LM, Gutierrez AL, Gaebler-Spira DJ. Health-related quality of life and functional outcome measures for children with cerebral palsy. *Dev Med Child Neurol* 2001; Sep;43(9):601-8.

155. Murphy CC, Yeargin-Allsopp M, Decoufle P, Drews CD. Prevalence of cerebral palsy among ten-year-old children in metropolitan Atlanta, 1985 through 1987. *J Pediatr* 1993; Nov;123(5):S13-20.

156. Yeargin-Allsopp M, Murphy CC, Oakley GP, Sikes RK. A multiple-source method for studying the prevalence of developmental disabilities in children: the Metropolitan Atlanta Developmental Disabilities Study. *Pediatrics* 1992; Apr;89(4 Pt 1):624-30.

157. Grimes DA, Schulz KF. Bias and causal associations in observational research. *Lancet* 2002; Jan 19;359(9302):248-52.

158. Grimes DA, Schulz KF. Cohort studies: marching towards outcomes. *Lancet* 2002; Jan 26;359(9303):341-5.

159. Mamdani M, Sykora K, Li P, Normand SL, Streiner DL, Austin PC, et al. Reader's guide to critical appraisal of cohort studies: 2. Assessing potential for confounding. *BMJ* 2005; Apr 23;330(7497):960-2.

160. Normand SL, Sykora K, Li P, Mamdani M, Rochon PA, Anderson GM. Readers guide to critical appraisal of cohort studies: 3. Analytical strategies to reduce confounding. *BMJ* 2005; Apr 30;330(7498):1021-3.

161. Blair E, Watson L. Epidemiology of cerebral palsy. *Semin Fetal Neonatal Med* 2006; Apr;11(2):117-25.

162. Delgado-Rodriguez M, Llorca J. Bias. *J Epidemiol Community Health* 2004; Aug;58(8):635-41.

163. Smith GD, Ebrahim S. Data dredging, bias, or confounding. *BMJ* 2002; Dec 21;325(7378):1437-8.

164. Potential selection bias in hospital-based studies of perinatal outcome. *Paediatr Perinat Epidemiol* 2004; Mar;18(2):153.

165. Ellenberg JH. Selection bias in observational and experimental studies. *Stat Med* 1994; Mar 15-Apr 15;13(5-7):557-67.

166. Evans DJ, Levene MI. Evidence of selection bias in preterm survival studies: a systematic review. *Arch Dis Child Fetal Neonatal Ed* 2001; Mar;84(2):F79-84.

167. Ventura SJ, Hamilton BE, Mathews TJ, Chandra A. Trends and variations in smoking during pregnancy and low birth weight: evidence from the birth certificate, 1990-2000. *Pediatrics* 2003; May;111(5 Part 2):1176-80.

168. Mayer JP, Hawkins B, Todd R. A randomized evaluation of smoking cessation interventions for pregnant women at a WIC clinic. *Am J Public Health* 1990; Jan;80(1):76-8.

169. Kramer MS. Determinants of low birth weight: methodological assessment and meta-analysis. *Bull World Health Organ* 1987;65(5):663-737.

170. Steyn K, de Wet T, Saloojee Y, Nel H, Yach D. The influence of maternal cigarette smoking, snuff use and passive smoking on pregnancy outcomes: the Birth To Ten Study. *Paediatr Perinat Epidemiol* 2006; Mar;20(2):90-9.

171. Altman DG, Bland JM. Generalisation and extrapolation. *BMJ* 1998; Aug 8;317(7155):409-10.

172. Britton A, McKee M, Black N, McPherson K, Sanderson C, Bain C. Threats to applicability of randomised trials: exclusions and selective participation. *J Health Serv Res Policy* 1999; Apr;4(2):112-21.

173. *Evidence-based medicine : how to practice and teach EBM.* New York ; Edinburgh: Churchill Livingstone; 1997.

