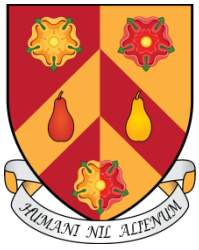
Prediction of Neurodevelopmental Outcome in Children Born Extremely Preterm

Louise Linsell

University of Oxford
Nuffield Department of Population Health



Wolfson College

Thesis submitted for the degree of Doctor of Philosophy
Trinity Term 2017

Abstract

Background: The survival rate of preterm children has risen steadily due to advances in obstetric and neonatal intensive care. Children born extremely preterm (≤28 weeks of gestation) are at high risk of long term developmental problems, including cerebral palsy, motor and cognitive impairment, visual and auditory deficits and behavioural problems. This can have serious implications for their quality of life and that of their family and carers. These children take up a disproportionate amount of neonatal intensive care unit resources and overall costs, and as they grow up are more likely to require additional health and social care services beyond routine care to compensate for their functional limitations. The early identification and management of factors that mediate long term outcome is necessary to assist healthcare professionals in selecting appropriate treatment pathways, and to develop, target and evaluate interventions. Many risk factor analyses for neurodevelopmental impairment have been published in preterm populations, but this vast literature has not been formally summarised. Furthermore, there is a dearth of studies reporting longitudinal analysis of neurodevelopmental trajectories from early childhood to adulthood.

Objectives: The first aim of this thesis was to perform a comprehensive systematic review of the world literature over the last two decades, to consolidate the evidence about the prognosis of neurodevelopmental outcome in children born very preterm or with very low birth weight. The second aim was to conduct a longitudinal analysis of a cohort of extremely preterm participants followed up into early adulthood to investigate the trajectories of long term sequelae over time, and to examine the association of neurodevelopmental course in relation to the predictive factors identified in the systematic review.

Methods: A systematic review was conducted using MEDLINE, EMBASE and PyscINFO databases to identify studies published between January 1 1990 and June 1 2014 reporting multivariable prediction models for the neurodevelopment of children born ≤32 weeks of gestation or with a birth weight ≤1250 grams (protocol registration number CRD42014006943). Seventy-eight studies reporting 222 risk factor models for neurodevelopmental outcome were identified. Two independent reviewers extracted key information about study design, outcome definition, risk factor selection, model development, reporting, and conducted a risk of bias assessment. To address the second objective of the study, a longitudinal analysis of cognitive and behavioural trajectories was conducted using a prospective, population-based cohort study in the United Kingdom and the Republic of Ireland. Three hundred and fifteen surviving infants born less than 26 completed weeks of gestation recruited at birth in 1995 and 160 term-born classroom peers recruited at age six were followed-up to 19 years. Participants were invited for up to four standardized, blinded cognitive assessments and the parent-completed Strengths and Difficulties Questionnaire was used to assess behavioural problems.

Results: The systematic review of risk factors for motor impairment in children born very preterm or with very low birth weight provided strong evidence that neonatal brain injury is a robust prognostic factor for cerebral palsy, and some evidence that the use of postnatal steroids increases the risk and the use of antenatal steroids reduces the risk of cerebral palsy. There was moderate evidence that male sex was prognostic for motor impairment at school age in children free of major disability. The systematic review of risk factors for cognitive impairment identified male sex, non-white ethnicity, lower levels of parental education and lower birth weight as significant predictors of global cognitive dysfunction in early infancy,

with parental education having a sustained impact after five years of age. There was also evidence that male sex was predictive of delayed language development in early infancy. Gestational age was found to be of limited use as prognostic factor for cerebral palsy, motor and cognitive impairment in cohorts restricted to ≤32 weeks of gestation. There was a dearth of good quality studies investigating risk factors for behavioural problems and psychiatric disorders and the findings of this review were inconclusive. The only factors that appeared to be consistent predictors of general behavioural problems were markers of socio-economic deprivation, neurodevelopmental or cognitive delay, and an abnormal behavioural screen in early infancy.

In the longitudinal analysis of the prospective, population-based cohort of extremely preterm children, cognitive trajectories were stable in both the extremely preterm and term-born groups over time with persistent deficit in the extremely preterm group of 25.2 IQ points (95% CI: -27.8 to -22.6, p<0.001) and only minimal catch-up over time. Participants with neonatal brain injury and of male sex had the largest deficits, but a lower level of maternal education and earlier gestational age at birth were also associated with reduced IQ scores. Behavioural problems were also more prevalent among the extremely preterm participants who had a mean Total Difficulties Score of 4.81 points above their term-born peers (95% CI: 3.76 to 5.87, p<0.001) and which persisted over the time period. Behavioural difficulties were mainly due to hyperactivity, inattention and peer problems and were strongly associated with a positive behavioural screen in early infancy.

Conclusions: The most robust predictors of poor neurodevelopmental outcome identified by the systematic review were neonatal brain injury, male sex, and markers of social

disadvantage. The unclear findings for many risk factors may reflect differences in study design, study population, methodological quality and lack of standardization of measures. Or it may simply reflect the fact that prognostic modelling in such a heterogeneous population is challenging and complex, with multiple risk factors acting sequentially over time, and often with the existence of multiple impairments within the same individual. The main conclusions from the longitudinal analysis of children born extremely preterm is that being born too soon appears to place limits on brain plasticity and function which is not recovered over time; with the most vulnerable being males and those with evidence of brain injury early in life. These structural abnormalities may disturb neurodevelopmental processes and impede the brain from maintaining a normal developmental trajectory. If extremely preterm children fail to achieve optimum levels of cognitive function and are still experiencing behavioural problems once they have reached maturity, then this has implications for health and well-being in later adulthood and old age. Cognitive test scores in infancy and early childhood reflect early adult outcomes and a positive behavioural screen in infancy is strongly associated with early adult behavioural outcomes.

Recommendations: The systematic review revealed some shortcomings in methodology and reporting that could be improved in future studies, and confirmed that that there is a dearth of properly designed and well-conducted prognostic modelling studies in this field. The findings and recommendations of this critical review should be used as a basis for the design, analysis and reporting of future studies seeking to develop multivariate risk factor or prognostic models in this population. There is an urgent need for larger population cohorts followed up routinely beyond two years as subtle outcomes such as impairment of executive function and fine motor skills cannot be reliably assessed at this age, and the natural course

of some disorders may have their onset later in childhood. Studies with larger sample sizes and greater power are needed for studying less common conditions in preterm populations and there should be more standardisation of outcome and risk factor measurements, particularly with the use of standard diagnostic evaluations to assess psychiatric disorders. Future studies should include a term-born comparison group and adopt appropriate statistical analysis techniques to analyse longitudinal outcome data and the impact of risk factors on these trajectories. Additional research is required to improve the prediction of individual differences, and to identify the neuropathological differences underlying different developmental trajectories and their interaction with environmental influences over time.

Acknowledgements

I first moved to Oxford in 2002, taking up a Research Fellow post in the Department of Public Health and Primary Health Care (now the Nuffield Department of Population Health) and joined the National Perinatal Epidemiology Unit (NPEU) in 2009. I've always felt at home in Oxford, both living and working here, and haven't left since. It has always been a personal ambition to complete a DPhil and it has been a rewarding journey, largely thanks to the collaborators involved along the way. The NPEU has provided a very supportive and flexible environment, allowing me to pursue these aspirations while continuing to work within the NPEU Clinical Trials Unit. It's a very friendly, vibrant and fun unit to work for, which I feel honoured to be part of.

I am particularly indebted to my supervisors; thank you all so much for your support and enthusiasm and for helping me to make this happen. Neil Marlow was open and engaging when I first approached him with my ideas for the project and has been a great inspiration, and also willing to help with anything at the drop of a hat. In addition to academic support, Jenny Kurinczuk has provided solid support as a mentor and always available if needed – I have always felt in very safe pair of hands for which I am very grateful. Thank you to Joan Morris for her support on all matters statistical, her constructive criticism and advice, and for also being a very nice person to work with. I feel like I've had a very solid team behind me and I will miss our friendly and relaxed supervisory meetings.

Sam Johnson, it's been a pleasure to work with you, you're so knowledgeable in your field and also good fun. I'm glad I've had the opportunity to collaborate with you, and hopefully it won't be the last time. I am very grateful to Enid Hennessy for all her hard work preparing the

EPICure datasets, even though she was supposed to be enjoying retirement. I'd also like to thank the other members of the EPICure team – Kate Bennett, Laura McCormack and Helen O'Reilly for always being so ready and willing to help.

I'd like to thank Nia Roberts, the outreach librarian at the Bodleian, for sharing her knowledge and expertise and helping devise the search strategies for the systematic review. I'm not sure what I would have done without her help and useful tips. And Reem Malouf – thank you - it was a monster of a systematic review, but you made it bearable. I miss our lively afternoons – thanks for all the time you devoted to this.

Thanks to Maggie Redshaw for your support and always being interested, and for assessing my Transfer and Confirmation of Status. Also thanks to Susan Mallet and Sarah Lewington for agreeing to be assessors and for their helpful input.

This work was funded by a National Institute of Health Research Doctoral Fellowship grant. I am grateful to have been given this opportunity to develop and broaden my horizons.

On a personal note, I'd like to thank all my family and friends for their continual interest and support. In particular, Mum and Molly, Lynne for being my sounding board - cheers mate (!), and John for being 100% supportive throughout all the ups and downs - you kept me sane.

I'd like to dedicate this work to my niece Maddy who was born in 2012, the same year as I started on this project. To Maddy, Adee and Lisa who have had similar experiences to many of the families who participated in the EPICure study.

Preface

This thesis describes the results of a series of collaborative studies between the National Perinatal Epidemiology Unit (NPEU) in the Nuffield Department of Population Health (NDPH) at the University of Oxford and the Institute of Women's Health at the University College London (UCL). Supervision was provided by:

Jennifer Kurinczuk	Professor of Perinatal Epidemiology and Director of the NPEU, NDPH, University of Oxford
Neil Marlow	Professor of Neonatal Medicine, Institute of Women's Health, UCL
Joan Morris	Professor of Medical Statistics, Queen Mary University of London

The work reported in this thesis has resulted in the following publications:

1. Linsell L., Johnson S., Morris J., Wolke D., Kurinczuk JJ., Marlow N. Neurobehavioral trajectories into adolescence following extremely preterm birth: a longitudinal cohort study (submitted).
2. Linsell L., Johnson S., Wolke, D., O'Reilly, H., Morris J., Kurinczuk JJ., Marlow N. Cognitive trajectories from infancy to early adulthood following birth before 26 weeks of gestation: a prospective population based cohort study (accepted *Archives of Disease in Childhood Fetal and Neonatal Edition*).
3. Linsell L., Malouf R., Morris J., Kurinczuk JJ., Marlow N. Risk factor models for neurodevelopmental outcome in children born very preterm or with very low birth weight: a systematic review of methodology and reporting. *The American Journal of Epidemiology* 2017: 185(7); 601-612.
4. Linsell L., Malouf R., Morris J., Kurinczuk JJ., Marlow N. Prognostic factors for cerebral palsy and motor impairment in children born very preterm or very low birth weight: A systematic review. *Developmental Medicine & Child Neurology* 2016: 58(6); 554-569.
5. Linsell L., Malouf R., Johnson S., Morris J., Kurinczuk JJ., Marlow N. Prognostic factors for behavioural problems and psychiatric disorders in children born very preterm or very low birth weight: A systematic review. *Journal of Developmental and Behavioral Pediatrics* 2016: 37(1); 88-102.
6. Linsell L., Malouf R., Morris J., Kurinczuk JJ., Marlow N. Prognostic factors for poor cognitive development in children born very preterm or very low birth weight: A systematic review. *JAMA Pediatrics* 2015: 169(12); 1162-1172.
7. Linsell L., Kurinczuk JJ., Marlow N., Morris J., Malouf R. What factors predict the severity of neurodevelopmental outcome in infants born very preterm? PROSPERO 2014: CRD42014006943. Available from http://www.crd.york.ac.uk/PROSPERO/.

Table of contents

List of tables

List of figures

List of abbreviations

ADHD	Attention deficit hyperactivity disorder
ASD	Autism spectrum disorder
ADOS	Autism Diagnostic Observation Scales
BITSEA	Brief Infant-Toddler Symptom Checklist
BPD	Bronchopulmonary dysplasia
BRS	Behaviour Rating Scale from the Bayley Scales of Infant Development-II
BW	Birth weight
BSID	Bayley Scales of Infant Development
CBCL	Child Behaviour Checklist
CP	Cerebral palsy
DAWBA	Development And Well Being Assessment
DCD	Developmental coordination disorder
DQ	Developmental quotient
DSM-IV-TR	Diagnostic and Statistical Manual of Mental Disorders (4th edition)
ELBW	Extremely low birth weight (<1000 grams)
EPT	Extremely preterm (<28 weeks of gestation at birth)
EPV	Events per variable
FSIQ	Full scale intelligence quotient from the Wechsler Abbreviated Scale of Intelligence
GA	Gestational age
IQ	Intelligence quotient
IQR	Interquartile range
ITSEA	Infant-Toddler Symptom Checklist
IVH	Intraventricular haemorrhage
KABC	Kaufman Assessment Battery for Children
LRT	Likelihood ratio test
MABC	Movement Assessment Battery for Children
MDI	Mental Development Index from the Bayley Scales of Infant Development-II
MI	Multiple imputation
MPC	Mental Processing Composite from the KABC
NDI	Neurodevelopmental impairment
NEC	Necrotising enterocolitis
NICHD NRN	National Institute of Child Health & Human Development Neonatal Research Network
NICU	Neonatal intensive care unit
OR	Odds ratio

PDA	Patent ductus arteriosus
PDI	Psychomotor Development Index from the Bayley Scales of Infant Development-II
PH	Pulmonary haemorrhage
PPROM	Premature preterm rupture of the membranes
PRISMA	Preferred reporting items for systematic reviews and meta-analyses
PVL	Periventricular leukomalacia
QUIPS	Quality in Prognostic Studies
RCT	Randomised controlled trial
RDS	Respiratory distress syndrome
ROP	Retinopathy of prematurity
SD	Standard deviation
SDQ	Strengths and Difficulties Questionnaire
SDS	Standard deviation score
SES	Socio-economic status
SGA	Small for gestational age
UK	United Kingdom
US	United States
VABSS	Vineland Adaptive Behavior Scales Screener
VLBW	Very low birth weight (<1500 grams)
VPT	Very preterm (<32 weeks of gestation at birth)
WASI	Wechsler Abbreviated Scale of Intelligence
WPPSI-R	Wechsler's Preschool and Primary Scale of Intelligence-Revised

Chapter 1: Introduction

1.1 Background

The incidence of preterm delivery is increasing and the survival rate of preterm children has risen steadily due to advances in obstetric and neonatal intensive care.[1,2] Surviving children born very preterm (VPT) or with very low birth weight (VLBW) are at high risk of long term developmental problems, including cerebral palsy (CP), impairment in motor and cognitive function, visual and auditory deficits and behavioural problems.[3] This can have serious implications for their quality of life and that of their family and carers. These children take up a disproportionate amount of neonatal intensive care unit (NICU) resources and overall costs, and as they grow are more likely to require additional health care services beyond routine care to compensate for their functional limitations.[4] The early identification and management of factors that mediate long term outcome is necessary to assist healthcare professionals in selecting appropriate treatment pathways, and to develop, target and evaluate interventions.

1.1.1 Epidemiology of preterm birth

1.1.1.1 Gestational age

Preterm deliveries are defined as those occurring before 37 weeks of gestational age (GA). The gold standard assessment of GA is early ultrasound in the first trimester together with fetal measurements. Prior to the use of ultrasound, GA assessment was based on the date of the last menstrual period and assumed that conception occurred 14 days after its onset. This method is considered to be less accurate due to errors in recall, the variable length of

menstrual cycle among women, and using an approximate date of conception. However, even ultrasound performed before 14 weeks can only date the pregnancy to within seven days and accuracy diminishes thereafter.[5] Accurate assessment of GA is critical at the limits of viability because mortality and morbidity outcomes vary greatly with each additional week of gestation, and guidelines for the clinical management of infants born at 25 weeks of gestation depend on it.[6,7]

In the past, there have been inconsistencies in the definitions of the categories of prematurity and birth weight (BW) with researchers using different truncation points for each.[8] Before the routine use of ultrasound, cohorts were generally defined by BW due to the certainty of BW and the unreliability of GA assessment. In a policy statement by the American Academy of Pediatrics, the conventional definitions of age during the perinatal period were reviewed and the use of standard terminology was recommended,[9] which has resulted in a more uniform classification of degree of prematurity and BW (Table 1.1).[3] The World Health Organisation has no published definitions of VPT or extremely preterm birth (EPT) but recommends which groupings to be used when reporting GA, which are consistent with those in Table 1.1. GA is expressed as completed weeks, therefore an infant born at 27 weeks and 6 days gestation is considered as a 27 week fetus, as opposed to rounding up to 28 weeks.

1.1.1.2 Prevalence

In a systematic analysis of preterm birth rates in the year 2010, the mean preterm birth rate (<37 weeks GA) in developed regions, including North America, Europe, Russia, Japan, Australia and New Zealand was estimated to be 8.6%.[10] The rate was notably higher in the United Stated (US) at 12%. This is in accordance with a previous review which reported rates of between 5% to 9% in Europe and other developed countries with higher

Table 1.1: Classification of gestational age and birth weight categories

Gestational age	
Post term	≥43 weeks
Term	37 to 42 weeks
Late preterm	34 to 36 weeks
Very preterm	<32 weeks
Extremely preterm	<28 weeks
Birth weight	
Low birth weight	<2500g
Very low birth weight	<1500g
Extremely low birth weight	<1000g

rates of 12 to 13% in the US.[2] The estimated rate of VPT (<32 weeks GA) in developed regions in 2010 was 1.3% and for EPT (<28 weeks GA) it was 0.4%.

The prevalence of preterm birth has increased in developed countries due to advances in obstetric and perinatal care and the effects of assisted reproductive technology, though increases may also be due to changes in data collection and registration.[1] In developed regions, the rate has increased from 7.2% in 1990 to 8.6% in 2010.[10] Of the 65 developed countries included in the report by Blencowe et al, 48 countries had increasing rates, 14 countries had stable rates (<0.5% annual change) and only 3 countries had declining preterm birth rates.

1.1.1.3 Causes of preterm delivery

The most common causes of a preterm delivery are unexplained spontaneous preterm labour, multiple pregnancy and premature preterm rupture of the membranes (PPROM).[2,11] PPROM

is defined as spontaneous rupture of the membranes prior to 37 weeks of gestation at least one hour before contractions begin. Risk factors for PPROM appear to be similar to spontaneous preterm labour with intact membranes, including infection and smoking during pregnancy.[12,13] Other common precursors to preterm labour are maternal or fetal indications for induction or Caesarean section, hypertensive disorders, severe intrauterine growth restriction, antepartum haemorrhage and cervical incompetence.[1,11]

Multiple pregnancy carries a tenfold increased risk of preterm birth compared to singleton pregnancies due largely to uterine over distension, and accounts for 12-27% of all preterm births.[11,14] Over the past few decades, the incidence of multiple pregnancy has increased with increasing maternal age, the more widespread use of assisted conception and drugs for subfertility.[11] Singleton pregnancies resulting from in-vitro fertilisation are also at an increased risk of preterm birth.[15] In developed countries, the rate of preterm birth differs between race, maternal age and socio-economic groups; it is higher among black women, older mothers and in women with a lower socio-economic status (SES).[11,14] To what extent and how these correlated factors interact with the mechanisms that lead to a preterm labour is unclear.

1.1.2 Neonatal care: a historical perspective

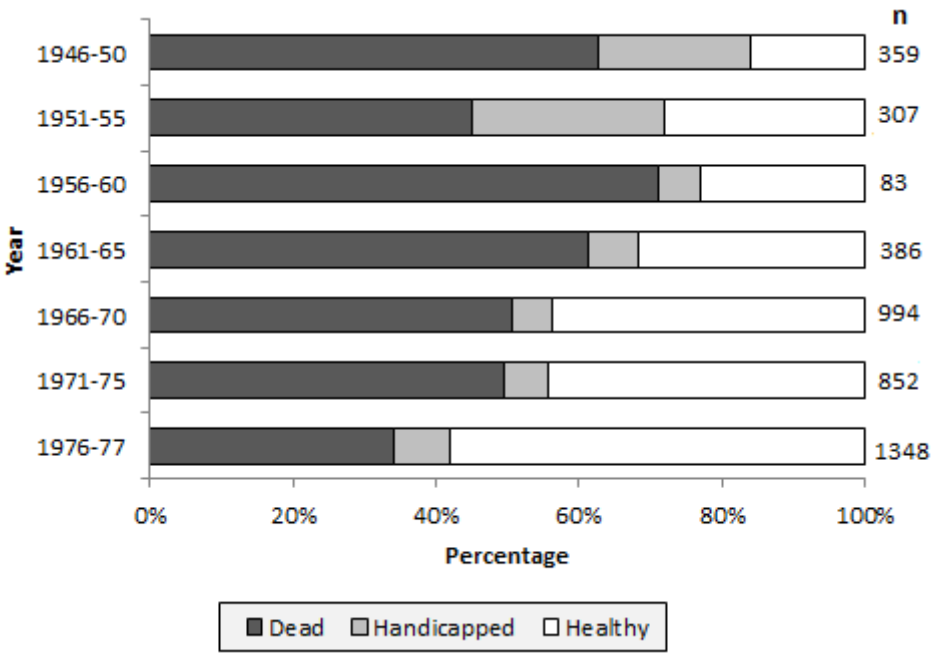
The first incubators used for live infants were exhibited in 1896; at this time the neonatal mortality rate in the US and the United Kingdom (UK) was high, at around 40 per 1,000 live births. This rate declined steeply to 15 per 1,000 even before neonatal intensive care was routinely available in the 1970s due to a number of basic improvements in care. These included the increasing use of incubators, the introduction of tube feeding, the use of Penicillin to treat infection and changes in the coordination of working practices around the

care of the newborn. After the development and proliferation of NICUs during the 1970s there were further advances in obstetric and neonatal medicine which resulted in a further reduction in the neonatal mortality rate to less than 4 per 1,000 in developed countries. These advancements included the routine introduction of assisted ventilation, surfactant therapy, prophylactic infection control, the use of antenatal steroids and enhancements in nutrition. The major causes of worldwide neonatal mortality today relate to preterm birth complications.[16]

The first comparisons of outcomes between infants born preterm and infants born at term in the 1930s and 1940s suggested that preterm birth did not alter the course of physical and mental development. However, these reports were based on highly select populations and by 1960 opinion had changed.[14] While the survival rate of preterm infants continued to increase, the link between immaturity and developmental impairment was recognised, and there was concern about the burden this growing group of disabled preterm survivors would have on the community. The first important systematic review of the trends in survival and disability in VLBW infants showed that despite increasing survival rates, the proportion of survivors with disability remained constant at 6-8% in each five year period from the mid-1950s to late 1970s.[17] This showed that the burden of disability in this population was far outweighed by the gain in healthy survivors (Figure 1.1). A more recent systematic review of studies published between 1970 and 1997 in EPT infants concluded that the incidence of disabilities had not changed in surviving EPT children, but the prevalence has risen due to increasing survival.[8]

Recent trends in the rate of survival and neurodevelopment impairment from 1990s into the 2000s are mixed, with variable reports from different countries. In the UK, the number of

Figure 1.1: Outcome studies for babies ≤1500g birth weight pooled by quinquennium to show the percentage of babies who died and survived with and without handicap (redrawn from Stewart et al 1981[17])



admissions for neonatal care of babies born between 22 and 25 weeks of gestation in England increased by 44% from 666 in 1995 to 959 in 2006, and survival to hospital discharge increased from 40% to 53% over the same period.[18] However, the proportion of survivors with major neonatal morbidity or neurodevelopmental impairment (at three years) remained similar, so the absolute number of surviving EPT infants at risk of developing future long-term developmental problems has risen.[18,19] Similarly, in France survival to discharge in infants born VPT increased between 1997 and 2011, but there has been no improvement in survival without neonatal morbidity in infants born <25 weeks.[20] However, the proportion of infants surviving without severe morbidity among higher gestations increased by 14.4% at 25 to 29 weeks and 6% at 30 to 31 weeks. Studies from the US comparing the outcomes of surviving infants born VPT or with VLBW between different epochs in the 1990s and 2000s found no significant increases in survival without neonatal or longer term morbidity.[21,22] However, a recent US study reported an increase in survival without major morbidity from 1993 to 2012

in infants born 25 to 28 weeks GA, with no change in infants born <25 weeks.[23] In population-based cohorts from the state of Victoria in Australia, the survival rate for extremely low birth weight (ELBW) infants plateaued in the 2000s, though neurological outcome is reported to have improved.[24] What does appear to be consistent is that the threshold for viability appears to have been reached; survivors below 23 weeks of gestation remain rare. Also the sharp rises in survival rates in the 1980s and early 1990s seen in many developed countries appear to have slowed down, and in some counties, levelled off completely.

1.1.3 Outcomes of a very preterm birth

1.1.3.1 Survival

The probability of surviving a preterm birth increases with each additional week of gestation and declines sharply below 28 weeks (Table 1.2).[25] In England and Wales in 2012, the survival rates to 28 days (as a percentage of all live births) were 4% in infants born at 22 weeks of gestation, 35% at 23 weeks, 66% at 24 weeks and over 82% at 25 weeks. The majority of deaths occur within the first week after birth and the probability of surviving to 28 days for infants still alive on the 7th day is over 80%, even for infants born at 22 and 23 weeks.

Multiple births account for around a quarter of all preterm births in England and Wales and for 16% of all neonatal deaths (Table 1.3). The pattern of survival rates in preterm multiples appear broadly similar and slightly better compared to those of singletons born at the same GA, though previous reports have been conflicting;[11] the opposite is true for term births. In a recent systematic review of prediction models for mortality in VPT infants,[26] eight factors that frequently predicted survival were identified, in addition to GA and BW:

Table 1.2: Neonatal and infant mortality rates per 1000 live births, England and Wales, 2012

Gestational age (weeks)	Live births[a]	Neonatal deaths[b] n (%)		Neonatal mortality ratio[c]	Infant deaths[d] n (%)		Infant mortality ratio[e]
All	726,572	1,971	(0.27)	2.7	2,808	(0.39)	3.9
<24	749	633	(84.5)	845.1	657	(87.7)	877.2
24-27	2,474	434	(17.5)	175.4	571	(23.1)	230.8
28-31	5,693	206	(3.62)	36.2	275	(4.83)	48.3
32-36	43,993	229	(0.52)	5.2	386	(0.88)	8.8
37-42	646,069	454	(0.07)	0.7	895	(0.14)	1.4
≥43	27,594	15	(0.05)	0.5	24	(0.09)	0.9

[a] Live births with known gestational age (n=2740 unknown).
[b] Death before 28 days.
[c] Number of neonatal deaths x 1000 / total number of live births.
[d] Death before 1 year of chronological age.
[e] Number of infant deaths x 1000 / total number of live births.

Source: Office for National Statistics gestation-specific infant mortality statistics.

Table 1.3: Neonatal and infant mortality rates per 1000 live births, England and Wales, 2012

Gestational age (weeks)	Live births[a]	Singleton births n (%)		Multiple births n (%)		Neonatal deaths[b] n (%)				Infant deaths[c] n (%)			
						Singletons		Multiples		Singletons		Multiples	
All	726,572	703,824	(96.9)	22,748	(3.13)	1,662	(0.24)	309	(1.36)	2,415	(0.34)	393	(1.73)
<24	749	610	(81.4)	139	(18.6)	526	(86.2)	107	(76.9)	545	(89.3)	112	(80.6)
24-27	2,474	1,869	(75.5)	605	(24.5)	319	(17.1)	115	(19.0)	430	(23.0)	141	(23.3)
28-31	5,693	4,169	(73.2)	1,524	(26.8)	161	(3.86)	45	(2.95)	211	(5.06)	64	(4.2)
32-36	43,993	33,427	(75.9)	10,566	(24.0)	201	(0.60)	28	(0.26)	338	(1.01)	48	(0.45)
37-42	646,069	636,171	(98.5)	9,898	(1.5)	440	(0.07)	14	(0.14)	867	(0.14)	28	(0.28)
≥43	27,594	27,578	(99.9)	16	(0.06)	15	(0.05)	0	0	24	(0.09)	0	0

[a] Live births with known gestational age (n=2740 unknown).
[b] Death before 28 days.
[c] Death before 1 year of chronological age.

Source: Office for National Statistics gestation-specific infant mortality statistics.

average size for GA, female sex, non-white ethnicity, absence of serious congenital malformations, use of antenatal steroids, higher 5-minute Apgar score, normal temperature on admission and better respiratory status.

There are large variations in the survival rates of infants born EPT between countries, particularly those born near the threshold of viability. This may reflect the differences in the policies and management of care, with interventionist practices and proactive management associated with higher survival rates and more selective management approaches resulting in lower survival rates.[27,28] Considerable differences in survival rates can also exist between different regions and centres within the same country, mainly due to the organisation of care and registration practices. Being born in a centre equipped with a higher level of intensive care, and the centralisation of neonatal intensive care, which involves the transfer of high-risk infants or mothers from local hospitals to specialised tertiary care centres, has been shown to be beneficial in several studies.[29] However, care should be taken when making comparisons of survival rates as they depend on the numerator and denominator used, and this could explain some of the apparent variation. For example, the numerator could be defined as survival to 28 days, discharge or two years and the denominator could be based on all births including stillbirths and terminations, all live births or admissions to NICU.

1.1.3.2 Neonatal morbidity

Infants born before 32 weeks of gestation are highly susceptible to a number of complications during the neonatal period which can impact on long-term neurodevelopment. This is because rapid organ growth and development occurs between 20 to 30 weeks and main bodily systems such as the respiratory system and central nervous system have not reached full maturity. Although neural migration in the brain is largely complete by this stage,

important cortical structures are still rapidly evolving during this critical period. These crucial stages of development that normally take place in utero are interrupted by the processes that surround a preterm birth and the following period of clinical care. The harsh NICU environment and other external influences can shape neuronal activity and connectivity and alter the course of development.[30]

1.1.3.2.1 Intraventricular haemorrhage

The most common type of brain injury among preterm infants is intraventricular haemorrhage (IVH) in the germinal matrix, which is the site from which neurons and glial cells migrate out during brain development. The germinal matrix is comprised of an intricate network of fragile thin-walled blood vessels and IVH occurs when a lack of cerebral blood flow (due to inability to self-regulate) causes these blood vessels to rupture, leading to bleeding. The majority of haemorrhages occur before 72 hours after birth and lesions are identified using early cranial ultrasound or magnetic resonance imaging scans. The Papile classification[31] is the most widely used system to grade the severity of the lesion. Grade I is defined as bleeding confined to the germinal matrix, grade II as bleeding that extends to the ventricle cavity and resolves without enlargement, grade III as ventricle enlargement due to accumulated blood, and grade IV as enlargement progressing to hydrocephalus. The incidence and severity of IVH is strongly associated with GA and is more common among male infants, with more severe injury. In recently recruited cohort populations the incidence of severe IVH (grade III-IV) was reported to be 5% in VPT infants and 13% in EPT infants.[18,20]

1.1.3.2.2 Periventricular leukomalacia

The other important type of brain injury in newborn premature infants is focal or diffuse white matter damage around the cerebral ventricles known as periventricular leukomalacia (PVL). The white matter of the brain consists largely of nerve fibres and its main role is to transmit nerve impulses, which modulate actions and coordinate communication between one area of the brain and another. Grey matter is comprised mainly of neuron cells from which these impulses originate and is primarily associated with processing and cognition, whereas changes in white matter morphology are thought to affect how the brain learns and functions. The disturbance of white matter development can occur during the antenatal, perinatal or postnatal period, thus the timing of the insult is more variable than for IVH. PVL has been graded on cranial ultrasound images by de Vries et al[32] and can be used as a guide for severity and prognosis. Grade I is defined as periventricular echogenicities that last for seven days or more, grade II as transient periventricular echogenicities that evolve into small localised cysts, grade III as evolution into extensive periventricular cystic lesions, and grade IV as extensive cystic lesions extending deep into the white matter. In VPT infants, the incidence of cystic PVL (grade II and above) is around 2%.[33]

1.1.3.2.3 Retinopathy of prematurity

Retinopathy of prematurity (ROP) is a disease of the eye in premature infants, caused by the disorganised growth of retinal blood vessels. The retina is the light sensitive membrane at the back of the eye that converts images produced by the lens into signals transmitted to the brain via the optic nerve. The retina can be scarred and damaged if the development of these blood vessels is interrupted by preterm birth. Maturation of the retina normally takes place in utero from around 16 weeks with full vascularisation completed a few weeks before term.

In term infants, oxygen is carefully regulated by the placenta, but premature exposure to oxygen (therapy required for other organs to develop properly outside the womb) can cause the constriction of healthy retinal blood vessels. This can then trigger the production of chemicals that produce new abnormal blood vessels, which are fragile and proliferate quickly, resulting in bleeding and scarring. This pulls the retina out of position causing damage to vision, and in severe cases, to become completely detached from the eye, resulting in blindness.

The retina is divided into three zones and the severity of retinopathy within each zone is graded I-V.[34] Stage I is characterised by a thin demarcation line between vascularised and non-vascular retina; stage II by an elevated ridge of tissue; stage III by the proliferation of abnormal blood vessels from the ridge; stage IV by partial retinal detachment; and stage V by total detachment. Severe ROP can be treated with laser therapy, which has largely replaced cryotherapy and involves the ablation of the non-vascularised retina, and anti vascular endothelial growth factor injections. Most cases of ROP occur in infants born <28 weeks of gestation, with the acute stages occurring between 32-38 weeks. In two recent cohorts, the incidence of treated ROP grade III and above has been reported to be between 6% and 16% among infants born EPT and 1% in infants born VPT.[18,20]

1.1.3.2.4 Pulmonary disease

Infants born prematurely are at increased risk of acute and chronic respiratory disease due to the immaturity of the lungs. Major lung development takes place between 17 and 26 weeks of gestation when branching of the respiratory system occurs and arteries and veins develop alongside it. After 27 weeks there is further branching and from 32 weeks aveoli, which are the tiny sacs that allow oxygen and carbon dioxide to move between the lungs and

bloodstream, start to appear as shallow indentations and continue to develop post term. Aveolar cells start producing surfactant from about 24 week's gestation. Surfactant is a lubricant which lines the sac, reducing surface tension and the effort of breathing. A lack of surfactant reduces the lungs ability to expand and increases surface tension, which can lead to the collapse or closure of the lung at end expiration, resulting in respiratory distress syndrome (RDS). To prevent this occurring, antenatal corticosteroids are routinely prescribed to women at risk of a preterm birth, and VPT infants are intubated and given surfactant therapy within the first hour of birth.

Bronchopulmonary dysplasia (BPD) is an important complication of RDS and is defined as oxygen dependency for 28 days or longer. Infants born VPT are graded as mild, moderate or severe based on their respiratory support requirement at 36 weeks or discharge home. A mild diagnosis is given if they are breathing air, moderate if they require less than 30% supplementary oxygen and severe if they require more than 30% and/or positive-pressure ventilation or continuous positive airway pressure.[35] There is an inverse relationship between the incidence of BPD and GA; recent studies report rates of 26% and 41% for severe BPD in infants born 22 to 26 weeks and 5% in infants born 27 to 31 weeks.[18,20]

Pulmonary haemorrhage (PH) is another condition which affects infants born VPT/VLBW. This occurs when red cells and capillary vessels filtrate into the lungs, typically onset is declared by bloody secretions from the nose and endotracheal tube. In some cases, this can lead to other respiratory complications, cerebral bleeding and seizures.

1.1.3.2.5 Necrotising enterocolitis

NEC is an acute inflammatory disease and the most common gastrointestinal condition occurring in neonates. Although it can affect term infants, it is more commonly found in VPT infants, and represents a significant clinical problem. It occurs when the bowel wall becomes infected and inflamed, usually due to the presence of bacteria and/or insufficient blood flow to the gut. This can lead to ulceration, haemorrhaging and tissue death in the gastrointestinal tract which may result in an intestinal perforation or irreversible bowel damage which can be fatal. Abdominal distension, bloody mucusy stools and bile stained aspirates or vomit are the classic gastrointestinal signs, but initial signs can be non-specific, making the condition difficult to detect. The variable clinical presentation also makes management of the condition challenging, although staging criteria developed by Bell et al[36] provide useful guidelines. A course of antibiotics is the recommended treatment for early stages of NEC, but once a perforation has occurred, surgery may be required.

Preterm infants are at particular risk of developing NEC because the mucous membrane of the gut is immature at birth. A deficiency in mucus and Immunoglobulin A can increase permeability to micro-organisms and toxins which may make the gut more vulnerable to NEC. Immature patterns of motility and a reduced capacity to digest may also predispose the preterm infant to NEC. The management of enteral nutrition, such as the delay and rate of feeding as well as the type of milk, can have an important impact on the risk of developing of NEC. In most preterm infants, NEC usually develops in the second or third week following the introduction of enteral feeds. In a recent population-based cohort, the incidence of severe NEC (Bell's stage II or III) was been reported to be 5.1% in infants born EPT and 3.7% in infants born VPT.[20]

1.1.3.2.6 Sepsis

Neonatal infections are an important cause of morbidity and mortality in infants born VPT/VLBW. As well as having less well developed immune systems, the use of invasive procedures such as catheters, ventilation tubes and long lines, along with prolonged hospital stay also exposes these vulnerable infants to a greater risk of acquiring infection. Early-onset sepsis is usually defined as bacteraemia or meningitis within 48-72 hours of birth and the primary routes of infection are maternal transmission or from the postnatal environment. Late-onset sepsis is typically defined as sepsis becoming clinically evident more than 48-72 hours after birth and is usually hospital acquired. The clinical signs of sepsis are abnormal temperature, irritability, poor skin tone, jaundice, mild diarrhoea, tachycardia and respiratory distress. Early diagnosis and treatment with antibiotics is crucial as progression from mild symptoms to death can occur rapidly. In the England, between 2006-2008, the incidence of infection (defined as positive culture and at least five days of antibiotic treatment) was 4.1% per 1,000 live births, and 71% occurred in infants born ≤32 weeks of gestation.[37]

1.1.3.2.7 Patent ductus arteriosus

In utero, the ductus arteriosus connects the pulmonary artery to the aorta, allowing blood from the right ventricle to bypass the lungs which are compressed and filled with fluid. In term babies, this vessel normally closes around the time of delivery to enable blood to be pumped into the lungs. This process is often delayed in preterm babies, and if closure does not occur within 72 hours, a diagnosis of patent ductus arteriosus (PDA) is made. In many cases, the duct eventually closes spontaneously and doesn't cause any clinical complications, and in some cases medication such as indomethacin is prescribed or surgery performed to close the duct. If the PDA persists untreated, complications such as metabolic acidosis, intracranial

haemorrhage, necrotising enterocolitis (NEC), pulmonary haemorrhage and congestive heart failure can develop.

1.1.3.3 Long term neurodevelopmental outcomes

Prematurity and its associated neonatal morbidities have a pervasive effect on neurodevelopment, leading to: CP, visual and auditory deficits, impairments in global and executive cognitive function, learning disabilities and behavioural problems. Studies report that 14% of VPT and up to 50% of EPT survivors have moderate to severe neurodevelopmental disabilities at school age, predominantly cognitive impairment, compared to around 1-3% of infants born at term.[38,39] Furthermore, the mean intelligence quotient (IQ) of school age children born preterm without severe disability has consistently found to be lower than their term-born peer group, and related to GA at birth.[40] VPT children are also at increased risk of a range of behavioural problems, including attention deficit/hyperactivity disorders (ADHD), autistic spectrum disorders (ASD), emotional and social problems and internalizing behavioural disorders.[41-43] With this array of cognitive and behavioural issues, many VPT survivors are at high risk of poor academic attainment and reduced lifelong earning potential and life chances. A significant proportion require full time specialist education, and the majority of those in mainstream education require specialist academic, health or behavioural support services to aid their transition through school.[39,44]

1.2 Rationale and aims of thesis

1.2.1 The burden of extremely preterm birth

The incidence of preterm birth in England in 2012 was 7.3% (n=52,909) and rising, with VPT births accounting for 1.2% (n=8,916) and EPT births accounting 0.4% (n=3,223) of all live births

respectively.[25] Neonatal interventions at extremely low GA are becoming more frequent, leading to an increasing number of admissions into NICUs. These infants take up disproportionate amount of NICU resources and overall costs, and as they grow up are more likely to require additional health care services beyond routine care and special educational support to compensate for their functional limitations and learning and behavioural difficulties. Not only does this impose a substantial financial burden on health care providers, special education and social services, but there is also the personal burden of disability incurred by the individuals affected and their families. The total cost of VPT and EPT birth to the public sector for an annual birth cohort to 18 years of age has been estimated at £242 and £989 million respectively. The incremental cost per VPT child surviving to 18 years compared with a birth at term is £61,781, and per EPT child £94,740.[4] Although late preterm (32 to 36 weeks) and VPT birth account for the largest majority of the total societal costs of preterm birth, the relative concentration of costs among EPT infants is noteworthy.

1.2.2 The importance of long term prognosis

The early identification and management of factors that mediate long term outcome is necessary to develop, target and evaluate interventions. The effects of preterm birth and associated perinatal care are frequently manifested many years later, and the evaluation of long term outcome is expensive and time-consuming. A way of overcoming this problem is to develop prognostic models using perinatal outcomes and other risk factors to predict future prognosis. If a combination of prognostic factors for later neurodevelopmental impairment can be identified, they can be used as a prognostic marker in randomised controlled trials (RCTs) testing the effectiveness of interventions. Longitudinal analysis of long term follow up provides a more robust and thorough understanding of the underlying processes that affect

development, which can inform decision making about the allocation of finite health and social care resources within the NHS. The ability to predict longitudinal course would provide a useful resource for clinicians and parents, who have reported a lack of consistency and sometimes conflicting information relating to the prognosis and care of extremely premature infants[45]. In summary, predicting future prognosis from key predictive assessments would be useful to (i) assist in healthcare and educational service planning and provision (ii) create risk groups to inform treatment or to stratify individuals by severity in epidemiological studies or clinical trials (iii) inform parents and the children about their future prospects.

1.2.3 Previous research on long term prognosis

Cohorts established during the early 1990s are maturing into early adulthood. To date, most analyses performed using these cohort data have used cross sectional techniques to analyse longitudinal follow-up data. For example, many studies have used regression analyses, correlation or pairwise t-tests to identify which factors at an earlier age predict outcomes during the next stage of childhood[46] or whether the outcomes of more recent cohorts have improved compared to older cohorts.[18,47] Such studies have yielded mixed results with some reporting increasing or decreasing problems over time and others reporting stable development. The problem with using cross-sectional statistical techniques to analyse longitudinal change, is that group prevalence of mild, moderate and severe disability within preterm cohorts remains fairly constant over time; however there can be considerable variation in individual trajectories, particularly in the non-severe groups, which is masked by these methods of analysis.[38]

Very few studies have used a longitudinal modelling approach to examine the continuity of neurodevelopment from early childhood to adulthood, and the changes within each child

using a 'growth' curve analysis approach. Those that have are characterised by several major shortfalls: failure to enrol an appropriate longitudinal comparison group; a limited set of outcome measures; the use of different tests at different time points; and short periods of follow-up.[48-51] None of these longitudinal studies, and no known studies to date, have developed and extensively validated a prognostic model for neurodevelopmental outcome in surviving VPT children using published guidelines.[52,53] A number of prognostic models have been developed and validated for survival[26], but the only known prognostic models developed for longer term outcomes are based on composite outcomes of survival and neurodevelopmental impairment (NDI) at 18-22 months.[54-56] While these are of great value, it is also important to untangle the factors that are prognostic in each developmental domain for surviving children, which may be very different to those predictive of a composite outcome.

1.2.4 Aims of the thesis

The aim of the work described in this thesis was to address these knowledge gaps, first by summarising the evidence on prognostic factors for poor neurodevelopment in surviving children born VPT/VLBW. Many risk factor analyses for neurodevelopmental impairment have been published, but this vast literature has not been formally summarised. A systematic review was performed on the world literature over the last two decades, to consolidate the evidence on prognosis to date for children born EPT in the five main developmental domains: motor function, cognition, hearing, vision and behaviour. The systematic review was extended to include VPT/VLBW children as the literature is more extensive in this group compared to EPT children alone. The results from the systematic review informed the second part of the project: a longitudinal analysis of the EPICure cohort of EPT participants born in

the UK in 1995 and followed up into early adulthood.[57] The trajectory of the key components of neurodevelopment in these EPT survivors over the course of childhood was examined in relation to the predictive factors identified in the systematic review, to explore whether developing prognostic models for neurodevelopment is possible and appropriate.

The research question addressed was: "Is it possible to predict the neurodevelopmental outcome in extremely preterm infants?" In summary, the objectives were to:

- Perform a systematic review of prognostic factors predicting neurodevelopmental impairment in surviving children born VPT/VLBW in the following outcome domains: motor function, cognition, hearing, vision and behaviour;
- Describe the trajectories of key components of neurodevelopment within individuals in the EPICure cohort from infancy to early adulthood and explore whether problems are transient or stable, and whether conditions deteriorate or improve over time;
- Investigate the effect of predictive factors identified in the systematic review on these patterns of neurodevelopment and to explore whether it is possible and appropriate to develop prognostic models for neurodevelopmental outcome in children born extremely preterm.

1.3 Structure of thesis

The methods and the main search results of the systematic review of prognostic factors for poor neurodevelopment in children born VPT/VLBW are described in Chapter 2. Details of the search strategy used, study eligibility criteria, data extraction, risk of bias assessment and data synthesis and reporting, as well as the search results and characteristics of the studies and

models retrieved are also reported in this chapter. The evidence on prognostic factors for each developmental domain is reported separately in Chapters 3-5. Chapter 3 summarises and discusses risk factor analyses published for motor and neurosensory impairment, including CP, general motor skills, hearing and vision. Chapter 4 consolidates the evidence on prognostic factors for cognitive impairment, encompassing IQ, language impairment, executive function and academic attainment. Chapter 5 reviews and discusses studies of prognosis for behavioural problems and psychiatric disorders, including ADHD and ASD. The chapters which follow describe the longitudinal analysis of the EPICure cohort study. Chapter 6 describes the study population and follow-up assessments, defines the outcome and explanatory variables, and outlines the general statistical methods and analysis used. Chapter 7 reports the findings from a longitudinal analysis of cognitive outcomes and Chapter 8 present the results from a longitudinal analysis of behavioural outcomes. The discussion in Chapter 9 provides an overview of the main findings and conclusions of the project, and suggests directions for future research.

Chapter 2: Systematic review of prognostic factors for poor neurodevelopment in children born VPT or with VLBW: methods and search results

2.1 Introduction

There is a large literature reporting risk factor analyses for poor neurodevelopment in the VPT/VLBW population which to date has not been formally summarised. Few studies have advanced beyond risk factor analyses to develop and validate risk prediction models for use in routine care and to assist with the planning and provision of healthcare. A number of prognostic models have been developed and validated for survival outcome in this population[26] and two web-based tools based on a composite outcome of survival and NDI at 18-22 months have been developed and validated by the National Institute of Child Health and Human Development Neonatal Research Network (NICHD NRN).[54-56] However, no known prognostic models have been developed for surviving VPT/VLBW children predicting outcomes in each separate developmental domain.

The objective of this systematic review was to consolidate the evidence on factors for poor neurodevelopment in surviving children born VPT/VLBW to inform future prognostic research. The aim was to identify all studies reporting a multivariable outcome prediction model for either motor function, cognition, hearing, vision and/or behavioural problems in VPT or VLBW children. The risk factors examined and reported by these studies, and the strength of their association with outcome was summarised for each outcome domain.

This chapter describes the methods of the systematic review and presents a summary of the search results. Characteristics related to the design and conduct of each study and the development, validation and reporting of risk prediction models are reported for all the articles identified by the search and included for data extraction.

2.2 Methods

The methods for the overall systematic review were published in a protocol (http://www.crd.york.ac.uk/PROSPERO/), registration number CRD42014006943 (Appendix 1).

2.2.1 Search strategy

Three electronic search strategies were devised in the MEDLINE, EMBASE and PsycINFO databases (Appendix 2) using the National Institutes of Health Medical Subject Headings (NIH MeSH). The searches identified any journal articles published from 1st January 1990 to 1st June 2014 reporting a multivariable risk prediction model for a neurodevelopmental outcome assessed after the age of 18 months in VPT/VLBW children. No language restrictions were applied. The search was confined to published journal articles and reports and excluded conference abstracts, letters, theses, dissertations, other unpublished sources and the grey literature. This was to ensure that the risk factor analyses included in the review had undergone peer review and passed the minimum standards of quality required by a journal for publication. Conference abstracts, letters and other short correspondence were not included, even if published, as they do not contain enough information to assess the quality of the work and the likelihood of selective reporting. The list of articles identified by the search strategies were exported to Endnote X6 for screening. The bibliographies of all articles included for full data extraction were hand-searched for further eligible articles.