174. Cole TJ, Freeman JV, Preece MA. British 1990 growth reference centiles for weight, height, body mass index and head circumference fitted by maximum penalized likelihood. *Stat Med* 1998; Feb 28;17(4):407-29.

175. Kramer MS, Platt RW, Wen SW, Joseph KS, Allen A, Abrahamowicz M, et al. A new and improved population-based Canadian reference for birth weight for gestational age. *Pediatrics* 2001; Aug;108(2):E35.

176. Macfarlane A. *Birth counts : statistics of pregnancy and childbirth; volume 1, text.* 2nd ed.The Stationery Office; 2000.

177. Macfarlane A. *Birth counts : statistics of pregnancy and childbirth; volume 2, tables.* 2nd ed.The Stationery Office; 2000.

178. Johnson AH, Peacock JL, Greenough A, Marlow N, Limb ES, Marston L, et al. High-frequency oscillatory ventilation for the prevention of chronic lung disease of prematurity. *N Engl J Med* 2002; Aug 29;347(9):633-42.

179. Dorling JS, Field DJ. Follow up of infants following discharge from the neonatal unit: structure and process. *Early Hum Dev* 2006; Mar;82(3):151-6.

180. Marlow N, Greenough A, Peacock JL, Marston L, Limb ES, Johnson AH, et al. Randomised trial of high frequency oscillatory ventilation or conventional ventilation in babies of gestational age 28 weeks or less: respiratory and neurological outcomes at 2 years. *Arch Dis Child Fetal Neonatal Ed* 2006; Sep;91(5):F320-6.

181. Dorling J, D'Amore A, Salt A, Seward A, Kaptoge S, Halliday S, et al. Data collection from very low birthweight infants in a geographical region: methods, costs, and trends in mortality, admission rates, and resource utilisation over a five-year period. *Early Hum Dev* 2006; Feb;82(2):117-24.

182. Salt A, D'Amore A, Ahluwalia J, Seward A, Kaptoge S, Halliday S, et al. Outcome at 2 years for very low birthweight infants in a geographical population: risk factors, cost, and impact of congenital anomalies. *Early Hum Dev* 2006; Feb;82(2):125-33.

183. Kamoji VM, Dorling JS, Manktelow BN, Draper ES, Field DJ. Extremely growth-retarded infants: is there a viability centile?. *Pediatrics* 2006; Aug;118(2):758-63.

184. Arnold CC, Kramer MS, Hobbs CA, McLean FH, Usher RH. Very low birth weight: a problematic cohort for epidemiologic studies of very small or immature neonates. *Am J Epidemiol* 1991; Sep 15;134(6):604-13.

185. Figueras F, Figueras J, Meler E, Eixarch E, Coll O, Gratacos E, et al. Customized birthweight standards accurately predict perinatal morbidity. *Arch Dis Child Fetal Neonatal Ed* 2007; Jan 24;.

186. Gardosi J. Customized fetal growth standards: rationale and clinical application. *Semin Perinatol* 2004; Feb;28(1):33-40.

187. Clausson B, Gardosi J, Francis A, Cnattingius S. Perinatal outcome in SGA births defined by customised versus population-based birthweight standards. *BJOG* 2001; Aug;108(8):830-4.

188. Gardosi J, Chang A, Kalyan B, Sahota D, Symonds EM. Customised antenatal growth charts. *Lancet* 1992; Feb 1;339(8788):283-7.

189. Gardosi J, Mongelli M, Wilcox M, Chang A. An adjustable fetal weight standard. *Ultrasound Obstet Gynecol* 1995; Sep;6(3):168-74.

190. Lemeshow S, Le Gall JR. Modeling the severity of illness of ICU patients. A systems update. *JAMA* 1994; Oct 5;272(13):1049-55.

191. Laptook AR, O'Shea TM, Shankaran S, Bhaskar B, NICHD Neonatal Network. Adverse neurodevelopmental outcomes among extremely low birth weight infants with a normal head ultrasound: prevalence and antecedents. *Pediatrics* 2005; Mar;115(3):673-80.