2.2.2 Eligibility criteria and rationale

Articles were included in the review if they satisfied the following eligibility criteria:

- Contained original data;
- Study population was born after 1st January 1990;
- Study population was ≤32 weeks GA or with BW ≤1250g and not a highly select group based on other clinical criteria;
- One objective was to perform a multivariable risk factor analysis (>2 variables) of a neurodevelopmental outcome assessed after 18 months of age.

2.2.2.1 Study design

Most studies reporting risk factor data are observational and include cohort, case-control and cross sectional designs. Some RCTs report the neurodevelopmental follow-up of trial participants receiving a particular treatment or intervention. All study designs were included, but the strength of study design was taken into consideration in a risk of bias assessment. Prospective cohort studies with a representative population are the strongest design for risk factor analysis and can provide the absolute risk of an event. Case-control studies may be subject to bias, arising from the selection of cases and controls and the retrospective nature of data collection about prognostic factors, and can only provide relative risk estimates in the form of odds ratios. Cross-sectional studies can only report factors associated with the outcome, and cannot provide information about the causal pathway or prediction. However, both these study designs can provide useful information, particularly for rarer outcomes, and when the required duration of follow-up is long, e.g. for behavioural and higher order cognitive outcomes. RCTs may be based on a selective population with stringent eligibility criteria, but are likely to have robust follow-up data. The trial treatment allocation should be

adjusted for in any study presenting a risk factor analysis using the combined control and treatment arms of a trial population, even if the treatment did not impact on the primary outcome, as it could have affected the risk of other secondary outcomes.

2.2.2.2 Year of birth

The year 1990 was chosen as a cut-off date because significant advances were made in treatment in the mid-1980s and early 1990s. Surfactant therapy was introduced in the mid-1980s, but was only adopted routinely in clinical care from the early 1990s. This was a time of transition from the 'pre-surfactant' era of high mortality and morbidity to the 'surfactant era' of improved survival and prognosis.[3,58] The increased use of assisted ventilation, prophylactic infection control and antenatal steroid therapy around this time also improved the outcomes of preterm infants. After 1990, no major new medical advancements have had significant impact on survival and developmental outcome. Survival to discharge in infants born between 22 and 25 weeks of gestation in England has increased from 40% to 53% over the period 1995 to 2006.[18] Despite this, the overall rate of disability and impairment remained similar over the same period, so the prognosis of surviving infants born in the 1990's is likely to be fairly similar to the prognosis of surviving infants born after 2000.

2.2.2.3 Gestational age and birth weight criteria

One aim of the systematic review was to inform the longitudinal analysis of a cohort of EPT children in the second part of this study. However, the literature based on EPT children is not as extensive as it is for VPT and VLBW children, therefore studies based on children born ≤32 weeks GA or with BW ≤1250g were also included. The prognosis of these slightly older and more mature babies should be informative for the more extreme group as GA and BW are on a continuous spectrum and the mechanisms determining severity of outcome are likely to be

similar. The GA range included 32 completed weeks, even though the conventional definition of a VPT birth is <32 weeks, so as not to exclude studies based on the large EPIPAGE cohort, which have made important contribution to research in this area. The BW truncation point of ≤1250g was chosen to exclude the subset of more mature but extremely growth restricted infants included in the typical <1500g VLBW cohort which can cause heterogeneity and lead to confounding bias when examining the relationship between risk factors and outcome.[59] Studies based on a highly select population of VPT/VLBW children using other clinical criteria were excluded, such as those with those with brain injury or those judged to be at high risk, to ensure that the results were more widely representative and thus applicable to the general VPT/VLBW population.

2.2.2.4 Type of risk prediction model

Two types of risk prediction studies can be distinguished in the literature; explanatory prognostic factor studies and multivariable outcome prediction studies.[60,61] Explanatory prognostic factor studies investigate the causal pathway between a single prognostic factor and an outcome, ideally adjusted for confounders, and estimate effect size. For example, a study which hypothesises that BPD is linked to cognitive delay, and then explores this relationship, adjusting for potential confounding factors associated with the BPD and cognitive delay. In these types of study, other risk factors are included based on the change in the regression coefficient of the prognostic factor under study. Other types of study considered as belonging to this category included those where the focus of the research question was to use early outcome assessments, neonatal scoring systems, brain scan data or genetic markers to predict later neurodevelopmental outcomes.

The second type of risk prediction study, which has the aim of developing a multivariable outcome prediction model, starts with a list of candidate risk factors which are included in the final model based on their predictive ability of the outcome. For example, a study that identifies several variables that are possible candidate risk factors from the literature, and then uses a univariable screening process to eliminate poor predictors and a multivariable model building process to determine the most robust predictors of the outcome of interest. Current guidelines recommend not combining these two distinct types of study as their objectives and model building strategies differ which, when synthesised, could lead to biased results.[60] Therefore only multivariable outcome prediction models are included in this review because the primary aim is to identify a robust set of predictors, rather than focus on the causal effect of individual predictors separately.

2.2.2.5 Neurodevelopmental outcome assessed 18 months and above

Studies were included if one objective was to perform a multivariable risk factor analysis for any aspect of neurodevelopmental outcome related to motor function, cognitive function, hearing, vision or psychiatric problems at 18 months or older. The search strategy targeted outcomes that have been reported previously in the literature, and examples are given below:

- Motor function - CP, motor coordination, gait, mobility, movement disorder, dystonia, neurological disorder;
- Cognitive function - executive function, memory, language, attention, verbal skills, IQ, phonological skills, reading, writing, maths, learning disabilities;
- Hearing - hearing impairments, use of aids, deafness;
- Vision - visual impairments, use of glasses, blindness, refractive errors, myopia;

- Behavioural problems and psychiatric disorders - ADHD, ASD, emotional or behavioural disorder, depression, internalising or externalising behaviours, anxiety.

Studies reporting outcomes before 18 months were not included as assessments in very early infancy have been found to have limited utility in predicting future outcome,[62,63] even in healthy term-born children.[64] Assessments at 18-24 months are not always reliably predictive, particularly for non-severe conditions,[65-67] however as very few cohorts are followed up beyond two years this age group was not excluded. The data synthesis was stratified by age of assessment to address this issue.

2.2.3 Data extraction

All articles identified by the search strategies were screened on title and abstract for definite exclusions and duplicates (screen 1). For the remaining articles, the full text was retrieved and the inclusion criteria were applied (screen 2). The two screens were performed by the author (LL) in the first instance, but if there was uncertainty about the eligibility of an article, it was screened independently by a second reviewer Reem Malouf (RM). If a decision could not be reached it was referred to the rest of the author review team Jennifer Kurinczuk (JK), Neil Marlow (NM) and Joan Morris (JM). Non-English articles included in the review were fully translated. Multiple articles based on the same cohort of children underwent a panel review (LL, RM and NM). Those reporting the same outcome domain (cognitive, motor, behaviour, hearing, vision) at the same age of assessment (<5 years and ≥5 years) were assessed on relevance to the review, and only one article was selected for data extraction. The age bands specified in the original protocol (18-24 months; 3-5 years; 6-12 years; >13 years) were collapsed to <5 years and ≥5 years as there were too few articles in the older age bands.

For all articles eligible for inclusion, both reviewers (LL and RM) independently completed a full data extraction form and risk of bias assessment on a customised MS Access 2010 database. Every single item entered was manually cross-checked for discrepancies at a face-to-face meeting and referred to the rest of the author review team if agreement could not be reached. The data items extracted are listed in Box 2.1. If critical information was missing or unclear the corresponding author was contacted once by email for clarification. The data extraction forms used can be found in Appendix 3.

Box 2.1: Data items extracted from articles included in the systematic review

study design
participant setting
centre selection
study location
year of birth
gestational age
birth weight
age at assessment
selection criteria of study population
sample size
completeness of data at follow-up
details of outcomes assessed
number of candidate risk factors assessed
variable selection
treatment of continuous variables
adjustment for confounders
method of analysis
model assumptions checked
missing data analysis
presentation of multivariable model
details of risk factors included in final model
strength of association
statistical validation
clinical validation

2.2.4 Types of risk factors

All types of risk factors were included, provided the focus of the research question was to fit a multivariable outcome prediction model for a neurodevelopmental outcome. Risk factors can be classified into several categories (examples are provided below but these lists are not exhaustive):

- Infant characteristics - GA, BW, sex, ethnicity;
- Antenatal - multiple pregnancy, parity, antenatal steroids, chorioamnionitis;
- Delivery - complications during labour and delivery, mode of delivery;
- Neonatal - base deficit, bradycardia, brain scan or ultrasound, Clinical Risk Index for Babies Score, Apgar score, type of milk, duration in neonatal intensive care;
- Neonatal complications - IVH, PVL, BPD, PH, ROP, PDA, NEC;
- Anthropometry - weight, height, length, head circumference;
- Assessments after discharge - neurological, motor, cognitive, behavioural screen;
- Environmental - SES, maternal/parental education, parental employment, parent mental health status, maternal interaction.

2.2.5 Risk of bias assessment

Several reviews have evaluated the quality of published articles that describe the development or validation of prediction models, and overwhelming evidence shows that conduct and reporting are poor,[68] leading to the development of quality assessment tools specific for these types of study. In this review, the quality of studies was assessed according to a modified version of the Quality in Prognostic Studies (QUIPS) tool, which is a standardised set of criteria recommended for use in reviews of prognosis[69] (Appendix 4). The authors advise tailoring these criteria to each individual review, hence slight modifications were made

to the criteria based on knowledge of the studies in this research field and suggestions from other sources.[70-72] The tool focuses on six areas of potential bias pertinent to studies of prognosis: study participation, study attrition, prognostic factor measurement, outcome measurement, confounding measurement and account, and statistical analysis. The Cochrane Handbook for Systematic Reviews of interventions[73] recognises four sources of potential systematic bias in randomised and non-randomised comparisons of healthcare interventions, including: selection bias, performance bias, attrition bias and detection bias. The QUIPS tool encompasses the usual four sources of bias assessed, but has adapted them for more meaningful and practical use in studies of prognosis. Studies were graded as [yes/partly/no] for each domain and classified as having a low-moderate risk of bias if they were graded as [yes] or [partly] in all six bias domains and moderate-high risk of bias otherwise.

The domain for confounding measurement and account is more relevant for explanatory prognostic factor studies which focus on investigating the causal pathway between a single prognostic factor and an outcome, adjusted for confounders. Confounding per se is not an issue when developing a multivariable prediction model, as the aim is to find the factors that maximise the prognostic power of the model. However, there are important variables that one would expect to be included in the list of candidate factors at the outset, based on previous epidemiological work: GA or BW, sex, multiple pregnancy and markers of social disadvantage. Therefore to meet the full criteria in this domain, a study had to consider all these factors in the initial screening. If the study sample was drawn from a randomised controlled trial (RCT) population, the study was penalised if trial allocation was not adjusted for in the risk factor analysis as a treatment intervention could potentially affect the risk of an outcome.

2.2.6 Data synthesis and reporting

Results were presented in accordance with the PRISMA guidelines.[74] Risk factors that were statistically significant ($p<0.05$) in the final model were reported for each study in a table. Studies were grouped according to type of outcome studied and according to age of assessment; <5 years and ≥5 years. The studies were stratified by age as outcome assessments are typically more unreliable before the age of five. Also assessments in early infancy are more crude measurements of development; cognitive assessment relies to some extent on motor function and vice versa, whereas assessment in later childhood measures higher order cognitive functions and finer motor skills. The results were stratified to take some account of these differences between age groups, as risk factors may differ. A risk factor was presented graphically if it was statistically significant in the final model in at least one study at low-moderate risk of bias and included in the final model of at least two other studies (including studies at moderate-high risk of bias) within the same outcome domain. The plots display the number and quality of all studies that entered that risk factor in the final model and whether it was reported as a significant predictor or not significant. The decision to only include risk factors entered into the final model in at least three studies was pragmatic as multitudes of risk factors were considered across studies, and if all of them were displayed the graphs would have a very long tail showing risk factors that were only considered by one or two studies. As a consequence it would not be possible to make any clear statements about the prognostic value of any factors with such a lack of evidence, so the graphs were truncated at the point at which they were considered to be non-informative.

Some articles contained multiple models for the same outcome domain, which has implications for the independence of observations; if all models were included in the data

synthesis for that outcome domain then some study populations would be included more than once. To ensure that a single study population was not over-represented, one model was selected per outcome per study for the purposes of data synthesis. The reason for selection and details of the models omitted are referenced in the relevant table footnote for individual studies. The general approach was as follows: the full/principle model presented for the latest assessment/global score was selected for inclusion and any further subgroup or sensitivity analyses were not included. If joint models were presented including exactly the same participants, for example, the same outcome with a different model parameterisation, then risk factors were only included in the graphical summary if both models agreed. If separate models were presented for independent subgroups, for example by GA group, with no overall model presented then risk factors were included in the graphical summary if significant in any of the models.

2.3 Results

2.3.1 Search results

The searches retrieved 44,500 articles, and after removing duplicates, the first screen on title and abstract was performed on 32,283 articles (Figure 2.1). For 29,999 the title or abstract clearly indicated that the topic of the article was not relevant to the review question or did not satisfy one of the inclusion criteria. The remaining 2,284 articles were screened on full text, applying the full set of eligibility criteria. Eligibility was unclear in 136 (6%) and these were reviewed by the second independent reviewer (RM), or the author was contacted (where uncertainty was due to missing information). After applying the eligibility criteria, 91

articles (from 48 cohort populations[a]) containing multivariable risk factor analyses were eligible for inclusion. Following panel review (LL, RM and NM), a further 13 articles were excluded as they reported the same outcome domain in the same age group of assessment in the same cohort as another article with a more relevant objective.[75-87] The remaining 78 articles,[41,43,46,55,75,88-160] reporting 222 risk prediction models, were included in the data extraction. Twenty-eight articles reported a risk factor analysis for motor outcomes, 31 for cognitive outcomes, 15 for behavioural problems or psychiatric disorders, three for visual outcomes and no articles for hearing outcomes. Twenty-seven articles reported a risk factor analysis for a composite outcome comprised of any combination of the other neurodevelopmental domains, and some studies reported risk factor models in more than one domain. No further articles were identified in the hand-search of bibliographies.

2.3.2 Study design

The characteristics of the 78 studies that were included for data extraction are shown inTable 2.1. The main study design was cohort (91%) and there were five RCT populations, one case-control study and one cross-sectional study. Of the 71 prospective cohorts, half (n=36) were ascertained from all live births in a geographically defined region and half (n=35) were recruited from NICU admissions. Forty-five percent (n=35) of studies recruited from a single centre and 55% (n=43) from multiple centres. Seventy-nine percent (n=62) collected risk factor and outcome data prospectively, but 17% (n=13) extracted data retrospectively from hospital records or a database, and in 27% (n=21) of studies the outcome data was collected

[a] Studies based in any centre participating in the NICHD NRN follow-up programme were classified as belonging to the same cohort.

Figure 2.1: Search and selection process used to identify published studies that developed a multivariable risk factor model for neurodevelopmental outcome children born VPT or with VLBW

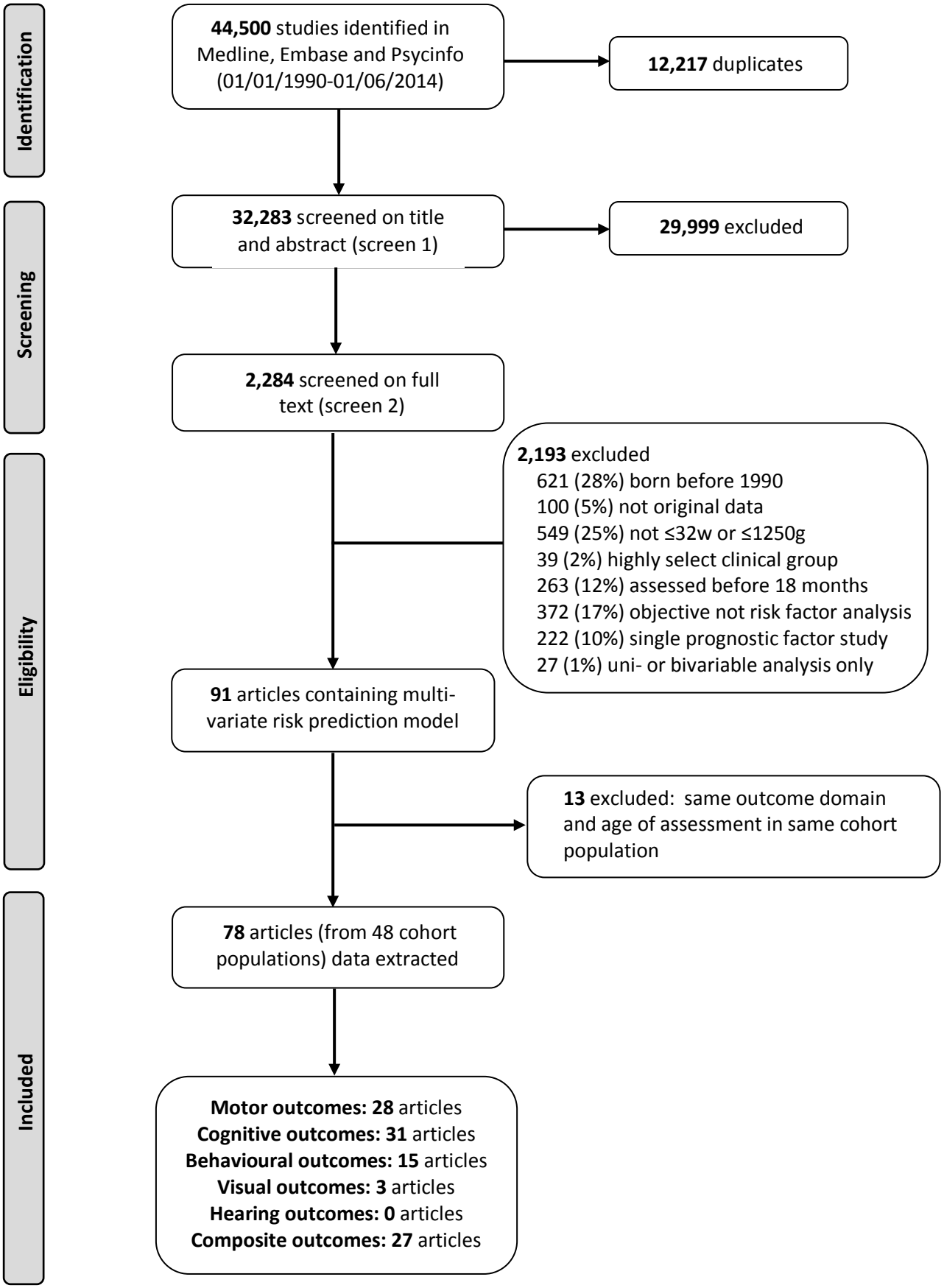


Table 2.1: Designs of studies presenting a multivariable risk factor model for neuro-developmental outcomes in VPT/VLBW participants included in the systematic review

	Number (%) of studies (n=78)	
Study design:		
Cohort	71	(91%)
RCT	5	(6.4%)
Case-control	1	(1.3%)
Cross-sectional	1	(1.3%)
Participant selection:		
Live births	36	(46%)
Admissions to NICU	35	(45%)
RCT participants	5	(6.4%)
Selected from a database	1	(1.3%)
Not clear	1	(1.3%)
Centre selection:		
Single centre	35	(45%)
All centres in a geographical -defined region	30	(38%)
Multicentre (non-random)	13	(17%)
Data extraction:		
Prospective	62	(79%)
Retrospective	13	(17%)
Not clear	3	(4%)
Routine follow-up:		
No	57	(73%)
Yes	21	(27%)
Continent of study:		
Europe	46	(59%)
North America	17	(22%)
Australia and New Zealand	12	(15%)
Japan	2	(2.6%)
International	1	(1.3%)
Start year of recruitment:		
1990-1995	25	(32%)
1996-2000	33	(42%)
2001-2010	20	(26%)
GA and BW inclusion criteria:		
VPT (<30 to ≤32 weeks)	24	(31%)
EPT (<25 to <29 weeks)	19	(24%)
ELBW (<1000g)	16	(21%)
Other GA/BW combination	19	(24%)
Exclusion criteria:		
Congenital anomalies	34	(44%)
Severity of disability	22	(28%)
Multiples	7	(9%)
Age of assessment (years)		
1.5 – 2	34	(44%)
3 – 5	21	(27%)
6 – 12	17	(22%)
Other	6	(8%)
Sample size (number assessed at follow-up):		
Median {IQR}	219 {141, 461}	
[Range]	[40, 7632]	

Fifty-nine percent of studies were conducted in Europe, 22% in North America, 15% in Australia and New Zealand, 3% in Japan and one study was based on an international RCT population. Most of the European studies were conducted in France (n=11; 7 from the EPIPAGE cohort[161]), followed by the UK (n=9; 4 from the EPICure cohort[57]) and the Netherlands (n=8). From North America, 15 were conducted in the US and two in Canada; 11 of the US studies were based on centres participating in the NICHD NRN programme that routinely follows up infants to 18-22 months. Fifty-five percent of studies defined the study cohort using GA only and 21% used BW only, the remaining studies used some combination of the two. European studies used solely GA criteria more commonly than other areas of the world (70%) and North American studies tended to favour the use of BW criteria; the NICHD NRN cohorts have an inclusion criteria of <1000g. Australian studies were more likely to use a combination of GA and BW than the other two main regions.

The most common age of assessment was 18-24 months (44%, n=34), but 27% studies followed up children to between 3-5 years and 22% to between 6-12 years. Six studies had an age range that spanned over more than one of these categories. No studies were found that performed a risk factor analysis beyond the age of 12 years in the VPT (≤32 weeks)/VLBW (≤1250g) population as defined by this review. The median sample size was 219 {IQR: 141 to 461} and 11 studies had more than 1000 participants. No studies performed a sample size or power calculation in relation to the risk factor analysis, and only one study referred to the number of events per variable in the modelling process.[114]

2.3.3 Model development

The median number of risk factor models presented per study was two and 25% of studies presented four or more models. To avoid the over-representation of studies presenting multiple models, the characteristics of the first model presented in the article was selected to summarise model characteristics. Most studies used the same techniques to develop their models and reported them the same way, so selecting just one model from each article to summarise model building, reporting and validation was thought to be a good reflection of practices adopted in the literature.

A summary of the model building techniques used is presented in Table 2.2. The median number of candidate risk factors considered at the outset was 17 [range: 3 to 51] and in seven studies it was unclear what the initial list of factors were. Seventy two percent (n=56) of studies provided a rationale for their choice of candidate factors, or used a comprehensive list with wide coverage (≥20 factors), but 28% (n=22) studies with less than 20 candidate factors gave no rationale at all. Sixty nine percent (n=54) of studies categorised some or all of the risk factors or outcomes that were measured on a continuous scale. In many cases there was a clinical rationale or a widely adopted convention was used, such as a threshold of 3 SD below the mean for IQ score to denote severe cognitive impairment, but the use of arbitrary thresholds with no justification was also widespread practice. The derivation and coding of outcomes and risk factors were described clearly in 85% (n=66) of studies and the main method of analysis used was linear or logistic regression.

Table 2.2: Model development of the first model presented in each study included in the systematic review of risk factor models for neurodevelopmental outcomes in children born VPT or with VLBW

	Number (%) of studies (n=78)
Number of candidate risk factors at outset:	
Median {IQR}	17 {11, 24}
[Range]	[3, 51]
Unclear	7 (9%)
Rational for candidate risk factors	
Yes	29 (37%)
No, but wide coverage (≥20)	27 (35%)
No	22 (28%)
Treatment of continuous risk factors:	
All left as continuous	17 (22%)
Some categorised	33 (42%)
All categorised	21 (27%)
No continuous risk factors	1 (1%)
Unclear	6 (8%)
Coding of risk factors and outcome clearly described:	66 (85%)
Statistical model	
Linear/logistic regression	74 (95%)
Multinomial/ordinal logistic regression	2 (3%)
Multilevel modelling	1 (1%)
Unclear	1 (1%)
Method of initial screening of candidate risk factors:	
No screening - all included	33 (43%)
P-value from univariable model	22 (28%)
P-value from multivariable model	6 (8%)
Other	7 (9%)
Unclear	10 (13%)
Model building strategy:	
All candidates/screened candidate risk factors included	30 (39%)
Stepwise selection	27 (35%)
Backward selection	9 (12%)
Forward selection	4 (5%)
Other	3 (4%)
Unclear	5 (6%)
Any interaction terms fitted:	2 (3%)
Any non-linear terms fitted:	0 (0%)
Any factors forced into the final multivariable model:	15 (19%)
Multicollinearity of risk factors mentioned/discussed:	11 (14%)
Model fit/assumptions checked:	15 (19%)
Number of risk factors in final multivariable model:	
Median {IQR}	5 {4, 9}
[Range]	[0, 25]
Unclear	1 (1.0%)

Forty-three percent (n=33) of studies entered all of the candidate variables into the multivariable model, 28% (n=22) only included candidate factors with a p-value below a set threshold in a univariable test of association with outcome and 8% (n=6) screened variables using multivariable analysis. The most popular method of model building after initial screening (or no screening) was to simply include all factors (39%, n=30) regardless of statistical significance, and stepwise selection (35%, n=27). Two studies attempted to fit some interaction terms and no studies reported fitting non-linear terms, and less than 20% discussed the issue of multicollinearity between risk factors or checked model assumptions and fit. Nineteen percent of studies (n=15) forced risk factors into the final model regardless of statistical significance; sometimes this was justified by the authors, for example based on other literature, but sometimes no explanation was given. The median number of risk factors included in the final prediction model was 5 {IQR: 4 to 9}.

A summary of the reporting and validation (if performed) of models is presented in Table 2.3. In studies where infants were lost to follow-up between hospital discharge and the outcome assessment, 30% (n=21) did not include a comparison of baseline characteristics between those lost to follow-up and those assessed. Generally the reporting of study attrition was quite poor, making it difficult to assess how representative the children assessed were of the original study sample recruited. Only 25% (n=19) of studies presented the amount of missing data for each variable included in the final model and only 31% (n=24) reported the number of infants included in the final model, so in most studies it was impossible to assess how representative the final analysis population was. Nearly three-quarters (n=57) of studies reported the point estimates, confidence interval or standard errors, and the p-value or significance level for all model coefficients included in the final

Table 2.3: Reporting and model validation in studies included in a systematic review of risk factor models for neurodevelopmental outcomes in children born VPT or with VLBW

	Number (%) of studies (n=78)	
Comparison of lost versus not lost to follow-up:		
No	21	(30%)
Yes	48	(70%)
N/A – all followed up	9	
Missing data presented for each variable:		
No	51	(67%)
Yes	19	(25%)
For some variables	6	(8%)
N/A – none missing	2	
Number of participants included in final model reported:	24	(31%)
Point estimates of model coefficients reported:		
No	8	(10%)
Yes – all variables	57	(73%)
Significant variables only	7	(9%)
Selected variables only	6	(8%)
Confidence intervals/standard errors of model coefficients reported:		
No	11	(14%)
Yes – all variables	56	(72%)
Significant variables only	5	(6%)
Selected variables only	6	(8%)
P-value/significance level of model coefficients reported:		
No	5	(6%)
Yes – all variables	55	(71%)
Significant variables only	11	(14%)
Selected variables only	7	(9%)
Any statistical validation performed:	4	(5%)
If yes, type of validation		
Discrimination	4	
Split sample	1	
Published in format for clinical use:	2	(3%)

multivariable model. However, 27% (n=21) of studies only reported these estimates for selected variables, or failed to report them at all.

Four studies[55,107,117,145] assessed the performance of the final model using statistical validation techniques. All four reported the area under the receiver operating curve which measures the discriminatory ability of the model to differentiate between individuals by severity of

outcome. One study[117] produced a decision tree/algorithm based on their model and another[55] produced a web based tool to predict either death or NDI (a composite of CP, deafness, blindness, motor or cognitive delay) for use in clinical practice (see the Outcome Trajectory Estimator at https://neonatal.rti.org/OTEstimator/). In this study the model was developed by randomly splitting the dataset into a development set (70%) and tested the model in a validation set (30%). The calculator produced can predict the probability of death, death or NDI, and NDI alone. The Outcomes Estimator developed by Tyson et al[54] that predicts death and death/NDI combined using five factors (on the same website) has been externally validated in independent populations,[56,162], however the model for NDI only (included in this review) has not yet been externally validated.

2.3.4 Risk of bias assessment

The overall results for the risk of bias assessment (Appendix 4) for all 78 studies are shown in Figure 2.2; the individual results for each study included in the synthesis of motor, cognitive and behavioural outcomes are reported in Chapters 3-5. Only 36% of studies fully satisfied the criteria for the selection of study population domain. To meet the full criteria, the study population had to be a prospective cohort of all live births in three or more centres with a sample size of at least 200 at discharge. Also sufficient information about the flow of participants from recruitment to discharge had to be reported to demonstrate that the study sample was representative of the general population of VPT/VLBW infants. Among the 49% of studies that did not meet the criteria in this domain, the most common reason was because recruitment was based in a single centre NICU.

To fully satisfy the criteria for the study attrition domain, a study had to have a follow-up rate of at least 80% in children up to five years of age, and at least 70% in children aged over

Figure 2.2: Risk of bias assessment using a modified version of the QUIPS tool

Study population: The study sample represents the population of interest on key characteristics sufficient to limit potential bias to the results.

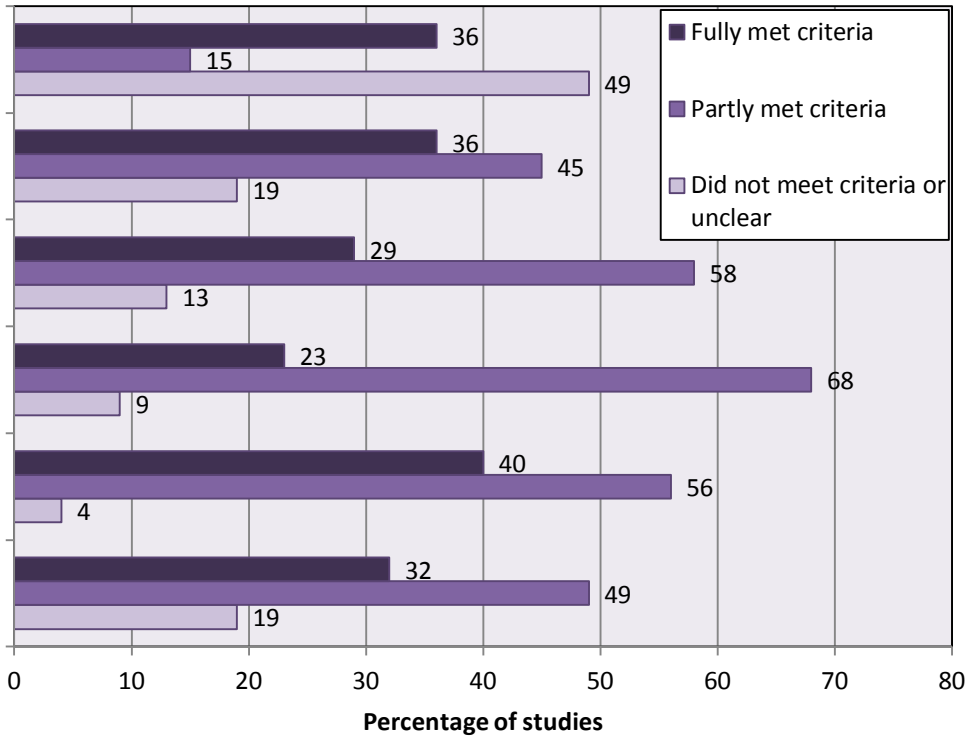
Study attrition: Loss to follow-up is not associated with key characteristics, sufficient to limit potential bias i.e. the study data at assessment adequately represent the original sample.

Outcome measurement: The outcomes of interest are adequately measured in study participants to sufficiently limit potential bias.

Prognostic factor measurement: The risk factors of interest are adequately measured in study participants to sufficiently limit potential bias.

Confounding measurement and account: Important potential confounders are considered as candidate factors.

Statistical analysis and reporting: The analysis is appropriate for the design of the study, limiting potential for presentation of invalid results, and the methods and results are well reported.



five years. Also the study needed to fully report the number of children lost to follow-up with reasons, perform a comparison of lost versus not lost to follow-up and address any important differences found. Only 36% of studies met all of these criteria, and generally the reporting of study attrition was quite poor, making it difficult to assess how representative the children assessed were of the original study sample recruited. Some studies included a diagram showing the flow of participants which was very useful, particularly when there were multiple assessments over time with dropouts and exclusions occurring at different stages.

The outcome measurement domain was fully satisfied by 23 (29%) of studies and partly satisfied by 45 (58%) of studies. If the study reported more than one outcome, then the worst graded outcome was used for the scoring. The most common reasons for a study being downgraded were unblinded assessment, blinding not reported, and outcome based on parent report. Apart from the lack of or non-reporting of blinding, outcomes were generally measured using comprehensive, well-validated tests or assessed using standardised diagnostic criteria. The 13% of studies that did not satisfy the criteria in this domain collected outcome data retrospectively, used data from routine clinical assessments with no standard protocol or strict diagnostic criteria, or used selected items from a validated instrument rather than the whole tool.

For the prognostic factor measurement domain, 68% of studies partially met the criteria. In general, risk factors were fairly well defined and clearly coded. Although most studies categorised some or all continuous variables, these were frequently based on widely used standard clinical criteria. The majority of studies also used a wide range of factors from different domains or provided some rationale for the inclusion of candidate variables. The most inadequately reported item in this domain was missing data for each prognostic factor

which is why only 23% of studies met the full criteria. Most studies did not report missing data separately for each outcome and risk factor and few reported the number of infants in the final model.

Thirty-one studies (40%) included GA or BW, sex, multiple pregnancy and at least one marker for social disadvantage in the list of candidate factors, but 44 (56%) of studies omitted at least one of these factors at the outset. The most frequently omitted factor was multiple pregnancy.

Twenty-two studies (35%) met the full criteria for the statistical analysis and reporting domain which required an appropriate statistical model for the study design and data, a model building strategy, with statistical methods and results reported fully and clearly. The most common reason for marking a study down was including all the candidate factors in a model with no model building strategy, a lack of clarity about the model building process (if there was one), and selective reporting of results.

2.3.5 Data synthesis of prognostic factors for poor neurodevelopmental outcome

The data synthesis and discussion of findings for motor/neurosensory, cognitive and behavioural outcomes are reported in Chapters 3-5. Twenty seven studies were based on composite outcomes comprising two or more outcome domains. The data from these models were not synthesised as the composites outcomes were too heterogeneous to be comparable. Furthermore, the findings from the review of motor, cognitive and behavioural outcomes suggested that different mechanisms underlie development in these domains, hence identifying risk factors for a composite outcome, would not have added any further clarity. A brief summary of the composite outcomes used is provided below.

2.3.6 Summary of composite outcomes

Twenty-seven studies reported a risk factor analysis for a composite outcome comprising two or more of the outcome domains; motor or cognitive function, hearing, vision, behaviour and death after discharge.[55,89,91,94-96,99,102-105,108,116,122,127,129,131,137-140,142,144, ,149,151,153,155] The domains used were summarised by age of assessment. The most common combination of outcomes used was from the motor, cognitive, vision and hearing domains. Two studies in this category also included death after discharge[142] and death after 36 weeks post menstrual age[104] as part of the composite, and two other studies in this category were based on the Amiel-Tison assessment.[163] The second most common combination was motor function, cognition and behaviour; two studies in this category used the Brunet-Lezine Revised assessment[164] and the two other studies used the Griffiths Mental Development Scale.[165] Both of these scales include items relating to the three components and combine them to produce a global developmental quotient (DQ) score.

Table 2.4: Outcome components used in the 27 studies reporting a risk factor analysis for a composite outcome by age of assessment

	Age at assessment			
Components	1.5-2 years	3-5 years	6-12 years	**Total**
Motor/Cognition/Vision/Hearing[a]	11	2	3	16
Motor/Cognition/Behaviour	3	1	0	4
Motor/Cognition/Vision/Hearing/Behaviour	1	1	1	3
Motor/Vision/Hearing	0	1	2	3
Motor/Cognition/Vision	1	0	0	1
Total	16	5	6	27

[a] 2 studies also included death after discharge or >36 weeks post menstrual age as part of the composite.

Apart from the four studies using a DQ score and the two studies based on the Amiel-Tison assessment, all other studies included in Table 2.4 used separate criteria for each component and children were classified as having NDI if they fulfilled the criteria for at least one

component. The most common criterion used in the motor domain was a diagnosis of CP (n=18) followed by having a motor score from the Bayley Scales of Infant Development (BSID)[166,167] below a predefined threshold (n=6). Similarly, having a cognitive score from the BSID below a predefined threshold was the most common criterion used to define cognitive delay (n=10), followed by IQ from another validated measurement tool (n=6). In the vision domain, blindness was used as the criterion in 11 studies and visual impairment (including blindness) was used in eight studies. In the hearing domain, deafness was used in seven studies and hearing impairment (including deafness) was used in 11 studies. In the three studies that included a behavioural domain (and did not use a global DQ score), a positive screen for ADHD was used as the criterion.

2.4 Discussion

The aim of the systematic review was to synthesise the available evidence from multivariable prediction models of neurodevelopmental outcome in children born VPT or VLBW aged over 18 months. The search resulted in 78 eligible articles that were included for data extraction in the following outcome domains: motor (n=28 articles), cognitive (n=31 articles), behaviour (n=15 articles), vision (n=3 articles), hearing (n=0 articles) and a composite outcome (n=27 articles). All studies were conducted in developed countries, which was not surprising as as a basic level of intensive care is required to keep VPT/VLBW infants alive in the first place. As many of these babies do not survive in developing countries, the developmental outcome of the few who do has not yet been highlighted as a major public health issue.

The risk of bias assessment highlighted a number of strengths and weakness in the research methodology and reporting of the studies identified. Also the data extracted on study design

and conduct, model development, validation and reporting revealed a number of issues which future studies should try to address. For each of the main issues, recommendations are made based on the latest literature in the field, and discussed in relation to the findings of the review.

2.4.1 Study design

2.4.1.1 Selection of participants

The optimal study design for prognostic research in the VPT/VLBW population is a prospective cohort of all live births in all centres in a geographically defined region, where the sample represents the source population and exposure to risk factors can be measured prior to an outcome occurring.[168] However, only 32% of studies in the review fulfilled the optimal study design and many were conducted in a single centre NICU. Studies based in a single centre, while convenient, have less generalisability due to centre differences in service provision, management practices and the socio-economic/ethnic characteristics of the local population. Also, single centre NICU populations tend not to be representative as they generally have a different case-mix compared with the wider VPT/VLBW population, containing a higher proportion of infants with poorer outcomes due to referrals of higher-risk infants from lower-level centres. Data from RCTs can be used to study prognosis, provided the treatment allocation is included as a predictor variable, though such studies may have restricted generalisabillity if strict trial eligibility criteria were used.

There was a lack of consistency in the GA and BW criteria used nationally and internationally to define cohorts, with some studies using prematurity or BW alone and other using a combination of both. Before the routine use of ultrasound, cohorts were generally defined by

BW due to the unreliable measurement of GA. Although there has been a shift from using BW defined cohorts in the last two to three decades, even studies conducted more recently used varying criteria. The typical VLBW cohort can be a fairly heterogeneous group due to the inclusion of the more mature but extremely growth restricted infants, and it is recommended that epidemiological studies of very small or immature newborn infants should be based on GA rather than BW criterion.[59] This is because the confounding of fetal growth status and maturity could lead to a distorted relationship between potential risk factors and outcome.

2.4.1.2 Sample size and events per variable

It is generally recommended that there should be a minimum of 10 outcome events per predictor variable (EPV) when using logistic regression models for predictive modelling, and models where the EPV is less than five should be interpreted with caution.[169] As the number of EPV declines below 10, simulation studies have shown an increase in bias and variability, unreliable confidence interval coverage and problems with model convergence.[170] Overfitting is a particular problem when small samples are used for predictive modelling, which can lead to the performance of the model being overoptimistic in the dataset from which it was developed.[171] The median sample size of studies included in the review was 219 and the median number of predictors in the final models was five, which means that, on average, studies lacked the power to detect associations in any outcome with an event rate of less than 20%. It is difficult to calculate a suitable sample size for risk prediction studies, and risk factor analysis is often a secondary aim when planning a cohort study, however, it is advisable not stray too far below the recommended number of events per variable.

2.4.2 Model development

2.4.2.1 Definitions of outcomes and risk factors

The measurement and assessment of outcome were generally clearly defined and robust in the studies included in this review, although it would be helpful to have more international agreement and standardisation, for example in the diagnostic criteria used by studies to define CP. There was more consistency across studies in the tests used to measure motor and cognitive outcomes than those used in the other outcome domains, however the use of different cut-points for impairment and the use of continuous scores makes it difficult to compare findings. Few studies reported whether assessments were blinded to previous medical history. For risk factor variables, some conditions such as BPD and IVH grading were defined quite consistently across studies, whereas other factors such as sepsis and NEC varied in definition or were not clearly defined at all. It is recommended that outcomes should be assessed prospectively using comprehensive, well-validated tests or using standard diagnostic criteria with a strict protocol, blinded to previous medical history. The definitions of outcomes and risk factors should be described clearly in sufficient detail if the model is to be correctly interpreted and applied by other researchers.

2.4.2.2 Coding and modelling of continuous variables

The factors retained in the final model, and the value of their coefficients are strongly influenced by the coding and methods used to model continuous and categorical variables.[172] The categorisation of continuous predictors should be avoided as this results in a loss of information and statistical power. It also results in the classification of individuals who are close but on either side of the chosen cut-point as apparently having very different levels of risk.[173] Using cut-points that are data-driven, such as the sample median, are problematic

when the model is transported to a different study population.[174] Forty-two percent of studies included in the review categorised some of the continuous risk predictors used in the modelling process, and 27% categorised them all. Assessing whether the relationship between a continuous predictor and an outcome is linear or nonlinear is also important; in the latter case the use of splines and fractional polynomials is recommended.[172] It is unclear how often this is done in practice, but none of the studies appeared to consider the fitting of non-linear terms.

2.4.2.3 Selection of risk predictors

There is no overall consensus on the best strategy to select variables for inclusion in a prognostic model, however some approaches are not recommended. This includes screening candidate factors using univariable (sometimes called bivariable) tests of association with outcome, as the correlation with other risk factors is not controlled for. This can result in the rejection of important predictors that only become prognostic after adjustment for other factors.[170,175] Twenty-eight percent (n=22) of studies adopted this approach and in 13% (n=10) the screening strategy was unclear. A further 43% (n=33) of studies included all candidate variables, but 25 of these studies started with fewer than 20 variables at the outset, so it is likely that some preselection process was applied but not reported. It is important to report any procedure used to reduce the number of candidate variables in sufficient detail so that the degree of coverage (of factors considered) can be assessed.

Automated variable selection processes, such as forward selection, backward elimination or stepwise approaches, are cautioned against as they are data-driven and ignore clinical plausibility. This can lead to poorly performing models with biased regression

coefficients.[171,176,177] Some important risk factors that are well-established in the literature may not always be statistically significant in a particular dataset, but it is advisable to include these in the development process. The preferred approach is to start with the full model and eliminate factors (by reviewing the effects of removing variables one at a time, taking both statistical significance and clinical relevance into consideration) as this avoids overfitting and selection bias and provides correct estimates of standard error.[178] The level of significance has a major effect on the number of variables retained; a 1% level will almost always result in a more parsimonious model than a 5% level. However, this approach can be problematic when there are a large number of candidate factors, as is usually the case when predicting neurodevelopmental outcome in the VPT/VLBW population, due to the vast amount of data usually collected during the neonatal period and the many environmental factors that can affect development following discharge. Six studies[55,105,119,120,125,158] in the review adopted a sequential multivariable approach to prediction, fitting candidate factors in stages according to the time frame in which they occurred or according to themes, such as clinical and socio-demographic. This approach seems reasonable, given that prognosis in these children is a complex and dynamic process, with environmental factors potentially superseding the influence of early biological factors as the child grows up.

2.4.3 Reporting and model validation

2.4.3.1 Attrition and missing data

One area that could be improved is the reporting of study ascertainment and attrition. While this was excellent in some studies, it was lacking in many others which made it difficult to determine how representative the study populations were. In prospective cohorts where children are followed up at multiple time points, the numbers and reasons for exclusions and

dropouts at each stage should be clearly reported in a flow diagram. In retrospective studies, the number of infants assessed for inclusion should be reported, in addition to the number selected for inclusion, which is important for assessing the risk of selection bias. There is evidence of selective drop-out in studies of preterm children, with those lost to follow-up more likely to be severely impaired or have mothers with a lower level of education,[38,179,180] therefore it is important to report and assess the potential impact of drop-out on the results. While the majority of studies reported that a comparison of children lost versus not lost to follow-up had been conducted, the data relating to this was not always fully reported in the article (or provided as supplemental material). If further participants are excluded from the final model due to missing data, the representativeness of analysis population should also be assessed, however the majority of studies did not even report the number of participants included in the final model after excluding those with missing outcome or risk factor data. The number of participants and the number of outcome events contributing to the data for each risk factor should be clearly reported, as well as the total number of participants contributing to the final model.

2.4.3.2 Model reporting and validation

Some studies only reported the results for selected variables in the final model, or only provided p-values or the regression coefficients with no confidence intervals. The absolute minimum that should be reported are the regression coefficients with confidence intervals for all factors retained in the final model, but it is also helpful to report the results of any intermediate stages of model development (as supplemental material if space is not permitted). If researchers wish to go further and develop a prognostic model, its performance can be assessed for calibration (comparing observed and predicted outcomes for groups),

accuracy (comparing the observed and predicted outcome for individuals) and discrimination (ability to distinguish between individuals at low and high risk of developing an outcome). These tests can be carried out on the data used to derive the model, but should also be carried out on new data, either from the same source (internal validation) or ideally on data from an independent source (external validation) to evaluate the transportability of the model. Methods of internal validation include split-sample (splitting the sample randomly into a 'training' set for model development and a 'test' set for model validation), cross-validation (developing a model in each set and testing it on the other set) and bootstrapping (resampling with replacement). The latter two resampling methods are more effective than the split-sample approach, which is inefficient and results in a loss of power, and therefore are the preferred approach for internal validation.[168,181] Only four studies[55,107,117,145] in the review assessed the performance of the model, and in three this was limited to a single measure of discrimination. Only one study[55] went further and used the split-sample method of internal validation and produced a tool available online that could be used in clinical practice.

2.4.4 Strengths and limitations of the review

The search filter used in this review was intentionally broad at the expense of accuracy in order to capture all studies with risk factor analyses, which resulted in a large number of articles retrieved. This approach is recommended for reviews in fields in which clinical prediction models are largely underdeveloped, rather than a more specific search filter, which would have had a high false negative rate, leading to many articles being missed.[182] No language restrictions were imposed and no further articles were identified in the hand-search of bibliographies of all studies included, so it is unlikely that there were any major omissions. The birth year of study cohorts spanned over a 24 year period and some of the factors

affecting outcome in the early 1990s may not be so relevant in current preterm populations. Also, some studies in the review were published before the proliferation of comprehensive guidelines on the conduct and reporting of research studies (http://www.equator-network.org/), and may not reflect the standards of more current work. Furthermore, many of them preceded the publication of the large literature that now exists on risk prediction modelling. Specific reporting guidelines for this field have only just emerged.[68]

A major challenge in this review was the number of studies from the same cohort population that overlapped in time period of recruitment, type of outcome assessed and age of assessment. For example 20 studies based in centres from the NICHD NRN were eligible for inclusion in this review, of which 11 were included[55,88,89,110,128,132,133,145,148,149,155] and nine were omitted.[76,78-85] Including them all would have led to an over-representation of this population and a lack of independence if the same participants were included in more than one study. A panel was convened to review such studies before reviewing the results, and selected studies based on relevance and objective, though it was not possible to apply a strict set of criteria to each set of overlapping studies. Another source of multiplicity was when a single study reported more than one model for in the same outcome domain. Again, it was difficult to apply a standard set of criteria for inclusion, but bias was minimised by making selection decisions before the synthesis was performed.

It was not possible to conduct a quantitative synthesis of the size of effect for any robust prognostic factors identified by the review, as explanatory prognostic factor studies that investigate the causal relationship between a single risk factor and outcome were not included in the review. For example, to perform a meta-analysis to estimate the relative risk

of CP for each grade of IVH, one would have to include all the articles that focus solely on IVH as a risk factor for CP, in either a univariable analysis or adjusting for confounders.