192. Vohr BR, Wright LL, Dusick AM, Mele L, Verter J, Steichen JJ, et al. Neurodevelopmental and functional outcomes of extremely low birth weight infants in the National Institute of Child Health and Human Development Neonatal Research Network, 1993-1994. *Pediatrics* 2000;105(6):1216-26.

193. Msall ME, Buck GM, Rogers BT, Duffy LC, Mallen SR, Catanzaro NL. Predictors of mortality, morbidity, and disability in a cohort of infants < or = 28 weeks' gestation. *Clin Pediatr (Phila)* 1993; Sep;32(9):521-7.

194. Wood NS, Costeloe K, Gibson AT, Hennessy EM, Marlow N, Wilkinson AR, et al. The EPICure study: associations and antecedents of neurological and developmental disability at 30 months of age following extremely preterm birth. *Arch Dis Child Fetal Neonatal Ed* 2005; Mar;90(2):F134-40.

195. Arad I, Durkin MS, Hinton VJ, Kuhn L, Chiriboga C, Kuban K, et al. Long-term cognitive benefits of antenatal corticosteroids for prematurely born children with cranial ultrasound abnormalities. *Am J Obstet Gynecol* 2002; Apr;186(4):818-25.

196. Agarwal R, Chiswick ML, Rimmer S, Taylor GM, McNally RJ, Alston RD, et al. Antenatal steroids are associated with a reduction in the incidence of cerebral white matter lesions in very low birthweight infants. *Arch Dis Child Fetal Neonatal Ed* 2002; Mar;86(2):F96-F101.

197. Baud O, Foix-L'Helias L, Kaminski M, Audibert F, Jarreau PH, Papiernik E, et al. Antenatal glucocorticoid treatment and cystic periventricular leukomalacia in very premature infants. *N Engl J Med* 1999; Oct 14;341(16):1190-6.

198. Davis S, Wright PW, Schulman SF, Hill LD, Pinkham RD, Johnson LP, et al. Participants in prospective, randomized clinical trials for resected non-small cell lung cancer have improved survival compared with nonparticipants in such trials. *Cancer* 1985; Oct 1;56(7):1710-8.

199. Schmidt B, Gillie P, Caco C, Roberts J, Roberts R. Do sick newborn infants benefit from participation in a randomized clinical trial?. *J Pediatr* 1999; Feb;134(2):151-5.

200. Wasson JH, Sox HC, Neff RK, Goldman L. Clinical prediction rules. Applications and methodological standards. *N Engl J Med* 1985; Sep 26;313(13):793-9.

201. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982; Apr;143(1):29-36.

202. Hack M, Fanaroff AA. Outcomes of children of extremely low birthweight and gestational age in the 1990's. *Early Human Development* 1999/1/1;53(3):193-218.

203. Jefferis BJ, Power C, Hertzman C. Birth weight, childhood socioeconomic environment, and cognitive development in the 1958 British birth cohort study. *BMJ* 2002; Aug 10;325(7359):305.

204. Hack M, Breslau N, Aram D, Weissman B, Klein N, Borawski-Clark E. The effect of very low birth weight and social risk on neurocognitive abilities at school age. *J Dev Behav Pediatr* 1992; Dec;13(6):412-20.

205. Thompson RJ,Jr, Goldstein RF, Oehler JM, Gustafson KE, Catlett AT, Brazy JE. Developmental outcome of very low birth weight infants as a function of biological risk and psychosocial risk. *J Dev Behav Pediatr* 1994; Aug;15(4):232-8.

206. Leonard CH, Clyman RI, Piecuch RE, Juster RP, Ballard RA, Behle MB. Effect of medical and social risk factors on outcome of prematurity and very low birth weight. *J Pediatr* 1990; Apr;116(4):620-6.

207. Ment LR, Vohr B, Allan W, Katz KH, Schneider KC, Westerveld M, et al. Change in cognitive function over time in very low-birth-weight infants. *JAMA* 2003; Feb 12;289(6):705-11.