2.4.5 Conclusions

This systematic review of 78 published articles reporting multivariable risk factor models for neurodevelopmental outcome in surviving VPT/VLBW children revealed some shortcomings in methodology and reporting that could be improved in future studies, and confirmed that that there is a dearth of properly designed and well-conducted prognostic modelling studies in this field. Modelling long-term outcome in such a heterogeneous population is challenging; often with the existence of multiple impairments within the same individual and with multiple risk factors acting sequentially over time. In the three published reviews of risk factors for cognitive and motor impairment and behavioural problems that were based on these studies, the evidence for most risk factors was mixed or unclear (Chapters 3-5). This may be due to the difficulty of modelling prognosis in this population, but also due to differences in study design, study population, methodological quality and lack of standardization of measures. The findings and recommendations of this critical review should be used as a basis for the design and analysis of future studies seeking to develop multivariate risk factor or prognostic models in this population. The reporting of future prognostic studies should follow the recent TRIPOD guidelines on the transparent reporting of prognostic research.[68]

Most of the studies did not follow up children beyond two years, and none of the studies beyond 12 years. There is an urgent need for larger population cohorts, followed up routinely beyond two years, as assessments are not always reliably prognostic at this age, particularly for non-severe conditions,[45,66] and the prediction of less common outcomes needs greater power. The use of meta-analysis to determine the strength of association of the prognostic

factors identified by the review, including the results of single risk factor studies, would also be a valuable future direction.

Chapter 3: Prognostic factors for motor and neurosensory impairment in children born VPT or with VLBW: systematic review results

3.1 Introduction

Surviving infants born VPT or with VLBW are at high risk of developing motor and neurosensory deficits in early childhood. The prevalence of CP is higher in preterm children and inversely related to GA.[154,183,184] In a Swedish study, the prevalence of CP per 1000 live births during 2003 to 2006 was: 71.4 for births <28 weeks GA, 39.6 for 28-31 weeks GA, 6.4 for 32-36 weeks GA compared to 1.4 in term born infants >36 weeks GA.[184] Even in children without CP or obvious neurological impairment, subtle abnormalities may be present, such as difficulties with balance, co-ordination and fine motor function. These relatively minor motor difficulties in otherwise normally functioning children are known as developmental coordination disorder (DCD),[185] and are often not diagnosed until school age. The prevalence DCD is around 5-6% in the general population,[185] but has been reported to be much higher in VPT/VLBW children with prevalence rates of between 30 and 40%.[90,109,186]

Visual impairment is more common among VPT/VLBW children compared to their term born peers, due to incomplete vascularisation and unstable oxygenisation at the time of birth. The more preterm the infant, the greater the risk of ROP during the neonatal period, which is a major cause of childhood blindness in developed countries.[187] Blindness and severe visual impairment is inversely related to GA, with rates of 4-8% reported in infants born 25 weeks GA and below and 1-2% in infants born 26-27 weeks.[65,155,188] Opthalmic morbidity (defined as

visual acuity below 0.0 log units, the presence of strabismus, myopia, colour vision defect or visual field defect) has been reported to be as high as 51% in low BW children aged 10-12 years compared to 20% in children not born with low BW.[189] The need for prescription glasses is related to GA, with 24% of EPT children requiring them by the age of six years compared to 4% of term controls.[65] This large disparity persists into adulthood, with rates of 64% in ELBW adults versus 37% in normal BW controls.[190]

VPT/VLBW children are also at increased risk of hearing impairments during early childhood as the rapid audiological development which occurs before 32 weeks is disrupted. Neonatal hearing loss increases with decreasing GA, ranging from 7.5% in infants born at 24 weeks of gestation to 1.2% in infants born at 31 weeks.[191] In a cohort of six year old EPT children, 6% required hearing aids and 4% had mild hearing loss compared to 1% of term controls.[65] In addition, a more subtle hearing abnormality known as disordered central auditory processing, has been reported to be more common in preterm children when assessed at eight years of age.[192]

Among the 78 articles that were identified in the systematic review and included in the data extraction, 28 contained risk factor analyses for motor outcomes.[89,90,93,96,98,101,107,109,111,117,118,123,126,128,129,132,134,140,141,147,149-151,154,155,158-160] These studies were based on 19 cohort populations; five other articles based on motor outcomes were excluded due to cohort overlap.[79,81,83-85] Three articles of the 78 reported risk factor analyses for visual impairment,[104,113,114] and no articles were found for hearing impairment. In this chapter, the results from the 28 articles on the risk of CP and/or motor impairment and the 3 articles on the risk of visual impairment are reported and synthesised where possible.

3.2 Data synthesis of prognostic factors for motor impairment

3.2.1 Study characteristics

The main study design was prospective cohort (n=26); there was also one case-control study[126] and one RCT population.[128] Of the 26 prospective cohorts, 12 were ascertained from all live births in a geographically defined region[90,93,96,98,111,123,129,147,150,151,154,158] and nine were recruited from a single centre NICU.[101,107,109,117,118,132,149,159,160] Studies were conducted in 15 countries: US (n=5), Australia (n=5), France (n=3), Netherlands (n=3), Canada (n=2), Finland (n=2) and 1 study each from Denmark, Estonia, Germany, Japan, Norway, Sweden, Switzerland and the UK & Republic of Ireland. The median sample size was 236 [range: 48 to 3785]. Three studies had more than 1000 participants[90,93,155] and the remaining studies had 545 or less. Only two studies were restricted to extremely preterm children; <28 weeks[140] and <26 weeks[158]. Two studies examining risk factors for CP excluded multiple births.[89,126] The risk of bias assessment classified 13 (46%) studies as having a low-moderate risk of bias and 15 (54%) studies as having a moderate-high risk of bias (Figure 3.1).

From the 28 studies identified, 44 risk factor models for motor outcome were included in the review. The models were divided into the following outcome domains for data synthesis: CP diagnosis, motor impairment measured by scales and neurological dysfunction not diagnosed as CP. Models based on motor impairment measured by scales were further divided into those that included the whole cohort representing the whole spectrum of disability, and those that excluded children with major disability; typically defined as a CP diagnosis and/or severe neurosensory or cognitive impairment. Several studies presented models for more than one type of motor outcome domain or reported multiple subtests within a domain. The median

Figure 3.1: Risk of bias assessment of the 28 studies reporting motor outcomes

	Risk of bias domain						
	Study Participation	Study Attrition	Prognostic Factor Measurement	Outcome Measurement	Study Confounding	Analysis and Reporting	Low-moderate risk of bias
Andrews (2008)							
Arnaud (2007)							✓
Beaino (2010)							✓
Charkaluk (2010)							✓
Davis (2007)							✓
Dewey (2011)							
Franz (2009)							
Goyen (2009)							
Hansen (2004)							✓
Janssen (2008)							
Janssen (2009)							
Leversen (2011)							✓
Matsuda (2000)							
Messinger (2010)							
Mikkola (2005)							
Orchinik (2011)							
Pin (2010)							
Schlapbach (2011)							✓
Schmidhauser (2006)							✓
Stoelhorst (2003)							✓
Taylor (2006)							
Tommiska (2003)							
Toome (2013)							✓
Vincer (2006)							✓
Vohr (2005)							✓
Wood (2005)							✓
Zanudin (2013a)							
Zanudin (2013b)							

Risk of bias: Low Moderate High

number of candidate risk factors considered at the outset was 17 [range: 6 to 45]. For the initial screening of candidate factors to be entered into the final model, 25 (57%) included them all and 13 (30%) included those with a p-value below a set threshold after initial screening. The most popular method of model building after initial screening was to include all factors screened 16 (36%) and stepwise selection 13 (30%). Only 16 (36%) models reported the number of participants included in the final model presented. Two studies[107,117] assessed model discrimination using the area under the receiver operating curve and one also presented the model as a decision tree for clinical application,[117] but apart from that no other studies performed any type of statistical or clinical validation.

3.2.2 Prognostic factors for cerebral palsy

Twelve studies contained a risk factor analysis for CP; seven studies assessed outcome between 1.5-2.5 years[126,140,150,151,154,155,158] (Table 3.1) and five studies between 5-8 years (Table 3.2).[89,93,107,111,129] Risk factors that were found to be significant in at least one study at low-moderate risk of bias and examined in the final model of at least two other (independent) studies are presented in Figure 3.2 by age of assessment (<5 years and ≥5 years).

There was strong evidence that brain injury diagnosed during the neonatal period was predictive of CP; in 11 studies out of the 12 that entered this factor in the final model, it was statistically significant in 9 (OR range: 2.3 to 43).[89,93,107,111,140,151,154,155,158] In the nine studies that found brain injury significant, four combined IVH grade 3-4 and periventricular leukomalacia PVL into a single variable;[107,111,140,151] one study used IVH grade 3-4 only;[154] one study defined brain injury as parenchymal pathology and/or ventriculomegaly,[158] and three studies entered IVH and PVL as separate variables.[89,93,155] In the latter three studies, IVH grade 3-4 was significant in all models and PVL was significant in the two largest studies.[93,155]

Table 3.1: Summary of studies reporting risk factor analyses for cerebral palsy in children born VPT or with VLBW aged <5 years

Study reference [identifier]	Country and recruitment period	Age of assessment (years)	GA (weeks) and/or BW (grams)	Design and participants	Number (%) of survivors assessed[a]	Definition and assessment of cerebral palsy as reported	Significant risk factors (p<0.05) in final model
Tommiska (2003)[150] [A]	Finland 1996-1997	1.5	<1000g	PC of all live births in Finland enrolled for the national routine FUP.	208 (100%)	Non-progressive motor impairment with spastic or dystonic muscle tone, brisk tendon reflexes, positive Babinski's sign and primitive reflexes. Clinical examination by a neurologist at local centre.	No AN steroids, hospital area, vaginal delivery.
Vohr (2005)[155] [B]	US 1993-1998	1.5-1.8	<33w and <1000g	PC of infants admitted to the NICU of 12 centres participating in the multi-centre NICHD NRN routine FUP.	3785 (80%)	Moderate-severe CP defined as non-ambulatory or requiring an assistive device for ambulation. Neurological examination based on Amiel-Tison[193] by a trained, blinded developmentalist.	No AN steroids, BPD, IVH 3-4, male sex, PN steroids, PVL.
Schlapbach (2011)[140] [C]	Sweden 2000-2007	1.5-2	<28w	PC of infants admitted to the NICU of 9 Swiss perinatal centres participating in the neonatal network routine FUP.	541 (77%)	SCPE classification. Assessed by developmental paediatricians at 1 of the 9 follow-up centres.	IVH 3-4/PVL, proven sepsis.
Vincer (2006)[154] [D]	Canada 1993-2002	2	<31w	PC of all live births of mothers resident in the province of Nova Scotia and enrolled in the perinatal FUP.	545 (97%)	Definition complied by Bax.[194] Assessed in a general physical and neurodevelopmental examination. Diagnosis of infants with suspected CP confirmed by a paediatric neurologist.	No AN steroids, era of birth, GA<28w, IVH 3-4, PDA, PN steroids, RDS, resuscitation in delivery room.
Toome (2013)[151] [E]	Estonia 2007	2	<32w	PC of all live births in Estonia enrolled in the national neonatal research routine FUP.	155 (99%)	SCPE classification. Physical assessment by a paediatrician and neurological examination by a child neurologist in 1 of 2 centres.	IVH 3-4/PVL 2-4.
Matsuda (2000)[126 b] [F]	Japan 1992-1996	2-2.5	<31w	Unmatched case-control study of all deliveries in a single centre (Kagoshima). Multiple births excluded.	192 (22 cases, 170 controls)	Chronic disability of CNS origin, characterised by aberrant control of movement or posture and non-progressive. All physical findings on presumed CP cases were reviewed retrospectively by a developmental paediatrician.	Intrauterine infection.
Wood (2005)[158] [G]	UK and Republic of Ireland 1995	2.5	<26w	PC of all live births in the UK and Republic of Ireland (EPICure Study).	283 (92%)	Non-progressive disorder of movement and posture. A structured neurological examination based on Amiel-Tison[163] was performed and Classified retrospectively based on Bax.[195]	Enteral feeding by day 7, male sex, moved hospital <24h, parenchymal pathology and/or ventriculomegaly, PN steroids >8w.

[a] Percentage of survivors assessed for outcome; not all included in the final model.

[b] 2 models for CP reported. Model using the total sample of 192 children included. Model excluding children with antenatal complications not included.

Abbreviations: AN antenatal; BPD; bronchopulmonary dysplasia; CNS central nervous system; CP cerebral palsy; FUP follow up; GA gestational age; IVH intraventricular haemorrhage; NICU neonatal intensive care unit; NICHD NRN National Institutes of Child Health and Human Development Neonatal Research Network; PC prospective cohort; PDA patent ductus arteriosus; PN postnatal; PVL periventricular leukomalacia; RDS respiratory distress syndrome; SPCE Surveillance of Cerebral Palsy in Europe.[196]

Table 3.2: Summary of studies reporting risk factor analyses for cerebral palsy in children born VPT or with VLBW aged ≥5 years

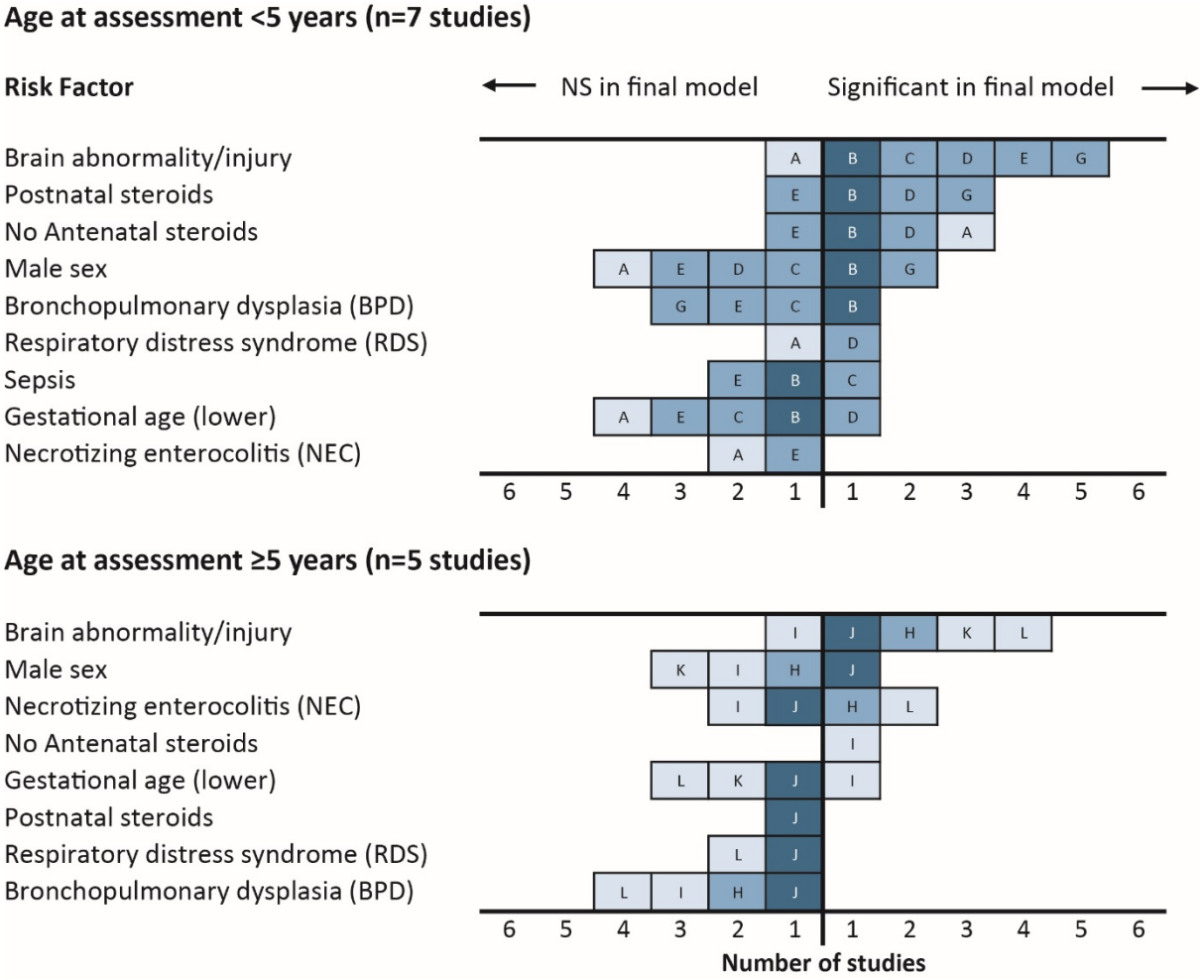
Study reference [identifier]	Country and recruitment period	Age of assessment (years)	GA (weeks) and/or BW (grams)	Design and participants	Number (%) of survivors assessed[a]	Definition and assessment of cerebral palsy as reported	Significant risk factors (p<0.05) in final model
Hansen (2004)[111] [H]	Denmark 1994-1995	5	<28w or <1000g	PC of all live births in Denmark ascertained from all 18 neonatal care units and the Danish Medical Birth Register (ETFOL Study).	252 (94%)	SCPE classification. One physician blinded to neonatal course conducted all assessments. Formal neurological assessment only performed if abnormal tone, gait or posture.	IVH 3-4/PVL, NEC (surgery or x-ray diagnosed).
Mikkola (2005)[129] [I]	Finland 1996-1997	5	<1000g	PC of all live births in Finland enrolled for the national routine FUP.	193 (94%)	A neurological examination using a modified partial Touwen test.[197] Results were classified according to Hadders-Algra.[198]	No AN steroids, earlier GA, hospital area.
Beaino (2010)[93] [J]	France 1997	5	<33w	PC of all live births in 9 French regions comprising one third of all births (EPIPAGE Study).	1817 (75%)	SCPE classification. Children with abnormal neurological examination underwent expert review to validate and classify diagnosis.	Cystic PVL/ IPH, IVH 2-3 echo-densities/VD, male sex.
Franz (2009)[107 b] [K]	Germany 1996-1999	4.6-7	<30w and <1500g	PC study of Infants admitted to a single centre level-3 NICU (Ulm University).	219 (83%)	Gross Motor Functioning Classification Scale (GMFCS)[199] ≥1.	IVH/PVH ≥3, lower maternal education, ROP ≥3, smaller HC SDS gain (birth to discharge).
Andrews (2008)[89] [L]	US 1996-1999	5-8	<32w	PC of consecutive live births in a single centre (Alabama). Multiple births excluded.	261 (70%)	Defined as abnormal muscle tone in at least 1 extremity and abnormal control of movement and posture. General physical and neurological examination by a certified nurse practitioner under supervision of a developmental paediatrician.	African-American race, history of seizures, IVH 3-4, NEC.

[a] Percentage of survivors assessed for outcome; not all included in the final model.

[b] 2 models for CP reported. Same perinatal factors fitted with change in weight variables added to 1st model and change in head circumference variables added to 2nd model. Perinatal factors reported as significant if p<0.05 in both models and non-significant if p≥0.05 in both models.

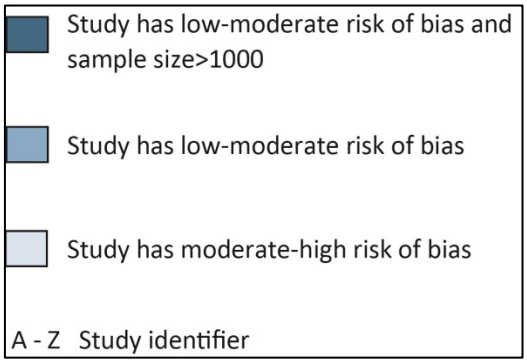
Abbreviations: AN antenatal; CP cerebral palsy; FUP follow up; GA gestational age; HC SDS head circumference standard deviation score; IPH intraparenchymal haemorrhage; IVH intraventricular haemorrhage; NEC necrotising enterocolitis; NICU neonatal intensive care unit; PC prospective cohort; PVH periventricular haemorrhage; PVL periventricular leukomalacia; ROP retinopathy of prematurity; SPCE Surveillance of Cerebral Palsy in Europe;[196] VD ventricular dilatation.

Figure 3.2: Evidence synthesis of risk factors for cerebral palsy in children born VPT or with VLBW



Note: Risk factor presented if significant ($p<0.05$) in the final model of at least one study at low-moderate risk of bias and entered into the final model of at least 3 studies (across all ages).
Brain abnormality/injury: IVH grade 2-4 [A]; IVH grade 3-4 [D/I]; IVH grade 3-4 and/or PVL [C/E/H/K]; IVH grade 3-4 and PVL separately [B/J/L]; Parenchymal pathology and/or ventriculomegaly [G].
BPD: Oxygen requirement at 36 weeks GA [B/C/E/G/H/I/J]; Not specified [L].
NEC: Perforated NEC [A/I]; Surgical or x-ray diagnosed [H]; NEC stage 2-3 [E]; Not specified [J/L].
RDS: Severe hyaline membrane disease (ventilator dependent) [D]; Not specified [A/J/L].
Sepsis: Confirmed with positive blood culture [C/E]; Early or late [B].

Key



The two studies at moderate-high risk of bias that did not find brain injury diagnosed in the neonatal period significantly predictive of CP in the final model used IVH grade 2-4[150] and grade 3-4.[129]

There was some evidence that postnatal steroids administered in the neonatal period was associated with a diagnosis of CP at around two years of age with three studies at low-moderate risk of bias reporting significance in the final model (OR range: 2 to 5)[154,155,158] and one study at low-moderate risk of bias reporting non significance.[151] However, postnatal steroid use was not found to be prognostic in the single large study based on diagnosis at five years.[93] In the two studies that reported non-significant findings, the prevalence of treatment with postnatal steroids was much lower (5% and 18% of the cohort population compared to 30-60% in the significant studies). Wood et al[158] examined the effect of the duration of postnatal steroid use and found that only those treated for more than eight weeks were at substantially increased risk of CP.

The use of antenatal steroids were found to be significantly protective for CP in two studies at low-moderate risk of bias[154,155] and one study at moderate-high risk of bias[150] based on diagnosis of CP at around two years of age (OR range: 0.3 to 0.66) and not significant in one study at low-moderate risk of bias.[151] One study at moderate-high risk of bias,[129] based on the same cohort as a study included in the <5 year age band,[150] corroborated this result at five years. Most of these studies reported that around 80% of the study population had been exposed to antenatal steroids.

The relationship between male sex and CP was conflicting in both age groups; three studies at low-moderate risk of bias,[93,155,158] including the two largest studies, found that males were

significantly more likely to develop CP, however four other studies at low-moderate risk of bias[111,140,151,154] and three studies at moderate-high risk of bias[107,129,150] did not support this finding. The range of ORs in the significant studies was 1.4 to 2.3 and only of borderline statistical significance.

Earlier GA was found not to be predictive of CP in seven of the nine studies including this as a factor in the final model.[89,93,107,140,150,151,155]

3.2.3 Prognostic factors for motor impairment measured by scales

Six studies presented a risk factor analysis for motor impairment in children of all levels of disability (Table 3.3)[128,132,134,147,149,155] and 10 studies in children free of major disability (Table 3.4).[96,98,101,109,111,117,118,123,158-160] The most common assessment used before five years was the Psychomotor Development Index (PDI) from the BSID-II.[166] This is a comprehensive, rigorously validated test which is widely used as a standardised assessment of key developmental milestones in infants up to 42 months. The PDI assesses gross and fine motor development, involving some aspects of psychological functioning. All studies assessing motor development in children free of major disability after five years used the Movement Assessment Battery for Children (MABC).[200] The MABC is a validated norm-referenced test that measures fine and gross motor skills in three key areas: manual dexterity, ball skills and static/dynamic balance. A total score <5th centile is indicative of definite DCD, and a total score <15th centile represents a borderline motor impairment.

The risk factors emerging from the four studies reporting outcomes prior to five years for children of all levels of disability[128,134,147,155] are summarised in Figure 3.3. There were only two studies at low-moderate risk of bias in this grouping and the results are dominated by the

largest study.[155] There was evidence that the use of postnatal steroids were associated with motor impairment in three studies,[134,147,155] but no other clear patterns emerged from this small sample of studies. The results from the 10 studies assessing motor development in children free of major disability[20, 21, 31-37, 40] are also summarised in Figure 3.3. In the seven studies using the total score from the MABC after five years of age, the score was treated as continuous in four studies,[101,111,123,160] dichotomised at <15th centile in two studies[109,117] and at <5th centile in one study.[98] Male sex emerged as the only risk factor for which there was evidence of an association with motor impairment in children free of major disability, in both age groups. All four studies at low-moderate risk of bias[98,111,123,158] and three studies at moderate-high risk of bias[117,118,160] that included male sex in the final model found a significant association.

3.2.4 Prognostic factors for neurological dysfunction not diagnosed as CP

Seven studies presented risk factor analysis for general neurological dysfunction[90,107, 129,141,150,158,159] (Table 3.5). The outcome measures were heterogeneous and there were only one or two studies within each age and level of disability stratum so it was not possible to synthesise the results in any meaningful way, however the risk factors that were significant in the final models are listed for completeness. Similarly to the results for CP, IVH and/or PVL appeared to be a prominent feature in the medical history of children developing later signs of neurological dysfunction.

Table 3.3: Summary of studies reporting risk factor analyses for motor impairment in children of any level of disability born VPT or with VLBW

Study reference [identifier]	Country and recruitment period	Age of assessment (years)	GA (weeks) and/or BW (grams)	Design and participants	Number (%) of survivors assessed[a]	Outcome measure (continuous (cts) unless otherwise specified)	Significant risk factors for poorer outcome (p<0.05) in final model
Age of assessment <5 years							
Pin (2010)[134] [M]	Australia 2006-2007	1.5	<30w	PC of infants admitted to 4 tertiary NICUs in Melbourne.	54 (76%)	Gross motor development.Total score from Alberta Infant Motor Scale (AIMS).[201] Blinded assessment.	PN steroids.
Vohr (2005)[155] [B]	US 1993-1998	1.5-1.8	<33w and <1000g	PC of infants admitted to the NICU of 12 centres participating in the multicentre NICHD NRN routine FUP.	3785 (80%)	Fine and gross motor skills. PDI score from BSID-II (<70 vs. ≥70). Blinded assessment.	Adjusted age, no AN steroids, lower BW, BPD, IVH 3-4, male sex, lower maternal education, multiple pregnancy, PVL, PN steroids, sepsis (early or late).
Stoelhorst (2003)[147 b] [N]	Netherlands 1996-1997	2	<32w	PC of all live births in 3 Dutch health regions comprising 9% of the population.	144 (61%)	Fine and gross motor skills. PDI score from BSID-I.[202] Blinded assessment.	PN steroids, SGA.
Messinger (2010)[128 c] [O]	US 1999-2001	2.5	<1000g	Infants admitted to the NICU of 12 centres participating in the multi-centre NICHD NRN routine FUP and enrolled in a glutamine supplementation RCT.	539 (47%)	Fine and gross motor skills. PDI score from BSID-II.	Lower BRS at 18m, BW ≤750g, male sex, lower maternal education, lower maternal income, lower PDI at 18m.
Age of assessment ≥5 years							
Orchinik (2011)[132 d] [R]	US 2001-2003	5	<28w or <1000g	PC of Infants admitted to a single centre NICU (Ohio) participating in the multicentre NICHD NRN routine FUP.	133 (67%)	Fine and gross motor skills. T-score from BOT-2 short form version[203] <10th centile. Blinded assessment.	Neurosensory disorder and/or MDI<70 at 20m.
Taylor (2006)[149 e] [S]	US 1992-1995	8	<1000g	PC of infants admitted to a single centre NICU (Ohio) participating in the multicentre NICHD NRN routine FUP.	204 (86%)	Fine and gross motor skills. T-score from BOT short form version[204] (cts and <1SD below mean of control group). Blinded assessment.	Model 1 (T-score<1SD): BPD, longer neonatal hospital stay. Model 2 (cts T-score): BPD, BW <750g, NEC, PN steroids, NRI>3, longer neonatal hospital stay.

[a] Percentage of survivors assessed for outcome measure specified; not all included in the final model.

[b] 2 models for motor skills reported. Model based on 2 year outcome included. Model based on 1.5 year outcome not included.

[c] Several models for motor skills fitted. Full model adjusting for 18 month PDI and BRS total score included.

[d] Each risk factor was fitted separately and adjusted for sex, race, parental SES and months in school at testing (the article did not report results for the adjustment factors).

[e] 2 models for motor skills reported; one based on dichotomous outcome and one based on continuous outcome. Risk factors reported separately for each model in the table. Each risk factor was fitted separately and adjusted for sex, race, parental SES, family stressors and family resources (the article did not report results for the adjustment factors).

Abbreviations: AN antenatal; BPD bronchopulmonary dysplasia; BRS Behaviour Rating Scale from BSID; BOT Bruininks-Oseretsky Test of Motor Proficiency; BSID Bayley Scales of Infant Development;[166] BW birth weight; FUP follow up; GA gestational age; IVH intraventricular haemorrhage; MDI Mental Developmental Index from BSID; NEC necrotising enterocolitis; NICU neonatal intensive care unit; NICHD NRN National Institutes of Child Health and Human Development Neonatal Research Network; NRI neonatal risk index; PC prospective cohort; PDI Psychomotor Developmental Index from BSID; PN postnatal; PVL periventricular leukomalacia; RCT randomised controlled trial; SGA small for gestational age; SD standard deviation.

Table 3.4: Summary of studies reporting risk factor analyses for motor impairment in children free of major disability born VPT or with VLBW

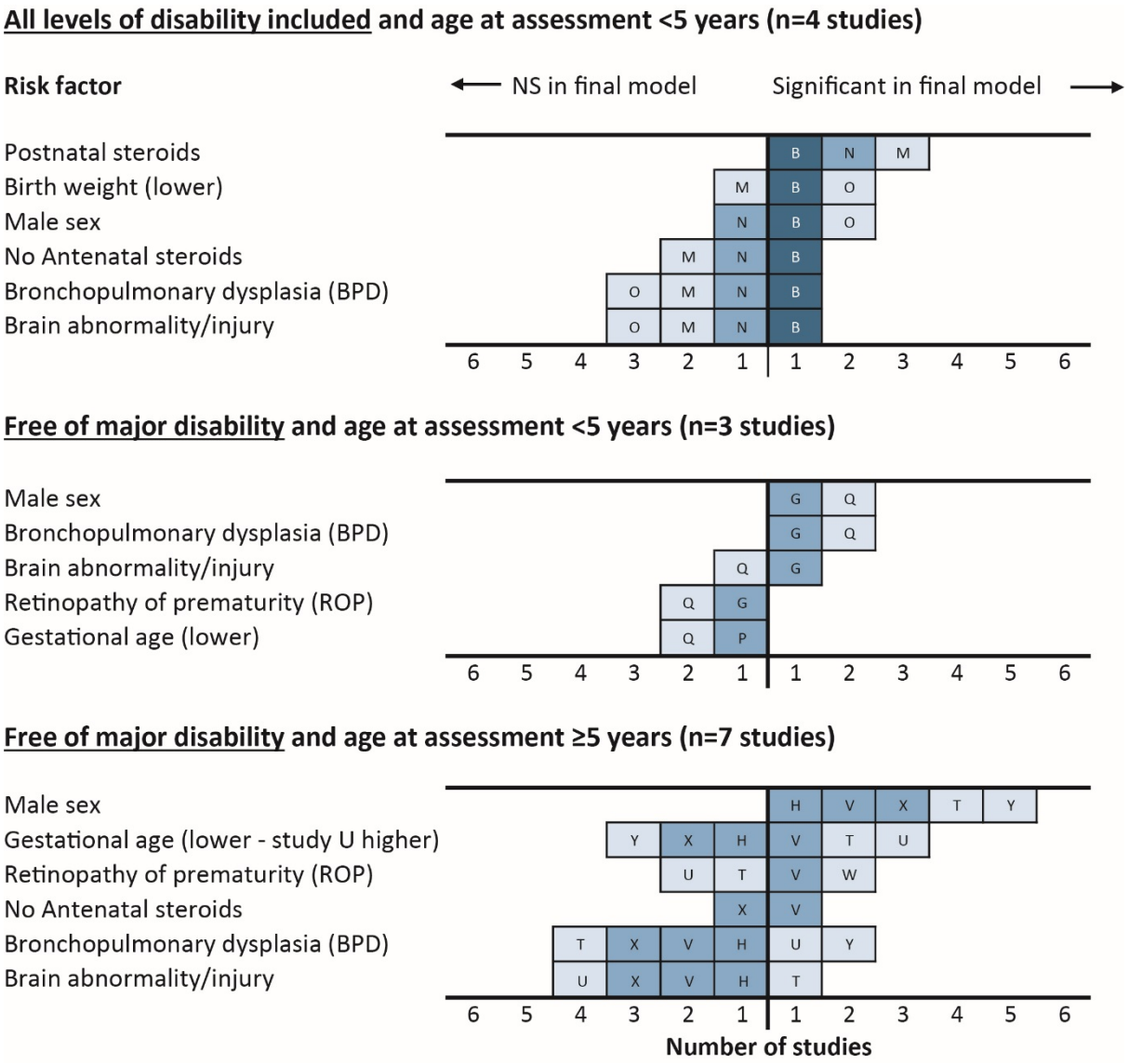
Study reference [identifier]	Country and recruitment period	Age of assessment (years)	GA (weeks) and/or BW (grams)	Design and participants	Exclusion criteria for major disability	Number (%) of survivors assessed[a]	Outcome measure (continuous (cts) unless otherwise specified)	Significant risk factors for poorer outcome (p<0.05) in final model
Age of assessment <5 years								
Charkaluk (2010)[96] [P]	France 1997	2	<33w	PC of all live births in 1 French region (Nord-Pas-de-Calais, EPIPAGE Study).	CP or severe neurosensory impairment (n=45).	347 (64%)	Fine motor domain of Brunet-Lezine scale (revised).[164]	Longer intubation, lower parental occupation.
Wood (2005)[158] [G]	UK and Republic of Ireland 1995	2.5	<26w	PC of all live births in the UK and Republic of Ireland (EPICure Study).	PDI score <55 from BSID-II or functional motor disability (n=47).	197 (64%)	Fine and gross motor skills. PDI score from BSID-II.	Breast milk in hospital, BPD, non vaginal breech delivery, male sex, multigravida, parenchymal pathology/ventriculomegaly
Janssen (2008)[118] [Q]	Netherlands 1996-2001	2-3	<33w	PC of infants admitted to a single centre NICU (Nijmegen) and enrolled for the Dutch neonatal routine FUP.	Under multidisciplinary treatment for disability (n=23).	437 (56%)	Fine and gross motor skills. PDI score from BSID-II (<85 vs. ≥85). Blinded assessment.	BPD, male sex, neonatal convulsions, lower maternal education.
Age of assessment ≥5 years								
Hansen (2004)[111] [H]	Denmark 1994-1995	5	<28w or <1000g	PC of all live births in Denmark (ETFOL Study).	CP or visual disability (n=23)	137 (51%)	Motor development. Total score from MABC. Blinded assessment.	Male sex.
Janssen (2009)[117] [T]	Netherlands 1996-2001	5	<33w	PC of infants admitted to a single centre NICU (Nijmegen) and enrolled for the Dutch Neonatal routine FUP.	CP, visual disability or under multidisiplinary treatment for disability (n=31).	371 (47%)	Motor impairment. Total score <15th percentile from MABC. Blinded assessment.	Male sex, GA <30w, IVH, PDI <90 at 30m, BRS motor quality <26 at 30m.
Dewey (2011)[101] [U]	Canada 1996-2000	5	<1000g	PC study of Infants admitted to a single centre NICU (Calgary).	CP, IQ<70 or visual impairment (n=44).	103 (52%)	Motor development. Total score from MABC.	BPD, later GA, PN steroids.
Leversen (2011)[123 b] [V]	Norway 1999-2000	5	<28w or <1000 g	PC of all live births in Norway.	CP or severe sensory deficit (n=33).	261 (70%)	Motor development. Total score from MABC.	Model 1 (GA <28w): No AN steroids, earlier GA, male sex, ROP >2, SGA. Model 2 (GA ≥28w): male sex, ROP 1-2.
Goyen (2009)[109] [W]	Australia 1992-1995	8	<29w or <1000g	Infants admitted to a tertiary referral centre (Sydney) receiving high-risk referrals from regional centres and enrolled for the routine FUP.	CP, visual/hearing impairment, IQ<85, not attending mainstream school (n=64).	50 (24%)	Motor impairment. Total score <15th percentile from MABC.	PROM, ROP.
Davis (2007)[98] [X]	Australia 1991-1992	8	<29w or <1000 g	PC of all live births in the state of Victoria.	CP or IQ>2SD below mean (n=45).	210 (70%)	Motor impairment. Total score <5th percentile from MABC. Blinded assessment.	Male sex.
Zanudin (2013a)[160] [Y]	Australia 1992-1994	12	<1000 g	Retrospective longitudinal study of infants admitted to a single centre NICU (Brisbane).	CP, neurological impairment, IQ >2SD below mean at 4 years (n not reported)	48 (<50%)	Motor development. Total score from MABC.	BPD, male sex.

[a] Percentage of survivors assessed for outcome measure specified; not all included in the final model.

[b] 2 models for motor skills reported for each gestational age group (<28 weeks and ≥28 weeks). Risk factor considered as significant in Figure 4 if p<0.05 in either model.

Abbreviations: AN antenatal; BPD bronchopulmonary dysplasia; BRS Behaviour Rating Scale from BSID; BSID Bayley Scales of Infant Development;[166] BW birth weight; CP cerebral palsy; FUP follow up; GA gestational age; IQ intelligence quotient; IVH intraventricular haemorrhage; MABC Movement Assessment Battery for Children;[200] NEC necrotising enterocolitis; NICU neonatal intensive care unit; NICHD NRN National Institutes of Child Health and Human Development Neonatal Research Network; NRI neonatal risk index; PC prospective cohort; PDI Psychomotor Developmental Index from BSID; PROM prolonged rupture of membranes; ROP retinopathy of prematurity; SGA small for gestational age; SD standard deviation.

Figure 3.3: Evidence synthesis of risk factors for motor impairment in children born VPT or with VLBW



Note: Risk factor presented if significant ($p<0.05$) in the final model of at least one study and entered into the final model of at least 3 studies (across all ages).
Brain insult/abnormality: IVH any grade [T]; IVH grade 3-4 [M]; IVH (or PVH) and/or PVL [H/O/U/V]; IVH grade 3-4 and PVL separately [B/N/X]; Parenchymal pathology and/or ventriculomegaly [G]; Intracranial haemorrhage, post-haemorrhage ventricular dilation and/or ischaemic cerebral abnormality [Q].
BPD: Oxygen requirement at 36 weeks GA [B/G/H/M/N/Q/T/U/V/Y]; Undefined [O/X].
ROP: Any grade [Q/T/V/W]; Grade 3-4 [U]; Treated [G].

Key

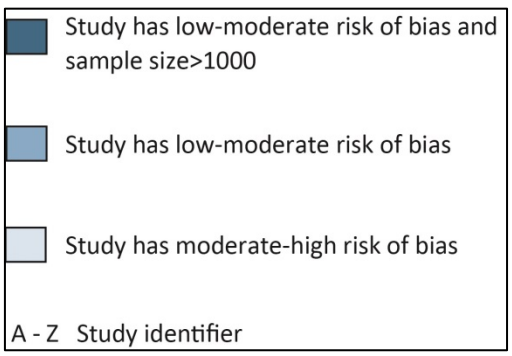


Table 3.5: Summary of studies reporting risk factor analyses for neurological dysfunction not diagnosed as CP in children born VPT or with VLBW

Study reference	Country and recruitment Period	Age of assessment (years)	GA (weeks) and/or BW (grams)	Design and participants	Exclusion criteria for major disability	Number (%) of survivors assessed[a]	Outcome measure	Significant risk factors (p<0.05) in final model
Children of all level of disability								
Age of assessment <5 years								
Wood (2005)[158]	UK and Republic of Ireland 1995	2.5	<26w	PC of all live births in the UK and Republic of Ireland (EPICure Study).	N/A	283 (92%)	Severe motor disability in one or more functional domains: unable to walk, sit, use hands together or control head movements.	Moved hospital <24h, parenchymal pathology/ ventriculomegaly, PN steroids >6w, PH.
Age of assessment ≥5 years								
Schmidhauser (2006)[141 b]	Switzerland 1992-1994	6	<1250g	PC of all live births in a single centre (Zurich).	N/A	88 (87%)	Z-scores from 5 components of Zurich Neuromotor Assessment (NMA)[205]: Pure Motor (PM); Adaptive Fine Motor (AFM); Adaptive Gross Motor (AGM); Static Balance (SB); Associated Movements (AS). Blinded assessment.	PM: Multiple births AFM: PVL any grade. AGM: IVH any grade. SB: No factors significant. AS: No factors significant.
Franz (2009)[107 c]	Germany 1996-1999	4.6-7	<30w and <1500 g	PC study of Infants admitted to a single centre level-3 NICU (Ulm University).	N/A	219 (83%)	Abnormal neurological examination. Mild (minor neurological signs such as broad gait, dysdiadocho-kinesis or dysmetria) to severe (paresis with or without spasticity, CP or ataxia). Blinded assessment.	IVH/PVL ≥3, lower maternal education, ROP ≥3, smaller HC SDS gain to discharge, smaller weight SDS gain to discharge.
Restricted to children free of major disability								
Age of assessment <5 years								
Tommiska (2003)[150]	Finland 1996-1997	1.5	<1000g	PC of all live births in Finland enrolled for the national routine FUP.	CP (n=23)	185 (89%)	Other motor impairment (not CP) defined as delay in motor development, hypotonia, hypertonia with a specific cause, or variable abnormalities in muscle tone	No AN steroids, IVH 2-4, male sex, no maternal infection.
Zanudin (2013b)[159]	Australia 1992-1994	4	<1000g	Retrospective longitudinal study of infants admitted to a single centre NICU (Brisbane).	CP, neurological impairment, IQ >2SD below mean at 4 years (n not reported)	48 (<50%)	Mild-moderate motor impairment (sensory-motor function, posture, balance, gross and fine motor performance) using the Neurosensory Motor Develop-ment Assessment (NSMDA).[206] Blinded assessment.	None significant.
Age of assessment ≥5 years								
Mikkola (2005)[129]	Finland 1996-1997	5	<1000g	PC of all live births in Finland enrolled for the national routine FUP.	CP (n=28)	165 (80%)	Simple or complex minor neurological dysfunction as defined by Hadders-Algra.[198] Assessed using a modified partial Touwen test.[197]	Hospital area, male sex.
Arnaud (2007)[90 d]	France 1997	5	<33w	PC of all live births in 9 French regions comprising one third of all births (EPIPAGE Study).	CP, IQ<50 or severe bilateral sensory impairment (n=150).	1817 (80%)	Mild and moderate minor neuromotor dysfunction (posture and muscle tone, reflexes, coordination and balance, motor behaviour of face and eyes). Assessed by Touwen examination (short form).[197]	Acute fetal distress, IVH 1-3/ white matter abnormalities/ IPH, multiple pregnancy, PN steroids.

[a] Percentage of survivors assessed for outcome; not all included in the final model.

[b] 5 models reported for each domain of the Zurich Neuromotor assessment. Risk factors reported separately for each model in the table.

[c] 2 models reported. Same perinatal factors fitted with change in weight variables added to 1st model and change in head circumference variables added to 2nd model. Perinatal factors reported as significant if p<0.05 in both models.

[d] 2 models for neuromotor dysfunction reported; moderate vs. no minor neuromotor dysfunction and mild vs. no minor neuromotor dysfunction. Risk factors reported as significant if p<0.05 in either model.

Abbreviations: AN antenatal; CNS central nervous system; CP cerebral palsy; FUP follow up; GA gestational age; IPH intraparenchymal haemorrhage; IVH intraventricular haemorrhage; HC head circumference; N/A not applicable; NICU neonatal intensive care unit; PC prospective cohort; PH pulmonary haemorrhage; PN postnatal; PVL periventricular leukomalacia; ROP retinopathy of prematurity; SD standard deviation; SDS standard deviation score.

3.3 Data synthesis of prognostic factors for visual impairment

One low-moderate[114] and two moderate-high[104,113] risk of bias studies reported risk factors for visual impairment. The risk of bias assessment is shown in Figure 3.4.

Figure 3.4: Risk of bias assessment of the 3 studies reporting visual outcomes

	Risk of bias domain						
	Study Participation	Study Attrition	Prognostic Factor Measurement	Outcome Measurement	Study Confounding	Analysis and Reporting	Low-moderate risk of bias
Farooqi (2008)							
Hok-Wikstrand (2010)							
Holmstrom (2014)							✓

Risk of bias:  Low  Moderate  High

The study at low-moderate risk of bias by Holstrom et al[114] was based on a prospective cohort of all live births <27 weeks GA in Sweden between 2004 and 2007. At 30 months, 411 surviving children underwent an ophthalmic examination and the outcomes used in the study were manifest strabismus (defined as esotropia, exotropia or hypertropia/hypotropia) and a composite score of eye and visual problems (defined as visual impairment of the worse eye, strabismus, or refractive errors (myopia <-3D, hypermetropia >+3D, astigmatism >=2D and/or anisometropia). For both outcomes, nine candidate variables (sex, GA, BW, small for gestational age (SGA), BW standard deviation score (SDS), weight SDS at 36 weeks, ROP, CP and cognitive disability) were screened in a univariable analysis then entered into a

multivariable model and eliminated using backward selection. Three variables were significant in both models for strabismus and the composite score; severe treated ROP stage 3-5, CP and cognitive disability, in addition BW SDS was significantly associated with manifest strabismus and BW with the composite score.

The study at moderate-high risk of bias by Farooqi et al[104] was based on a prospective cohort of all live births <26 weeks GA and with <1000g BW in Sweden between 1990 and 1992. The primary caregivers of 88 surviving children at age 11 years, were asked the question, "Is (name) blind, nearly blind or does he/she have difficulty seeing?" and, "Do regular eye glasses or contact lenses completely correct the problem?" If the response to the first question was affirmative and not corrected by glasses, the child was classified as having reduced vision. Six factors were entered into the model: BPD, severe ROP (≥Stage 3), brain injury (IVH grade 3-4 or cystic PVL), sex, GA and social risk. Only severe ROP and brain injury were statistically significant, but were not significant when children free of neurosensory impairment were removed from the analysis.

Lastly, the study at moderate-high risk of bias by Hok-Wikstrand et al[113] was also carried out in Sweden, and based on 66 infants born <32 weeks GA in the years 1999 to 2000 in a single hospital (Sahlgrenska University). The children were assessed at age 4-6 years for monocular visual acuity with best correction using the KM visual acuity chart (7 children with developmental delay were given less demanding tests). Visual impairment was defined as visual acuity ≤0.3 (logMAR 0.5). Twenty peri- and postnatal variables were examined in a univariable analysis, but it was not clear which factors were selected for entry into a stepwise regression model. The only factor that remained significant in the final model was head circumference SDS. The factors that were significant in at least one of the other two vision

studies reported in this systematic review; ROP, BW SDS, brain injury and CP were included in the univariable analysis (ROP/BW SDS were not significant and brain injury/CP were), but it is not known if these were tested in the multivariable model.

3.4 Discussion

3.4.1 Neonatal brain injury and CP

There was strong evidence that IVH grade 3-4 alone or in combination with PVL was predictive of CP. The importance of IVH and PVL as predictors of CP has been long established, with large well-conducted studies that have examined these factors in isolation reporting strong linear trends with grade of IVH and PVL,[207-210] although the evidence for the influence of lower grade 1-2 IVH remains mixed.[211-213] The two studies that did not find brain injury significant were based on the same cohort population and used routine data gathered from the national healthcare system, which may have been less standardised and more liable to bias and misclassification.[129,150] PVL alone was found to be significant in the two largest[93,155] of the three studies that tested this as a separate factor to IVH. The non-significant result in the smaller third study[89] is likely to be explained by a lack of power with fewer cases of PVL included in the analysis. Since severe PVL is highly predictive, but also quite rare, it would seem sensible to combine it with IVH for the purposes of prognostic modelling, to avoid it being dropped from models based on small samples with few cases.

3.4.2 Postnatal steroids and CP

In studies where postnatal steroids were commonly prescribed there was moderate evidence of an association with CP, whereas in two studies where the prevalence of treatment with postnatal steroids in the cohort population was much lower[93,151] the findings were non-

significant. Several Cochrane reviews of RCTs have been performed on the use of postnatal corticosteroids in the preterm population. Early use (<8 days) was found to increase the incidence of CP (Relative risk (RR) 1.45, 95% CI 1.06 to 1.98),[214] though late use (>7 days) was found to be non-significant (RR 1.22, 95% CI 0.84 to 1.77).[215] Studies that have focused solely on the relationship between postnatal steroids and increased risk of CP have also found a positive association.[216,217] It is unclear whether postnatal steroids themselves are the causal factor, or whether the reasons underlying the use of these drugs (typically in sicker, more immature infants) are responsible for the increased risk. Since the use of postnatal steroids has declined greatly since the 1990s, its utility as a prognostic factor is probably much less relevant in current preterm populations.

3.4.3 Antenatal steroids and CP

There was moderate evidence that the use of antenatal steroids was a protective factor for CP. The one study that found it not significant[151] was based on children born around a decade later than the other studies. A Cochrane review of RCTs on the use of antenatal corticosteroids among women at risk of preterm birth concluded that it was a non-significant protective factor for developing CP (RR 0.60, 95% CI 0.34 to 1.03).[218] In four observational studies examining the effect of antenatal steroids on CP exclusively, two found a significant association[219,220] and two no significant association,[221,222] though the RR was <1 in the latter two studies, suggesting a non-significant protective effect. Recent evidence has emerged on the neuroprotective effect of magnesium sulphate and the subsequent reduction in the incidence of CP, when prescribed as a tocolytic or to prevent preeclamptic convulsions,[223,224] however the studies included in this review predate this finding.

3.4.4 Gestational age and CP

The inverse relationship between the prevalence of CP and GA is well established,[184,225,226] hence the lack of association in this review was surprising. One explanation is the restriction to cohorts <33 weeks; the prevalence of CP declines quite steeply after 32 weeks, hence discrimination below this threshold may become less apparent.[226] Another explanation is that GA is associated with other critical risk factors so its power as an independent predictive factor may be reduced due to multicollinearity with factors that are more strongly causally related to CP, such as neonatal brain injury.

3.4.5 Participant sex and motor impairment

Male sex was found to be of limited use as a prognostic factor for CP, but there was fairly strong evidence that it was predictive of later motor impairment in children free of major disability. Other studies that have focussed solely on the association of infant sex with motor function in VPT children have reported mixed findings,[227-230] however these were not restricted to children free of major disability. One small imaging study found that at 12 years, VPT males without evidence of IVH or PVL on neonatal ultrasound had widespread reductions in regional white and grey matter volumes compared to VPT females, who did not differ from term born controls.[231] This implies that VPT males continue to demonstrate abnormal brain development into childhood, in the absence of any brain injury in the neonatal period.

3.4.6 Other risk factors and CP or motor impairment

There were mixed or inconclusive findings for the other risk factors included in the data synthesis, which may be due to heterogeneity among the study cohorts which span a 24 year period and represent diverse international populations with differing methods of ascertainment, inclusion criteria, clinical practices and environmental influences. Also studies

did not all consider the same set of candidate factors in the outset, so some may have been significant had they been tested.

3.4.7 Outcome measures

There was some uniformity in the outcome measures used across studies, as most studies used the PDI from the BSID-II to assess motor skills in infancy, and the MABC to assess motor impairment at school age. However, there were still variations on reporting and analysis with the use of continuous scores versus dichotomous thresholds, and the methods of dealing with missing scores (imputation or exclusion). The assessment of neurological dysfunction was less standardised, which made it difficult to combine findings.

3.4.8 Conclusion

In the VPT/VLBW population there was strong evidence that IVH and PVL are robust prognostic factors for CP, and some evidence that the use of postnatal steroids increases the risk and the use of antenatal steroids reduces the risk of CP. There was moderate evidence that male sex is prognostic for DCD and other motor impairments at school age in children free of major disability. Gestational age was found to be of limited use as prognostic factor for CP, probably due to a reduced discriminatory power in cohorts restricted to earlier GAs and confounding with other important clinical events which are more strongly causally related.

There was a lack of large, well-conducted studies examining risk factors for visual and hearing impairment and further research is required in these areas. The three Swedish studies included in the review of visual outcomes were of variable quality and used definitions of visual impairment that were not directly comparable. There was some suggestion from these

studies that severe ROP, brain injury, CP and measures of growth may be possible predictors of later visual dysfunction, but further evidence is necessary to corroborate these findings.

Chapter 4: Prognostic factors for cognitive impairment in children born VPT or with VLBW: systematic review results

4.1 Introduction

While CP and neurosensory disorders such as deafness and blindness can have a severe impact on development, cognitive impairment is the most common neurological outcome in infants born VPT or with VLBW and cognitive delay has been reported to be as high as 40% at school age among EPT children born <28 weeks of gestation.[65,188,232] Poor cognitive test scores at school age are strongly related to low GA at birth, even children without severe disability have consistently been found to have lower scores than their term-born peers.[40]

In addition to being at increased risk of global cognitive impairment, VPT/VLBW children are more likely to perform less well in tests of attention and executive function compared to their full-term peers,[233] even after adjusting for IQ.[45,234,235] They also have a higher rate of language problems, in both the expressive and receptive domain, that persist into middle childhood.[236] Problems with cognitive and language development mean that many VPT/VLBW survivors are at high risk of poor academic attainment and reduced lifelong earning potential and life chances. A significant proportion require full time specialist education, and the majority of those in mainstream education require specialist academic, health or behavioural support services to aid their progress through school.[44]

There is likely to be a complex relationship between cognitive function, biological and environmental factors, and clinical events both during and after the perinatal period for a VPT birth. This relationship may change as the child grows up, as social factors potentially supersede the influence of early biological factors, and therefore prognosis is likely to be a dynamic process. Factors such as parent-infant interactions, maternal distress, hospitalisation, parental socio-economic status and education, as well as access to resources may all play an important role in modifying brain development.[30] To help promote optimal development, the contribution of all these factors to risk needs to be disentangled. The focus of this chapter is to identify risk factors that are robust predictors of impaired cognitive function, including language skills, executive function and academic attainment as well as global IQ.

Among the 78 articles that were identified in the review and included in the data extraction, 31 contained risk factor analyses for cognitive outcomes.[46,75,88,89,92,96,97,106,107,111,112,115,121,123-125,128,129,132,135,136,139,144,147-151,155,156,158] These studies were based on 21 cohort populations; five other articles based on cognitive outcomes were excluded due to cohort overlap.[73,79,81-83] In this chapter, the results from the 31 articles reporting the risk of cognitive impairment are summarised and synthesised where possible.

4.2 Data synthesis of prognostic factors for cognitive impairment

4.2.1 Study characteristics

The main study design was prospective cohort (n=27); there was also one cross-sectional study[106] and three randomised controlled trial populations.[124,125,128] Of the 27 prospective cohorts, 12 were ascertained from all live births in a geographically defined

region[46,92,96,97,111,123,129,144,147,150,151,158] and 10 were recruited from a single centre NICU.[75,107,115,132,135,136,139,148,149,156] Studies were conducted in 12 countries: US (n=9), UK (n=4), Netherlands (n=4), Germany (n=3), Australia (n=2), Finland (n=2), France (n=2) and one study each from Austria, Denmark, Estonia, Italy and Norway. The median sample size was 219 [range: 45 to 3785] and three studies had more than 1000 participants.[88,92,155] Four studies were restricted to EPT children[46,112,144,158] and three studies excluded multiple births.[75,89,124] The risk of bias assessment classified 14 (45%) studies as having low-moderate risk of bias and 17 (55%) studies as having moderate-high risk of bias (Figure 4.1).

From the 31 studies identified, 98 risk factor models for cognitive outcome were included in the review. Many studies presented models for more than one type of cognitive outcome domain and some studies reported multiple subtests within a domain. The models were divided into the following groups for data synthesis: global cognitive function (measured by a general cognitive test or IQ score), executive function, academic attainment and language skills. The median number of candidate risk factors considered at the outset was 11 [range: 3 to 45]. In 20 (20%) models, all candidates were included in the final model with no screening or selection process, and in 33 (34%) models a forwards, backwards or stepwise selection process was applied. In 39 models (belonging to the same 3 studies)[132,148,149] each candidate risk factor was modelled individually, adjusting for several factors identified a priori. Only 14 (14%) of the 98 models reported the number of participants included in the final model presented and no studies performed any type of statistical or clinical validation.

Figure 4.1: Risk of bias assessment of the 31 studies reporting cognitive outcomes

	Risk of bias domain						
	Study Participation	Study Attrition	Prognostic Factor Measurement	Outcome Measurement	Study Confounding	Statistical Analysis and Reporting	Low-moderate risk of bias
Aarnoudse-Moens (2013)	High	High	Moderate	Moderate	Low	Moderate	
Adams-Chapman (2013)	Moderate	Low	Moderate	Low	Low	Moderate	✓
Andrews (2008)	High	Moderate	Moderate	Moderate	Low	Moderate	
Beaino (2011)	Low	Moderate	Moderate	Low	Low	Low	✓
Charkaluk (2010)	Low	Moderate	Low	Moderate	Moderate	Low	✓
Cooke (2005)	Low	Moderate	Moderate	Moderate	Moderate	Moderate	✓
Ford (2011)	High	Moderate	High	Moderate	Moderate	High	
Franz (2009)	High	Moderate	Low	Low	Low	Moderate	
Hansen (2004)	Low	Low	Low	Low	Moderate	Moderate	✓
Helderman (2012)	Low	Moderate	Moderate	Low	Moderate	Moderate	✓
Howard (2011)	High	Low	Moderate	Low	Low	Low	
Johnson (2011)	Low	Moderate	Low	Low	Moderate	Moderate	✓
Kiechl-Kohlendorfer (2013)	Moderate	Moderate	Low	Moderate	Low	Low	✓
Leversen (2011)	Low	Moderate	Moderate	Moderate	Low	Low	✓
Lowe (2009)	High	Moderate	Moderate	High	Low	Moderate	
Marston (2007)	High	High	Moderate	High	Moderate	Low	
Messinger (2010)	High	High	Moderate	Moderate	Moderate	Moderate	
Mikkola (2005)	Low	Low	High	Moderate	Low	Low	
Orchinik (2011)	High	Low	Moderate	Low	Moderate	High	
Potharst (2012)	High	Moderate	Moderate	Moderate	Moderate	Moderate	
Potharst (2013)	High	Moderate	High	Moderate	Moderate	Moderate	
Sansavini (2011)	High	High	Moderate	Moderate	Moderate	Low	
Stahlmann (2009)	Moderate	Low	Moderate	Low	Moderate	High	
Stoelhorst (2003)	Low	Moderate	Moderate	Moderate	Moderate	Moderate	✓
Taylor (2006)	High	Low	Moderate	Low	Moderate	High	
Taylor (2011)	High	Low	High	Low	Moderate	High	
Tommiska (2003)	Moderate	Low	Moderate	Moderate	Moderate	Moderate	✓
Toome (2013)	Moderate	Low	Moderate	Moderate	Low	Moderate	✓
Vohr (2005)	Low	Moderate	Moderate	Low	Low	Moderate	✓
Voss (2012)	High	Low	Moderate	Moderate	Moderate	High	
Wood (2005)	Low	Low	Moderate	Moderate	Low	Moderate	✓

Risk of bias: Low Moderate High

4.2.2 Prognostic factors for global cognitive impairment

Twenty studies contained a risk factor analysis for general cognitive function or IQ; eight studies assessed outcome between 1.5-2.5 years (Table 4.1)[88,112,128,147,150,151,155,158] and 12 studies between 5-13 years (Table 4.2).[89,92,97,107,111,123,129,132,135,144,149,156] The most common assessment used before five years was the Mental Development Index (MDI) from the BSID-II,[166] or the Cognitive Composite Score from version III[167] in the more recent studies. The BSID-II MDI assesses cognition through evaluation of sensory-perception, knowledge, memory, problem solving, and early language. The more recent version splits cognitive and language skills into separate domains. There was more variety in measurement scales used in assessments above 5 years, the most common being the Mental Processing Composite (MPC) score from the Kaufman Assessment Battery for Children (KABC)[237] and the full scale IQ from Wechsler's Preschool and Primary Scale of Intelligence-Revised (WPPSI-R).[238] Risk factors that were found to be significant in at least one study at low-moderate risk of bias and examined in the final model of at least two other (independent) studies are presented in Figure 4.2 (<5 years) and Figure 4.3 (≥5 years).

In studies where the age of assessment was <5 years (Figure 4.2), the two largest studies at low-moderate risk of bias[88,155] and at least one other study at low-moderate risk of bias found the following factors to be predictive of poorer cognitive development: male sex, non-white ethnicity, lower parental education, lower BW and brain injury diagnosed during the neonatal period. However, the other studies that also examined these risk factors sometimes contradicted these findings,[128,147,150,151] with the exception of ethnicity. There was also some evidence that the failure to give antenatal steroids and birth at an earlier GA were not predictive of poorer cognitive function <5 years of age.

Table 4.1: Summary of studies reporting risk factor analyses for global cognitive impairment in children born VPT or with VLBW at asessed <5 years of age

Study reference [identifier]	Country and recruitment period	Age of assessment (years)	GA (weeks)/ BW (grams)	Design and participants	Number (%) of survivors assessed[a]	Outcome measure (continuous (cts) unless otherwise specified)	Method for dealing with untestable children	Significant risk factors for poorer outcome (p<0.05) in final model
Tommiska (2003)[150] [A]	Finland 1996-1997	1.5	<1000g	PC of infants born and treated in a single centre NICU (Helsinki) and enrolled for the national routine FUP.	78 (94%)	MDI score from BSID-II.	Excluded if severe developmental problem (n=3) or exhaustion (n=2).	Number of days from Jan 1st 1996 to birth.
Vohr (2005)[155] [B]	US 1993-1998	1.5-1.8	<33w and <1000g	PC of infants admitted to the NICU of 12 centres participating in the multi-centre NICHD NRN routine FUP.	3785 (80%)	MDI score from BSID-II (<70 vs. ≥70). Blinded assessment.	Excluded if test not completed (n=118).	Birth epoch (93-94 vs. 95-96) , lower BW, BPD, any high frequency ventilation, IVH 3-4, male sex, lower maternal education, no private insurance, multiple pregnancy, non-white ethnicity, outborn, PVL, PN steroids.
Adams-Chapman (2013)[88] [C]	US 2006-2008	1.5-1.8	<1000g	PC of infants admitted to the NICU of 20 centres participating in the multi-centre NICHD NRN routine FUP.	1477 (91%)	Cognitive composite score from BSID-III. Blinded assessment.	Assigned a score of 54 if severely delayed (n=39).	Lower BW, black ethnicity, dysfunctional feeding, GMFCS ≥2, non-English speaking, male sex, lower maternal education, MV days, multiple pregnancy, NEC 2-3, no private insurance, IVH 3-4/ PVL.
Stoelhorst (2003)[147 b] [D]	Netherlands 1996-1997	2	<32w	PC of all live births in 3 Dutch health regions comprising 9% of the population.	146 (62%)	MDI score from BSID-I.	Assigned a score of 50 if severely disabled (n=3), otherwise excluded (n=5).	Male sex, younger maternal age, non-Dutch, PN steroids, SGA.
Helderman (2012)[112] [E]	US 2002-2004	2	<28w	PC of all live births in 14 centres in 5 states (ELGAN Study).	921 (77%)	MDI score from BSID-II (<55 and 55-69 vs. ≥70). Blinded assessment.	Excluded if GMFCS ≥1 (n=83).	BW <-2SD, BMI>30, male sex, lower maternal education, non-white ethnicity.
Toome (2013)[151] [F]	Estonia 2007	2	<32w	PC of all live births in Estonia enrolled in the national neonatal research routine FUP.	155 (99%)	Cognitive composite score from BSID-III (<70 vs. ≥70).	Assigned a score of -4SD below the mean (n not reported).	IVH 3-4/PVL 2-4, NEC 2-3.
Wood (2005)[158] [G]	UK and Republic of Ireland 1995	2.5	<26w	PC of all live births in the UK and Republic of Ireland (EPICure Study).	196 (64%)	MDI score from BSID-II.	Excluded if MDI<55 or functional motor disability (n=52).	Afro-Caribbean ethnicity, no AN steroids, BPD, male sex, lower maternal education, PROM.
Messinger (2010)[128 c] [H]	US 1999-2001	2.5	<1000g	Infants enrolled in NICHD NRN routine FUP (12 centres) and recruited to a glutamine supplementation RCT.	539 (47%)	MDI score from BSID-II.	Excluded if test uncompleted (n not reported).	BW ≤750g, higher maternal income, higher MDI at 18m.

[a] Percentage of survivors assessed for outcome measure specified.

[b] 2 models for motor skills reported. Model based on 2 year outcome included. Model based on 1.5 year outcome not included.

[c] Several models for cognitive function fitted. Full model adjusting for 18 month MDI and BRS total score included.

Abbreviations: AN antenatal; BMI body mass index; BPD bronchopulmonary dysplasia; BSID Bayley Scales of Infant Development[166]; BW birth weight; GA gestational age; GMFCS Gross Motor Functional Classification System[199]; IVH intra-ventricular haemorrhage; MDI Mental Developmental Index from the BSID; MV mechanical ventilation; NEC necrotising enterocolitis; NICU neonatal intensive care unit; NICHD NRN National Institutes of Child Health and Human Development Neonatal Research Network; PN post natal; PC prospective cohort; PROM prolonged rupture of membranes; PVL periventricular leukomalacia; RCT randomised controlled trial; SGA small for gestational age; SD standard deviation.

Table 4.2: Summary of studies reporting risk factor analyses for global cognitive impairment in children born VPT or with VLBW at assessed ≥5 years of age

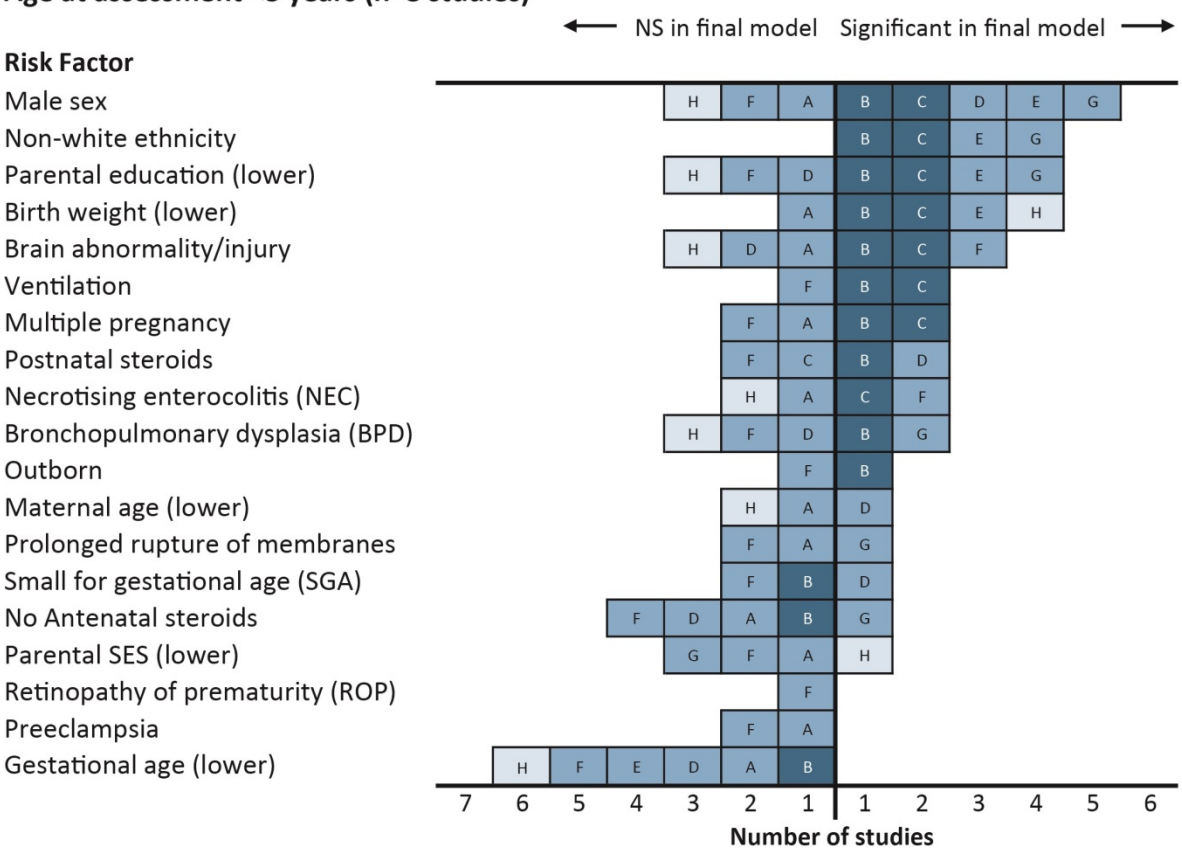
Study reference [identifier]	Country and recruitment period	Age of assessment (years)	GA (weeks)/ BW (grams)	Design and participants	Number (%) of survivors assessed[a]	Outcome measure (continuous (cts) unless otherwise specified)	Method for dealing with untestable children	Significant risk factors for poorer outcome (p<0.05) in final model
Franz (2009)[107] [b] [I]	Germany 1996-1999	4.6-7	<30w and <1500 g	PC study of Infants admitted to a single centre level-3 NICU (Ulm University).	219 (83%)	MPC from KABC. Blinded assessment.	Assigned a core of 30 if minimal speech and a score of 20 if minimal sensory or motor achievements elicited.	Lower BW SDS, smaller HC SDS gain (discharge to 5 years), IVH/ PVH ≥3, lower maternal education, MV days, PVL.
Hansen (2004)[111] [c] [J]	Denmark 1994-1995	5	<28w or <1000g	PC of all live births in Denmark ascertained from all 18 neonatal care units and the Danish Medical Birth Register (ETFOL Study).	247 (94%)	FIQ from WPPSI-R (cts score and <70 vs. ≥70). Blinded assessement.	Excluded if test not completed (n=5). Children with CP, visual disability, 1st language not Dutch or >27w GA excluded from analysis of continuous score (n=110).	**Model 1 (cts score):** BPD, lower parental education **Model 2 (<70 vs. ≥70):** BPD, IVH 3-4/PVL.
Mikkola (2005)[129] [c] [K]	Finland 1996-1997	5	<1000g	PC of all live births in Finland enrolled for the national routine FUP.	172 (83%)	FIQ from WPPSI-R (cts score and <70 vs. ≥70).	Excluded if cognitively impaired and unable to cooperate (n=9). Excluded if test not completed (n=12).	**Model 1 (cts score):** No AN steroids, IVH 3-4, male sex, multiparity, multiple pregnancy, lower parental SES, vaginal delivery. **Model 2 (<70 vs. ≥70):** No AN steroids, BPD, hospital area, perforated NEC.
Leversen (2011)[123] [d] [L]G	Norway 1999-2000	5	<28w or <1000 g	PC of all live births in Norway.	248 (67%)	FIQ from WPPSI-R.	Excluded if CP, blind, deaf or autistic (n=33). Excluded if test not completed (n=25).	**Model 1 (GA <28w):** Lower maternal education, preeclampsia, ROP>2. **Model 2 (GA ≥28w):** Male sex.
Potharst (2012)[135] [e] [M]	Netherlands 2002-2004	5	<30w or <1000g	PC study of Infants admitted to a single centre NICU (Amsterdam).	102 (68%)	FIQ from WPPSI-III	Excluded if too disabled to be tested (n=4).	Behaviour problems at 2 yrs, lower MDI at 2yrs, lower parental education, parental foreign country of birth, sepsis or meningitis.
Beaino (2011)[92] [N]	France 1997	5	<33w	PC of all live births in 9 French regions comprising one third of all births (EPIPAGE Study).	1503 (62%)	MPC from KABC (<70 and 70-84 vs. ≥85). Blinded assessment.	Excluded if moderate to severe neuro-sensory disability (n=70). Excluded if test not completed (n=239).	Breast fed at discharge, cystic PVL or IPH, GA ≤28w, IVH grade 3/echo-densities/VD, lower parental SES, SGA, ≥3 siblings.
Orchinik (2011)[132] [f] [O]	US 2001-2003	5	<28w or <1000g	PC of Infants admitted to a single centre NICU (Ohio) participating in the multicentre NICHD NRN routine FUP.	142 (72%)	WJ-III COG Brief Intellectual Ability <10th centile. Blinded assessment.	Assigned a score of 40 if too low functioning to comply with test demands.	GA <25w, infection (sepsis, NEC or meningitis), IVH 3-4/PVL/ VD, neurosensory disorder and/or MDI<70 at 20m.

Table 4.2: Summary of studies reporting risk factor analyses for global cognitive impairment in children born VPT or with VLBW asessed at ≥5 years of age (continued)

Study reference [identifier]	Country and recruitment period	Age of assessment (years)	GA (weeks)/ BW (grams)	Design and participants	Number (%) of survivors assessed[a]	Outcome measure (continuous (cts) unless otherwise specified)	Method for dealing with untestable children	Significant risk factors for poorer outcome (p<0.05) in final model
Andrews (2008)[89 c] [P]	US 1996-1999	5-8	<32w	PC of infants admitted to a single centre NICU (Alabama) participating in the multicentre NICHD NRN routine FUP. Multiple births excluded.	259 (69%)	WISC-IV or DAS if <6 years or unable to complete the WISC-IV (cts score and <70 vs. ≥70).	Excluded if test not completed (n=2).	**Model 1 (cts score):** Earlier GA, PVL. **Model 2 (<70 vs. ≥70):** No history of PROM, earlier GA.
Cooke (2005)[97] [Q]	England 1991-1992	7	<32w	PC of all live births in all 8 hospitals in the Liverpool postal district.	280 (77%)	WISC-III (<89 vs. ≥89; mean of the group).	Excluded if not free of major disability and not attending mainstream school (n=29).	Earlier GA, smaller HC at 7yrs, PDA.
Taylor (2006)[149 c g] [R]	US 1992-1995	8	<1000g	PC of infants admitted to a single centre NICU (Ohio) participating in the multicentre NICHD NRN routine FUP.	204 (86%)	MPC from KABC (cts and <1SD below mean of control group). Blinded assessment.	Excluded if untestable due to severe developmental impairments (n=10).	**Model 1 (cts score):** Longer neonatal hospital stay, NEC, NRI>3, PVL, VD. **Model 2 (<70 vs. ≥70):** Longer neonatal hospital stay, NRI>3, PVL, VD.
Stahlman (2009)[144] [S]	Germany 1997-1999	7-9	<27w	PC of all live births in all 8 perinatal centres in Schleswig-Holstein.	75 (82%)	MPC from KABC or equivalent[h] (<70 vs. ≥70). Blinded assessment.	Assigned a score of <70 if untestable due to extremely limited capacities.	IVH 3-4/PVL.
Voss (2012)[156] [T]	Germany 1993-1998	10-13	<1000g	PC study of Infants admitted to a single centre NICU (Hannover).	148 (87%)	HAWIK-III composite IQ score.	Assigned a score of 39 (40 is the lowest possible score).	HC increase <6mm/w, IVH 3-4/PVL, immigrant status, parenteral nutrition >41 days.

[a] Percentage of survivors assessed for outcome measure specified.
[b] 2 models for cognitive function reported. Same perinatal factors fitted with change in weight variables added to 1st model and change in head circumference variables added to 2nd model. Perinatal factors included in Figure 3b as significant if p<0.05 in both models and non-significant if p≥0.05 in both models, else not included.
[c] 2 models for cognitive function reported; one based on dichotomous outcome and one based on continuous outcome. Risk factors included in Figure 3b as significant if p<0.05 in both models and non-significant if p≥0.05 in both models, else not included.
[d] 2 models for cognitive function reported for each gestational age group (<28 weeks and ≥28 weeks). Risk factor considered as significant if p<0.05 in either model.
[e] 2 models for FIQ at 5 years reported; one including 2 year developmental assessments and one including 3 year developmental assessments. The former model is reported as 2 year assessments are more routine in general practice.
[f] Each risk factor was fitted separately and adjusted for sex, ethnicity, parental SES and months in school at testing (the article did not report results for the adjustment factors).
[g] Each risk factor was fitted separately and adjusted for sex, ethnicity, parental SES, family stressors and family resources (the article did not report results for the adjustment factors).
[h] 7 children were tested with an equivalent instrument; HAWIK, Snijders-Oomen Nonverbal Intelligence Test (SON-R) or the Culture Fair Intelligence Tests (CFT).
Abbreviations: AN antenatal; BPD bronchopulmonary dysplasia; BSID Bayley Scales of Infant Development;[166] BW birth weight; CP cerebral palsy; DAS Differential Ability Scales;[239] GA gestational age; FIQ Full-scale IQ from WPPSI-R; HC head circumference; IQ intelligence quotient; HAWIK Hamburg Wechsler Intelligence Test for Children;[240] IPH Intraparenchymal haemorrhage; IVH intraventricular haemorrhage; KABC Kaufman Assessment Battery for Children;[237] MDI Mental Developmental Index from the BSID; MPC Mental Processing Composite from the KABC; MV mechanical ventilation; NEC necrotising enterocolitis; NICU neonatal intensive care unit; NRI Neonatal Risk Index; NICHD NRN National Institutes of Child Health and Human Development Neonatal Research Network; PC prospective cohort; PDA patent ductus arteriosus; PROM prolonged rupture of membranes; PVH periventricular haemorrhage; PVL periventricular leukomalacia; ROP retinopathy of prematurity; SGA small for gestational age; SD standard deviation; SDS standard deviation score; SES socio-economic status; VD ventricular dilatation; WISC Wechsler Intelligence Scale for Children;[241] WJ-III COG Woodcock Johnson Tests of Cognitive Abilities, 3rd edition;[242] WPPSI-R Wechsler's Preschool and Primary Scale of Intelligence-Revised.[238]

Figure 4.2: Evidence synthesis of risk factors for global cognitive impairment in children born VPT or with VLBW assessed at <5 years of age



Note: Risk factor presented if significant (p<0.05) in the final model of at least one study at low-moderate risk of bias and entered into the final model of at least 3 studies (across all ages).
Brain abnormality/injury: IVH and/or PVL [B/C/D/F/H]; IVH grade 2-4 [A].
BPD: Oxygen requirement at 36 weeks GA [B/D/F/G]; Not specified [H].
NEC: Perforated NEC [A]; NEC stage 2-3 [C/F]; Not specified [H].
PROM: >24 hours before labour [G]; Not specified [A/F].
Ethnicity: Non-white [B/E]; Black [C]; Afro-Caribbean [G].
ROP: ROP stage ≥3 with laser therapy [F]
Ventilation: Any High-frequency [B]; Mechanical ventilation days [C/F].

Key

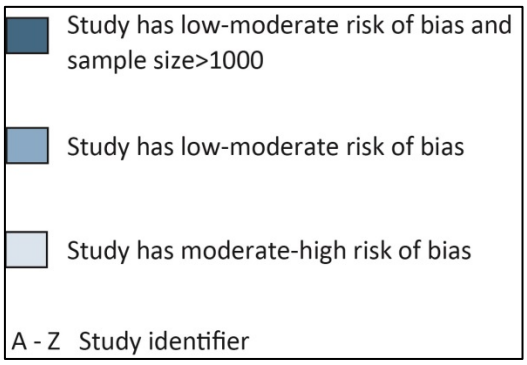
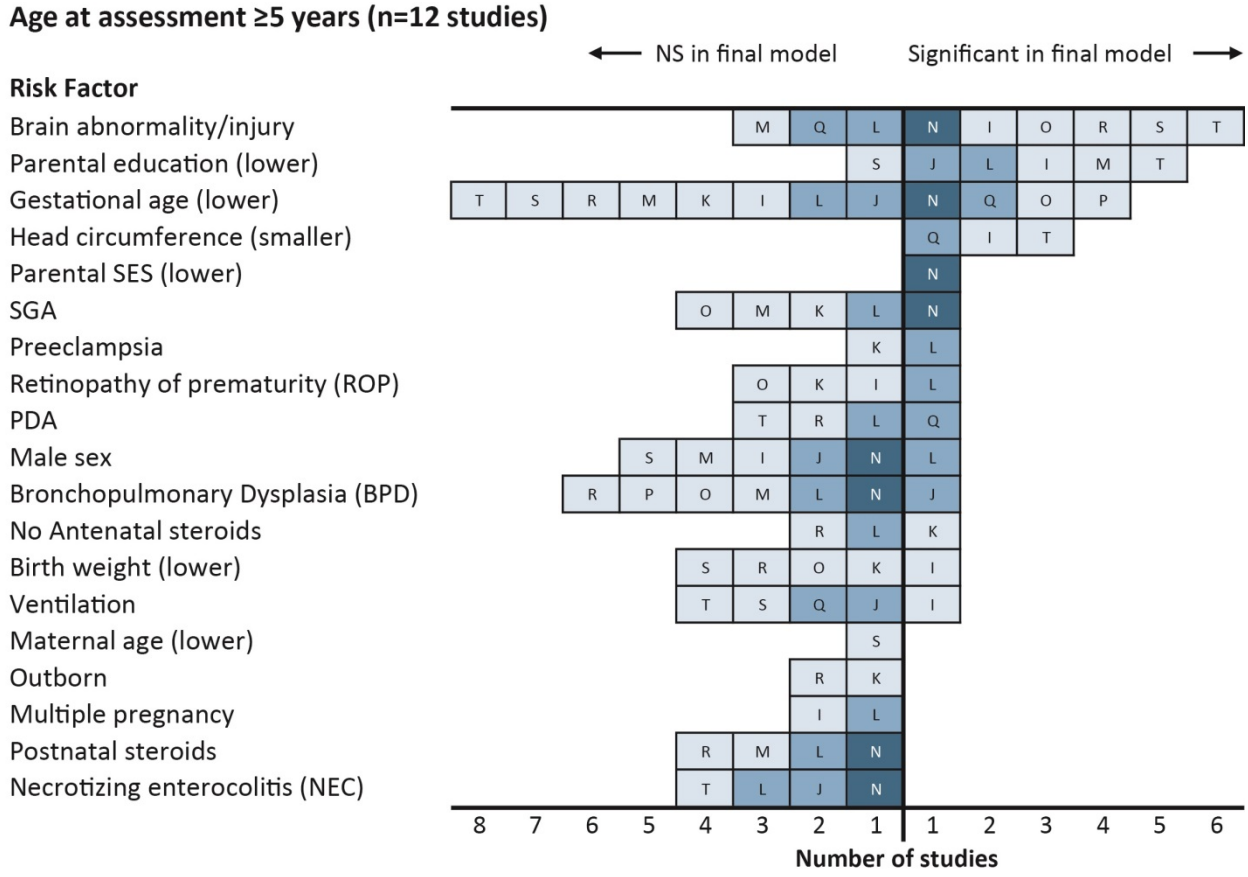
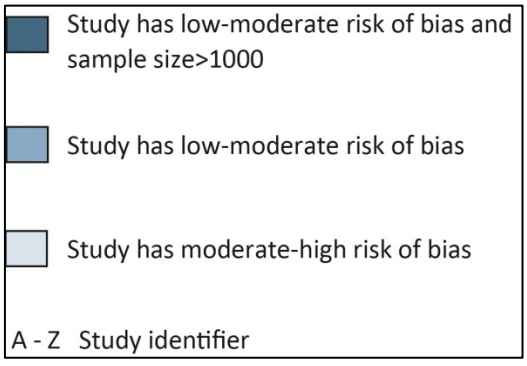


Figure 4.3: Evidence synthesis of risk factors for global cognitive impairment in children born VPT or with VLBW assessed at ≥5 years of age



Note: Risk factor presented if significant (p<0.05) in the final model of at least one study at low-moderate risk of bias and entered into the final model of at least 3 studies (across all ages).
Brain abnormality/injury: IVH and/or PVL [I/L/M/O/S/T]; PVL or VD [R]; Parenchymal lesion [Q];
IVH grade 1-3/echodensities/ventricular dilatation (VD)/cystic PVL/intraparenchymal haemorrhage [N].
BPD: Oxygen requirement at 36 weeks GA [J/L/M/N/O/R]; Not specified [P].
Head circumference (HC): Increase in HC from discharge to 5 years [I]; Occipitofrontal circumference 7yr centile [Q]; Increase in HC <6mm/wk [T].
NEC: Surgical or x-ray diagnosed [J]; Bowel perforation or NEC [T]; Not specified [L/N].
ROP: ROP stage 3-4 [I/K/L]; ROP stage 4-5 or treatment with cryotherapy/laser therapy [O].
Ventilation: Any mechanical ventilation [J]; Mechanical ventilation days [I/Q/S/T].

Key



Most of the studies examining cognitive function ≥5 years were at moderate-high risk of bias (Figure 4.3). The association of parental educational level and cognitive impairment was also evident in this age group, but the association with male sex was greatly diminished. Ethnicity was not entered into the final model in any of the studies in older children, or not reported when it was used as an adjustment factor,[132,149] so it was not possible to determine whether the influence of this factor prevailed into middle childhood. Most studies in this age group also found that birth at earlier GA had little prognostic value in a multivariable prediction model.

4.2.3 Prognostic factors for language impairment

Risk factor analyses for language development was conducted in eight studies; five studies assessed outcome between 1.5-3 years (Table 4.3)[88,96,125,139,151] and three studies assessed outcome between 5-8 years (Table 4.4), but were at moderate-high risk of bias.[115,132,149] There was more heterogeneity in the types of test used to measure language skills compared to global cognition. Figure 4.4 displays the risk factors that were found to be significant in at least one study at low-moderate risk of bias and included in the final model of at least two other studies.

All five studies conducted in children <5 years included male sex in the final model and reported that it was predictive of poor language development.[88,96,125,139,151] It was not possible to comment on the effect of male sex in middle childhood as two of the three studies conducted after five years adjusted for it but did not report the results for adjustment factors[132,149] and the third study did not include sex into the final model as it was not significant in the univariable analysis.[115] Three studies reported that lower parental educational level was associated with poor language development[88,96,115] whereas two

Table 4.3: Summary of studies reporting risk factor analyses for language impairment in children born VPT or with VLBW asessed at <5 years of age

Study reference [identifier]	Country and recruitment period	Age of assessment (years)	GA (weeks)/ BW (grams)	Design and participants	Number (%) of survivors assessed[a]	Outcome measure (continuous (cts) unless otherwise specified)	Method for dealing with untestable children	Significant risk factors for poorer outcome (p<0.05) in final model
Age of assessment <5 years								
Adams-Chapman (2013)[88] [b] [C]	US 2006-2008	1.5-1.8	<1000g	PC of infants admitted to the NICU of 20 centres participating in the multi-centre NICHD NRN routine FUP.	1477 (91%)	Language Composite, Receptive Language and Expressive Language from BSID-III. Blinded assessment.	Assigned a Language Composite score of 46 if severely delayed (n=39).	Black ethnicity, Feeding problems, GMFCS ≥2 at 18-22m, hearing impairment, male sex, lower maternal education, multiple pregnancy, MV days, non-English speaking, no private insurance.
Charkaluk (2010)[96] [U]	France 1997	2	<33w	PC of all live births in 1 French region (Nord-Pas-de-Calais, EPIPAGE Study).	347 (64%)	Language domain of Brunet-Lezine scale (revised).[164]	Excluded children with CP or severe neurosensory impairment (n=45).	Earlier GA, intubation days, male sex, lower parental education, lower parental occupation, SGA.
Sansavini (2011)[139] [c] [V]	Italy 2003-2008	2	<33w	PC study of Infants admitted to a single centre NICU (Bologna).	150 (Not known)	Lexical delay <10th centile and absence of word combination from MB-CDI. Parent report.	Excluded children with major cerebral damage, hearing or visual impairment and if bilingual (n not reported).	**Lexical delay:** BPD, male sex. **Word combination:** Male sex.
Toome (2013)[151] [F]	Estonia 2007	2	<32w	PC of all live births in Estonia enrolled in the national neonatal research routine FUP.	155 (99%)	Language Composite from BSID-III (<70 vs. ≥70).	Assigned a score of -4SD below the mean (n not reported).	IVH 3-4/PVL 2-4, male sex.
Marston (2007)[125] [W]	UK, Republic of Ireland and Australia 1998-2001	1.8-3	<29	Infants requiring endotracheal intubation from birth and enrolled in a multicentre high-frequency ventilation RCT.	288 (49%)	Number of words spoken from 100 word checklist from MB-CDI (short form). Parent report.	N/A.	Any disability, male sex, longer neonatal hospital stay, lower weight SDS at 12m.

[a] Percentage of survivors assessed for outcome measure specified.

[b] The results for the Language Composite are reported, which is comprised of the Receptive and Expressive subscales.

[c] 2 models for language delay reported. Risk factors included in Figure 4a as significant if p<0.05 in both models and non-significant if p≥0.05 in both models, else not included.

Abbreviations: BPD bronchopulmonary dysplasia; BSID Bayley Scales of Infant Development;[167] BW birth weight; CP cerebral palsy; FUP follow up; GA gestational age; GMFCS Gross Motor Functional Classification System;[199] IVH intraventricular haemorrhage; MB-CDI MacArthur-Bates Communicative Development Inventories;[243] MV mechanical ventilation; N/A not applicable; NICU neonatal intensive care unit; NICHD NRN National Institutes of Child Health and Human Development Neonatal Research Network; PC prospective cohort; PVL periventricular leukomalacia; SGA small for gestational age; SD standard deviation; SDS standard deviation score.

Table 4.4: Summary of studies reporting risk factor analyses for language impairment in children born VPT or with VLBW assessed at ≥5 years of age

Study reference [identifier]	Country and recruitment period	Age of assessment (years)	GA (weeks)/ BW (grams)	Design and participants	Number (%) of survivors assessed[a]	Outcome measure (continuous (cts) unless otherwise specified)	Method for dealing with untestable children	Significant risk factors for poorer outcome (p<0.05) in final model
Howard (2011)[115] [b] [X]	Australia 2001-2003	5	<30w or <1250g	PC study of Infants admitted to a single centre NICU and enrolled in Victorian Infant Brain Studies (Melbourne).	187 (83% of enrolled)	Expressive and Receptive Language Skills Standard Scores from KSEALS. Blinded assessment.	Excluded if significant medical issue, CP or neurosensory impairment. (n=12)	**Expressive Language:** Lower BW, poorer communication skills at 24m, lower parental education, WMA. **Receptive Language:** Poorer communication skills at 24m, lower parental education.
Orchinik (2011)[132] [c] [O]	US 2001-2003	5	<28w or <1000g	PC of Infants admitted to a single centre NICU (Ohio) participating in the multicentre NICHD NRN routine FUP.	142 (72%)	Word knowledge from WJ-III COG Verbal Comprehension subtest <10th centile. Blinded assessment.	Assigned a score of 40 if too low functioning to comply with test demands.	IVH 3-4/PVL/ VD, neurosensory disorder and/or MDI<70 at 20m.
Taylor (2006)[149] [d] [R]	US 1992-1995	8	<1000g	PC of infants admitted to a single centre NICU (Ohio) participating in the multicentre NICHD NRN routine FUP.	204 (91%)	Picture Vocabulary from the WJ-III (cts and <1SD below mean of control group). Blinded assessment.	Excluded if untestable due to severe developmental impairments (n=10).	**Model 1 (cts score):** Longer neonatal hospital stay, outborn. **Model 2 (<70 vs. ≥70):** Outborn.

[a] Percentage of survivors assessed for outcome measure specified.

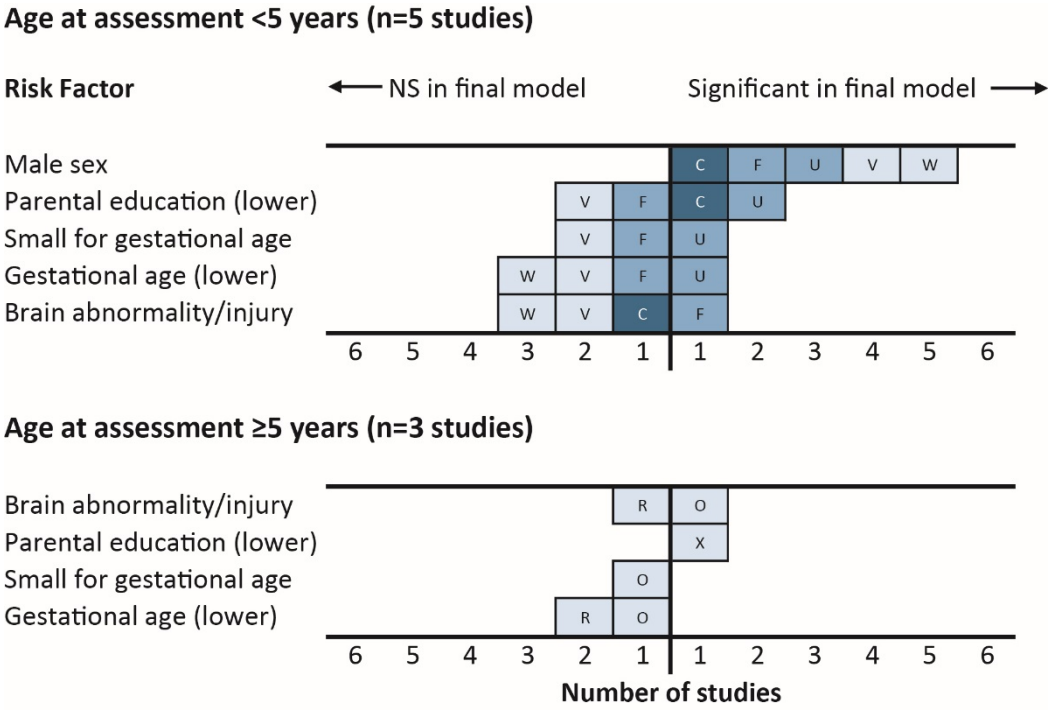
[b] 2 models for reported, one for Expressive Skills and one for Receptive Skills. Risk factors included in Figure 4b as significant if p<0.05 in both models and non-significant if p≥0.05 in both models, else not included.

[c] Each risk factor was fitted separately and adjusted for sex, ethnicity, parental SES and months in school at testing (the article did not report results for the adjustment factors).

[d] Each risk factor was fitted separately and adjusted for sex, ethnicity, parental SES, family stressors and family resources (the article did not report results for the adjustment factors). 2 models for cognitive function reported; one based on dichotomous outcome and one based on continuous outcome. Risk factors included in Figure 4b as significant if p<0.05 in both models and non-significant if p≥0.05 in both models, else not included.

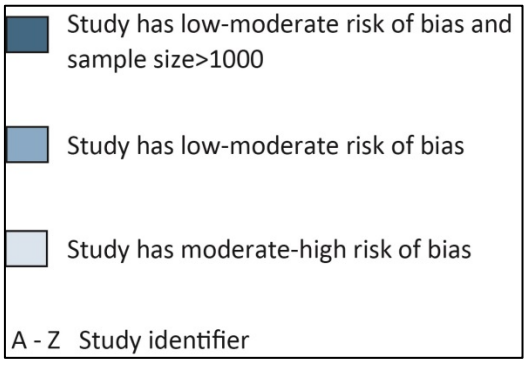
Abbreviations: BW birth weight; CP cerebral palsy; FUP follow up; GA gestational age; IVH intraventricular haemorrhage; KSEALS Kaufman Survey of Early Academic and Language Skills;[244] MDI Mental Developmental Index from the BSID;[167] NICU neonatal intensive care unit; NICHD NRN National Institutes of Child Health and Human Development Neonatal Research Network; PC prospective cohort; PVL periventricular leukomalacia; SD standard deviation; VD ventricular dilatation; WJ-III Woodcock Johnson Tests of Achievement, 3rd edition;[245] WJ-III COG Woodcock Johnson Tests of Cognitive Abilities, 3rd edition;[242] WMA white matter abnormality.

Figure 4.4: Evidence synthesis of risk factors for language impairment in children born VPT or with VLBW



Note: Risk factor presented if significant ($p<0.05$) in the final model of at least one study at low-moderate risk of bias and entered into the final model of at least 3 studies (across all ages).
Brain abnormality/injury: IVH or PVL or ventricular dilation[C/F/O/R]; Persistent hyperechogenicity of white matter (non-major) [V]; Major cranial ultrasound abnormality [W].

Key



studies reported no association.[151,139] There were also mixed findings for the prognostic value of being SGA. It was not possible to draw any conclusions about brain injury diagnosed in the neonatal period as a prognostic factor, because studies used different methods to deal with children with severe neurosensory impairment for whom standard assessments could not be used; some imputed the lowest possible score and others excluded this group completely. As with cognition, there was evidence from multivariable models that GA was not a strong predictor of language development.

4.2.4 Prognostic factors for impaired executive function

Seven studies at moderate-high risk of bias presented risk factor analyses for different aspects of executive function, all of which (except one) were based on age of assessment at 5-12 years. The median number of tests administered within each study was five and the maximum was 13. The risk factors reported in Table 4.5 were significant in at least one of the final models. It was difficult to combine these results in any meaningful way due to the small number of studies using a wide variety of tests to measure interrelated cognitive processes.

4.2.5 Risk factors for poor academic attainment

Four studies (two at low-moderate risk of bias[46,121] and two at moderate-high risk of bias[148,149]) performed risk factor analyses for academic attainment (Table 4.6), all based on age of assessment between 5-12 years. All four presented a model of mathematical ability, two based on letter and word identification and one based on reading scores. Again, there were too few studies and not enough overlap in the risk factors included in the final models to combine the results and draw any meaningful conclusion.