208. Dammann O, Leviton A. Biomarker epidemiology of cerebral palsy. *Ann Neurol* 2004; Feb;55(2):158-61.

209. Dammann O, Leviton A. Maternal intrauterine infection, cytokines, and brain damage in the preterm newborn. *Pediatr Res* 1997; Jul;42(1):1-8.

210. Yoon BH, Park CW, Chaiworapongsa T. Intrauterine infection and the development of cerebral palsy. *BJOG* 2003; Apr;110 Suppl 20:124-7.

211. Ambalavanan N, Baibergenova A, Carlo WA, Saigal S, Schmidt B, Thorpe KE, et al. Early prediction of poor outcome in extremely low birth weight infants by classification tree analysis. *J Pediatr* 2006; Apr;148(4):438-44.

212. MacKendrick W. Understanding neurodevelopment in premature infants: Applied chaos theory. *The Journal of Pediatrics* 2006/4;148(4):427-9.

213. Donders AR, van der Heijden GJ, Stijnen T, Moons KG. Review: a gentle introduction to imputation of missing values. *J Clin Epidemiol* 2006; Oct;59(10):1087-91.

214. Zupancic JA, Richardson DK, Horbar JD, Carpenter JH, Lee SK, Escobar GJ, et al. Revalidation of the Score for Neonatal Acute Physiology in the Vermont Oxford Network. *Pediatrics* 2007; Jan;119(1):e156-63.

215. van Baar AL, van Wassenaer AG, Briet JM, Dekker FW, Kok JH. Very preterm birth is associated with disabilities in multiple developmental domains. *J Pediatr Psychol* 2005; Apr-May;30(3):247-55.

216. Schmidt B, Asztalos EV, Roberts RS, Robertson CM, Sauve RS, Whitfield MF, et al. Impact of bronchopulmonary dysplasia, brain injury, and severe retinopathy on the

outcome of extremely low-birth-weight infants at 18 months: results from the trial of indomethacin prophylaxis in preterms. *JAMA* 2003; Mar 5;289(9):1124-9.

217. Hack M, Wilson-Costello D, Friedman H, Taylor GH, Schluchter M, Fanaroff AA. Neurodevelopment and predictors of outcomes of children with birth weights of less than 1000 g: 1992-1995. *Arch Pediatr Adolesc Med* 2000; Jul;154(7):725-31.

218. Pinto-Martin JA, Riolo S, Cnaan A, Holzman C, Susser MW, Paneth N. Cranial ultrasound prediction of disabling and nondisabling cerebral palsy at age two in a low birth weight population. *Pediatrics* 1995; Feb;95(2):249-54.

219. Papile L, Burstein J, Burstein R, Koffler H. Incidence and evolution of subependymal and intraventricular hemorrhage: A study of infants with birth weights less than 1,500 gm. *The Journal of Pediatrics* 1978/4;92(4):529-34.

220. Wolke D, Sohne B, Ohrt B, Riegel K. Follow-up of preterm children: important to document dropouts. *Lancet* 1995; Feb 18;345(8947):447.

221. Castro L, Yolton K, Haberman B, Roberto N, Hansen NI, Ambalavanan N, et al. Bias in reported neurodevelopmental outcomes among extremely low birth weight survivors. *Pediatrics* 2004; Aug;114(2):404-10.

222. Callanan C, Doyle L, Rickards A, Kelly E, Ford G, Davis N. Children followed with difficulty: how do they differ?. *J Paediatr Child Health* 2001; Apr;37(2):152-6.

223. Hand D.J., Henley W.E. Statistical Classification Methods in Consumer Credit Scoring: a Review. *Journal of the Royal Statistical Society* 1997;160(3):523-41.

224. Ridley S. Severity of illness scoring systems and performance appraisal. *Anaesthesia* 1998; Dec;53(12):1185-94.