Table 4.5: Summary of studies reporting risk factor analyses for poor executive function in children born VPT or with VLBW

Study reference	Country and recruitment period	Age of assessment (years)	GA (weeks)/ BW (grams)	Design and participants	Number (%) of survivors assessed[a]	Outcome measure (continuous (cts) unless otherwise specified)	Method for dealing with untestable children	Significant risk factors for poorer outcome (p<0.05) in any of the final models.
Lowe (2009)[124]	US 2001-2003	1.5-1.8	<1000g	Infants mechanically ventilated (12-48 hrs) and enrolled in a multicentre hydrocortisone therapy RCT. Excluded multiple births.	233 (80%)	Object permanence measured using 3 items from the BSID-II MDI scale.	Not specified.	Male sex, lower emotional regulation from the BSID-II MDI scale.
Orchinik (2011)[132] [b]	US 2001-2003	5	<28w or <1000g	PC of Infants admitted to a single centre NICU (Ohio) participating in the multicentre NICHD NRN routine FUP.	148 (75%)	13 tests of executive function from a variety of scales. Blinded assessment.	Assigned a score of 40 for WJ-III-COG tests and lowest possible raw score for other tests if too low functioning to comply with test demands.	BW <750g, GA<25w, IVH 3-4/PVL/ VD, neurosensory disorder and/or MDI<70 at 20m.
Potharst (2012)[135] [c]	Netherlands 2002-2004	5	<30w or <1000g	PC study of Infants admitted to a single centre NICU (Amsterdam).	102 (68%)	VIQ, PIQ and PSQ domains from WPPSI-R Dutch version.	Excluded if too disabled to be tested (n=4).	IVH 3-4/PVL 2-4/PHH, male sex, lower MDI at 2yrs, lower PDI at 2 yrs, lower parental education, parental foreign country of birth, sepsis or meningitis.
Potharst (2013)[136]	Netherlands 2002-2004	5	<30w or <1000g	PC study of Infants admitted to a single centre NICU (Amsterdam).	102 (68%)	5 tests of executive function from a variety of scales.	Excluded if too disabled to be tested (n=4).	BPD, lower parental education, parental foreign country of birth, SGA.
Taylor (2006)[149] [d]	US 1992-1995	8	<1000g	PC of infants admitted to a single centre NICU (Ohio) participating in the multicentre NICHD NRN routine FUP.	204 (86%)	NEPSY overall score (cts and <1SD below mean of control group). Blinded assessment.	Excluded if untestable due to severe developmental impairments (n=10).	BW <750g, IVH 1-2, IVH 3-4, NEC, longer neonatal hospital stay, NRI>3, PVL, PN steroids, any US abnormality, VD.
Ford (2011)[106]	Australia 1996-2000	7-9	<32w or <1000g	Cross sectional study of infants identified from the routine database of a single centre (Brisbane).	45 (not a PC)	5 tests of executive function from a variety of scales.	Excluded if not attending mainstream school, a physical or neurological disability or GCI ≤85 on the MSCA at 4yrs.	Lower BW, interaction between neonatal risk score and maternal education.[e]
Aarnoudse-Moens (2013)[75]	Netherlands 1996-2004	4-12	<31w	PC study of Infants admitted to a single centre NICU (Rotterdam). Excluded multiple births.	200 (20%)	5 tests of executive function from a variety of scales.	Excluded if severe disabilities requiring physical assistance to perform daily activities.	Absence of NEC 2-3 or meningitis, lower parental education, lower postnatal growth SDS at 6w CA.

[a] Percentage of survivors seen at assessment; some children had missing data for some executive function outcomes/tests.

[b] Each risk factor was fitted separately and adjusted for sex, ethnicity, parental SES, months in school at testing and age at assessment (the article did not report results for the adjustment factors).

[c] 2 models for each domain at 5 years reported; the first including 2 year and the second including 3 year developmental assessments. The former model is reported as 2 year assessments are more routine in general practice.

[d] Each risk factor was fitted separately and adjusted for sex, ethnicity, parental SES, family stressors and family resources (the article did not report results for the adjustment factors).

[e] Adverse effect of high neonatal risk score ameliorated among children with more highly educated mothers.

Abbreviations: BPD bronchopulmonary dysplasia; BSID Bayley Scales of Infant Development;[166] BW birth weight; CA corrected age; GA gestational age; GCI General Cognitive Index; IVH intraventricular haemorrhage; MSCA McCarthy Scales of Children's Abilities; MDI Mental Developmental Index from the BSID; NEC necrotising enterocolitis; NEPSY A Developmental NEuroPSYchological Assessment;[246] NICU neonatal intensive care unit; NRI Neonatal Risk Index; NICHD NRN National Institutes of Child Health and Human Development Neonatal Research Network; PC prospective cohort; PHH posthaemorrhagic hydrocephalus; PIQ Performance IQ; PN post natal; PSQ Processing Speed Quotient; PVL periventricular leukomalacia; SGA small for gestational age; SD standard deviation; SDS standard deviation score; US ultrasound; VD ventricular dilatation; VIQ Verbal IQ; WJ-III COG Woodcock Johnson Tests of Cognitive Abilities, 3rd edition;[242] WPPSI-R Wechsler's Preschool and Primary Scale of Intelligence-Revised.[238]

Table 4.6: Summary of studies reporting risk factor analyses for poor academic attainment in children born VPT or with VLBW

Study reference	Country and recruitment period	Age of assessment (years)	GA (weeks)/ BW (grams)	Design and participants	Number (%) of survivors assessed[a]	Outcome measure (continuous (cts) unless otherwise specified)	Method for dealing with untestable children	Significant risk factors for poorer outcome (p<0.05) in final model
Kiechl-Kohlendorfer (2013)[121]	Austria 2003-2006	5	<32w	PC study of all live births in a single centre NICU (Innsbruck) serving the whole state of Tyrol.	161 (60%)	TEDI-MATH[247] Sum T-score (<40 vs. ≥40).	Excluded if untestable due to severe disabilities (n=10). Imputed score of <40 for children with severe delay in numerical abilities (n=9).	BPD, ICH, smoking in pregnancy.
Taylor (2011)[148 b]	US 2001-2003	5-6	<28w or <1000g	PC of infants admitted to a single centre NICU (Ohio) participating in the multicentre NICHD NRN routine FUP.	148 (78%)	Letter Word Identification, Spelling, Calculation and Applied Problems subtests from the WJ-III (<85 vs. ≥85). Blinded assessment.	Imputed a score of <85 for children too low functioning to be tested (n≥6).	**Letter Word:** Abnormal US, neurosensory disorder and/or MDI<70 at 20m. **Spelling:** Abnormal US, neurosensory disorder and/or MDI<70 at 20m, lower parental SES. **Calculation:** Neurosensory disorder and/or MDI<70 at 20m. **Applied Problems:** Abnormal US, neurosensory disorder and/or MDI<70 at 20m, lower parental SES.
Taylor (2006)[149 c]	US 1992-1995	8	<1000g	PC of infants admitted to a single centre NICU (Ohio) participating in the multicentre NICHD NRN routine FUP.	204 (86%)	Academic Skills Cluster, Letter Word Identification, Spelling and Calculation subtests from the WJ-III. Blinded assessment.	Excluded if untestable due to severe developmental impairments (n=10).	**Letter Word:** Longer neonatal hospital stay, PVL. **Spelling:** BPD, PN steroids. **Calculation:** NEC, Longer neonatal hospital stay, PN steroids. **Academic Skills Cluster:** BPD, IVH 3-4, NEC, longer neonatal hospital stay, NRI>3, PN steroids, PVL.
Johnson (2011)[46 d]	UK and Republic of Ireland 1995	10-12	<26w	PC of all live births in the UK and Republic of Ireland (EPICure Study).	219 (71%)	Reading and Mathematics Composite Scales from the WIAT-II. Blinded assessment.	Imputed a score of 39 for children with severe cognitive deficit (n=18).	**Reading scores:** no breast milk received, smaller HC SDS at 30m, lower MDI at 30m, lower parental SES, PROM. **Mathematics scores:** smaller HC SDS at 30m, lower MDI at 30m, NEC, lower parental SES, lower PDI at 30m.

[a] Percentage of survivors seen at assessment; some children had missing data for some academic tests.

[b] Each risk factor was fitted separately and adjusted for sex, ethnicity and parental SES (the article did not report results for the adjustment factors).

[c] The Academic Skills Cluster is a composite score based on the Letter Word Identification, Spelling and Calculation subtests. 2 models for the composite score and each subtest was reported; one based on continuous outcome and one based on dichotomous outcome (<1SD below mean of control group). Risk factor reported as significant in this table if p<0.05 in both models. Each risk factor was fitted separately and adjusted for sex, ethnicity, parental SES, family stressors and family resources.

[d] 2 models for reading and mathematics scores reported. Models including neonatal and 30m outcomes as risk factors included. Models including 6 year neuropsychological outcomes only as risk factors excluded.

Abbreviations: BPD bronchopulmonary dysplasia; BSID Bayley Scales of Infant Development;[166] GA gestational age; HC head circumference; ICH intracerebral haemorrhage; IVH intraventricular haemorrhage; MDI Mental Developmental Index from the BSID; NEC necrotising enterocolitis; NICU neonatal intensive care unit; NRI Neonatal Risk Index; NICHD NRN National Institutes of Child Health and Human Development Neonatal Research Network; PC prospective cohort; PDI Psychomotor Developmental Index from the BSID; PN post natal; PROM prolonged rupture of membranes; PVL periventricular leukomalacia; SDS standard deviation score; SES socioeconomic status; US cranial ultrasonography; WIAT-II Wechsler Individual Achievement Test, 2nd edition;[248] WJ-III Woodcock Johnson Tests of Achievement, 3rd edition.[245]

4.3 Discussion

4.3.1 Infant characteristics and cognitive/language impairment

For the VPT/VLBW population, there was fairly strong evidence that male sex was a prognostic factor both for poorer cognitive development and language skills in early infancy, a finding which is supported by other studies that have focussed exclusively on the association of infant sex with cognitive function.[227,230,249] However, in the studies conducted later in childhood that were included in this review, the influence of sex on general cognition was largely diminished. It was not possible to comment on whether this was also true for language development as there were a lack of studies assessing children above the age of five. There were similar findings for non-white ethnicity and lower BW in relation to cognitive dysfunction; both factors were clearly prognostic in early infancy, but with no evidence available in middle childhood for ethnicity, and a lack of association in later years for BW.

4.3.2 Parental education and cognitive/language impairment

There was evidence that a lower level of parental education was predictive of cognitive dysfunction, supported by a recent study in an EPT population that focussed solely on this hypothesis.[250] Unlike factors related to infant characteristics, the influence of parental education appeared to persist into middle childhood. Evidence for the prognostic value of parental education in relation to language development was weak.

Research has shown links between non-white ethnicity, lower BW and parental education/ socio-economic status,[251,252] so it is interesting that these factors were independent predictors in four studies in the <5 age group.[88,112,155,158] Other studies have found that the effect of ethnicity is strongly mediated by markers of deprivation. Duncan et al[253] reported

that black and white Hispanic EPT children in the US performed worse in language and cognition tests at 18-22 months compared to white EPT children, although the differences in cognition were no longer significant once other medical factors and the primary carer's education level were taken into account. There were similar findings in another US study by Brooks-Gunn et al,[254] which examined the ethnic differences in IQ at five years and the mediating role of the environment in a group of preterm low BW children. The IQ scores among black children were around one SD below those of white children and the difference halved when adjusted for markers of poverty; however there was no further reduction after adjusting for maternal education and verbal ability. In the review reported here, parental education emerged as prognostic factor of cognitive outcome whereas parental SES did not. This may be due to multicollinearity, or possibly a single marker of parental SES such as income or occupation, as used in most studies in the review, is not enough to capture an accurate measure of social disadvantage. Combining a range of social markers into a composite score may be a more effective modelling strategy.

4.3.3 Gestational age and cognitive/language impairment

There was strong evidence that GA was not a robust predictor of cognitive and language development in infancy or in middle childhood in the VPT/VLBW population, which was also the case in the review of risk factors for CP and motor impairment. Although the relationship between birth at later GA and improved cognition is well established across the whole spectrum of GA from 25 to 40 weeks,[40] it appears to be a less powerful predictor when restricted to the most preterm subgroups. In the four (of five) studies where GA was found to be predictive of cognitive or language development, the GA inclusion criterion was less restrictive than in studies that did not observe an association. It is also worth noting that

although a strong positive relationship with GA is seen when survival without neurodevelopmental impairment is calculated as a function of all live births, when the denominator becomes survivors at discharge, as with all the studies included in this review, the association weakens. This is because the proportion of surviving children rises steeply with increasing GA while the proportion of impaired survivors does not.

4.3.4 Neonatal brain injury and cognitive impairment

Many studies that have focussed exclusively on the relationship between brain injury diagnosed in the neonatal period and later cognitive function have reported strong linear trends with grade of injury.[207,208,210,255] However, the prognostic value of brain injury in the multivariable models reported in this review was mixed. This is possibly because cognitive and language development are multifactorial, unlike a diagnosis of CP which is more directly related to focal brain injury, so that when other factors are entered into a model, the influence of perinatal factors become less pronounced. The unclear findings may also reflect the different modelling strategies adopted by the studies; some excluded children with CP or other neurosensory impairment, some imputed lowest scores and others adjusted for motor disability. Marret et al[210] observed that the trend between cognitive delay and the severity of cerebral lesions incurred during the neonatal period was strongest when accompanied by motor deficiencies. The incidence of isolated cognitive dysfunction in the absence of motor impairment was similar among children with moderate, mild or no cerebral lesions, and less common in those with major cerebral lesions. Combined motor and cognitive impairment may have a different aetiology to cognitive dysfunction in isolation, so the way in which motor or neurosensory disability is dealt with in a prediction model, will clearly have an impact on the results. It has also been noted that a considerable number of VPT children without

clinically defined white matter pathology go on to develop later cognitive dysfunction compared to their full-term peers, which points to the important influence of disturbances to cerebral growth and structural alterations in the preterm brain, in the absence of a discrete injury.[255]

4.3.5 Outcome measures

One difficulty in the review of cognitive outcomes was the sheer variety of assessments used, particularly in children aged five and above. This was a particular problem for the language, executive function and academic attainment domains, compounded by the smaller number of studies in these areas compared to the global cognitive domain. Studies of executive function typically performed a whole battery of tests, or reported several models based on different subdomains of a global measure of cognition. It is challenging trying to synthesise the results in these circumstances without over representing one particular study population, particularly when the constructs being measured are interrelated and likely to be highly correlated.

Within the global cognitive domain, there were also issues with different versions of the same tool being used over the 24 year study period. For example, one study used the BSID-I when other studies of the same era adopted the BSID-II, so cognitive delay in this study was defined using a different normative population. Prior to 2006, the BSID-II was the standard tool for assessing outcomes in the VPT/VLBW population before it was superseded by the BSID-III. However, the MDI in version II is not directly comparable to the Cognitive or Language Composite scores in version III; it is based on a combination of cognitive and language abilities, thus a low MDI score could represent a delay in cognition, language or both. The Cognitive Composite score in version III is a purely cognitive domain and scores are elevated

compared to those obtained in version II, so the definition of cognitive delay is more stringent in the latest version.[256]

4.3.6 Assessing cognitive function infants with other severe disability

Another major issue in this review was the heterogeneous nature of the study groups included in the risk factor modelling, due to the lack of uniformity in inclusion criteria. While some studies used the whole cohort, representing the entire spectrum of disability, and imputed values for those who's functioning was too impaired to be tested by standard assessments, other studies excluded children with CP, major disability or those not attending mainstream schools at the outset. Some studies fell somewhere between these two approaches, excluding only children if they were untestable which could have been due to a variety of reasons including behavioural problems, motor difficulties or low cognitive functioning. It is difficult to isolate cognitive and language outcomes from motor and neurosensory impairments, so the importance of stating the precise research question and using the correct study design and modelling strategy to mirror the proposed aims cannot be understated.

4.3.7 Conclusion

In conclusion, there was evidence that male sex, non-white ethnicity, lower levels of parental education and lower BW were significant predictors of global cognitive dysfunction in children aged 18-30 months who were born VPT or VLBW. After the age of five years, the impact of infant sex and BW diminished, parental education was still influential, and there was no evidence on the sustained effect of ethnicity. It is highly unlikely that ethnicity itself is a causal factor for cognitive impairment; other research has demonstrated a strong correlation between ethnicity, poverty and social disadvantage. There was evidence that male sex was

predictive of impaired language development in early infancy, but no evidence that this was sustained into childhood. There were mixed findings on the prognostic value of brain injury diagnosed during the neonatal period on language and cognition, but this may reflect the heterogeneous selection criteria and methods of dealing with missing data related to severe disability across the studies. There was evidence that within the VPT/VLBW population, GA had little value as a prognostic factor in multivariable models predicting the risk of cognitive or language development at any age over 18 months. The findings of this review suggest that the impact of perinatal risk factors diminish over time as other environmental and social factors become more influential.

Chapter 5: Prognostic factors for behavioural problems and psychiatric disorders in children born VPT or with VLBW: systematic review results

5.1 Introduction

In addition to cognitive impairments, other long term sequelae in surviving VPT/VLBW children include behavioural problems and psychiatric disorders. Studies using behavioural screening questionnaires have shown that children born VPT or with VLBW are at increased risk of social, emotional and attention problems and internalizing problems (anxiety/depression) compared with term-born controls.[42] Clinically significant behaviour problems on screening questionnaires have been reported in 13% to 46% of VPT/VLBW children.[257] However, screening tools are designed to have a high rate of sensitivity, in order to identify children who are at risk of developing a psychiatric disorder and for whom further assessment would be beneficial[258-260] and thus the rates of disorders are typically lower. Studies using diagnostic evaluations have reported an excess of ADHD, ASD and psychiatric disorders in children born VPT or with VLBW compared to term-born controls.[40,41] A recent review of clinical cohort studies reported that the prevalence of DSM-IV-TR (Diagnostic and Statistical Manual of Mental Disorders - 4th edition)[185] based ADHD diagnoses ranged between 16% to 19% in VPT/VLBW children, with an increase in odds of 2 to 3 compared to term-born peers.[257] ASD is less common, with a median prevalence of 0.6% in the general population, but two studies have reported that 3.6% of ELBW children[110] and 8% of EPT

children,[119] respectively, met the diagnostic criteria for ASD when assessed between 8-11 years. Behavioural problems in VPT/VLBW children have been shown to persist into adolescence,[261,262] and there is evidence that the risk of being diagnosed with psychiatric disorders in adulthood increases with decreasing GA at birth.[263,264]

The pattern of behavioural problems observed in VPT/VLBW children has been shown to be similar across different countries, despite cultural differences and disparity in neonatal care, implicating some underlying biological mechanism.[265] It has been suggested that a 'preterm behavioural phenotype' may exist, characterised by socio-communicative and emotional problems and inattention.[257] The mechanisms underlying this neurobehavioural profile are unclear, although several explanations have been proposed.[30] The VPT/VLBW newborn brain is extremely vulnerable and clinical and environmental factors that disturb a critical period of brain development that normally takes place in utero may be highly influential. Exposure to prolonged hospitalisation and therapeutic interventions may disrupt normal neurodevelopment, even in the absence of focal brain injury. This is compounded by the stressful environment of a busy NICU with a high noise level, constant bright lighting, multiple monitoring devices and reduced opportunity for parent-infant interaction. Early exposure to such a sustained level of stress may have an adverse impact on brain development, akin to that observed in adults.[266] Later environmental influences in early infancy and childhood, such as parental mental health, caregiving style, or limited contact with peers and family due to prolonged periods of hospitalisation/illness may impede the development of coping strategies, emotional regulation, attachment and other social skills.[267] All of which are more likely to occur following VPT/VLBW birth.

Among the 78 articles that were identified in the review and included in the data extraction, 15 contained risk factor analyses for behavioural problems or psychiatric disorders.[41,43,100,110,119,120,130,133,143-146,149,152,157] These studies were based on nine cohort populations; two other articles based on cognitive outcomes were excluded due to cohort overlap.[96,128] In this chapter, the results from the 15 articles reporting the risk of behavioural problems and psychiatric disorders are summarised.

5.2 Data synthesis of prognostic factors for behavioural problems and psychiatric disorders

5.2.1 Study characteristics

The main study design was prospective cohort (n=14), and there was one RCT population.[133] Of the 14 prospective cohorts, seven were ascertained from all live births in a geographically defined region,[41,43,100,119,130,144,146] five were recruited from a single centre NICU,[110,120,143,149,152] and three from multiple NICUs.[133,145,157] Studies were conducted in seven countries: US (n=4), UK or England (n=4), Australia (n=2), France (n=2) and one study each from Germany, Netherlands and New Zealand. The median sample size was 219 (range 75 to 1228) and two studies had more than 1000 participants.[43,100] Five studies were restricted to extremely preterm children; <27 weeks[130,144,145] and <26 weeks,[41,119] and two studies excluded multiple births.[43,100] The risk of bias assessment classified four studies as having low-moderate risk of bias and 11 studies as having moderate-high risk of bias (Figure 5.1).

The 15 studies included in the review comprised 47 risk factor analyses which divided into two broad groups: behavioural problems measured by a screening tool and psychiatric

Figure 5.1: Risk of bias assessment of the 15 included studies reporting behavioural outcomes

Risk of bias domain							
	Study Participation	Study Attrition	Prognostic Factor Measurement	Outcome Measurement	Study Confounding	Analysis and Reporting	Low-moderate risk of bias
Delobel-Ayoub (2006)							✓
Delobel-Ayoub (2009)							
Hack (2009)							
Johnson (2010a)							✓
Johnson (2010b)							✓
Jones (2013)							
Moore (2012)							
Peralta-Carcelen (2013)							
Spittle (2009)							
Stahlmann (2009)							
Stephens (2012)							✓
Stoelhorst (2003)							
Taylor (2006)							
Treyvaud (2013)							
Wong (2014)							

Risk of bias: Low Moderate High

disorders. The psychiatric outcomes were further divided by method of assessment (screening tool versus diagnostic criteria) and by type of disorder (any, ASD and ADHD). Some studies reported a model for a global score and also models for each subdomain, whilst others reported additional models adjusting for concurrent factors such as cognition and language. The median number of candidate risk factors considered at the outset in each study was 16 [range: 7 to 42]. For the initial screening of candidate factors to be entered into the final model, five studies included them all and seven included those with a p-value below a set threshold after initial screening. The most popular method of model building after initial screening was to include all factors screened (n=6 studies) and stepwise selection (n=5 studies). Only six of the 15 studies reported the number of participants included in the final model presented. One study assessed model discrimination using the area under the receiver operating curve,[143,145] but apart from that no studies performed any type of statistical or clinical validation.

5.2.2 Prognostic factors for behavioural problems

Eight studies contained a risk factor analysis for general behavioural problems; five studies assessed outcome under five years of age (Table 5.1)[100,120,133,143,146] and three studies over five years of age (Table 5.2).[43,144,149] Six of the studies excluded and/or adjusted for neurodevelopmental delay or cognitive impairment.[43,100,120,133,144,146] All studies used validated, parent report behavioural screening questionnaires, the most common being the Strengths and Difficulties Questionnaire (SDQ)[268] (n=4).[43,100,120,144] The Total Difficulties Score comprises four (five-item) subscales: conduct problems, inattention-hyperactivity, emotional symptoms and peer problems. One study[146] used the Child Behaviour Checklist (CBCL),[269] which is a 99-item questionnaire with six syndrome scales that are combined to

Table 5.1: Summary of studies reporting risk factor analyses for general behavioural problems in children born VPT or with VLBW asessed at <5 years of age

Study reference	Country and recruitment period	Age of assessment (years)	GA (weeks)/ BW (grams)	Design and participants	Number (%) of survivors assessed[a]	Outcome measure (continuous (cts) unless otherwise specified)	Exclusion criteria and/or adjustment for concurrent neurodevelopmental delay	Significant risk factors for poorer outcome (p<0.05) in final model
Stoelhorst (2003)[146] [b]	Netherlands 1996-1997	2	<32w	PC of all live births in 3 Dutch health regions comprising 9% of the population.	160 (68%)	Total Problem score, Internalizing and Externalizing scale from the CBCL. Parent report.	**Adjusted for:** neurological abnormalities at 2 yrs.	**Total problem:** SGA. **Internalizing:** SGA. **Externalizing:** None significant.
Spittle (2009)[143]	Australia 2001-2003	2	<30w or <1250g	PC study of Infants admitted to a single centre NICU and enrolled in Victorian Infant Brain Studies (Melbourne).	188 (84%)	Externalizing, Internalizing, Dysregulation and Competence domains from ITSEA. Parent report.	None.	**Internalizing:** Higher social risk.[c] **Externalizing:** None significant. **Dysregulation:** None significant. **Competence:** Lower BW, PN steroids, female sex, moderate-severe WMA.
Peralta-Carcelen (2013)[133]	United States 1999-2001	2.5	<1000g	Infants admitted to the NICU of 15 centres participating in the multi-centre NICHD NRN routine FUP and enrolled in a glutamine supplementation RCT.	696 (60%)	Total Competence (≤15th vs. >15th centile) and Total Problem score (≥75th vs. <75th centile) from BITSEA Parent report.	**Excluded**: blind, deaf, syndrome (n=30). **Adjusted for:** CP, abnormal neurological exam, MDI<70 and PDI<70 from BSID-II at 2.5 yrs.	**Total Competence:** Hispanic or non-white ethnicity, MDI<70 and PDI<70 from BSID-II at 2.5 yrs. **Total Problem:** Female sex, lower household income, MDI<70 and PDI<70 from BSID-II at 2.5 yrs.
Delobel-Ayoub (2006)[100]	France 1997	3	<33w	PC of all live births in 9 French regions comprising one third of all births (EPIPAGE Study). Excluded multiples.	1228 (69%)	Total Difficulties score from SDQ (≤10th vs. >10th centile of control group). Parent report.	**Excluded:** blind, deaf, severe CP (n=63). **Adjusted for**: neurodevelopmental delay and health status at 3 yrs.	Hospitalisation in last yr, younger maternal age, lower maternal education, neurodevelopmental delay and poor health status at 3 yrs.
Jones (2013)[120]	New Zealand 1998-2000	4	<33w	PC of infants admitted to a single centre NICU (Christchurch).	105 (98%)	Social competence composite score.[d] Parent report.	**Excluded from all:** blind (n=1). **Model 1:** Risk factors to term. **Model 2:** Adjusted for family functioning and parenting 2-4 yrs. **Model 3:** Adjusted for factors in model 2 plus IQ at 4 yrs.	**Model 1:** Family SES, male sex, severity of neonatal WMA. **Model 2:** Male sex, higher maternal anxiety at 2-4 yrs, negative and intrusive parenting at 4 yrs. **Model 3:** GA<28w, higher maternal anxiety at 2-4 yrs, negative and intrusive parenting at 4 yrs, lower IQ at 4 yrs.

[a] Percentage of survivors assessed for outcome measure specified.

[b] 9 models reported in total; Total Problem score, Internalizing and Externalizing scales and the 6 syndrome scores that comprise the total score.

[c] Social risk was based on a composite measure of six social risk factors: family structure, education of primary caregiver, occupation and employment status of primary income earner, language spoken at home and maternal age at birth.

[d] Social competence composite score was derived by the authors by summing sub-scale scores across the following measures: SDQ, Behavioural Inventory of Executive Function – Preschool version (BRIEF-P), Emotional Regulation Checklist (ERC), Infant Toddler Symptom Checklist (ITSC), Emotional Regulation subscale from BSID-II, Penn Interactive Peer Play Scale (PIPPS). 3 models were reported; the full model adjusting for IQ is reported in this review.

Abbreviations: BITSEA Brief Infant-Toddler Social and Emotional Assessment;[270] BSID Bayley Scales of Infant Development;[166] BW birth weight; CBCL Child Behaviour Checklist;[269] CP cerebral palsy; FUP follow up; GA gestational age; IQ intelligence quotient; ITSEA Infant-Toddler Social and Emotional Assessment;[271] MDI Mental Developmental Index from the BSID-II; NICU neonatal intensive care unit; NICHD NRN National Institutes of Child Health and Human Development Neonatal Research Network; PC prospective cohort; PDI Psychomotor Developmental Index from the BSID-II; PN post natal; RCT randomised controlled trial; SES socio-economic status ; SDQ Strengths and Difficulties Questionnaire;[268] SGA small for gestational age; WMA white matter abnormality.

Table 5.2: Summary of studies reporting risk factor analyses for general behavioural problems in children born VPT or with VLBW assessed at ≥5 years of age

Study reference	Country and recruitment period	Age of assessment (years)	GA (weeks)/ BW (grams)	Design and participants	Number (%) of survivors assessed[a]	Outcome measure (continuous (cts) unless otherwise specified)	Exclusion criteria and/or adjustment for concurrent neurodevelopmental delay	Significant risk factors for poorer outcome (p<0.05) in final model
Delobel-Ayoub (2009)[43]	France 1997	5	<33w	PC of all live births in 9 French regions comprising one third of all births (EPIPAGE Study). Excluded multiples.	1102 (59%)	Total Difficulties score from SDQ (≤10th vs. >10th centile of control group). Parent report.	**Excluded**: blind, deaf, severe CP (n=63). **Adjusted for:** IQ and development (parent reported) at 5 yrs.	Hospitalisations in last 5 years, younger maternal age, poor maternal mental well-being in previous month, lower IQ and delayed development at 5yrs
Stahlmann (2009)[144]	Germany 1997-1999	7-9	<27w	PC of all live births in all 8 perinatal centres in Schleswig-Holstein.	75 (82%)	Total Difficulties score from SDQ. Parent report.	**Adjusted for:** IQ at 7-9 yrs.	Lower maternal education, IQ<70 at 7-9 yrs.
Taylor (2006)[149] [b]	United States 1992-1995	8	<1000g	PC of infants admitted to a single centre NICU (Ohio) participating in the multicentre NICHD NRN routine FUP.	204 (86%)	Adaptive Behaviour Composite score from the VABSS (cts and <1SD below mean of control group). Parent report.	Each risk factor was fitted separately and adjusted sex, race, parental SES, family stressors and family resources (p-values not reported).	**Model 1 (cts score):** Longer neonatal hospital stay, outborn. **Model 2 (<70 vs. ≥70):** Longer neonatal hospital stay, NRI>3.

[a] Percentage of survivors assessed for outcome measure specified.

[b] 2 models for Adaptive Behaviour Composite and its 3 domains reported; one based on dichotomous outcome and one based on continuous outcome.

Abbreviations: BW birth weight; CP cerebral palsy; FUP follow up; GA gestational age; IQ intelligence quotient; NICU neonatal intensive care unit; NICHD NRN National Institutes of Child Health and Human Development Neonatal Research Network; NRI neonatal risk index; PC prospective cohort; SES socio-economic status ; SDQ Strengths and Difficulties Questionnaire;[268] VABSS Vineland Adaptive Behaviour Scales Screener.[272]

give an overall Total Problem score: anxious/depressed, withdrawn, aggressive, destructive, sleep problems and somatic behaviour. The 169-item Infant-Toddler Symptom Checklist (ITSEA)[271] and its brief 42-item version (BITSEA)[270] were used by two studies.[133,143] Both checklists include items measuring internalizing and externalizing problems, dysregulation and socio-emotional competence. Both the SDQ and ITSEA/BITSEA have been shown to be highly correlated with the CBCL.[270,273] One study used the Vineland Adaptive Behaviour Scales Screener (VABSS)[272] which measures adaptive functioning in the domains of communication, socialisation and daily living skills.

There was only one study at low-moderate risk of bias (which also had a sample size >1000) among this group of eight studies examining general behavioural problems.[100] Factors that were found to be significant predictors for behavioural problems at age three years in this study were: hospitalisation after neonatal discharge, younger maternal age, lower level of maternal education, and neurodevelopmental delay/poor health status measured at the time of assessment. The later study in the same cohort at age five years[43] had similar findings. All eight studies entered some indicator of socio-economic deprivation into the final model, for example, parental education, income, social risk; and five found at least one of these factors significantly related to behavioural problems.[100,120,133,143,144] In the six studies that adjusted for neurodevelopmental delay or general cognitive ability at the time of assessment, five studies found a significant association between these factors and poorer behavioural outcomes.[43,100,120,133,144] Two studies reported that female sex and one study reported that male sex was significantly associated with behavioural problems, but five of the studies did not find sex significant in the final model. Overall, there was not enough overlap in the risk

factors identified for general behavioural problems in this small group of studies to provide any conclusive evidence about risk factors for prognosis.

5.2.3 Prognostic factors for psychiatric disorders

Seven studies reported risk factor analyses for psychiatric disorders (Table 5.3); two for any DSM-IV-TR[185] diagnosis,[41,152] five for ASD symptoms or diagnoses[110,119,130,145,157] and one for ADHD[110] (this study also reported a model for ASD).

Both studies examining the risk of any psychiatric disorder used the Development And Well Being Assessment (DAWBA),[274] which is a structured psychiatric evaluation administered to parents and teachers. In both studies, a DSM-IV-TR diagnosis was assigned by two blinded clinical psychologists aided by the DAWBA computer scoring algorithm for common childhood diagnoses, such as ADHD, ASD, emotional and conduct disorders. The DAWBA has good concurrent validity when compared with clinical diagnoses.[274] In the study at moderate-high risk of bias conducted at age seven years,[152] the prevalence of any DSM-IV-TR diagnosis was 24% in VPT children, which was similar to 23% prevalence rate reported by the study at low-moderate risk of bias conducted at age 10-12 years in EPT children.[41] The first study screened 11 candidate risk factors in a univariable analysis and retained four significant factors predictive of any DSM-IV-TR in the final model: brain abnormality at term, female sex, social-emotional problems at five years (SDQ) and higher familial social risk at seven years.[b] The second study entered 34 candidate risk factors into a multivariable forward stepwise regression model and retained five significant factors predictive of any DSM-IV-TR in the final model: necrotising enterocolitis, internalizing behaviour problems at 2.5 years (CBCL),

[b] Familial social risk was based on a composite measure of six social risk factors: family structure, education of primary caregiver, occupation and employment status of primary income earner, language spoken at home and maternal age at birth.

pervasive attentional and conduct problems at six years (SDQ) and serious neurodevelopmental disability at six years. These findings suggest that behavioural problems identified by screening tests in infancy and early childhood may help to identify children at risk of developing a psychiatric disorder in later childhood.

The five studies examining risk factors for ASD were divided into those that assessed ASD symptoms using dimensional measures,[119,157] the rate of positive screens using screening tools[110,130,145] and diagnoses made using a diagnostic evaluation[119] (Table 5.3). The two studies reporting risk factor analyses for ASD symptoms were not comparable with respect to age of assessment, outcome measure used, GA group, exclusion criteria, risk of bias, and had no significant risk factors in common.[119,157] However, similar to the findings for general behavioural problems and any psychiatric disorders, markers of social deprivation and poor language development,[157] and problems identified in earlier cognitive and behavioural assessments[119] were reported to be significantly associated with ASD symptoms later in childhood.

Of the three studies that presented risk factor analyses for a positive ASD screen, two were conducted in early childhood.[130,145] One had a low-moderate risk of bias[145] that defined cases as children with at least one positive screen from three different screening tests at 18 months (Pervasive Developmental Disorders Screening Test-II[275] and two items adapted from the Autism Diagnostic Observation Scales (ADOS)[276]). The second was study at study at moderate-high risk of bias[130] that used the 23-item Modified Checklist for Autism in Toddlers (M-CHAT)[277] at two years. The third study, also at moderate-high risk of bias, was conducted in later childhood at eight years[110] using the 12 items related to Autistic Disorder from the Parent Child Symptom Inventory.[278] Amongst these studies the prevalence of a positive

Table 5.3: Summary of studies reporting risk factor analyses for psychiatric disorders in children born VPT or with VLBW

Study reference	Country and recruitment period	Age of assessment (years)	GA (weeks)/ BW (grams)	Design and participants	Number (%) of survivors assessed[a]	Outcome measure (continuous (cts) unless otherwise specified)	Exclusion criteria and/or adjustment for concurrent neurodevelopmental delay	Significant risk factors for poorer outcome (p<0.05) in final model
Any psychiatric disorder: diagnosis								
Treyvaud (2013)[152]	Australia 2001-2003	7	<30w or <1250g	PC study of Infants admitted to a single centre NICU and enrolled in Victorian Infant Brain Studies (Melbourne).	177 (79%)	DAWBA. Parent report. DSM-IV-TR diagnosis assigned using scoring algorithm and clinical judgement of 2 blinded reviewers.	None.	Brain abnormality at term, female sex, social-emotional problems at 5 years (SDQ), higher familial social risk at 7 years.[b]
Johnson (2010a)[41]	UK and Republic of Ireland 1995	10-12	<26w	PC of all live births in the UK and Republic of Ireland (EPICure Study).	219 (71%)	DAWBA. Parent report. DSM-IV-TR diagnosis assigned using scoring algorithm and clinical judgement of 2 blinded reviewers.	None.	NEC, Internalizing behaviour problems at 2.5 yrs (CBLC), pervasive attentional and conduct problems at 6 yrs (SDQ), serious functional disability at 6 yrs.
Autism spectrum symptoms: dimensional measure								
Wong (2014)[157]	England 2010-2012	1.8-2.2	<33w	Infants attending routine FUP in 13 centres (London). Neonatal data extracted retrospectively.	141 (70%)	Q-CHAT score. Parent report.	**Excluded:** CP or severe neurosensory impairment (n=10). **Adjusted for**: Language Composite Score from BSID-III at 2 yrs.	Higher deprivation, non-white ethnicity, BSID-III Language Composite Score at 2 yrs.
Johnson (2010b)[119]	UK and Republic of Ireland 1995	10-12	<26w	PC of all live births in the UK and Republic of Ireland (EPICure Study).	219 (71%)	Total score from SCQ. Parent report.	None.	No breast milk, IQ<2SD at 6 yrs, pervasive attentional and peer problems at 6 yrs (SDQ), withdrawn (CBCL) at 2.5 yrs.
Autism spectrum disorder: positive screen								
Stephens (2012)[145]	United States 2008-2010	1.5-1.9	<27w	PC of infants admitted to the NICU of 15 centres participating in the multi-centre NICHD NRN routine FUP.	554 (74%)	1+ positive screen on 3 tests: PDDST-II (parent report), Response to Joint Attention and Response to Name (ADOS, direct observation).	**Excluded from all**: severe CP, blind, deaf (n=31). **Model 1:** Unadjusted **Model 2:** Adjusted for cognition and language at 18m. **Model 3**: Adjusted for cognition, language and behaviour at 18m.	**Model 1:** Lower BW, non-white ethnicity, male sex. **Model 2:** Male sex, lower Cognitive and Language Composite Score from BSID-III at 18m. **Model 3:** Lower Cognitive and Language Composite Score from BSID-III at 18m, higher Problem and lower Competence Score from BITSEA at 18m.
Moore (2012)[130 c]	England 2006	2	<27w	PC of all live births in England (EPICure-2 Study).	559 (54%)	Positive M-CHAT screen. Parent report.	**Model 1:** Full cohort **Model 2:** Excluded neuro-sensory impairment (n=72). **Model 3**: Excluded any disability (n=320)	**Model 1:** Severe BPD, any CUSS abnormality, PN steroids, positive blood culture ≥72 hrs, male sex. **Model 2:** Positive blood culture ≥72 hrs. **Model 3:** Any CUSS abnormality, positive blood culture <72hrs and ≥72 hrs, male sex.
Hack (2009)[110]	United States 1992-1995	8	<1000g	PC of infants admitted to a single centre NICU (Ohio) participating in the multicentre NICHD NRN routine FUP.	219 (97%)	CSI-4 based on DSM-IV-TR diagnostic criteria. Parent report.	Each risk factor was fitted separately and adjusted sex, race, parental SES (p-values not reported).	BPD.

Table 5.3: Summary of studies reporting risk factor analyses for psychiatric disorders in children born VPT or with VLBW (continued)

Study reference	Country and recruitment period	Age of assessment (years)	GA (weeks)/ BW (grams)	Design and participants	Number (%) of survivors assessed[a]	Outcome measure (continuous (cts) unless otherwise specified)	Exclusion criteria and/or adjustment for concurrent neurodevelopmental delay	Significant risk factors for poorer outcome (p<0.05) in final model
Autism spectrum disorder: diagnosis								
Johnson (2010b)[119]	UK and Republic of Ireland 1995	10-12	<26w	PC of all live births in the UK and Republic of Ireland (EPICure Study).	219 (71%)	DAWBA. Parent report. DSM-IV-TR diagnosis assigned using scoring algorithm and clinical judgement of 2 blinded reviewers.	None.	Cognitive impairment at 6 yrs, pervasive peer problems at 6 yrs (SDQ).
Attention deficit/hyperactivity disorder: positive screen								
Hack (2009)[110]	United States 1992-1995	8	<1000g	PC of infants admitted to a single centre NICU (Ohio) participating in the multicentre NICHD NRN routine FUP.	219 (97%)	CSI-4 based on DSM-IV-TR diagnostic criteria. Parent report.	Each risk factor was fitted separately and adjusted sex, race, parental SES (p-values not reported).	None significant for hyperactive, inattentive or combined type of ADHD.

[a] Percentage of survivors assessed for outcome measure specified.
[b] Familial social risk was based on a composite measure of six social risk factors: family structure, education of primary caregiver, occupation and employment status of primary income earner, language spoken at home and maternal age at birth.
[c] 5 further models were reported: 3 models using only risk factors known at birth for the full cohort, excluding all disability and excluding neurosensory disability; 2 models with more stringent criteria for a positive screen.
Abbreviations: ADOS Autism Diagnostic Observation Scales;[276] BPD bronchopulmonary dysplasia; BSID Bayley Scales of Infant Development;[166] BW birth weight; CBCL Child Behaviour Checklist;[269] CP cerebral palsy; CSI-4 Parent Child Symptom Inventory;[278] CUSS cranial ultrasound abnormality; DAWBA Development And Wellbeing Assessment;[274] DSM-IV-TR Diagnostic and Statistical Manual of Mental Disorders;[185] FUP follow up; GA gestational age; M-CHAT Modified Checklist for Autism in Toddlers;[277] NICU neonatal intensive care unit; NICHD NRN National Institutes of Child Health and Human Development Neonatal Research Network; NEC necrotising enterocolitis; PDDST Pervasive Developmental Disorders Screening Test;[275] PC prospective cohort; Q-CHAT Quantitative Checklist for Autism in Toddlers;[279] *SDQ Strengths and Difficulties Questionnaire;*[268] *SCQ Social Communication Questionnaire*.[280]

ASD screen varied greatly; 20%, 41% and 2% respectively likely reflecting differences in population denominators and screening tools. Bronchopulmonary dysplasia and male sex were significant risk factors in two out of three of these studies, but there were no other significant risk factors in common. The ASD screening study at low-moderate risk of bias[145] presented two additional models adjusting for language, cognition and social-emotional behavioural problems, at the same age of assessment, all of which were significant.

Only one study at low-moderate risk of bias conducted at age 10-12 years assigned ASD diagnosis based on standard diagnostic DSM-IV-TR criteria, using the DAWBA.[119] The prevalence of an ASD diagnosis was 8% (n=16 cases): 13 (6.5%) with autistic disorder and three (1.5%) with pervasive developmental disorder not otherwise specified. The risk prediction model was based on any ASD diagnosis and after entering 42 candidate variables sequentially into a multivariable stepwise model, only two factors significantly affected the prediction of any ASD diagnosis: cognitive impairment and pervasive peer problems at age six years (SDQ).

The only study that presented a risk factor analysis for a positive screen for ADHD[110] did not report any significant risk factors for either hyperactive, inattentive or the combined type of ADHD. The prevalence of a positive screen for ADHD was reported to be 17% (n=37) in this study.

5.3 Discussion

5.3.1 Summary of findings

The eight studies reporting risk factor analysis for general behavioural problems all had a moderate to high risk of bias, with one exception,[100] and the screening tools used were fairly heterogeneous with different sub-domains assessed. The modelling of outcome scores also varied with some studies reporting the proportion of children scoring above the cut-off for a positive screen and others analysing continuous scores. The studies also differed in the way children with neurodevelopmental delay or disability were handled in the design and analysis; some studies excluded them completely, some adjusted for motor and/or cognitive impairment and some adopted both or neither strategy. There was also a lack of commonality in the risk factors studied for prognosis, therefore it was difficult to synthesise the results and reach any meaningful conclusion. The only factors that appeared to be consistent predictors of general behavioural problems were markers of socio-economic deprivation and neurodevelopmental or cognitive delay, but apart from these there was no clear evidence about the prognostic value of any other risk factors studied.

Two studies examined the risk of developing a psychiatric disorder in later childhood[41,152] and reported that social-emotional and behavioural problems identified by screening questionnaires in infancy or early childhood were predictive of later disorders. This finding is supported by other studies that have examined the predictive validity of screening tests and the stability of diagnoses over time.[281,282] Early screening tests are known to identify a large number of false-positives, particularly in impaired populations with high rates of neurological and cognitive impairment,[258] so the positive predictive value for later psychiatric diagnoses is likely to be low. There is also a lack of evidence about how sensitive these general screening

tests are for predicting specific types of DSM-IV-TR disorder. However, there is evidence of enhanced specificity in prediction in VPT/EPT populations for both general disorders and specific behavioural outcomes.[283] Given the stability of neurodevelopmental and behavioural outcomes in VPT/VLBW children, early screening may thus have greater predictive validity and clinical utility in preterm populations.

The only factors consistently associated with ASD symptoms, positive screen or diagnosis, were cognitive or language impairment, and poor performance on a behavioural screening test earlier in childhood. Aside from these, no clear evidence emerged for any other risk factors. The number of cases with an ASD diagnosis was very small in the study conducted in later childhood using DSM-IV-TR based diagnostic criteria,[119] so a lack of power means that results should be interpreted with caution. Only one study presented a risk factor analysis for a positive screen for ADHD and no significant predictive factors were identified.[110]

5.3.2 Explanation of findings

An explanation for the inconclusive findings, beside the small number of studies examining each type of disorder and the lack of commonality in the candidate risk factors studied, is the several different strategies for dealing with confounding due to neurological and cognitive impairment. In some studies the whole cohort was included, representing the whole spectrum of disability in the VPT/VLBW population. In other studies, children with neurological and/or cognitive impairment were excluded from the modelling process to identify risk factors in a more homogeneous population and to eliminate the noise created by additional impairments. Other studies attempted to achieve this by adjusting for these factors in the analysis. However, the risk factors for a psychiatric disorder in the absence of any impairment may be very different to those factors which are prognostic for behavioural

difficulties accompanying profound impairment. Therefore the strategy for dealing with motor, neurosensory and cognitive impairment in risk factor analyses, in terms of exclusion and/or adjustment, will crucially affect the findings. Adjustment for cognitive impairment is particularly problematic as it is frequently associated with psychiatric conditions in the term population; adjustment in a VPT/VLBW population where cognitive delay is more common and part of the preterm phenotype is therefore likely to lead to overcorrection.[284]

VPT/VLBW children with motor or cognitive impairment have been reported to be at higher risk of developing behavioural and emotional problems, compared to VPT/VBLW children with no impairments.[216] It is possible that the challenge of living with a profound impairment could induce feelings low self-esteem, anxiety, insecurity and detachment which then manifests as a behavioural problem. It might also be due to physical or biological effects on brain function leading to an alteration emotional and behavioural regulation. However, the high rate of problems in children with neurodevelopmental impairments may also be related to measurement issues. In a cohort of two year old EPT children, Kuban et al[285] reported that increased odds of a positive screen for autism using the M-CHAT among those unable to sit or stand was 23-fold, 8-fold in those with a major vision or hearing impairment and 13-fold in those with severe cognitive impairment, compared to EPT children without such impairments. Moore et al[130] reported similar findings in a cohort of EPT children at two years; 16.5% of children without disability screened positive on the M-CHAT compared to 96% with severe motor impairment, 56% with cognitive impairment and all children with a significant vision or hearing impairment. However, such findings should be interpreted with caution, as many items on the M-CHAT rely on an intact motor, hearing and vision function which leads to an inflated false-positive rate among children with such disabilities. Indeed, a recent study has

shown that screening for autism using the M-CHAT questionnaire is especially confounded in a preterm population.[286] However, the rate of positive screens were still threefold higher among unimpaired EPT children compared to unselected populations,[285] hence neurological impairment cannot be the sole explanation for the differences observed. Even so, it is difficult to disentangle the aetiology of neurobehavioural disorders outside the context of the neurological sequelae that follow VPT birth.

5.3.3 Recommendations

This systematic review points to the need for further well-conducted research investigating risk factors for psychiatric disorders and behavioural problems in the VPT/VLBW population. Such conditions are common following VPT birth and can have an adverse impact on the lives of children and their families. In order to plan care for the VPT/VLBW population, it is important to determine the mechanisms underlying the 'preterm behavioural phenotype'; the clinical expression of disorders appear to differ in VPT children and thus risk factors may be different to those identified in the term population. As such, the identification of predictive factors is important for understanding aetiological mechanisms and for developing appropriate screening, intervention and treatment strategies.

Studies with larger sample sizes and greater power are needed for studying childhood psychiatric disorders in this population, particularly for less common conditions such as ASD or ADHD. Longer term follow-up with outcome evaluations beyond 18-24 months is also needed, as the risk for psychiatric disorders cannot be reliably assessed at this age and because of the natural course of some disorders which may onset later in childhood. Furthermore, prognosis is likely to be a dynamic process with social and environmental factors potentially superseding the influence of early clinical and biological factors as the child

grows up. It is recommended that future studies of the risk of behavioural problems use of standard diagnostic evaluations to assess outcome and evaluate of a broad range of biological and social risk factors over time. They should also make a clear statement and rationale as to the inclusion or exclusion of children with cognitive or neurological impairment since behavioural problems in the absence/presence of motor, cognitive and neurosensory impairment may be different in nature and aetiology.

5.4 Overall conclusions of the systematic review

The most robust prognostic factors that emerged from the systematic review of 78 published articles reporting multivariable risk factor models for neurodevelopmental outcome in surviving VPT/VLBW children were neonatal brain injury, male sex and parental education. These were investigated in further detail in Chapters 7 and 8 in relation to cognitive and behavioural trajectories into early adulthood. There were mixed findings and conflicting evidence for most other risk factors examined. This may be due to differences in study design, study population, methodological quality and lack of standardisation of measures. However, it could also be due to the difficulty and complexity of modelling long-term outcome in such a heterogeneous population; with multiple risk factors acting sequentially over time and with the existence of multiple impairments within the same individual. It is possible that different preterm phenotypes exist, for example individuals with isolated cognitive impairment may have a distinct pattern of development compared to individuals with both motor and cognitive problems. Further research is required on identifying these phenotypes before accurate prognostic models can be developed for use in routine clinical practice and research or policy evaluation.

The systematic review also revealed some shortcomings in methodology and reporting that could be improved in future studies, and confirmed that that there is a dearth of properly designed and well-conducted prognostic modelling studies in this field, particularly in older children in which the evidence base is quite weak.

Chapter 6: Longitudinal analysis of the EPICure cohort: methods

6.1 Introduction

The trajectories of neurodevelopment from infancy to early adulthood, and the effect of predictive factors identified in the systematic review on these patterns were investigated using the EPICure study; the largest prospective, population-based cohort of EPT infants followed up to adulthood. This longitudinal analysis explored whether cognitive, behavioural deficits ascertained in childhood were transient or stable, and whether conditions deteriorated or improved as the children reached maturity. Firstly, the change in neurodevelopment in EPT participants over time was compared to the term-born comparison group and the impact of sex and maternal education was investigated. Secondly, neurodevelopmental outcomes within the EPT group were examined in relation to brain injury during the neonatal period and GA at birth. Despite the lack of evidence for GA as a robust predictor in the prognostic models identified by the systematic reviews, it has been shown to be strongly related to neurodevelopmental outcome across the whole spectrum of GA at birth and was included for completeness.[40] In addition, the analysis of behavioural outcomes was stratified by cognitive and functional impairment to determine whether having other disabilities has an impact on behaviour.

In this chapter, the design and follow-up of the EPICure cohort study are described in detail, including the methods of assessment, data collection procedures and flow of participants through the study. The outcomes and explanatory factors included in the longitudinal analysis

are defined here, but further detailed descriptions of cognitive and behavioural assessments are provided in the methods sections of Chapters 7 and 8. The statistical methods used for the longitudinal analysis and the model development process is described, and the study limitations and their possible impact on the analysis are discussed.

6.2 Study population and follow-up

EPICure was one of the first and largest prospective, national, population-based cohort studies to identify and follow up a birth cohort of EPT infants born <26 weeks of gestation. The purpose of the study was to describe the survival, health and development of these infants and monitor their course through childhood into adolescence. Born in 1995, 306 survivors were contacted for follow up at 1, 2.5, 6, 11, 16 and 19 years of age. In the earlier pre-school studies the focus was on identifying disability, neurological problems, developmental delay and growth impairment. As the infants moved into childhood and adolescence, the focus shifted to the measurement of neurocognitive and behavioural outcomes. Figure 6.1 provides a summary of the assessments conducted at each stage of follow-up and Figure 6.2 displays the flow of participants through the study according to whether participants were born EPT or at term.

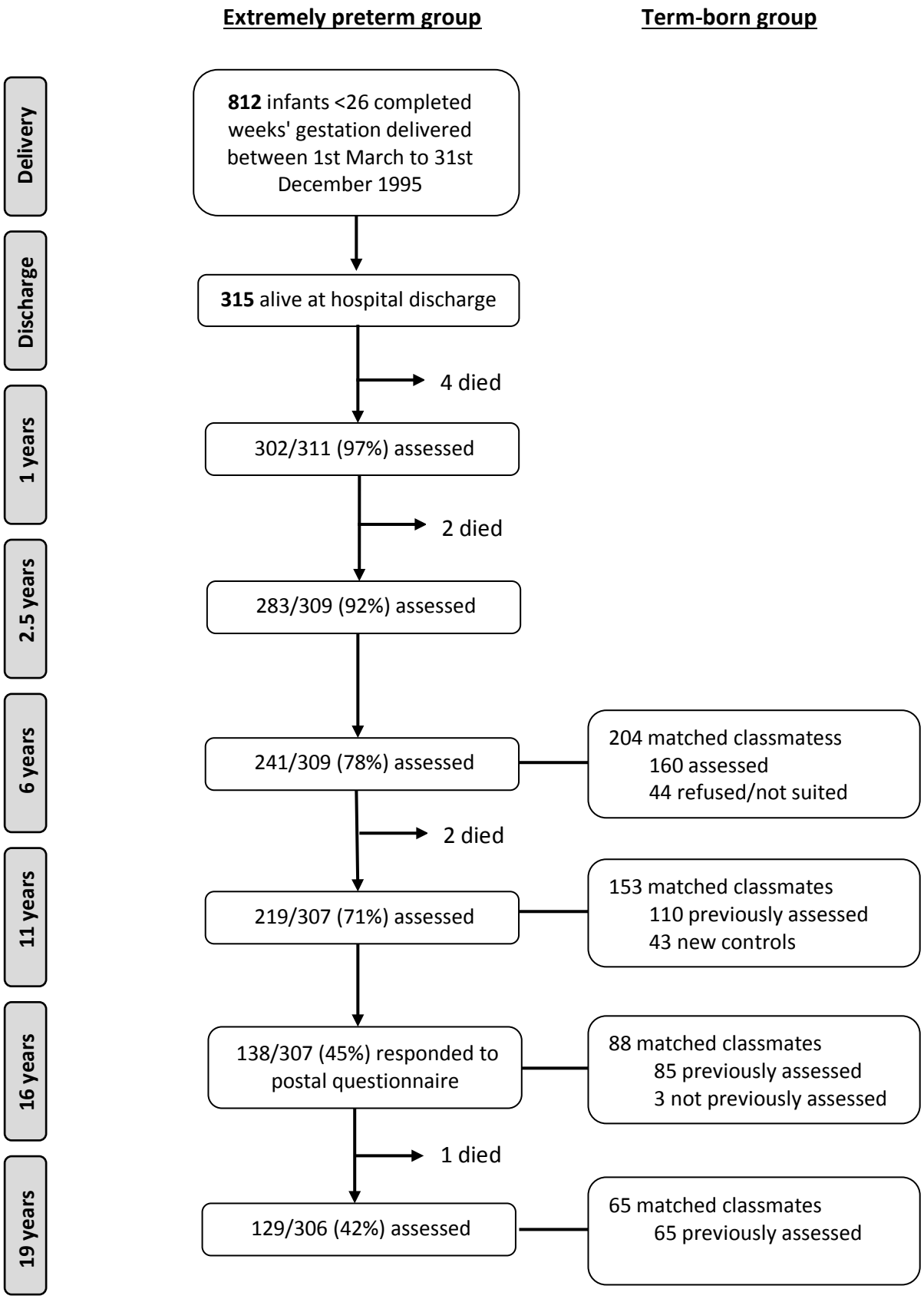
6.2.1 Delivery to 30 month assessment

The original EPICure cohort consists of all live births between 1st March to 31st December 1995 believed to be at 20 to 25 completed weeks of gestation at the time of birth from all 276 maternity units in the UK and Republic of Ireland. . The expected date of delivery was based on the last menstrual period of the mother or on the results of an ultrasound scan performed before 20 weeks' gestation, if only one or the other was available. If both were

Figure 6.1: Data collection schedule in the EPICure cohort at follow-up assessments

Domain	2.5 year (1997-98)	6 year (2001-02)	11 year (2006-07)	16 year (2011-12)	19 year (2014-15)
Functional disability	Paediatric assessment	Paediatric assessment	Paediatric assessment		Clinical assessment
General health	Parent questionnaire	Parent questionnaire	Parent questionnaire	Parent questionnaire	Participant Questionnaire
Health and social service use	Parent questionnaire	Parent questionnaire	Parent questionnaire		Life course interview
Health status	Parent questionnaire	Health Utilities Index III	Health Utilities Index III		Health Utilities Index III
Laterality	Paediatric assessment	Paediatric assessment	Paediatric assessment Edinburgh Handedness Inventory		Edinburgh Handedness Inventory
Motor skills	Bayley Scales of Infant Development II Psychomotor Developmental Index	NEPSY Sensorimotor NEPSY Visuospatial Processing	NEPSY Sensorimotor NEPSY Visuospatial Processing		Beery-Buktenica Test of Visual Motor Integration
Intelligence quotient	Bayley Scales of Infant Development II Mental Developmental Index	Kaufman Assessment Battery for Children (KABC)	Kaufman Assessment Battery for Children (KABC)		Wechsler Abbreviated Scale of Intelligence
Attention and executive function	Tester ratings of child behaviour	NEPSY Attention and Executive Function	NEPSY Attention and Executive Function		Attention Network Test Rapid Automatized Naming Test Verbal fluency task Digit span forward and backward Prospective memory task
Language		Preschool Language Scale-3			Examiner ratings of speech and language
Phonological abilities		Phonological Abilities Test	WIAT-II Non word reading		Test of Word Reading Efficiency
Academic attainment and learning difficulties		Test of Academic Attainment Skills KABC Achievement scale	Wechsler Individual Achievement Test II Reading and Mathematics Scales Magnitude Estimation Test	GCSE results	Test of Word Reading Efficiency Woodstock-Johnson III Test of Achievement Mathematics sub tests Life course interview
Special educational needs (SEN)		Teacher questionnaire	Teacher questionnaire	Parent questionnaire Participant questionnaire	Life course interview
Behaviour screen	Child Behaviour Checklist II	Strengths and Difficulties Questionnaire (parent report)	Strengths and Difficulties Questionnaire (parent report)	Strengths and Difficulties Questionnaire (parent and participant self-report)	Achenbach Youth Self Report
ADHD symptoms		Conner's Rating Scales	Du Paul Rating Scale IV	Du Paul Rating Scale IV	Adult ADHD Self Report Scale
Autism symptoms			Social Communication Questionnaire		Adults Autism Spectrum Quotient Broad Autism, Phenotype Questionnaire
Emotional symptoms					Beck Depression Inventory Anxiety from diagnostic interview
Psychotic symptoms					Psychotic Like Symptoms Inventory
Psychiatric disorders			Development and Wellbeing Assessment		Clinical Interview Schedule
Socio-economic status	Parent questionnaire	Parent questionnaire	Parent questionnaire	Parent questionnaire	Participant questionnaire
Self esteem					Rosenberg Self Esteem Scale Visual Analogue self-esteem scale
Quality of life					World Health Organisation Quality of Life Scale
Employment					Life course interview
Risk taking behaviours					Life course interview
Family, social and sexual relationships					Life course interview

Figure 6.2: Flow of participants through the EPICure study



available and the EDD derived from scans differed from that based on LMP by 14 days, the scan gestation was used. Eight hundred and twelve infants with a confirmed gestation of between 20 to 25 completed weeks were admitted, of which 315 were alive to be discharged home. Comprehensive clinical data were collected on all 812 infants up until death or hospital discharge, and a brief medical assessment was performed on 302 surviving infants at one year. The next comprehensive assessment was conducted at 30 months corrected age in 283 of the 309 infants who were still alive. This consisted of a structured neurological examination[163] and a developmental assessment using the BSID-II,[166] as well as a parent questionnaire collecting family and socio-demographic information.

6.2.2 Six year assessment

Of the 309 children known to have been alive at 30 months, all survived to 6 years. Fifteen were living outside the UK or Ireland, and of the remaining 294 children, 241 (82%) participated in the study at a median age of 6 years and 4 months (range: 5 years and 2 months to 7 years and 3 months). Of the 241 participants, 204 attended classes in a mainstream school, 34 attended a special-needs school, and three were being educated in a special-needs class attached to a mainstream school. For the 204 children attending classes in a mainstream school, three classmates born full-term and matched on age, sex and race were identified by the head teacher, and one was selected at random to participate in the study. A total of 160 children were evaluated as the comparison group and for the remaining 44, the head teachers or parents refused to participate or a suitable child could not be identified. The median age of assessment for the comparison group was 6 years and 2 months (range: 5 years and 1 month to 7 years and 2 months). The 207 EPT children attending mainstream schools were evaluated using a clinical examination including a

neuropsychological and cognitive assessment. Children with disabilities in special-needs schools were evaluated without the use of a comparison child and by means of an appropriate developmental assessment administered by a panel of 15 experienced and formally trained developmental paediatricians and psychologists. A parent and family questionnaire collected information on the home environment, socio-demographic variables and the child's behaviour. Additionally, a teacher questionnaire collected information on behaviour and the child's performance at school. All assessors were blinded to the EPT exposure status of the child and their neonatal course.

6.2.3 Eleven year assessment

Two further children died before the 11 year assessment, leaving 307 EPT survivors at 11 years of age. Eleven (4%) children were living outside the UK and Ireland and the parents of 18 (6%) children refused to participate, and a further 57 (19%) did not respond to study invitations. Thus a total of 219 EPT children were assessed with a median age of 10 years and 11 months (range: 10 years and 1 month to 12 years and 1 month). Of the 160 classmates selected at the six year assessment, 110 (69%) were reassessed at 11 years of age. Where the EPT child was at a different school to the original six year classmate comparator, or where the six year classmate declined to participate, a new classmate was selected at 11 years using the same method as before. This resulted in 43 new classmates joining the comparison cohort, giving a total of 153 classmates with median age of 10 years and 11 months (range: 9 years and 9 months to 12 years and 3 months). A similar battery of tests to the six year assessment were conducted at 11 years (Figure 6.1), with assessors blinded to EPT exposure status and neonatal course.

6.2.4 Sixteen year assessment

The assessment at 16 years was brief and questionnaire based. A similar parent questionnaire to previous years was sent out, collecting demographic and behavioural data, and in addition a self-report questionnaire on behaviour was sent to participants. The response rate for the parent questionnaire was 138 (45%) for the 307 surviving EPT participants and 87 (57%) for 153 classmate comparators, and the response rates for the participant questionnaire were 143 (47%) and 88 (58%) respectively.

6.2.5 Nineteen year assessment

One more child died before their 19th birthday, leaving 306 survivors at 19 years of age. A total of 127 (42%) EPT participants were assessed with a median age of 19 years and 4 months (range: 18 years and 5 months to 20 years and 6 months). Of the 153 classmates assessed at 11 years, 65 (42%) were reassessed at 19 years of age with a median age of 19 years and 2 months (range: 18 years and 1 month to 20 years and 1 month). The most thorough assessment was conducted at age 19; in addition to the usual neuromotor examination, cognitive and behaviour tests, more detailed tests were administered for attention and executive function, ASD, ADHD and other psychiatric symptoms and disorders. Additionally, interviews about life course such as employment, risk taking behaviours and family, social and sexual relationships were conducted.

6.3 Outcomes and explanatory variables

6.3.1 Outcomes

Key outcome data collected on cognition and behaviour were analysed separately (Table 6.1); measures of assessment will be described in detail in the subsequent chapters. It was

intended to analyse motor outcomes, however this was not possible as there was insufficient consistency in the assessments of motor function over time, both with respect to the aspects of motor function assessed and the scoring systems used.

Table 6.1 Cognitive and behavioural assessments used for the longitudinal analysis of the EPICure cohort

Year	Cognition	Behaviour (parent questionnaire)
2	BSID-II MDI	CBCL
6	KABC MPC	SDQ (parent and teacher)
10	KABC MPC	SDQ (parent and teacher)
16	-	SDQ (parent and self)
19	WASI FSIQ	SDQ (parent and self)

6.3.2 Explanatory variables

6.3.2.1 Participant sex

As determined at delivery, there were 160/315 (51%) males in the EPT group and 118/206 (57%) males in the comparison group.

6.3.2.2 Maternal qualifications

Maternal education was based on the highest qualification recorded for the mother over the study period and classified into two groups: up to O'level (or equivalent) versus A'level (or equivalent) and above. In the EPT group, 99/274 (36%) of mothers had an A'level qualification or above and 75/184 (41%) in the comparison group.

6.3.2.3 Brain injury diagnosed in the neonatal period

EPT participants were classified into five groups according to the severity of brain injury diagnosed during the neonatal period recorded on the worst scan (Table 6.2). For the analysis participants with moderate to severe brain injury in groups 3-5 were compared to participants with no or mild brain injury in groups 1-2. One infant was not classified.

Table 6.2: Severity classification of neonatal brain injury in EPT participants according to the worst scan

Severity	Group	Definition	Number (%)
None/mild	1	No pathology, haemorrhage or dilatation, or evidence of parenchymal injury or parenchymal haemorrhage.	103 (32.8)
	2	Subependymal/choroidal/intraventricular haemorrhage or ventricular size ≤4mm over 97th centile, and no parenchymal cysts or parenchymal haemorrhage.	140 (44.6)
Moderate/severe	3	Ventricular size >4mm over 97th centile but no parenchymal haemorrhage, cysts or cystic leukomalacia.	14 (4.5)
	4	Unilateral parenchymal problems (left or right), haemorrhage, cysts or cystic leukomalacia.	42 (13.4)
	5	Bilateral parenchymal lesions	15 (4.8)

6.3.3 Gestational age

GA was based on the last menstrual period of the mother or the results of an ultrasound scan before 20 weeks'. If both were available and the measures differed by more than 14 days then the scan gestation was used. Among the 315 EPT infants who survived to discharge: two (0.6%) were born at 22 weeks GA; 27 (8.6%) at 23 weeks; 100 (31.8%) at 24 weeks; and 25

(59.1%) at 25 weeks. For the analysis, GA was dichotomised into two groups; 25 weeks GA and <25 weeks GA.

6.3.4 Functional impairment

Participants were classified into groups based on their functional impairment, as reported at their last assessment using guidelines developed by a working group in the 1990s (Table 6.3).[222]

Table 6.3: Severity classification of functional impairment in EPT participants at last assessment

Severity	Definition	Number (%)
None	An IQ no more than 1 SD below the mean and no functional disabilities, reduced vision, hearing aids or psychiatric diagnosis.	45 (15.4)
Mild	One or more of the following: abnormal neurological signs with minimal functional consequences, an IQ score 1 to 2 SD below the mean, mild hearing loss not sufficient to require aids, squints or refractive errors.	97 (33.2)
Moderate	One or more of the following: ambulant CP with moderate impairment, an IQ score 2 to 3 SD below the mean, sensorineural hearing loss improved by aids, functionally impaired vision without blindness.	78 (26.7)
Severe	A disability causing the participant to be heavily dependent on caregivers, and may include one or more of the following: nonambulant CP, an IQ score more than 3 SD below the mean, profound sensorineural hearing loss not improved by aids, or blindness.	72 (24.7)

In the analysis of behavioural outcomes, cognitive impairment was dichotomised into two groups: moderate to severe (IQ score more than 2 SD below the mean) and none to mild (IQ score no more than 2 SD below the mean), and functional impairment was dichotomised and none to mild.

6.4 Statistical analysis

6.4.1 General approach

Hierarchical mixed modelling was used to examine the trajectories of neurodevelopment from early childhood to adulthood using Stata/SE 13.1 for Windows. The advantages of this approach compared with cross-sectional analysis techniques that simply compare group averages at each time point are that it:

- Incorporates within-individual correlation across repeated assessments and is a sensitive method for assessing change as well as group effects;
- Is possible to test for different patterns of growth (intercept, slope and curvature) in different subgroups;
- Uses maximum likelihood methods which allow individuals with incomplete longitudinal data to be included in the model and to contribute to estimation of model effects;
- Allows for unequally spaced measurement intervals, i.e. different ages at assessment;
- Allows the variance around the group mean to be fitted as a random effect and can test whether the variation in development between individuals widens or narrows over time.

Separate multilevel models were fitted for cognition and behaviour. Unique 'development' curves were constructed for each infant from infancy to early adulthood (age 19) and used to examine average patterns of development over time, taking into account the repeated measurements within each individual. Within-infant and between-infant variability was explored, and explanatory factors were added to determine which characteristics affect variations in development at different stages of life. Group mean developmental trajectories

were plotted with confidence intervals for each outcome, along with the level of individual variation over time. Plots were stratified by any explanatory factors found to be strongly prognostic. A two-sided 5% level of significance was used throughout for reporting confidence intervals.

6.4.2 Descriptive analysis

For each outcome, the distribution of data was examined by plotting histograms by exposure status at each time point along with summary statistics, including the mean [SD] with 95% confidence intervals and the median {IQR} with the range. A plot of outcome trajectories over time by exposure status were inspected for patterns in longitudinal group change. Group mean differences between EPT and comparison participants and 95% confidence intervals were calculated for each time point. These descriptive analyses were repeated, stratified by each pre-specified categorical explanatory variable.

6.4.3 Comparative analysis of outcome trajectories by exposure status

Throughout the analysis, the participants were defined as level two in the hierarchical data structure, and the repeated measures assessed on different occasions as level one. Level two variation represents between-subject variability (variation between individuals), and level one variation represents within-subject variability (variation between measurements within the same individual). A term in the model is described as a fixed effect if level one and level two variation is assumed to be constant among individuals and a random effect if individuals are allowed to vary about the overall group mean estimate. At each step of developing the model, the improvement of fit by adding an additional term(s) to the model, was assessed using the likelihood ratio test (LRT) for the model as a whole and the Wald test for individual coefficients.

6.4.3.1 Linear growth model with age as a fixed effect

The most basic model for each outcome variable assumes that outcome is a linear function of age. This model assumes that within-subject (level one) and between-subject (level two) variation is constant over time and fits a straight line with a common intercept α and fixed slope. This model can be expressed as:

$$y_{ij} = (\beta_0 + \mu_j) + \beta_1 Age_{ij} + e_{ij}$$

Fixed effects: $\beta_0 + \beta_1 Age_{ij}$ Random effects: $\mu_j + e_{ij}$

$$\mu_j \sim N(0, \sigma^2_\mu) \text{ and } e_{ij} \sim N(0, \sigma^2_e)$$

where the response variable y_{ij} denotes the score of individual j (j=1, 2 ... n) measured at occasion i (i=1, 2 ... m). The age term specifies the age of individual j at occasion i and the coefficient β_1 represents the increase (or decrease) in the mean score with year of age. The term μ_j represents the between-subject variability at level two, i.e. the difference of the jth individual from the overall mean line given by the fixed part of the model. The term e_{ij} represents the within-subject variability at level one, i.e. the departure of the jth individual's actual score from the predicted score at occasion i. It is assumed that random effects μ_j and e_{ij} are independent and Normally distributed with a mean of zero and a variance of σ^2_μ and σ^2_e respectively. The similarity between individuals is measured by the intraclass correlation coefficent:

$$\frac{\sigma_\mu^2}{\sigma_\mu^2 + \sigma_e^2}$$

This is the proportion of variation that is due to differences between individuals.

6.4.3.2 Linear growth model with age as a random effect

When age is fitted as a random effect, this allows the between-subject (level two) variance to become a function of age and allows for the correlation between an individual's score at different ages. This model can be expressed as:

$$y_{ij} = (\beta_0 + \mu_{0j}) + (\beta_1 + \mu_{1j})\, Age_{ij} + e_{ij}$$

Fixed effects: $\beta_0 + \beta_1 Age_{ij}$ Random effects: $\mu_{0j} + \mu_{1j} Age_{ij} + e_{ij}$

$$\begin{bmatrix}\mu_{0j} \\ \mu_{1j}\end{bmatrix} \sim N(0, \sigma^2_{\mu 0} + 2\,\sigma_{\mu 01} + \sigma^2_{\mu 1}) \text{ and } e_{ij} \sim N(0, \sigma^2_{e})$$

The parameter μ_{1j} represents the departure of the jth individual from the slope of the overall mean growth line β, allowing the rate of growth as well as the average level of growth to vary between individuals. The variance of μ_{1j} is denoted by $\sigma^2_{\mu 1}$ and the covariance of μ_{0j} and μ_{1j} by $\sigma_{01.}$ The variance function for between-individual variation can be calculated as follows:

$$var(\mu_{0j} + \mu_{1j} Age_{ij}) = var(\mu_{0j}) + 2cov(\mu_{0j}, \mu_{1j} Age_{ij}) + var(\mu_{1j} Age_{ij})$$

$$= var(\mu_{0j}) + 2cov(\mu_{0j}, \mu_{1j}) Age_{ij} + var(\mu_{1j}) Age^2_{ij}$$

$$= \sigma^2_{\mu 0} + 2\,\sigma_{\mu 01} Age_{ij} + \sigma^2_{\mu 1} Age^2_{ij}$$

6.4.3.3 Addition of EPT exposure status

A group term was added as a fixed effect specifying whether individual j was born EPT (EPT=1) or at term (EPT=0):

$$y_{ij} = (\beta_0 + \mu_{0j}) + (\beta_1 + \mu_{1j})\, Age_{ij} + \beta_2 EPT + e_{ij}$$

The coefficient β_2 represents the difference in the mean score of EPT participants compared to the term-born comparison group. This model fits a separate line for both groups, allowing

the intercept to vary but with the same slope. An interaction term between age and group can be added to test whether EPT and term-born participants vary on slope:

$$y_{ij} = (\beta_0 + \mu_{0j}) + (\beta_1 + \mu_{1j})\, Age_{ij} + \beta_2 EPT + \beta_3 Age_{ij} * EPT + e_{ij}$$

6.4.3.4 Non-linear growth model

A constant rate of development may not be likely in reality and this can be incorporated into the model by including non-linear functions of age, such as quadratic, cubic terms and other polynomials which allows for curvature in the fitted line. Since individuals in this study only have a maximum of four data points for cognitive and behavioural outcomes, with most classmates only having two or three data points, only a quadratic function of age was considered as a fixed effect. Introducing higher order non-linear functions would be likely to result in overfitting the data. It is possible to model level two variation as a non-linear function of age, allowing the slope of each individual's curve to vary about the overall mean development curve. However, due to small number of data points, and size of the sample, this additional complexity was not introduced to the model. Adding a quadratic function of age as a fixed effect, results in the model:

$$y_{ij} = (\beta_0 + \mu_{0j}) + (\beta_1 + \mu_{1j})\, Age_{ij} + \beta_2 EPT + \beta_3 Age_{ij} * EPT + \beta_4 Age^2_{ij} + e_{ij}$$

$$\text{Fixed effects: } \beta_0 + \beta_1 Age_{ij} + \beta_2 EPT + \beta_3 Age_{ij} * EPT + \beta_4 Age^2_{ij}$$

$$\text{Random effects: } \mu_{0j} + \mu_{1j} Age_{ij} + e_{ij}$$

$$\begin{bmatrix} \mu_{0j} \\ \mu_{1j} \end{bmatrix} \sim N(0, \sigma^2_{\mu 0} + 2\,\sigma_{\mu 01} + \sigma^2_{\mu 1}) \text{ and } e_{ij} \sim N(0, \sigma^2_e)$$

where the additional coefficient β_4 represents the increase (or decrease) in the mean score with year of age squared. Where a quadratic function of age is found to be significant, then

an interaction term between quadratic age and group was fitted to test whether EPT and term-born participants vary in the curvature of the fitted line.

6.4.3.5 Different Level 1 (within-individual) variance for EPT and comparison group

So far, all models described have assumed that the level one (within-individual) variance is constant for all participants, but it may be different within the EPT group compared to the term-born group. This can be tested by fitting group-specific parameters e_{0ij} and e_{1ij} for the EPT and comparison group, and comparing it to the model which assumes a common level one variance:

$$y_{ij} = (\beta_0 + \mu_{0j}) + (\beta_1 + \mu_{1j})\, Age_{ij} + (\beta_2 + e_{0ij})EPT + e_{1ij}Term$$

$$\text{Fixed effects: } \beta_0 + \beta_1 Age_{ij} + \beta_2 EPT$$

$$\text{Random effects: } \mu_{0j} + \mu_{1j}Age_{ij} + e_{0ij} \text{ for EPT, } \mu_{0j} + \mu_{1j}Age_{ij} + e_{1ij} \text{ for Term}$$

$$\begin{bmatrix}\mu_{0j} \\ \mu_{1j}\end{bmatrix} \sim N(0, \sigma^2_{\mu 0} + 2\,\sigma_{\mu 01} + \sigma^2_{\mu 1}) \text{ and } \begin{bmatrix}e_{0j} \\ e_{1j}\end{bmatrix} \sim N(0, \sigma^2_{e0} + \sigma^2_{e1})$$

This model includes a dummy variable for the comparison group, where Term=0 for EPT participants and Term=1 for term-born classmates. Note that the covariance between e_{0ij} and e_{1ij} is constrained to zero as this can only exist if some participants belong to both groups, which is not possible.

6.4.3.6 Different Level 2 (between-subject) variance for EPT and comparison group

It is also possible to allow the level-2 (between-subject) variance to vary between the EPT and comparison group. In this model, separate parameters were fitted for the random intercept variance, random-slope variance and their covariance for each group:

$$\text{Term: } y_{ij} = (\beta_0 + \mu_{00j}) + (\beta_1 + \mu_{01j})Age_{ij} + e_{1ij}$$

$$\text{EPT: } y_{ij} = (\beta_0 + \mu_{10j}) + (\beta_1 + \mu_{11j})\text{Age}_{ij} + \beta_2 + e_{0ij}$$

For the final model of age and exposure status, the variance functions for EPT participants and comparison group were plotted against age to examine the level and change in variation in each group over time. Residual and diagnostic plots were produced to assess the overall goodness of fit.

6.4.3.7 Addition of explanatory factors

The impact of participant sex on the final model of age and exposure status was examined by adding a fixed main effect ($\beta_5\text{Male}_{ij}$) and interaction effect with group ($\beta_6\text{Male*EPT}_{ij}$). This process was repeated for maternal education. The effect of brain injury diagnosed in the neonatal period and GA on outcome trajectories within the EPT group was tested on the basic model in 6.4.3.2, with the omission of the parameter for group.

6.4.3.8 Model fitting and covariate selection

The likelihood ratio test was used to assess the fit of one model compared to another. This is -2In(maximum likelihood in model 1/maximum likelihood model 2) on *k* degrees of freedom, where model 1 is the nested (null) model and model 2 has *k* additional parameter estimates. A two-sided p-value of <0.05 was considered as significant.

6.4.3.9 Analysis of categorical behavioural scores

The SDQ scores used for the analysis of behavioural trajectories can be classified as being in the normal and abnormal range (see Chapter 8). The risk ratio of the Total Difficulties Score, subscale scores and the Impact Score being in the abnormal range in EPT participants compared to the term-born group was calculated using random effects Poisson regression model (with fixed exposure time).

6.4.4 Multiple imputation of cognitive outcomes

For some children who attended for assessment, it was not possible to administer the planned cognitive test on the day, either due to the child's behaviour or because they were too cognitively impaired to be tested. However, it was possible to estimate the category of cognitive ability according to standard deviation from the mean (e.g. 70-84 (-2SD), 55-69 (-3SD), 40-54 (-4SD) and so on) either based on clinical observation or an alternative developmental test. For these children, values in the estimated IQ category were imputed using multiple imputation (MI). The following procedure was used to implement the MI analysis:

- Participants with missing IQ score were allocated to one of the following groups, based on their clinical assessment and/or other developmental tests: 85-114, 70-84, 55-69, 40-54 and 25-39;
- 10 sets of imputed values for the missing data points were generated randomly in the assigned range, thus creating 10 completed datasets;
- For each of the completed datasets 10 estimates for the parameters of interest were calculated, e.g. mean score, regression coefficient;
- The results from the 10 completed datasets were combined to obtain a single estimate for the parameters of interest, with standard errors that took into account the variability in results between the 10 imputed datasets.

One limitation with MI is that statistics which are not estimators, such as p-values and test statistics, cannot be combined in the usual way using Rubin's rules.[287] This makes model building and model checking difficult, as such statistics can only be generated separately for each MI dataset and vary slightly depending on the extent and nature of the missing data.

Various strategies have been suggested for overcoming the problem of variable selection;[288] the strategy used in this analysis was to assess model fit in each separate MI dataset using the LRT. For a parameter to be retained in the model, it was required to have a p-value<0.05 in the LRT in all 10 MI datasets. Predictive and diagnostic plots were generated from the first MI dataset as it is not possible to generate these from the combined MI model.

6.4.5 Missing data analysis due to drop out

The number of participants assessed for outcome was reported at each date of assessment along with the completeness of risk factor data. Infants were classified according to their pattern of missing data and selected baseline characteristics, other important outcomes and risk factor variables were tabulated by dropout status. Graphical displays were used to investigate the missing data mechanism, for example by plotting outcome against time of assessment by dropout status.

The hierarchical mixed modelling approach uses the method of maximum likelihood, which has the advantage of allowing incomplete longitudinal follow-up data to be considered in the estimation of developmental trajectories. This method of analysis fits a line through existing data points for each individual; so if there were two follow-up visits then the estimates for a straight line can be derived; for three data points a non-linear monotonic curve can be fitted; and with four or more data points, a non-linear and non-monotonic curve can be derived. An additional advantage is that the data points do not have to be equally spaced or consecutive, so an individual who attended the 30 month assessment at age three, missed the six and ten year assessment, but reappeared at 19 years can still be included in the analysis. Mixed-effects models for longitudinal data with maximum-likelihood estimations provide valid inferences in the presence of ignorable non-response, meaning that the reason for non-

response is related to either observed covariates or previous observed values of the dependent variable.[289]

Analyses were first conducted in all participants with data available at any time point, and then restricted to those with complete follow-up assessments. Imputed scores were counted as non-missing for all analyses because excluding the scores of participants who came for clinical assessment but who were unable to complete the planned cognitive tests would have increased the level of bias due to missing data.

Chapter 7: Longitudinal Analysis of Cognitive Outcome

7.1 Introduction

Low IQ scores in childhood are associated with reduced survival and poorer health later in life.[290-292] Studies show that cognitive ability remains relatively stable from middle childhood onward in the general population.[293,294] Recent evidence from VPT/VLBW cohorts suggests that deficits in cognitive function and academic attainment persist into early adulthood.[295-297] and that developmental scores from as early as 2 years remain stable into adulthood.[298] Little is known about the maturation of cognition over childhood and into adulthood for individual EPT survivors. Failure to catch up by early adulthood raises concerns about the future trajectory of cognitive function in later adult life.

A recent review of studies investigating longitudinal cognitive development in VP/VLBW survivors yielded mixed findings.[299] Some studies reported a decline in cognitive function over time,[300-304] some suggested stability,[50,305] and others reported improvement over time relative to term-born individuals.[48,306,307] However, much of this evidence has come from cross-sectional analyses of longitudinal data which may explain the inconsistent findings. There may be considerable variation in individual trajectories which is not detectable using such analytic methods.[38] The few studies that have adopted a longitudinal modelling approach are characterized by several major shortfalls, including: failure to enroll an appropriate longitudinal comparison group, selective drop out, small sample sizes and short periods of follow-up.[48-50]

This chapter reports the results from a longitudinal analysis of cognitive trajectory from infancy to early adulthood in the EPICure study; the largest prospective, population-based cohort of EPT births.[65,308] Changes in cognitive development in EPT participants were compared to those of a term-born comparison group. The effect of sex and maternal education on these trajectories was examined, the two main prognostic factors for cognitive impairment in VP/VLBW children,[309] and within the EPT group, the effect of GA and brain injury diagnosed in the neonatal period, which are also strongly related to neurodevelopmental outcome.[40,207,208,210]

7.2 Cognitive assessments

7.2.1 Bayley Scales of Infant Development II - Mental Development Index

The BSID-II,[166] is a comprehensive, rigorously validated test which is widely used as a standardised assessment of key developmental milestones in infants up to 42 months. It is a highly structured, individually administered test with a stringent protocol and requires examiners to be highly qualified with at least a degree or professional training. The BSID-II assesses developmental functioning in three domains using the following scales: Mental Development Index (MDI), Psychomotor Development Index (PDI) and the Behaviour Rating Scale (BRS). The MDI includes items that assess memory, habituation, problem-solving, early number concepts, generalisation, classification, vocalisations, language and social skills.

The BSID-II consists of item sets, ordered by their increasing difficulty, with the entry level depending on the infant's chronological age. The test takes between 25 and 60 minutes to administer and counts the cumulative (rather than consecutive) number of passes and failures within each item set, to establish the basal and ceiling limits for the child. Raw scores are converted into age-adjusted index scores (mean 100, SD 15) and developmental ages for both

MDI and PDI. Index scores of 85 and 115 correspond to 1 SD below and above the mean, scores of 70 and 130 are 2 SDs from the mean, and scores of 55 and 145 are 3 SDs from the mean based on the normative sample. Based on the Normal distribution, nearly all (99.9%) children should obtain scores within 3 SDs of the mean, and 95% within 2 SDs of the mean.

The BSID-II was updated and revised in 1993, and the normative sample was based on 1,700 US infants and sampling was stratified by age, sex, race/ethnicity, geographical region and parental educational level (based on the US Census 1988). Norms are provided in one month intervals for children aged between one to 36 months. However, normative data were not collected for children with disabilities. Assessing populations with impairment or developmental delay can be problematic because the test requires entry based on chronological age which can result in administering a larger number of items drawn from beyond the starting item set, thus increasing the number of necessary items and failure rate. The manual suggests using corrected age to determine both the entry point and corresponding norm table for children born preterm. The lowest index score is 50 and the BSID-II is not recommended for testing children with serious disabilities as developmental level cannot be quantified. In such cases, a nominal value can be assigned to reflect a level of functioning >3SD below the mean, or extrapolated norms for index scores below 50 can be used.[310]

7.2.2 Kaufman Assessment Battery for Children - Mental Processing Composite

The Kaufman Assessment Battery for Children (KABC)[237] is an individually administered test that measures the processing and cognitive abilities of children aged 2.5 to 12.5 years. It was one of the first intelligence tests to be developed from neuropsychological theory and its purpose was to evaluate pre-schoolers, minority groups and children with learning disabilities. The test takes between 35 and 85 minutes to complete and consists of four global test scores: sequential

processing; simultaneous processing; achievement; and mental processing composite (MPC) which is the combined score from the sequential and simultaneous processing scales. The primary aim of the sequential processing scale is to evaluate short term memory and comprises subtests that focus on problem solving skills, arranging items in sequential order and ability to recall. The simultaneous processing scale evaluates problem solving skills that involve several processes at once, for example facial recognition, selecting a picture that completes or is similar to another picture, arranging pictures in a meaningful order. The MPC represents a global estimate of a child's level of intellectual functioning and ability to solve unfamiliar problems simultaneously and sequentially.

Test norms were developed on a large sample of US children of the same age and sampling was stratified by age, sex, race/ethnicity, geographic region, community size, parental education and educational placement (normal versus special class). The MPC is standardised to have a mean of 100 and SD of 15 with the lowest score 25, and most children (99.9%) score between 55 and 145.

7.2.3 Wechsler Abbreviated Scale of Intelligence - Full Scale IQ

The Wechsler Abbreviated Scale of Intelligence (WASI-II)[311] is a short form version of the Wechsler Adult Intelligence Scale - Third Edition (WAIS-IV)[312] and was developed to provide quick, reliable method of measuring intellectual functioning. The four subtest version can be administered in 30 minutes and yields the Verbal Comprehension Index (VCI), Perceptual Reasoning Index (PRI) and Full Scale IQ scores (FSIQ), which is the sum of the VCI and PRI. The VCI consists of the Vocabulary subtest which measures word knowledge, verbal concept formation and fund of knowledge, and the Similarities subtest which evaluates verbal reasoning and concept formation. The PRI comprises of the Matrix Reasoning subtest which measures visual information processing and abstract reasoning skills, and the Block Design subtest which

measures the ability to analyse and synthesize visual stimuli, nonverbal concept formation, and visual-motor coordination. Subtest raw scores are converted to T-scores from which the VCI, PRI and FSIQ composite scores can be generated; all with a mean of 100 and SD of 15 and a lowest score of 25. The normative data were based on a sample of 2,300 nationally representative US children and adults aged 6 to 90 years, with appropriate representation from special groups with cognitive impairment.

In the longitudinal analysis reported here, a comparison group of term-born participants was used as reference data where possible, rather than historic normative data, to address the issue of secular trends in IQ over time.[313,314] The upward drift of standardised scores over time, known as the 'Flynn effect', may result in the overestimation of developmental level if older normative data are used as the reference.

7.3 Multiple imputation

At each assessment, a number of participants did not complete the standard cognitive tests, but were either clinically assessed or assessed using another developmental test. Where possible, these participants were assigned to a category of cognitive ability according to standard deviation from the mean. At the 2.5 year assessment, there were 34 EPT participants who did not complete the Bayley scales and a further 18 EPT participants with raw scores too low to be assigned an MDI index score which has a floor value of 50. For these 18 participants, values were imputed in the range 40 to 54 (-4SD below the mean). At the 6 year assessment there were 41 EPT participants and 1 term-born participant with no KABC MPC score and 18 EPT participants in year 11, for whom values were imputed. At 19 years there were 3 EPT participants who did not complete the WASI for whom FSIQ scores were imputed. Table 7.1 provides a breakdown of the number of

participants with imputed values at each assessment, with the category of cognitive ability they were assigned to.

Table 7.1: Summary of imputed cognitive test scores by age of assessment

Imputed range	n (%) of EPT participants with imputed values by assessment year			
	2.5 (n=283)	**6** (n=241)	**11** (n=217)	**19** (n=127)
85-114	2	2[a]	2	-
70-84	5	5	-	-
55-69	10	9	-	-
40-54	35	7	-	-
25-39	-	19	16	3
Total	52 (18.4)	42 (17.4)	18 (8.3)	3 (2.4)

[a] One term-born participant also had imputed values in the range 85-114 at age 6 years.

7.4 Missing data analysis due to dropout

The patterns of missing cognitive assessments at 2.5, 6, 11 and 19 years are shown in Table 7.2 for EPT participants. Assessments were counted as complete if a cognitive test score was imputed due to participants not completing the set cognitive test, based on clinical judgement during the assessment and/or an alternative developmental test. Assessments were defined as missing if the participant was not seen and no cognitive test score could be assigned. EPT participants were classified into two groups according to their pattern of missing assessments; completers (all four assessments) and non-completers (three or less assessments). Thirty-six percent (n=114) of EPT participants had a complete set of four cognitive test scores, 29.2% (n=92) had three scores, 27.3% (n=86) had one or two scores and 7.3% (n=23) had none. Maternal and infant characteristics collected at baseline, family SES and infant developmental outcomes at 2.5 years, mothers' highest educational qualification and neurodevelopmental

Table 7.2: Pattern of missing IQ scores among EPT participants (n=315)

Number of assessments	Assessment (year)					
	2.5	6	11	19	n	(%)
Four	+	+	+	+	114	(36.2)
Three	+	+	+		84	(26.7)
	+	+		+	5	(1.6)
	+		+	+	2	(0.6)
		+	+	+	1	(0.3)
Two	+	+			33	(10.5)
	+		+		11	(3.5)
	+			+	3	(1.0)
			+	+	2	(0.6)
		+	+		1	(0.3)
One	+				31	(9.8)
		+			3	(1.0)
			+		2	(0.6)
None					23	(7.3)

impairment at last assessment were compared between the two groups. Two-sided p-values were calculated using Fisher's Exact Test for binomial variables and the t-test for continuous variables (Table 7.3). Participants with complete follow-up assessments were more likely to be from a multiple pregnancy, have mothers who were older, of white ethnicity and better educated, and have fathers with a non-manual occupation. They also had higher BSID-II MDI and PDI scores at 2.5 years and less likely to have moderate to severe CP.

Table 7.3: Perinatal and neurodevelopmental characteristics of EPT participants according to completeness of cognitive assessments

	Completers: 4 assessments (n=114)	Non-completers <4 assessments (n=201)	P value[a]
Maternal/paternal factors			
Maternal age (years), mean [SD]:	29.4 [5.4] (n=113)	27.9 [6.1] (n=200)	0.03
Mother of non-white ethnicity, % (n):	15.9% (18/113)	27.4% (55/201)	0.03
Primigravida, % (n):	35.1% (40/114)	27.5% (55/200)	0.16
Mother smoked during pregnancy, % (n):	27.4% (31/113)	36.9% (62/168)	0.12
Antenatal steroids given, % (n):	81.6% (93/114)	77.4% (154/199)	0.47
Mother's highest educational qualification A 'level or above, % (n):[b]	49.6% (55/111)	27.0% (44/163)	<0.001
Father's occupation non-manual, % (n):[b]	51.0% (50/98)	26.5% (35/132)	<0.001
Infant perinatal factors			
Multiple birth, % (n):	32.5% (37/114)	21.0% (42/200)	0.03
Congenital anomaly present, % (n):	1.8% (2/114)	3.0% (6/200)	0.72
Gestational age 24 weeks or less, % (n):	43.0% (49/114)	39.8% (80/201)	0.63
Birthweight (grams), mean [SD]:	745 [127] (n=114)	749 [107] (n=201)	0.76
Male sex, % (n):	45.6% (52/114)	51.2% (103/201)	0.35
Moderate/Severe brain injury during neonatal period, % (n):[c]	18.4% (21/114)	25.0% (50/200)	0.21
Laser or cryotherapy for retinopathy of prematurity, % (n):	14.3% (16/112)	14.7% (29/198)	1.00
Laparotomy for necrotising enterocolitis, % (n):	1.8% (2/111)	3.6% (7/197)	0.50
Supplemental oxygen at 36 weeks, % (n):	71.1% (81/114)	75.6% (152/201)	0.42
Infant developmental outcomes at 2.5 years			
BSID-II Mental Developmental Index, mean [SD]:	82.7 [14.4] (n=114)	76.1 [17.2] (n=169)	0.001
BSID-II Psychomotor Developmental Index, , mean [SD]:	83.9 [17.4] (n=108)	79.0 [18.6] (n=154)	0.03
BSID-II Behaviour Rating Percentile, , mean [SD]:	36.4 [26.6] (n=108)	35.7 [24.5] (n=151)	0.82
CBCL Total Problem T-score, mean [SD]:	55.6 [9.4] (n=112)	55.8 [9.9] (n=158)	0.89
Neurodevelopmental impairment at last assessment[b]			
Cerebral palsy, % (n):	8.8% (10/114)	23.8% (46/193)	0.001
Moderate/severe cerebral palsy, % (n):	6.1% (7/114)	14.7% (28/190)	0.03
Moderate/severe cognitive impairment, % (n):	45.6% (52/114)	49.4% (88/178)	0.55
Moderate/severe visual impairment, % (n):	7.9% (9/114)	12.5% (24/192)	0.26
Moderate/severe hearing impairment, % (n):	0.9% (1/114)	2.6% (5/193)	0.42
Moderate/severe functional impairment, % (n):[d]	47.4% (54/114)	53.9% (96/178)	0.28
Severe functional impairment, % (n):[d]	20.2% (23/114)	27.5% (49/178)	0.17

[a] Two-sided p-values were calculated using Fisher's Exact Test for binomial variables and the t-test for continuous variables.

[b] Not collected at discharge; first collected at one or two year assessment.

[c] Parenchymal pathology and/or ventriculomegaly on worst cranial ultrasound scan before discharge home.

[d] Functional impairment includes cerebral palsy, cognitive, visual or hearing impairment.

The patterns of missing assessments at 6, 11 and 19 years are shown in Table 7.4 for the term-born classmates recruited at 6 years. Classmates were classified according to their pattern of missing assessments in a similar way as for EPT participants: twenty-six (n=53) had three cognitive assessments, 72.8% (n=150) had one or two and 1.5% (n=3) had none. There were a limited number of factors that could be compared for the comparison group as, unlike the EPT group, comprehensive data were not collected at baseline or 2.5 years. Infant sex, mother's highest educational qualification, and neurodevelopmental impairment were compared between the completers versus non-completers in the comparison group using the same methods as for the EPT participants (Table 7.5). Except for visual impairment at last assessment, there were no statistically significant differences between classmates with complete cognitive assessments and those with fewer visits, however given the small number of participants with impairments and the limited amount of data to make comparisons, it is not possible to rule out that differences may actually exist. The excess in the number of cases of visual impairment (and any functional impairment) in the complete group was due mainly to prescriptions for glasses recorded at the 19 year assessment, which only 11 of the non-completers attended, so this is likely to be a spurious finding.

Table 7.4: Pattern of missing assessments among term-born classmates (n=206)

Number of assessments	Assessment year				
	6	11	19	n	(%)
Three	+	+	+	53	(25.7)
	+	+		57	(27.7)
Two		+	+	11	(5.3)
	+			50	(24.3)
One		+		32	(15.5)
None				3	(1.5)

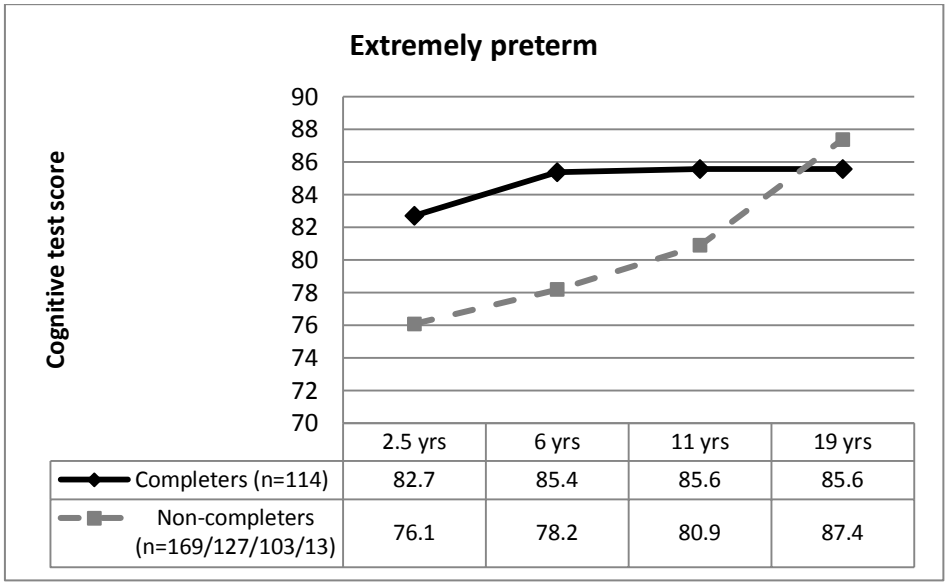
Table 7.5: Characteristics of term-born classmates according to completeness of follow-up assessment

	Completers: 3 assessments (n=53)	Non-completers <3 assessments (n=153)	P value[a]
Mother's highest educational qualification A 'level or above, % (n):	48.1% (25/52)	37.9% (50/132)	0.24
Male sex, % (n):	39.6% (21/53)	43.8% (67/153)	0.63
Any cognitive impairment at last assessment, % (n)	17.0% (9/53)	18.7% (28/150)	0.84
Moderate/severe cognitive impairment at last assessment, % (n):	3.8% (2/53)	2.0% (3/150)	0.61
Any visual impairment at last assessment, % (n):	39.6% (21/53)	14.1% (21/149)	<0.001
Any hearing impairment at last assessment, % (n):	1.9% (1/53)	0.0% (0/149)	0.26
Any functional impairment at last assessment, % (n):[b]	49.1% (26/53)	30.9% (46/149)	0.02

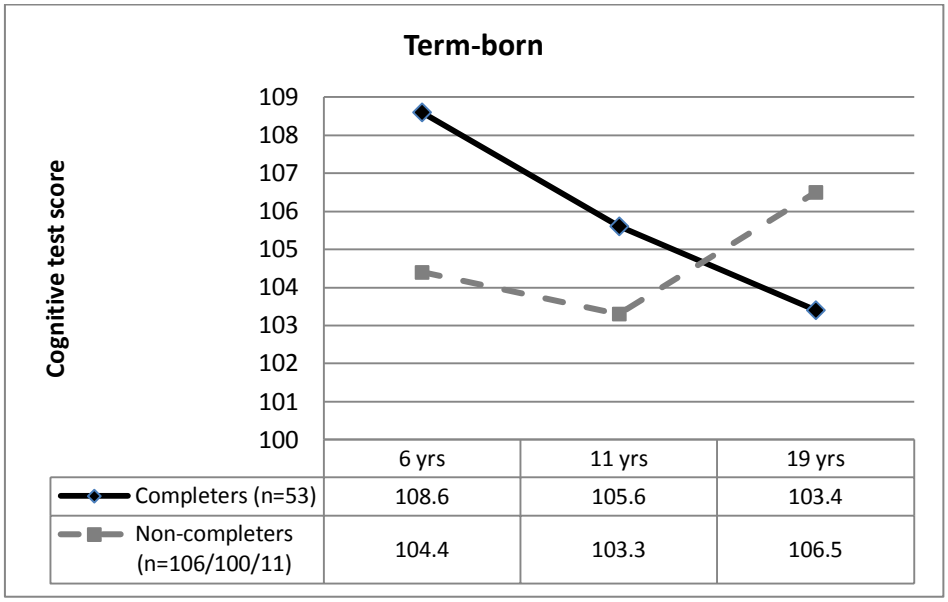
[a] Two-sided p-values were calculated using Fisher's Exact Test.
[b] Functional impairment includes cerebral palsy, cognitive, visual or hearing impairment.

Cognitive test scores were calculated at each time point separately by dropout group and plotted graphically to investigate the missing data mechanism in relation to outcome (Figure 7.1). The scores were lower among the non-completers compared to the completers in the EPT and term-born comparison group, except at age 19 years. However, very few participants in the non-completer group contributed to the estimates at age 19 (n=13 EPT, n=11 term-born) as few non-completers attended the last assessment; but it appears that those who did had higher cognitive scores than non-completers who didn't.

Figure 7.1: Cognitive test scores in the EPT and term-born group by dropout status



	2.5 yrs	6 yrs	11 yrs	19 yrs
Completers (n=114)	82.7	85.4	85.6	85.6
Non-completers (n=169/127/103/13)	76.1	78.2	80.9	87.4



	6 yrs	11 yrs	19 yrs
Completers (n=53)	108.6	105.6	103.4
Non-completers (n=106/100/11)	104.4	103.3	106.5

7.5 Primary analysis of all participants with at least one assessment

7.5.1 Descriptive analysis of cognitive test scores in the EPT and term-born group

The distributions of cognitive test scores for the EPT and term-born comparison group are displayed in Figure 7.2. They are approximately Normally distributed, and the EPT group have lower mean scores compared to the comparison group in years 6, 11 and 19. The spread of scores also appears to increase in the EPT group as the children grow older. The individual longitudinal trajectories for both groups are shown in Figure 7.3. They are generally quite flat and lower across all time points in the EPT group compared to the comparison group. The mean IQ scores and 95% confidence intervals at each age are presented in Table 7.6 and displayed graphically in Figure 7.4. The mean scores in the EPT group are around 20 points lower than the comparison group at each assessment and remain stable in both groups over time. There appears to be a slight increase in scores in the EPT group, but this is likely to be due to the higher rate of dropout in those with lower developmental scores in infancy as seen in Table 7.3. The number of standard deviations below the mean are also presented in Table 7.6 for each group. These are based on a UK term-born sample at the 2.5 year assessment,[308] and the distribution of scores in the EPICure term-born comparison group from age 6 onwards. Across all ages, around 85% of the comparison group achieved a score higher than one standard deviation below the mean, compared to around 30-40% of the EPT group. Between 15-20% of the EPT group had cognitive test scores less than 3 standard deviation below the mean at all time points after 2.5 years.

Figure 7.2: Distribution of cognitive test scores in the EPT and term-born group by age of assessment

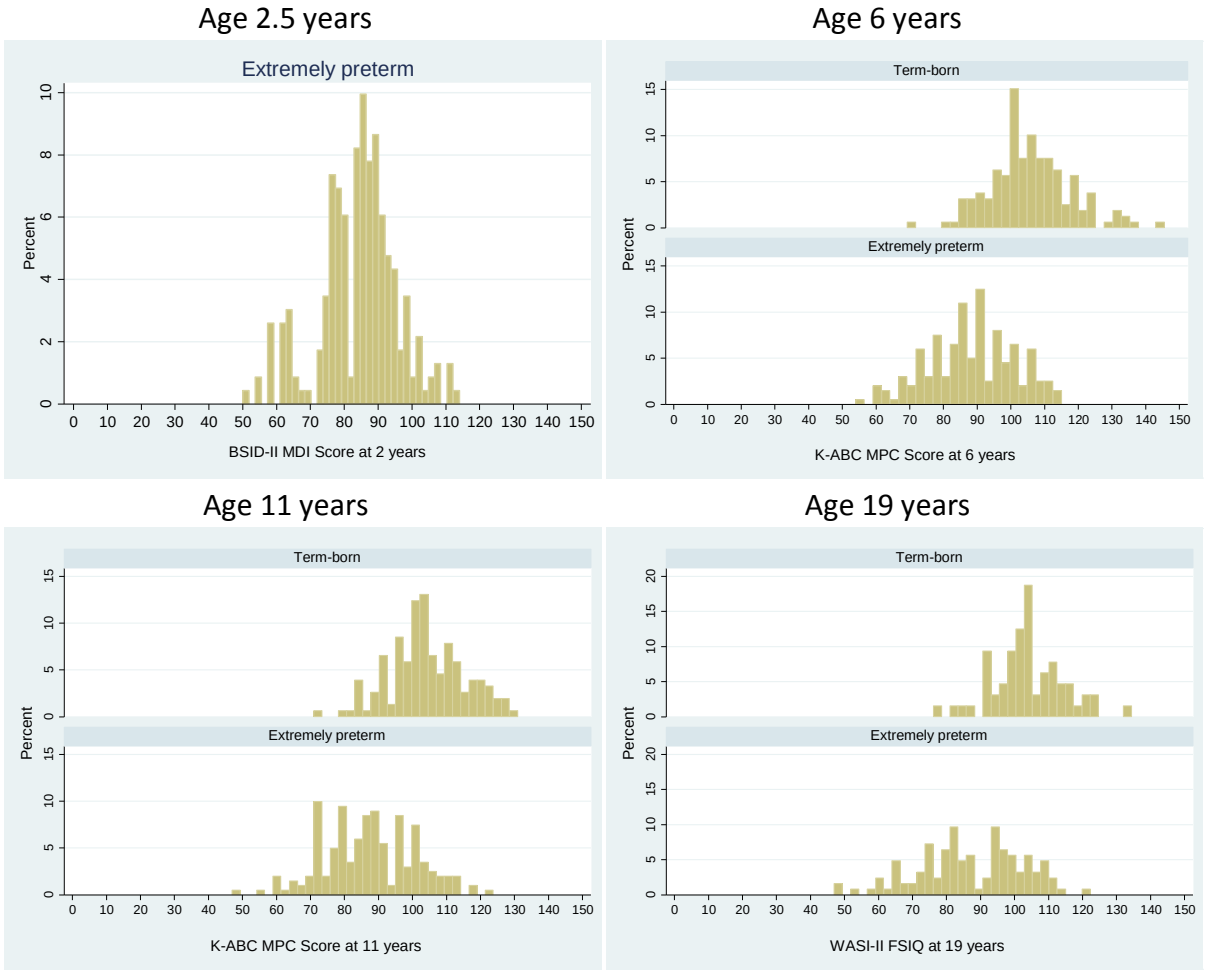
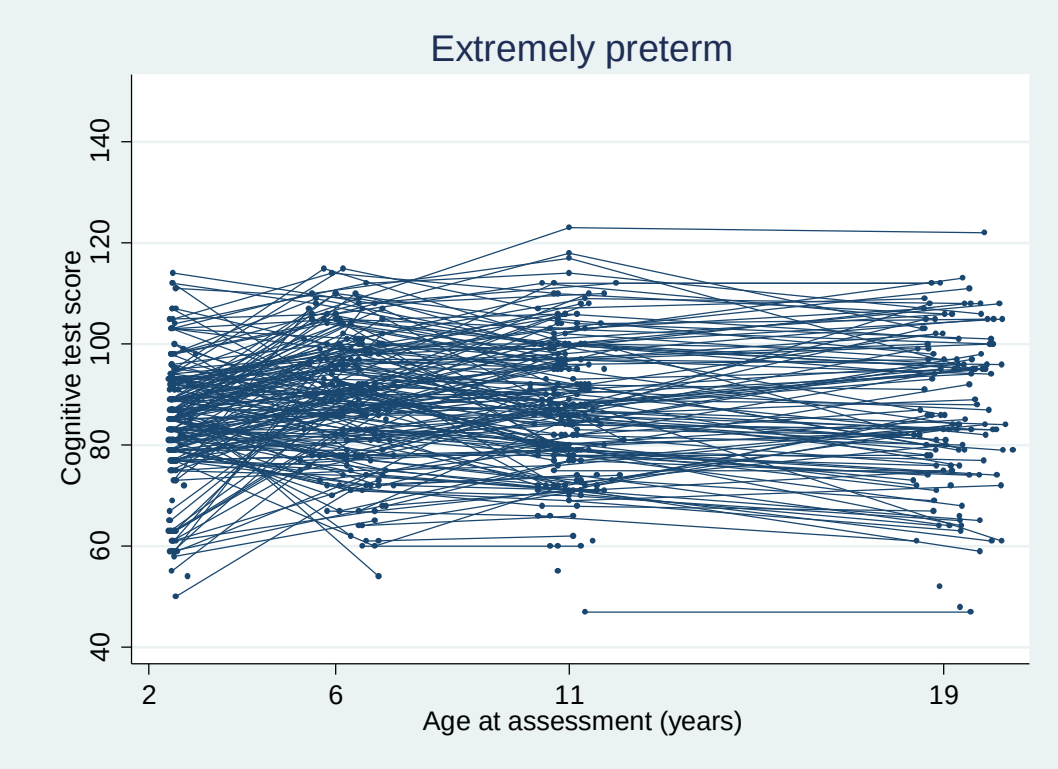


Figure 7.3: Longitudinal trajectories of observed cognitive test scores in the EPT and term-born group from age 2.5 to 19 years



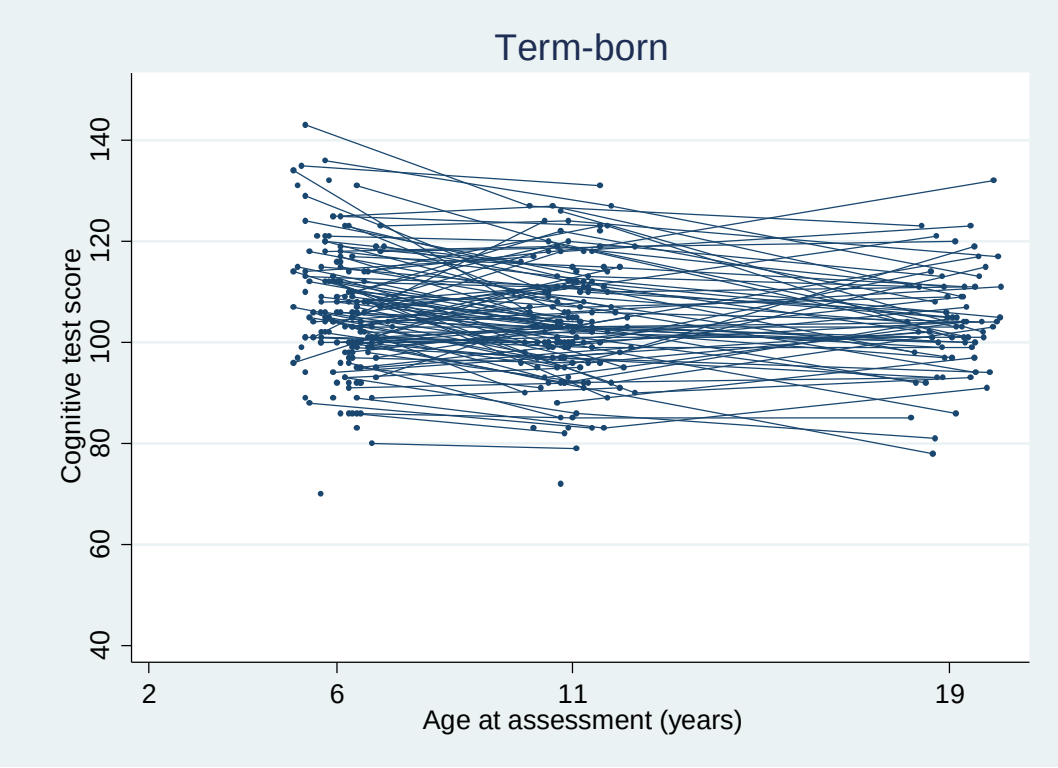


Table 7.6: Cognitive test scores in the EPT and term-born group by age of assessment

	Age 2.5 years	Age 6 years		Age 11 years		Age 19 years	
	EPT (n=283)	EPT (n=241)	Term (n=160)	EPT (n=219)	Term (n=153)	EPT (n=127)	Term (n=64)
Cognitive test administered:[a]	BSID-II MDI	KABC MPC		KABC MPC		WASI-II FSIQ	
Age at assessment:[b]							
Mean [SD]	2.5 [0.1]	6.3 [0.5]	6.1 [0.5]	11.0 [0.4]	11.0 [0.6]	19.3 [0.6]	19.2 [0.5]
Median {range}	2.5 {2.3, 3.3}	6.3 {5.2, 7.3}	6.2 {5.1, 7.2}	10.9 {10.2, 12.2}	10.9 {9.8, 12.3}	19.3 {18.4, 20.5}	19.2 {18.1, 20.1}
Cognitive test score:[c]							
Mean	78.7	81.6	105.7	83.4	104.1	85.7	103.9
(95% confidence interval)	(76.7 to 80.7)	(79.0 to 84.2)	(103.9 to 107.6)	(80.8 to 85.9)	(102.3 to 105.8)	(82.7 to 88.8)	(101.4 to 106.4)
Mean difference (95% confidence interval)		-24.2 (-27.7 to -20.6)		-20.7 (-24.1 to -17.3)		-18.1 (-22.8 to -13.5)	
Standard deviation from the mean, n (%):[c d]							
1SD below mean or higher	86 (30.4)	68 (28.2)	138 (86.3)	69 (31.5)	129 (84.3)	53 (41.7)	54 (84.4)
1-2SD below mean	114 (40.3)	75 (31.1)	20 (12.5)	63 (28.8)	22 (14.4)	17 (13.4)	8 (12.5)
2-3SD below mean	26 (9.2)	48 (19.9)	2 (1.3)	55 (25.1)	2 (1.3)	32 (25.2)	2 (3.1)
3SD below mean or lower	57 (20.1)	50 (20.8)	0 (0.0)	32 (14.6)	0 (0.0)	25 (19.7)	0 (0.0)

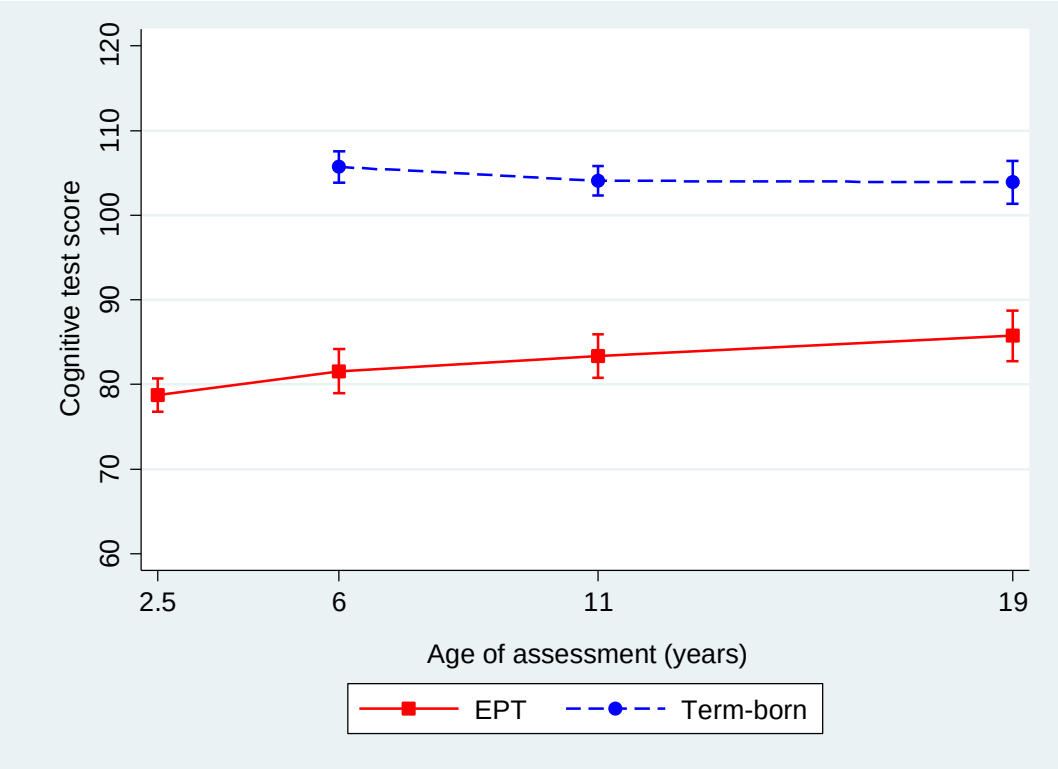
[a] BSID-II MDI: Bayley Scales of Infant Development (version 2) - Mental Development Index, KABC MPC: Kaufman Assessment Battery for Children - Mental Processing Composite, WASI-II FSIQ: Wechsler Abbreviated Scale of Intelligence (version 2) - Full Scale IQ.

[b] Corrected age at year 2.5 and chronological age in years 6, 11 and 19.

[c] Values multiply imputed within a clinically specified range for EPT participants seen at assessment but not tested with planned cognitive test: n=52 (year 2.5), n=41 (year 6), n=18 (year11), n=3 (year 19). Values multiply imputed for one term-born participant in year 6.

[d] Standard deviations based a UK term-born sample in year 2.5[308] and the distribution of scores in the EPICure term-born comparison group in years 6, 11 and 19.

Figure 7.4: Mean plus 95% confidence intervals for observed cognitive test scores in the EPT and term-born group at age 2.5, 6, 11 and 19



7.5.2 Comparative analysis of cognitive test scores in the EPT and term-born group

The key stages of model development are summarised in Table 7.7. This model was based on 495 participants (292 EPT participants and 203 term-born classmates) and 1247 observed cognitive test scores. Age was centred at age 6 years to make the intercept coefficient more meaningful; it represents the mean difference between the EPT and comparison group at age 6. The reference model (model one) for cognitive outcome fitted age as a random effect and various parameters were added in turn and retained if the fit of the model was significantly improved ($p \leq 0.05$ in the LRT). The equations, LRT statistics and corresponding p-values are displayed for each model. The MI analysis produced 10 LRT statistics and p-values, one for each of the 10 MI datasets, and a parameter was only retained if $p \leq 0.05$ in all 10 MI datasets.

The values provided for each model are the lowest LRT statistic and highest p-value within the set of 10, which is the most conservative estimate. In model two, EPT exposure status was added as a fixed effect, and in model three, a EPT*Age interaction was included. Both of these terms significantly improved the fit of the model and were retained. This implies that the cognitive trajectories of EPT participants and classmates have different intercepts and slopes; the EPT group have significantly different scores compared to the comparison group as they get older, and this difference changes with age. The curvature of the slopes was not significantly different between groups, assessed by fitting age as a non-linear quadratic function (model four). Heteroscedasticity was assumed to be constant within and between groups in models one to four, meaning that the estimates for the between- and within-individual differences were assumed to have constant variation for all participants. In model five the within-individual variance was allowed to differ between the EPT and comparison group, and in model six the between-individual variance was allowed to differ. Both changes resulted in models with a significantly improved fit.

The estimated coefficients, their standard errors and p-values (Wald test) for the fixed and random parts of the final model (model six) are presented in Table 7.8. The fixed part of the model is as follows:

Cognitive test score = 105.2 - 25.2EPT - 0.3(Age-6) + 0.5EPT*(Age-6)

On average, the cognitive test scores of EPT participants were 25.2 points below their term-born classmates at age 6 (95% CI: -27.8 to -22.6, $p<0.001$). Trajectories were similar between groups, though there was a small but statistically significant increase of 0.5 IQ

Table 7.7: Log-likelihood statistics for key model-fitting stages in the mixed model analysis of cognitive test scores of the EPT and term-born group

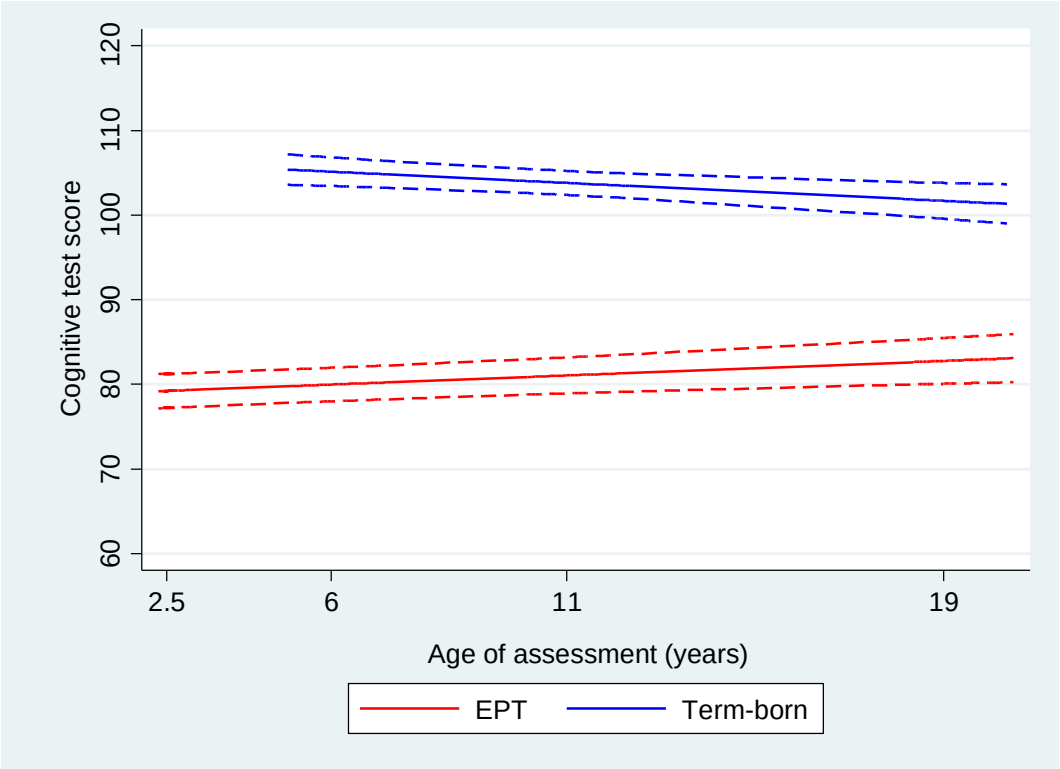
Model Number	Addition to model	Model	Likelihood ratio statistic (LRT)[a]	p-value[a]
1	Linear age	$y_{ij} = (\beta_0 + \mu_{0j}) + (\beta_1 + \mu_{1j})\, Age_{ij} + e_{ij}$	reference	
2	EPT	$y_{ij} = (\beta_0 + \mu_{0j}) + (\beta_1 + \mu_{1j})\, Age_{ij} + \beta_2 EPT + e_{ij}$	225.4 (1 d.f.)	<0.001
3	EPT*Age interaction	$y_{ij} = (\beta_0 + \mu_{0j}) + (\beta_1 + \mu_{1j})\, Age_{ij} + \beta_2 EPT + \beta_3 EPT*Age_{ij} + e_{ij}$	16.6 (1 d.f.)	<0.001
4	Non-linear age	$y_{ij} = (\beta_0 + \mu_{0j}) + (\beta_1 + \mu_{1j})\, Age_{ij} + \beta_2 EPT + \beta_3 EPT*Age_{ij} + \beta_4 Age^2_{ij} + e_{ij}$	2.4 (1 d.f.)	0.12
5	Different level-1 (within-individual) variance for group	$y_{ij} = (\beta_0 + \mu_{0j}) + (\beta_1 + \mu_{1j})\, Age_{ij} + (\beta_2 + e_{1ij})EPT + \beta_3 EPT*Age_{ij} + e_{0ij}Term$	29.0 (1 d.f.)	<0.001
6	Different level-2 (between-individual) variance for group	Term: $y_{ij} = (\beta_0 + \mu_{00j}) + (\beta_1 + \mu_{01j})Age_{ij} + e_{0ij}$ EPT: $y_{ij} = (\beta_0 + \mu_{10j}) + (\beta_1 + \mu_{11j})Age_{ij} + \beta_2 + \beta_3 Age_{ij} + e_{1ij}$	63.0 (3 d.f.)	<0.001

[a] Lowest LRT statistic and highest p-value reported from the 10 multiply imputed models. The LRT statistic reported for model 2 is compared to the reference model. The LRT statistics for subsequent models are compared to the last model that improved the fit with $p \leq 0.05$.

Table 7.8: Mixed model analysis of cognitive test scores in the EPT and term-born group

	Parameter	Estimated coefficient	95% CI	p-value
Fixed effects				
Constant	β_0	105.2	(103.5 to 106.9)	<0.001
EPT	β_1	-25.2	(-27.8 to -22.6)	<0.001
Age	β_2	-0.3	(-0.5 to -0.1)	0.006
EPT*Age interaction	β_3	0.5	(0.2 to 0.7)	<0.001
Random effects				
Level 1 (within-individual)				
Term	σ_{e0}	5.7	(4.9 to 6.7)	
EPT	σ_{e1}	8.7	(8.1 to 9.4)	
Level 2 (between-individual)				
Term intercept SD	$\sigma_{\mu 00}$	10.6	(9.2 to 12.2)	
Term slope SD	$\sigma_{\mu 01}$	0.7	(0.5 to 1.0)	
Term correlation(intercept, slope)	$\sigma_{\mu 00\,\mu 01}$	-0.6	(-0.8 to -0.3)	
EPT intercept SD	$\sigma_{\mu 10}$	16.3	(14.9 to 17.9)	
EPT slope SD	$\sigma_{\mu 11}$	0.5	(0.3 to 0.8)	
EPT correlation(intercept, slope)	$\sigma_{\mu 10\,\mu 11}$	0.3	(-0.01 to 0.6)	

Figure 7.5: Predicted mean cognitive test score and 95% confidence intervals for the fixed portion of the model for the EPT and term-born group



points per year in the EPT group relative to the comparison group; IQ scores fell by 0.3 points for each year after age 6 in the comparison group and increased by 0.2 points in the EPT group. The predicted mean trajectories and 95% confidence intervals for the fixed portion of the model are displayed in Figure 7.5. For the random part of the model, the estimated within-individual variation was 5.7 points for classmates and 8.7 points for EPT participants, so the test scores of EPT participants at different time points vary more around their average score than in control classmates. The estimated function for between-individual variation in classmates is given by:[c]

$$\text{var}(\mu_{00j} + \mu_{01j}\text{Age}_{ij}) = 10.6^2 - 2 \times 4.2(\text{Age-6})_{ij} + 0.7^2(\text{Age-6})^2_{ij}$$

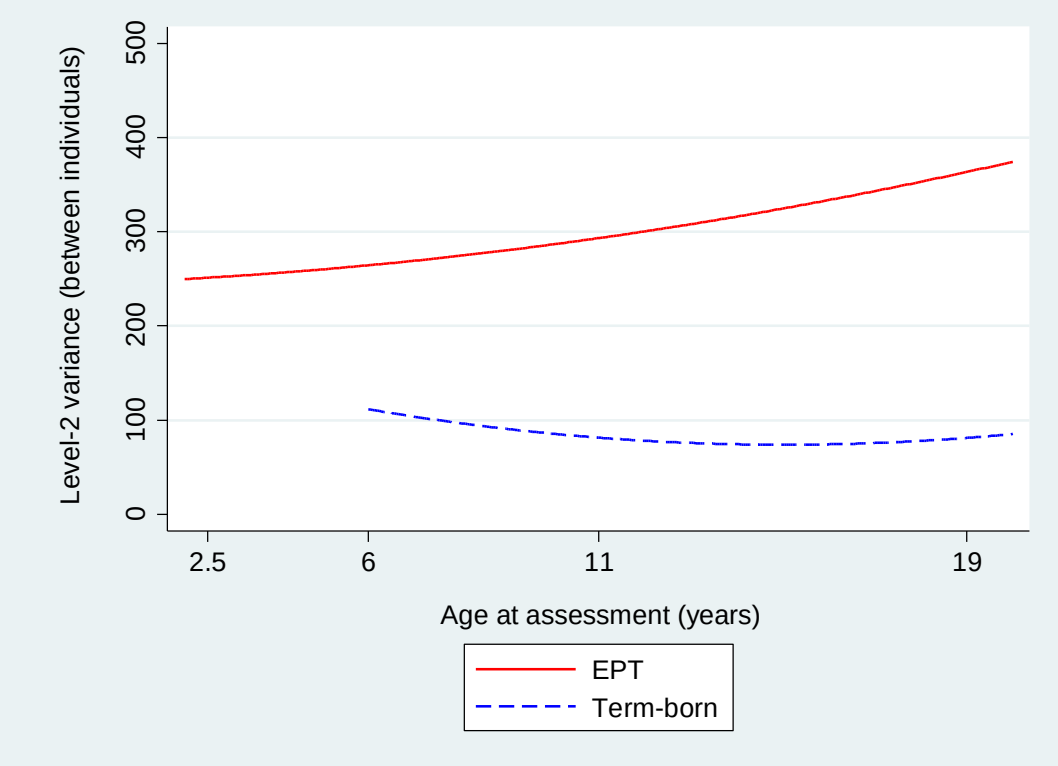
and for EPT participants,

$$\text{var}(\mu_{10j} + \mu_{11j}\text{Age}_{ij}) = 16.3^2 + 2 \times 2.7(\text{Age-6})_{ij} + 0.5^2(\text{Age-6})^2_{ij}$$

These functions are plotted in Figure 7.6 and indicate that the variance in IQ scores between EPT participants was higher than in classmates and increased as they got older, whereas the between-individual variance dipped slightly among classmates. The SD around the mean intercept was lower in classmates (10.6 versus 16.3 points), but the SD around the mean slope was higher (0.7 versus 0.5 points). The negative covariance of -4.2 in the comparison group implied that there was a negative correlation between random slopes and intercepts, so the higher a term-born individual's IQ score early on, the less steep the increase over time. Conversely, the positive smaller covariance of 2.7 in the EPT group implied that the higher an EPT individual's IQ score early on, the steeper the increase over time.

[c] Covariance(intercept, slope) = correlation(intercept, slope) x SD(intercept) x SD(slope)

Figure 7.6: Predicted between-individual variance function for the EPT and term-born group



7.5.3 Diagnostics and model fit

The standardised level one residuals were approximately Normally distributed (Figure 7.7), and were fairly randomly scattered when plotted against the predicted scores and age at assessment, although there was some evidence that in the comparison group higher scores tended to be slightly overestimated and lower scores slightly underestimated. Most of the standardised level one residuals had values <|2| with a few outliers with values <|3|. Large residuals sometimes occurred in the MI datasets when imputed values were at the low end of the assigned range, but this fluctuated with each imputed dataset. For example, the EPT child with the residual of <-4 in the graphs displayed here for the first imputed dataset had an imputed value of 25 which was bottom of the 25-39 range assigned, but in subsequent MI datasets imputed values were higher and the corresponding residuals less extreme.

The predicted random intercepts and slopes were also approximately Normally distributed (Figure 7.8), though the predicted random intercepts in the EPT group was somewhat negatively skewed indicating that a greater proportion of EPT participants had lower predicted intercepts than might be expected in a Normal distribution. The caterpillar plots in Figure 7.9 rank the level two residual for each child with 95% confidence intervals. Almost all of the confidence intervals overlapped zero in the predicted random slope plots, but in the predicted random intercept plots a small group of individuals at the lower and upper end of the plots (particularly in the EPT group) had confidence intervals for their residuals that did not overlap zero. Since these residuals represent individual departures from the overall average intercept predicted by the fixed parameters β_0 and β_1, this means that these are the individuals that differed significantly from the average at the 5% level.

Figure 7.7: Standardised level one residual diagnostic plots for cognitive test scores

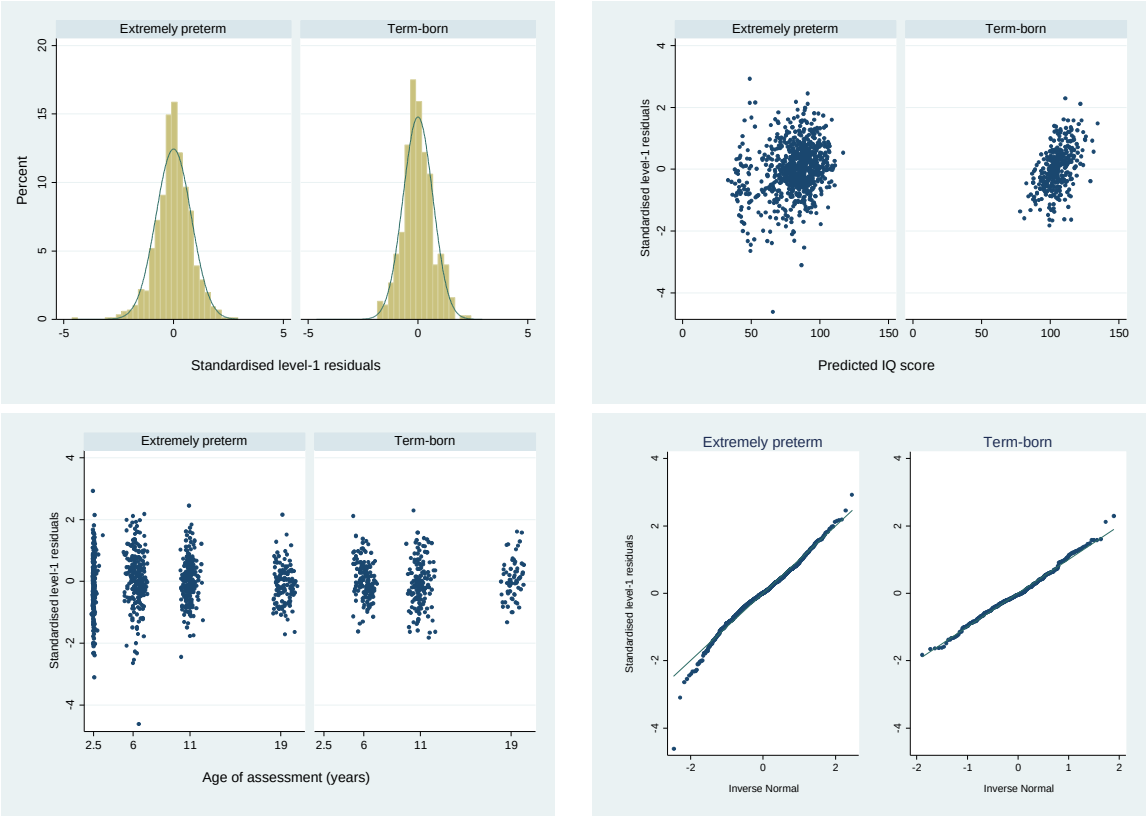


Figure 7.8: Predicted random intercepts and slopes for cognitive test scores

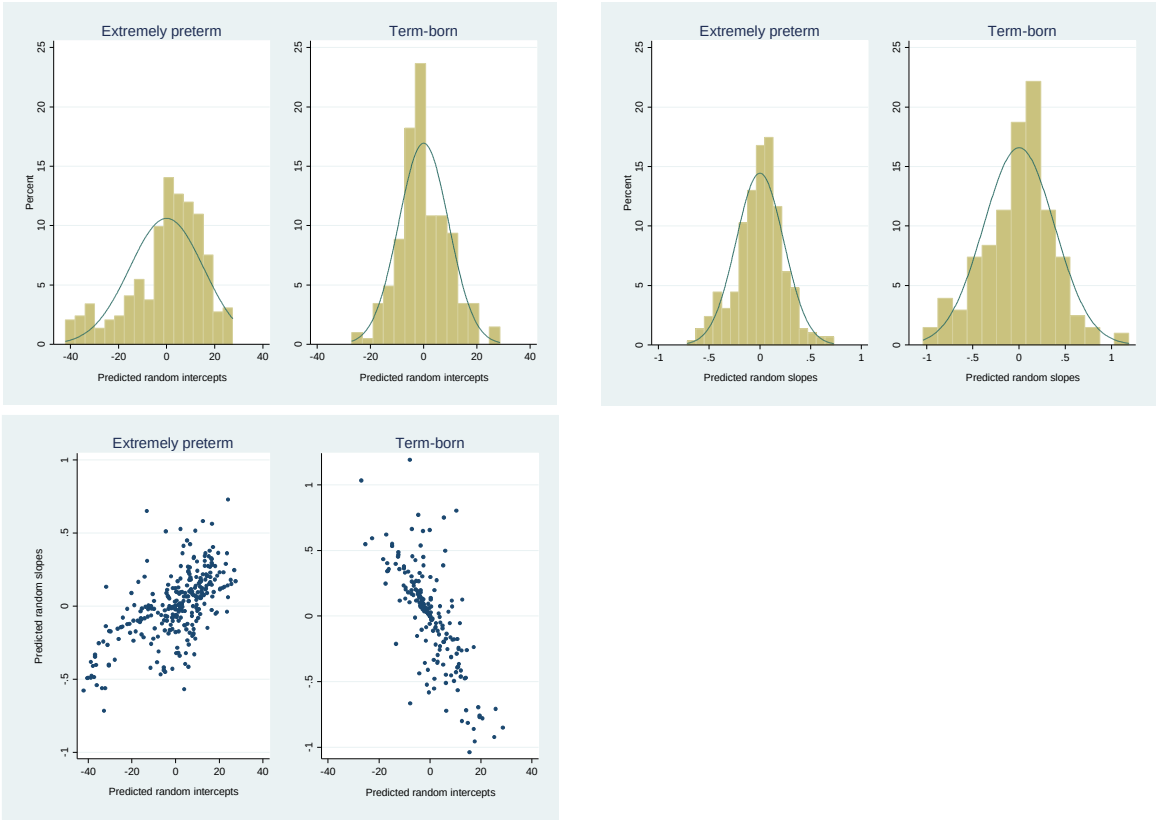
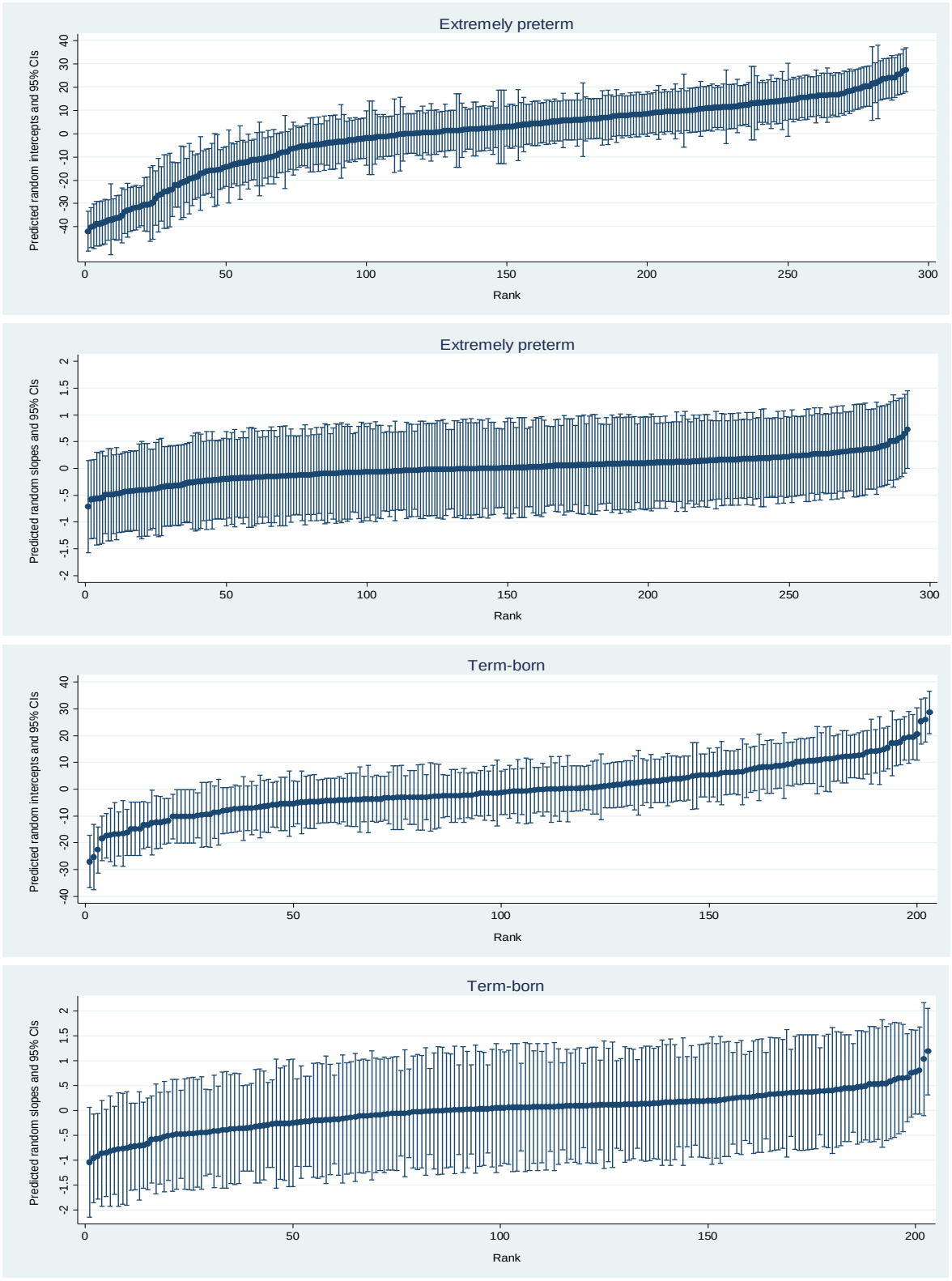


Figure 7.9: Caterpillar plots of random intercept and slope predictions and 95% confidence intervals versus ranking for cognitive test scores



7.5.4 Stratified analysis of cognitive test scores in the EPT and term-born group by participant sex and maternal education

The mean IQ scores and 95% confidence intervals at each age for EPT participants and term-born classmates, stratified by sex and maternal education, are presented in Table 7.9. Observed and predicted mean trajectories adjusted for sex and maternal education are displayed in Figure 7.10 . Adding sex as a main effect to the unadjusted model did not have a significant effect, but there was evidence of an interaction between sex and group (Table 7.10). The estimated IQ scores of EPT males were 8.8 points below EPT females (95% CI: -13.6 to -4.0, $p<0.001$), but there was no significant difference between males and females in the comparison group. The estimated IQ scores of participants with more highly educated mothers were 3.2 points higher in both the EPT and comparison group (95% CI: 0.8 to 5.7, $p=0.01$). The interaction between education and group was not significant (lowest LRT $p=0.92$) implying that the impact of maternal education on the rate of cognitive development was similar in both groups (Table 7.10).

Table 7.9: Cognitive test scores in the EPT and term-born group by age of assessment stratified by sex of participant and maternal education

	Age 2.5 years	Age 6 years		Age 11 years		Age 19 years	
	EPT (n=283)	EPT (n=241)	Term (n=160)	EPT (n=219)	Term (n=153)	EPT (n=127)	Term (n=64)
Cognitive test:[a-c]	BSID-II MDI	KABC MPC		KABC MPC		WASI-II FSIQ	
Infant sex							
Female							
No. (%)	148 (52.3)	119 (49.4)	89 (55.6)	118 (53.9)	89 (58.2)	68 (53.5)	39 (60.9)
Mean	82.4	87.2	105.5	87.6	104.2	88.6	103.6
(95% confidence interval)	(80.0 to 84.9)	(83.9 to 90.4)	(103.2 to 107.8)	(84.5 to 90.6)	(102.2 to 106.3)	(84.9 to 92.3)	(100.7 to 106.5)
Male							
No. (%)	135 (47.7)	122 (50.6)	71 (44.4)	101 (46.1)	64 (41.8)	59 (46.5)	25 (39.1)
Mean	74.6	76.1	106.0	78.5	103.8	82.4	104.4
(95% confidence interval)	(71.7 to 77.6)	(72.3 to 79.9)	(103.0 to 109.0)	(74.4 to 82.6)	(100.8 to 106.9)	(77.6 to 87.2)	(99.7 to 109.0)
Mean difference (95% CI)	-7.8 (-11.6 to - 4.0)	-11.1 (-16.2 to -6.1)	0.5 (-3.3 to 4.2)	-9.1 (-14.1 to -4.0)	-0.4 (-4.0 to 3.2)	-6.2 (-12.2 to -0.2)	0.8 (-4.5 to 6.1)
Maternal education							
Up to O'level (or equivalent)							
No. (%)	174 (64.7)	146 (63.8)	82 (58.2)	131 (63.0)	87 (60.0)	64 (51.6)	32 (50.8)
Mean	77.9	80.6	104.7	82.7	102.6	85.6	103.4
(95% confidence interval)	(75.4 to 80.4)	(77.4 to 83.8)	(102.1 to 107.2)	(79.4 to 86.1)	(100.3 to 104.8)	(81.7 to 89.4)	(99.4 to 107.3)
A'level (or equivalent) and above							
No. (%)	95 (35.3)	83 (36.2)	59 (41.8)	77 (37.0)	58 (40.0)	60 (48.4)	31 (49.2)
Mean	81.1	83.8	108.7	86.3	106.1	87.3	104.3
(95% confidence interval)	(77.9 to 84.3)	(79.0 to 88.7)	(105.6 to 111.7)	(82.1 to 90.6)	(103.0 to 109.1)	(82.7 to 91.8)	(101.1 to 107.5)
Mean difference (95% CI)	3.2 (-1.0 to 7.4)	3.2 (-2.4 to 8.9)	4.0 (-0.01 to 8.0)	3.6 (-1.8 to 9.0)	3.5 (-0.2 to 7.2)	1.7 (-4.3 to 7.7)	0.9 (-4.3 to 6.2)

[a] BSID-II MDI: Bayley Scales of Infant Development (version 2) - Mental Development Index, KABC MPC: Kaufman Assessment Battery for Children - Mental Processing Composite, WASI-II FSIQ: Wechsler Abbreviated Scale of Intelligence (version 2) - Full Scale IQ.

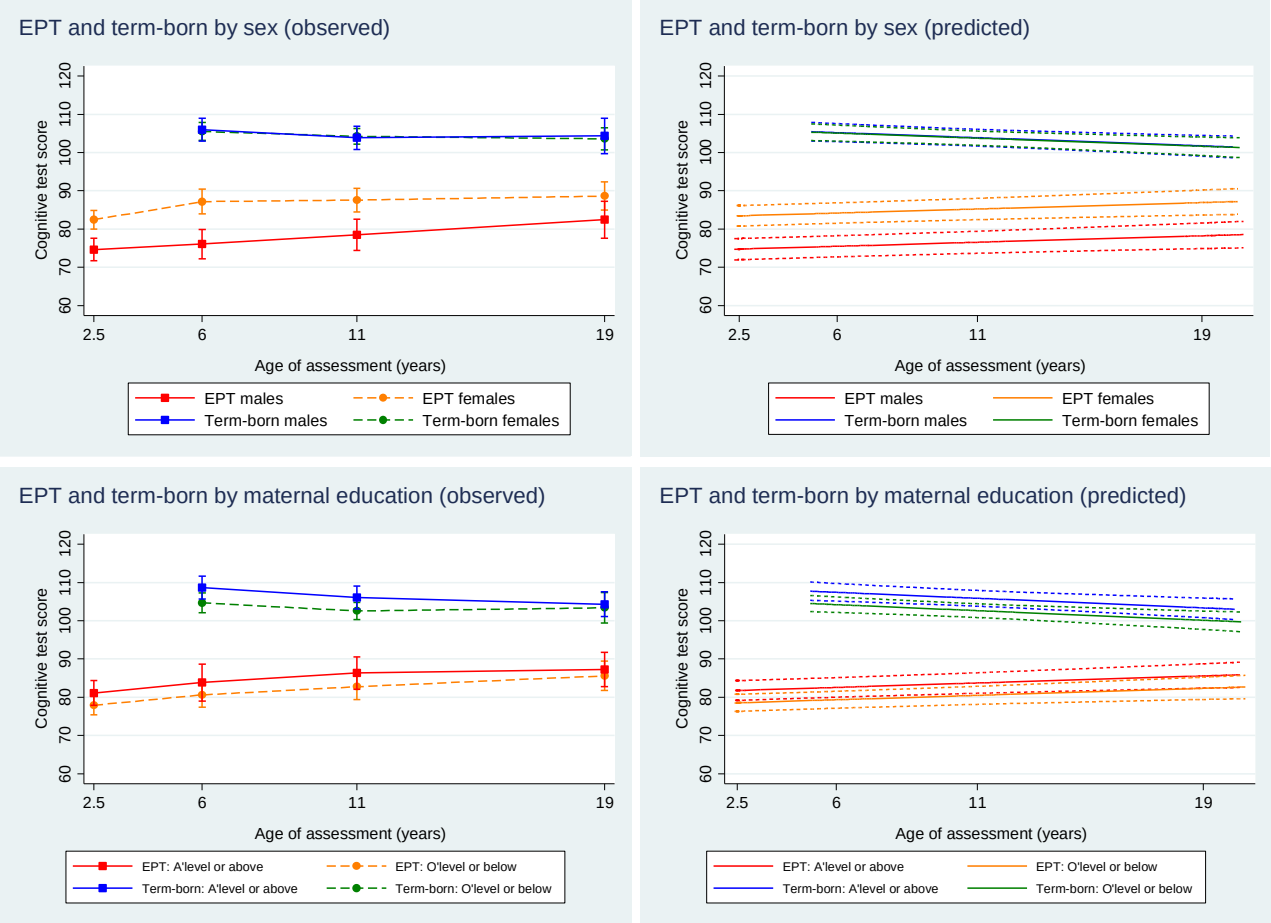
[b] Corrected age at year 2.5 and chronological age in years 6, 11 and 19.

[c] Values multiply imputed within a clinically specified range for EPT participants seen at assessment but not tested with planned cognitive test: n=52 (year 2.5), n=41 (year 6), n=18 (year11), n=3 (year 19). Values multiply imputed for one term-born participant in year 6.

Table 7.10: Mixed model analysis of cognitive test scores in the EPT and term-born group adjusted for sex and maternal education

	Adjusted for sex (n=495)		Adjusted for maternal education (n=456)	
	Estimate	**95% CI**	**Estimate**	**95% CI**
Fixed effects				
Constant	105.1	(103.0 to 107.2)	104.2	(102.2 to 106.3)
EPT	-21.0	(-24.4 to -17.6)	-25.0	(-27.7 to -22.2)
Age	-0.3	(-0.5 to -0.1)	-0.3	(-0.5 to -0.1)
EPT*Age interaction	0.5	(0.2 to 0.7)	0.5	(0.3 to 0.8)
Male	0.1	(-2.8 to 3.0)	-	-
EPT*Male interaction	-8.8	(-13.6 to -4.0)	-	-
Higher maternal education	-	-	3.2	(0.8 to 5.7)
Random effects				
Level 1 (within-individual)				
Term	8.7	(8.1 to 9.4)	8.8	(8.1 to 9.5)
EPT	5.7	(4.9 to 6.7)	5.6	(4.8 to 6.6)
Level 2 (between-individual)				
Term intercept SD	15.8	(14.3 to 17.3)	16.2	(14.7 to 17.9)
Term slope SD	0.5	(0.3 to 0.8)	0.5	(0.3 to 0.8)
Term correlation(intercept, slope)	0.4	(0.04 to 0.7)	0.3	(-0.02 to 0.6)
EPT intercept SD	10.6	(9.2 to 12.2)	10.6	(9.1 to 12.2)
EPT slope SD	0.7	(0.5 to 1.0)	0.7	(0.5 to 1.0)
EPT correlation(intercept, slope)	-0.6	(-0.8 to -0.3)	-0.6	(-0.8 to -0.3)

Figure 7.10: Observed and predicted mean cognitive test scores plus 95% confidence intervals in the EPT and term-born group stratified by sex and maternal education



7.5.5 Stratified analysis of cognitive test scores of EPT participants by neonatal brain injury and gestational age

The mean IQ scores and 95% confidence intervals at each age of assessment for EPT participants, stratified by GA and brain injury diagnosed during the neonatal period, are presented in Table 7.11. The observed and predicted mean trajectories adjusted for neonatal brain injury and GA are displayed in Figure 7.11. In the predicted model (Table 7.12), scores of EPT participants with moderate/severe neonatal brain injury were on average 10.9 points below participants with no/mild brain injury (95% CI: -15.5 to -6.3, p<0.001). The interaction term between brain injury and age was not significant (lowest LRT p=0.50), indicating that the rate of cognitive development between those with none/mild and moderate/severe brain injury was similar over time. On average, the IQ scores of participants born less than 25 weeks of gestation were 4.4 points lower than participants born at 25 weeks (95% CI: -8.4 to -0.4, p=0.03) and the interaction between GA and age was also not significant (lowest LRT p=0.78) (Table 7.12).

Table 7.11: Cognitive test scores in EPT participants by age of assessment stratified by neonatal brain injury and gestational age

	Age 2.5 years (n=283)	Age 6 years (n=241)	Age 11 years (n=219)	Age 19 years (n=127)
Cognitive test:[a-c]	BSID-II MDI	KABC MPC	KABC MPC	BSID-II MDI
Neonatal brain injury				
No/mild brain injury				
No. (%)	220 (77.7)	187 (77.6)	169 (77.5)	104 (82.5)
Mean (95% CI)	80.5 (78.3 to 82.6)	85.2 (82.7 to 87.7)	86.1 (83.6 to 88.5)	87.2 (84.2 to 90.2)
Moderate/severe brain injury				
No. (%)	63 (22.3)	54 (22.4)	49 (22.5)	22 (17.5)
Mean (95% CI)	72.7 (68.1 to 77.3)	68.9 (62.0 to 75.8)	73.8 (66.5 to 81.0)	78.4 (68.8 to 88.1)
Mean difference (95% CI)	-7.8 (-12.4 to -3.12)	-16.3 (-22.3 to 10.4)	-12.3 (-18.3 to -6.3)	-8.7 (-16.7 to -0.8)
Gestational age at delivery				
25 weeks				
No. (%)	167 (59.0)	144 (59.8)	126 (57.5)	75 (59.1)
Mean (95% CI)	79.8 (77.5 to 82.2)	84.3 (81.2 to 87.5)	86.1 (83.3 to 88.9)	87.6 (84.2 to 91.0)
<25 weeks				
No. (%)	116 (41.0)	97 (40.2)	93 (42.5)	52 (40.9)
Mean (95% CI)	77.1 (73.8 to 80.4)	77.4 (73.1 to 81.8)	79.7 (75.1 to 84.4)	83.1 (77.7 to 88.6)
Mean difference (95% CI)	-2.7 (-6.7 to 1.3)	-6.9 (-12.2 to 1.6)	-6.3 (-11.5 to 1.2)	-4.5 (-10.6 to 1.7)

[a] BSID-II MDI: Bayley Scales of Infant Development (version 2) - Mental Development Index, KABC MPC: Kaufman Assessment Battery for Children - Mental Processing Composite, WASI-II FSIQ: Wechsler Abbreviated Scale of Intelligence (version 2) - Full Scale IQ.

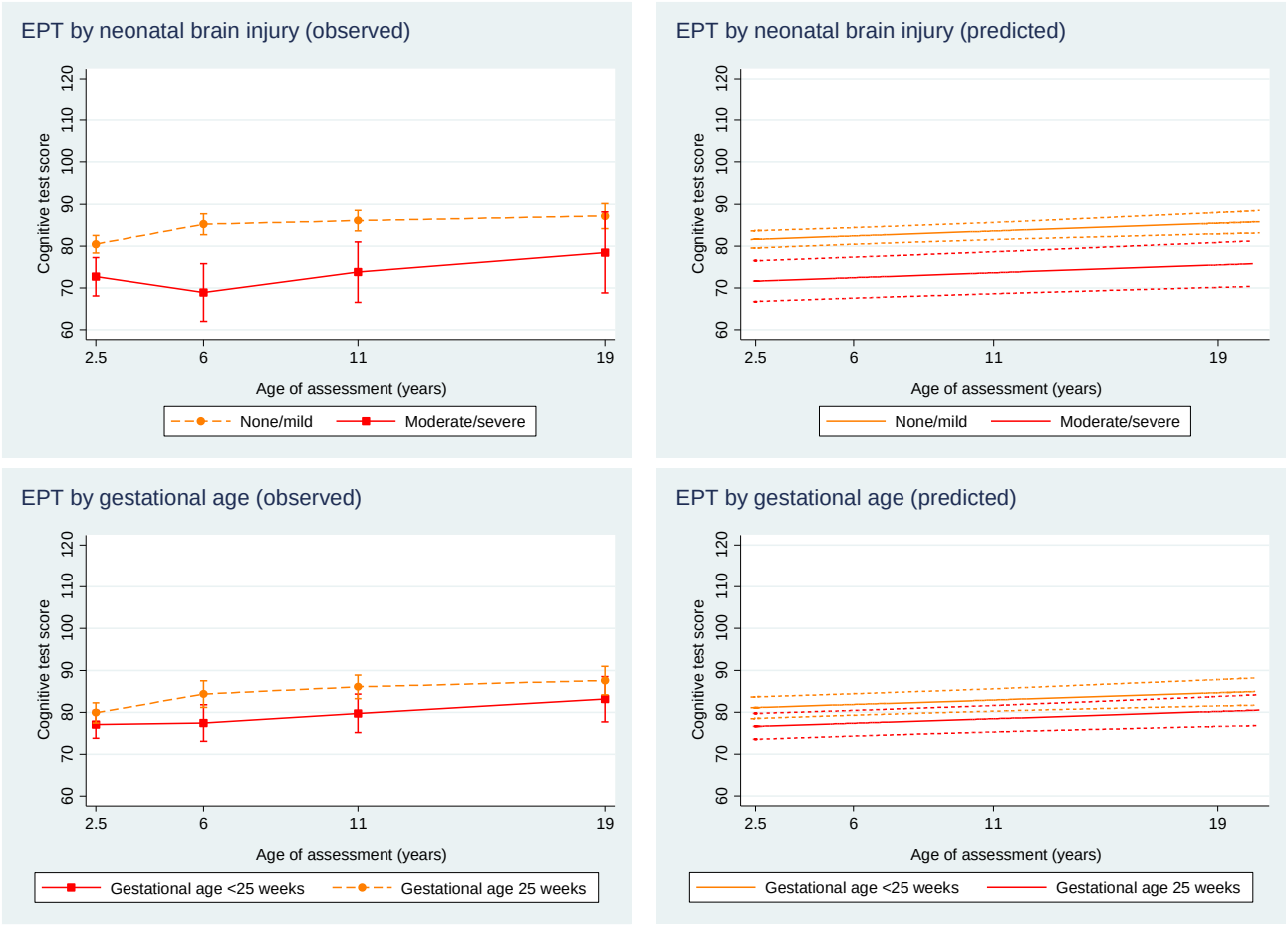
[b] Corrected age at year 2.5 and chronological age in years 6, 11 and 19.

[c] Values multiply imputed within a clinically specified range for EPT participants seen at assessment but not tested with planned cognitive test: n=52 (year 2.5), n=41 (year 6), n=18 (year11), n=3 (year 19).

Table 7.12: Mixed model analysis of cognitive test scores in EPT participants adjusted for neonatal brain injury and gestational age

Parameter	Adjusted for neonatal brain injury (n=291)		Adjusted for gestational age (n=292)	
	Estimate	**95% CI**	**Estimate**	**95% CI**
Fixed effects				
Constant	82.4	(80.2 to 84.5)	81.8	(79.2 to 84.4)
Age	0.2	(0.06 to 0.3)	0.2	(0.06 to 0.3)
Moderate-severe brain Injury	-10.9	(-15.5 to -6.3)	-	-
Gestational age <25 weeks	-	-	-4.4	(-8.4 to -0.4)
Random effects				
Level 1 (within-individual)	8.7	(8.1 to 9.4)	8.7	(8.1 to 9.4)
Level 2 (between-individual)				
Intercept SD	15.6	(14.2 to 17.2)	16.2	(14.7 to 17.7)
Slope SD	0.5	(0.3 to 0.8)	0.5	(0.3 to 0.8)
Correlation(intercept, slope)	0.3	(-0.03 to 0.6)	0.3	(-0.02 to 0.6)

Figure 7.11: Observed and predicted mean cognitive test scores plus 95% confidence intervals in EPT participants stratified by neonatal brain injury and gestational age



7.6 Complete case analysis of cognitive test scores in the EPT and term-born group

All analyses were repeated for the 167 completers (114 EPT and 63 classmates) and the results are shown in Table 7.13 and Figure 7.12-Figure 7.13. IQ scores were around 3 points higher on average in the complete-case analysis compared to the analysis including all participants. All of the differences reported between groups were of similar magnitude in the complete-case analysis, and the same results were statistically significant, except for a reduced effect of maternal education; difference in means for completers: 2.1 points (95% CI: -1.6 to 5.9, p=0.26).

Table 7.13: Complete case: mixed model analysis of cognitive test scores in the EPT and term-born group adjusted for sex, maternal education, neonatal brain injury and gestational age

Extremely preterm and term-born participants

	Unadjusted model (n=167)		Adjusted for sex (n=167)		Adjusted for maternal education (n=163)	
	Estimate	**95% CI**	**Estimate**	**95% CI**	**Estimate**	**95% CI**
Fixed effects						
Constant	108.2	(105.1 to 111.3)	108.5	(104.8 to 112.2)	107.4	(103.9 to 111.0)
EPT	-23.9	(-28.1 to -19.8)	-19.5	(-24.5 to -14.4)	-23.7	(-27.8 to -19.6)
Age	-0.4	(-0.6 to -0.2)	-0.4	(-0.6 to -0.2)	-0.4	(-0.6 to -0.2)
EPT*Age	0.5	(0.2 to 0.8)	0.5	(0.2 to 0.8)	0.6	(0.3 to 0.8)
Male	-	-	-0.8	(-5.9 to 4.4)	-	-
EPT*Male	-	-	-9.7	(-17.1 to -2.4)	-	-
Higher maternal education	-	-	-	-	2.1	(-1.6 to 5.9)
Random effects						
Within-individual						
Term SD	4.8	(4.0 to 5.8)	4.8	(4.0 to 5.8)	4.8	(4.0 to 5.8)
EPT SD	7.5	(6.8 to 8.3)	7.5	(6.9 to 8.4)	7.5	(6.8 to 8.3)
Between-individual						
Term Intercept SD	10.7	(8.5 to 13.3)	10.6	(8.5 to 13.3)	10.5	(8.4 to 13.2)
Term Slope SD	0.7	(0.5 to 0.9)	0.7	(0.5 to 0.9)	0.7	(0.5 to 0.9)
Term corr(intercept, slope)	-0.5	(-0.7 to -0.2)	-0.5	(-0.7 to -0.2)	-0.5	(-0.7 to -0.2)
EPT Intercept SD	14.3	(12.4 to 16.6)	13.3	(11.5 to 15.4)	13.7	(11.8 to 15.9)
EPT Slope SD	0.6	(0.5 to 0.8)	0.6	(0.5 to 0.8)	0.6	(0.5 to 0.8)
EPT corr(intercept, slope)	0.1	(-0.2 to 0.4)	0.1	(-0.2 to 0.4)	0.1	(-0.2 to 0.3)

Extremely preterm participants only

	Adjusted for neonatal brain injury (n=114)		Adjusted for gestational age (n=114)	
	Estimate	**95% CI**	**Estimate**	**95% CI**
Fixed effects				
Constant	86.1	(83.2 to 89.1)	87.1	(83.5 to 90.7)
Age	0.1	(-0.02 to 0.3)	0.1	(-0.02 to 0.3)
Moderate-severe brain Injury	-10.1	(-16.9 to -3.2)	-	-
Gestational age <25 weeks	-	-	-6.6	(-12.0 to -1.2)
Random effects				
Within-individual SD	7.5	(6.9 to 8.3)	7.5	(6.9 to 8.3)
Between-individual				
Intercept SD	13.8	(11.9 to 15.9)	14.0	(12.1 to 16.1)
Slope SD	0.6	(0.5 to 0.8)	0.6	(0.5 to 0.8)
corr(intercept, slope)	0.1	(-0.2 to 0.4)	0.1	(-0.2 to 0.4)

Figure 7.12: Complete case: observed and predicted mean cognitive test scores plus 95% confidence intervals in the EPT participants and the term-born group stratified by sex and maternal education

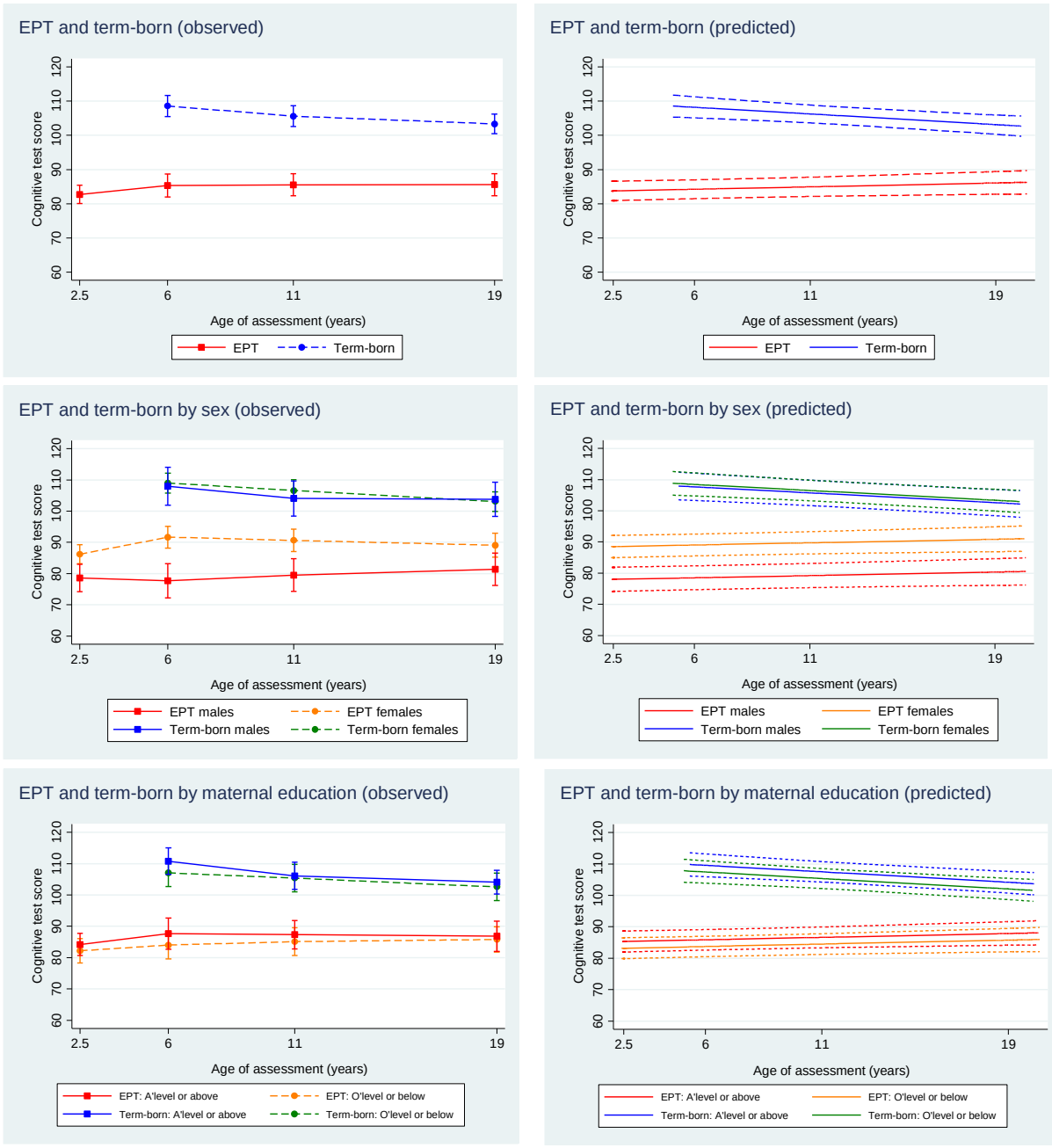
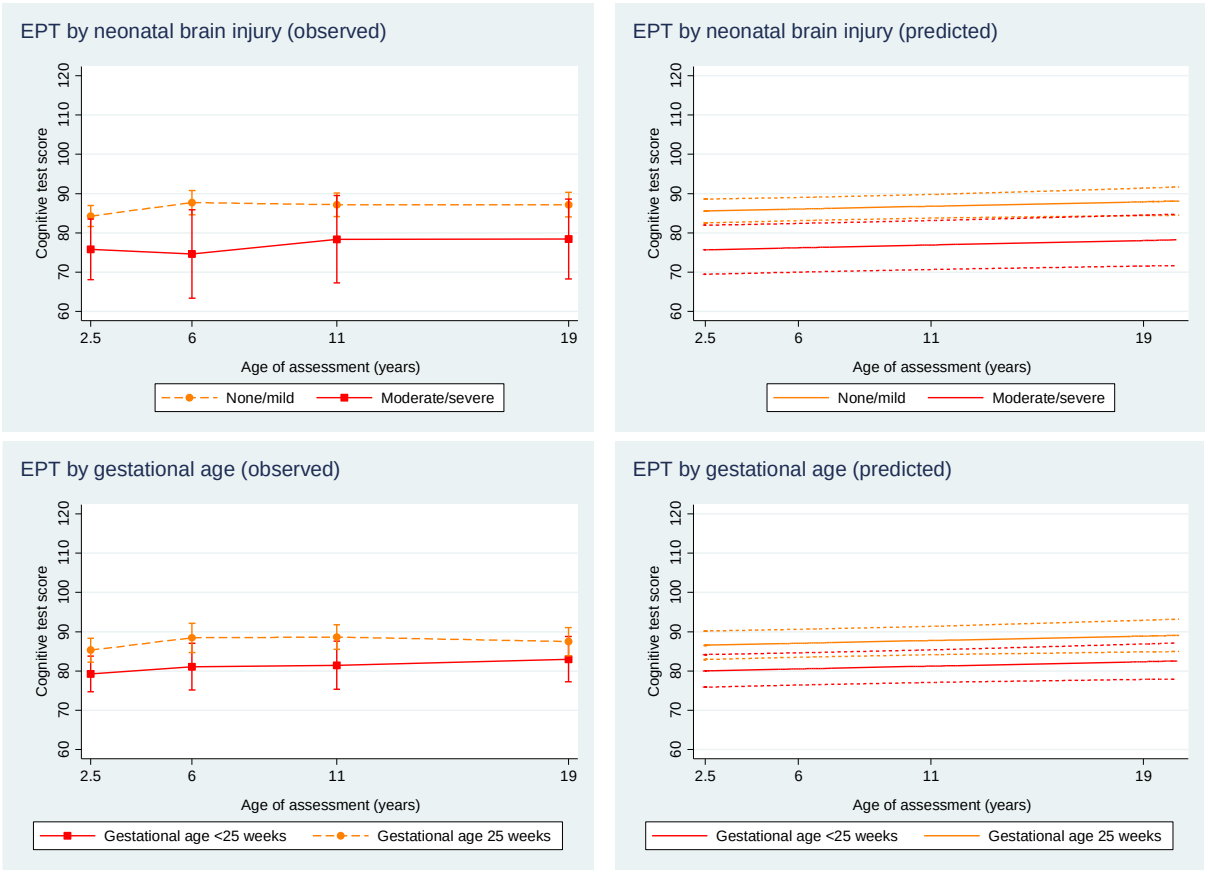


Figure 7.13: Complete case: observed and predicted mean cognitive test scores plus 95% confidence intervals in EPT participants stratified by neonatal brain injury and gestational age



7.7 Discussion

In this large dual nation, population-based study, impaired cognitive function in EPT children was found to persist into early adulthood with no evidence of substantial 'catch-up' with term-born peers, or further decline. IQ scores were on average 25 points lower in EPT individuals, and scores were more variable both within individuals and between individuals in the EPT group. While there were no differences between term-born males and females, being an EPT male had a detrimental effect on cognitive function, with scores over 8 points lower on average than EPT females, persisting across childhood and adolescence. There was some evidence that higher maternal education was marginally beneficial in both the EPT and term-born group. Having moderate to severe neonatal brain injury also had an adverse effect on outcome for EPT individuals, who had IQ scores 10 points lower on average compared to EPT individuals with no/mild brain injury. Survivors born at 25 weeks of gestation had significantly better cognitive function than those born less than 25 weeks, by a magnitude of about five points.

The findings in this contemporary EPT cohort followed prospectively from birth, were consistent with those from older VPT/VLBW cohorts in which cognitive scores were persistently lower by around 10 points compared to term-born individuals into early adulthood.[40,50,298,299,315] There is thus a significant decline in cognitive function with lower gestation as has been shown in childhood.[316] The association between level of maternal education and SES and cognitive outcome in both term and preterm populations has been frequently reported,[295,299,317] yet there was only weak evidence of this association in the present study. This may be due to the higher drop-out rate in younger, less well-educated mothers, which restricted the educational range included in the analysis, making it less sensitive to detect a difference. It is also possible that maternal education alone is not a sufficient marker of social disadvantage. The impact of sex on

the cognitive trajectory of EPT participants supports the findings from cross-sectional studies which have shown that EPT males are inherently disadvantaged and at greater risk of poorer neurodevelopmental outcome than their female counterparts.[227,230] Likewise, the adverse effect of brain injury diagnosed in the neonatal period on subsequent cognitive function is well documented by studies that report strong linear trends with severity of brain injury.[207,210,299] Widespread structural brain changes are still seen in adulthood and have been shown to be associated with IQ.[318] The larger between-individual differences among EPT participants may be partly explained by factors such as sex and brain insult and has been reported previously in VPT children.[299,304] The larger within-individual variance was mainly driven by the lack of comparison group data at 2.5 years; participants with severe cognitive impairment were more likely to have greater variation in scores over time due to the difference in the lower limits of the BSID-II and the later IQ tests.

Comparison of cognitive functioning over time is complicated by the use of different assessments at different ages; the range of neuropsychological tests that can be performed increases with age as the brain matures and executive functions develop. Cognitive tests during early childhood rely to some extent on motor function and provide a much cruder measure of development than those that can be performed in later childhood and adulthood, and may not accurately measure core cognitive ability.[319] Additionally, IQ scores have a tendency to improve over time,[313] which can be seen in the upward drift of scores in both preterm and term-born groups, however the impact of this on group differences at any particular time point was avoided by the using reference data from the contemporary comparison group.

In common with other longitudinal studies, the numbers of participants lost to follow-up increased over time, and was related to markers of social disadvantage and significant disability.

It is therefore likely that the true extent of cognitive impairment has been underestimated. The classmates recruited from mainstream schools are likely to represent a relatively healthy group, and therefore any differences found may have been exaggerated. However, this is believed to be minimal, as most children with special educational needs in the UK are integrated into mainstream schools and only 1.1% of children have special-school placements.[320] Hence the inclusion of classmates for EPT participants attending special schools (15% in this study) would have inappropriately biased the comparison group. The findings were strengthened by the analysis of individuals with complete follow-up which corroborated the main results. Although this subset of participants had slightly higher cognitive scores in general due to the higher rate of dropout in those with lower developmental scores in infancy, the same differences between groups were detected, and were of similar magnitude.

In conclusion, it appears that being born EPT places limits on brain plasticity and function which is not recovered over time with the most vulnerable being males and those who experienced brain injury early in life. Acquired brain lesions during the neonatal period have been shown to cause focal and diffuse structural abnormalities which may disturb neurodevelopmental processes and impede the brain from maintaining a normal developmental trajectory.[321,322] If EPT participants fail to achieve optimum levels of cognitive function once they have reached maturity, then this has implications for health and well-being in later adulthood and old age.[323]

Chapter 8: Longitudinal analysis of behavioural outcome

8.1 Introduction

Compared with term-born children, a higher prevalence of parent- and/or teacher-reported neurobehavioural problems, in particular emotional symptoms, inattention and peer relationship problems, are well documented in school-aged children born EPT.[42,232,324,325] Studies using diagnostic evaluation have also reported a significant excess of psychiatric disorders, in particular ADHD, ASD and emotional disorders in school aged children born EPT or VPT.[40,41,257,267] However, a recent systematic review of the social development of children born VPT highlighted the paucity of research in adolescence and early adulthood, a critical period of physical, emotional and social change, and a notable lack of longitudinal data.[326,327]

The few longitudinal studies that have investigated the prevalence of neurobehavioural symptoms or disorders in children and adolescents who were born preterm or low birth weight have shown that the increased prevalence of problems persists over time and may have greater stability in individuals born preterm compared with those born at term or with normal birthweight.[43,282,324,328,329] The significant excess of anxiety and peer relationship problems has been shown to persist from childhood through adolescence,[324,330,331] and although there may be some evidence of catch-up in attention regulation through adolescence[323] and a decrease in the between-group difference in ADHD symptoms,[324] there remains a significant excess of attention problems and disorders in young adults who were born preterm.[323]

Previous studies employed cross-sectional analysis techniques which are not able to detect variation in individual trajectories. Little is known about how such symptoms change over time following preterm birth, taking into account differences in the natural onset and course of different neurobehavioural disorders. In particular, the trajectories of neurobehavioural problems into adulthood for EPT survivors have not been studied. Failure to understand and provide support for the long-term development of these individuals will impact on their life-course opportunities and their social, occupational and family functioning in adulthood.

This chapter reports the results of a longitudinal analysis of the change in neurobehavioural problems in EPT survivors from childhood to early adulthood in the EPICure study. The main objective of this analysis was to investigate trajectories of neurobehavioural problems in EPT survivors from 6 to 19 years of age compared to those of a term-born comparison group; for neurobehavioural problems overall, and separately for emotional, conduct, hyperactivity/ inattention and peer relationship problems. As with the longitudinal analysis of cognitive outcome, the effect of sex and maternal education on these trajectories was investigated, and within the EPT group, the effect of GA and neonatal brain injury. In addition, the impact of a positive screen for behavioural problems in early infancy and the effect of moderate/severe cognitive and functional impairment on EPT trajectories were investigated.

8.2 Behavioural screening assessments

8.2.1 Child Behaviour Checklist

The CBCL[269] is a 100 item behavioural screening questionnaire completed by parents or carers. The items are checked from zero to two; parents are requested to circle a zero if an item is not true for their child, one if it is somewhat or sometimes true and two if it is often/very true. The

CBCL for two to three year olds includes six syndrome scales that are combined to give an overall Total Problem score: Anxious/depressed, Withdrawn, Aggressive, Destructive, Sleep problems and Somatic behaviour. The Anxious/depressed and Withdrawn scales are combined to calculate an Internalizing Problems score and the Aggressive and Destructive scales are used to generate an Externalizing Problems score. The Total Problem raw score is converted into a T-score (mean 50, SD 10) and can be prorated if the number of missing items is less than 10%. Scores above the 90th centile (T-score >60) are defined as clinically significant and scores from the 85th through the 90th centile (T-score 51-60) as borderline clinical. Scores below the 85th centile (T-score ≤ 50) are classified as normal.

8.2.2 Strengths and Difficulties Questionnaire

The Strengths and Difficulties Questionnaire (SDQ)[268] can be used as both a behavioural screening tool and dimensional measure[332,333] and has excellent psychometric properties for identifying children with behavioural and emotional difficulties in clinical and community populations.[334-336] including children born extremely preterm.[336] The questionnaire can be completed by parents and teachers for 3 to 16 year olds, and adolescents aged 11 to 16 years can complete the form themselves, reporting on perceived strengths and difficulties. The questionnaire is shorter than the CBCL, consisting of 25 items which are scored zero is if an item is not true, one if it is somewhat true and two if it is certainly true. The 25 items in the SDQ comprise five scales of five items each: Emotional Symptoms, Conduct Problems, Hyperactivity/Inattention, Peer Problems and Prosocial Behaviour. A Total Difficulties Score is calculated by combining the first four scales (excluding the Prosocial Scale) to give a total score of zero to 40, with higher scores indicating greater difficulties and a score of 17 and above indicating abnormal behaviour. Scale scores may be prorated if at least three of the five items

are completed. The presence of clinically significant problems in each domain and overall can be identified using published norms.[337] The normative data for the UK was based on a large national survey conducted by the Office for National Statistics in 1999 and included 10,298 parent and 8,208 teacher questionnaires for five to 15 year olds and 4,228 self-completed questionnaires from 11 to 15 year olds.

The SDQ also contains supplementary items on overall distress and social impairment which can be summed to generate an Impact Score that ranges from zero to 10 for the parent-completed version. The five questions include, "Do the difficulties upset or distress your child?" and "Do the difficulties interfere with your child's everyday life in the following areas? – home life; friendships; classroom learning; leisure activities" A score of zero is given if the response is not at all or only a little, one if quite a lot and two if a great deal. An overall impact score of two to 10 is classified as abnormal.

8.3 Missing data analysis due to dropout

The patterns of missing response to the parent questionnaires sent at six, 11, 16 and 19 years are shown in Table 8.1 for EPT participants who were classified into two groups according to their pattern of missing assessments; completers (all four assessments) and non-completers (three or less assessments). Twenty-six percent (n=82) of EPT participants had a complete set of four behavioural assessments, 19.7% (n=62) had three, 33.3% (n=105) had one or two and 21% (n=66) had no behavioural assessments.

Table 8.1: Pattern of missing parent-reported SDQ Total Difficulties Score among EPT participants (n=315)

Number of assessments	Assessment (year)				N	(%)
	6	11	16	19		
Four	+	+	+	+	82	(26.0)
Three	+	+	+		36	(11.4)
	+	+		+	19	(6.0)
		+	+	+	5	(1.6)
	+		+	+	2	(0.6)
Two	+	+			50	(15.9)
		+	+		4	(1.3)
		+		+	2	(0.6)
			+	+	2	(0.6)
	+			+	1	(0.3)
	+		+		1	(0.3)
One	+				29	(9.2)
		+			11	(3.5)
				+	3	(1.0)
			+		2	(0.6)
None					66	(21.0)

Maternal and infant characteristics collected at baseline, family SES at two years, mothers' highest educational qualification, and neurodevelopmental impairment at last assessment were compared between the two groups. Regression analysis was performed for each characteristic (linear for continuous variables and logistic for binary variables) to test for differences between the complete versus non-complete groups (Table 8.2). EPT participants with complete follow-up assessments were more likely to have mothers who were of white ethnicity, better educated and have fathers with a non-manual occupation. They also had higher BSID-II MDI and PDI scores at 2.5 years and were less likely to be diagnosed with moderate/severe CP. There was no evidence of a difference in behavioural problems at 2.5 years between completers and non-completers as assessed by parent-reported CBCL scores.

Table 8.2: Perinatal and neurodevelopmental characteristics of EPT participants according to completeness of behavioural screening assessments

	Completers: 4 assessments (n=82)	Non-completers <4 assessments (n=233)	P value[a]
Maternal/paternal factors			
Maternal age (years), mean [SD]:	29.4 [5.6] (n=81)	28.1 [6.0] (n=232)	0.10
Mother of non-white ethnicity, % (n):	12.4% (10/81)	27.0% (63/233)	0.01
Primigravida, % (n):	37.8% (31/82)	27.6% (64/232)	0.09
Mother smoked during pregnancy, % (n):	24.7% (20/81)	36.5% (73/200)	0.07
Antenatal steroids given, % (n):	81.7% (67/82)	77.9% (180/231)	0.53
Mother's highest educational qualification A 'level or above, % (n):[b]	47.5% (38/80)	31.4% (61/194)	0.01
Father's occupation non-manual, % (n):[b]	54.2% (39/72)	29.1% (46/158)	<0.001
Infant perinatal factors			
Multiple birth, % (n):	32.9% (27/82)	22.4% (52/232)	0.08
Congenital anomaly present, % (n):	2.4% (2/82)	2.6% (6/232)	1.00
Gestational age 24 weeks or less, % (n):	46.3% (38/82)	39.1% (91/233)	0.30
Birthweight (grams), mean [SD]:	733 [120] (n=82)	752 [112] (n=233)	0.19
Male sex, % (n):	41.5% (34/82)	51.9% (121/233)	0.12
Moderate/Severe brain injury during neonatal period, % (n):[c]	17.1% (14/82)	24.6% (57/232)	0.22
Laser or cryotherapy for retinopathy of prematurity, % (n):	13.4% (11/82)	14.9% (34/228)	0.86
Laparotomy for necrotising enterocolitis, % (n):	2.5% (2/79)	3.1% (7/229)	1.00
Supplemental oxygen at 36 weeks, % (n):	76.8% (63/82)	73.0% (170/233)	0.56
Infant developmental outcomes at 2.5 years			
BSID-II Mental Developmental Index, mean [SD]:	84.5 [12.7] (n=82)	76.5 [17.3] (n=201)	<0.001
BSID-II Psychomotor Developmental Index, , mean [SD]:	86.3 [15.1] (n=80)	78.7 [19.0] (n=182)	0.002
BSID-II Behaviour Rating Percentile, , mean [SD]:	39.2 [26.1] (n=79)	34.6 [24.9] (n=180)	0.18
CBCL Total Problem T-score, mean [SD]:	54.5 [9.2] (n=82)	56.2 [9.9] (n=188)	0.17
Neurodevelopmental impairment at last assessment[b]			
Cerebral palsy, % (n):	7.3% (6/82)	22.2% (50/225)	0.002
Moderate/severe cerebral palsy, % (n):	4.9% (4/82)	14.0% (31/222)	0.03
Moderate/severe cognitive impairment, % (n):	45.1% (37/82)	49.1% (103/210)	0.60
Moderate/severe visual impairment, % (n):	7.3% (6/82)	12.1% (27/224)	0.30
Moderate/severe hearing impairment, % (n):	1.2% (1/82)	2.2% (5/225)	1.00
Moderate/severe functional impairment, % (n):[d]	46.3% (38/82)	53.3% (112/210)	0.30
Severe functional impairment, % (n):[d]	18.3% (15/82)	27.1% (57/210)	0.13

[a] Two-sided p-values were calculated using Fisher's Exact Test for binomial variables and the t-test for continuous variables.

[b] Not collected at discharge; first collected at one or two year assessment.

[c] Parenchymal pathology and/or ventriculomegaly on worst cranial ultrasound scan before discharge home.

[d] Functional impairment includes cerebral palsy, cognitive, visual or hearing impairment.

The patterns of missing assessments at 6, 11, 16 and 19 years are shown in Table 8.3 for the term-born classmates, who were classified according to their pattern of missing assessments in the same way as for EPT participants: 18% percent (n=37) had four completed SDQ questionnaires, 18.9% (n=39) had three, 57.3% (n=118) had one or two and 5.8% (n=12) had no questionnaires.

Table 8.3: Pattern of missing parent-reported SDQ Total Difficulty Score among term-born classmates (n=206)

Number of assessments	Assessment (year)					
	6	11	16	19	n	(%)
Four	+	+	+	+	37	(18.0)
Three	+	+	+		26	(12.6)
	+	+		+	8	(3.9)
		+	+	+	4	(1.9)
	+		+	+	1	(0.5)
Two	+	+			33	(16.0)
		+	+		14	(6.8)
		+		+	4	(1.9)
	+			+	1	(0.5)
One	+				40	(19.4)
		+			22	(10.7)
			+		4	(1.9)
None					12	(5.8)

There were a limited number of factors that could be compared for the comparison group as, unlike the EPT group, comprehensive data were not collected at baseline or 2.5 years. Participant sex, mothers' highest educational qualification, and neurodevelopmental impairment were compared and there were no statistically significant differences between completers and non-completers in the comparison group. However, given the limited amount of data to make comparisons and the relatively small numbers, it is not possible to rule out that differences may actually exist.

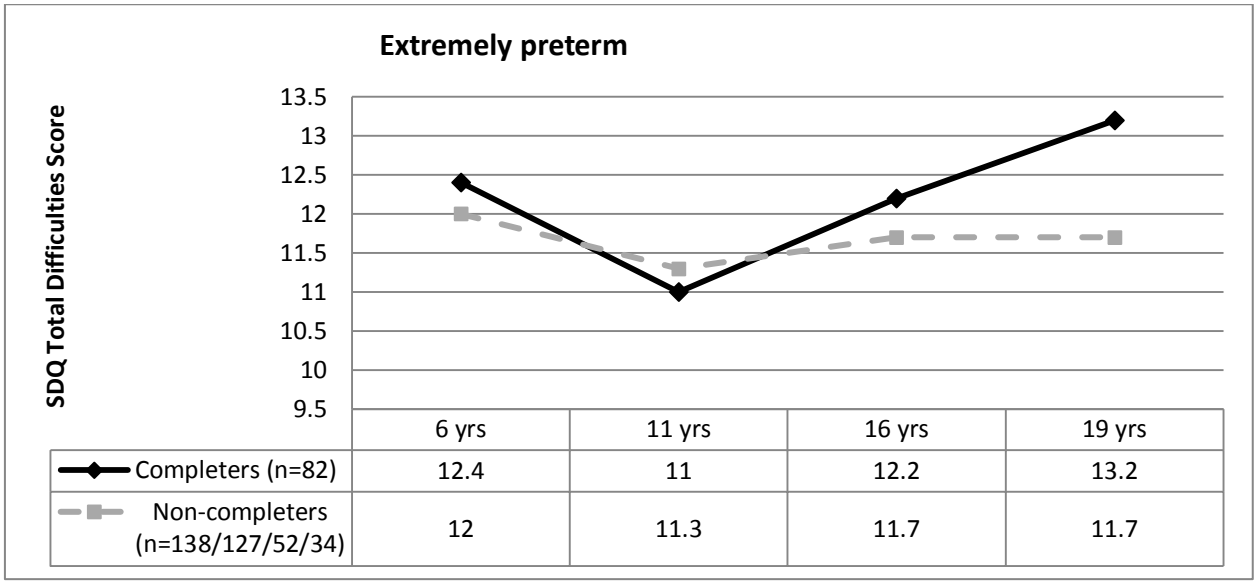
Table 8.4: Characteristics of term-born classmates according to completeness of behavioural screening assessments

	Completers: 4 assessments (n=37)	Non-completers <4 assessments (n=169)	P value[a]
Mother's highest educational qualification A 'level or above, % (n):	50.0% (18/36)	38.5% (57/148)	0.26
Male sex, % (n):	40.5% (15/37)	43.2% (73/169)	0.86
Any cognitive impairment at last assessment, % (n)	8.1% (3/37)	20.5% (34/166)	0.10
Moderate/severe cognitive impairment at last assessment, % (n):	0.0% (0/37)	3.0% (5/166)	0.59
Any visual impairment at last assessment, % (n):	32.4% (12/37)	18.2% (30/165)	0.07
Any hearing impairment at last assessment, % (n):	2.7% (1/37)	0.0% (0/165)	0.18
Any functional impairment at last assessment, % (n):[b]	40.5% (15/37)	34.6% (57/165)	0.57

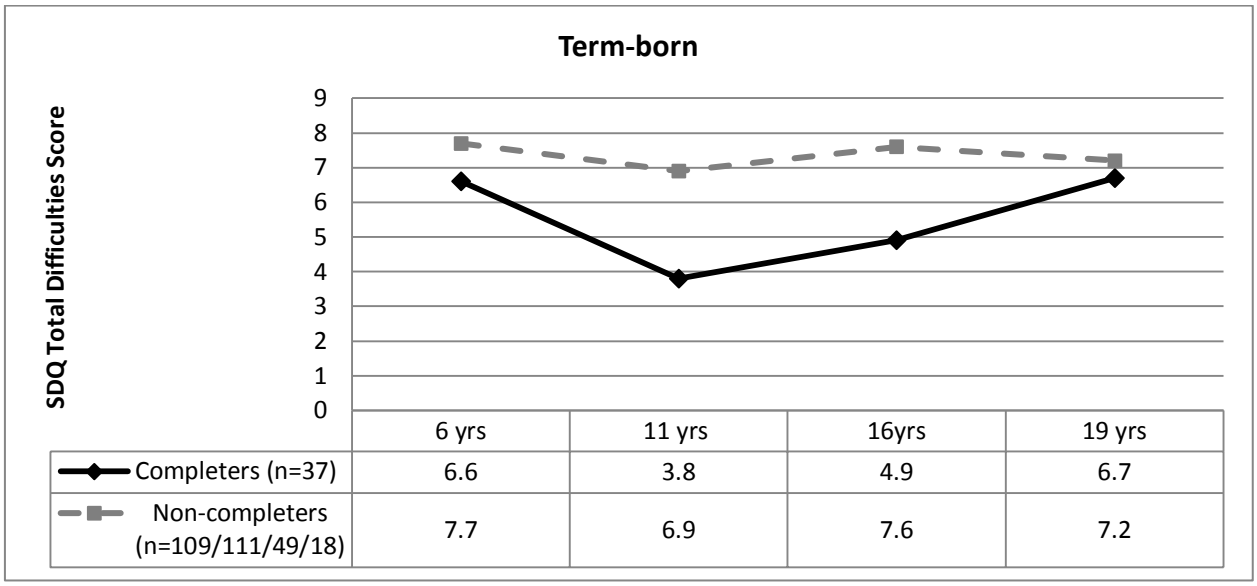
[a] Two-sided p-values were calculated using Fisher's Exact Test.
[b] Functional impairment includes cerebral palsy, cognitive, visual or hearing impairment.

The SDQ Total Difficulties score were calculated at each time point separately by dropout group and plotted graphically to investigate the missing data mechanism in relation to outcome (Figure 8.1). The scores were fairly similar in the completer and non-completer group among the EPT participants except at age 19 when the completers had slightly worse scores. Within the term-born comparison group, scores were higher in the non-completer group at all time points, though there was some convergence at 19 years.

Figure 8.1: Parent-report SDQ Total Difficulty Score in the EPT and term-born comparison group by dropout status



	6 yrs	11 yrs	16 yrs	19 yrs
Completers (n=82)	12.4	11	12.2	13.2
Non-completers (n=138/127/52/34)	12	11.3	11.7	11.7



	6 yrs	11 yrs	16yrs	19 yrs
Completers (n=37)	6.6	3.8	4.9	6.7
Non-completers (n=109/111/49/18)	7.7	6.9	7.6	7.2

8.4 Primary analysis of all participants with at least one assessment

8.4.1 Descriptive analysis of SDQ scores in EPT participants and term-born classmates

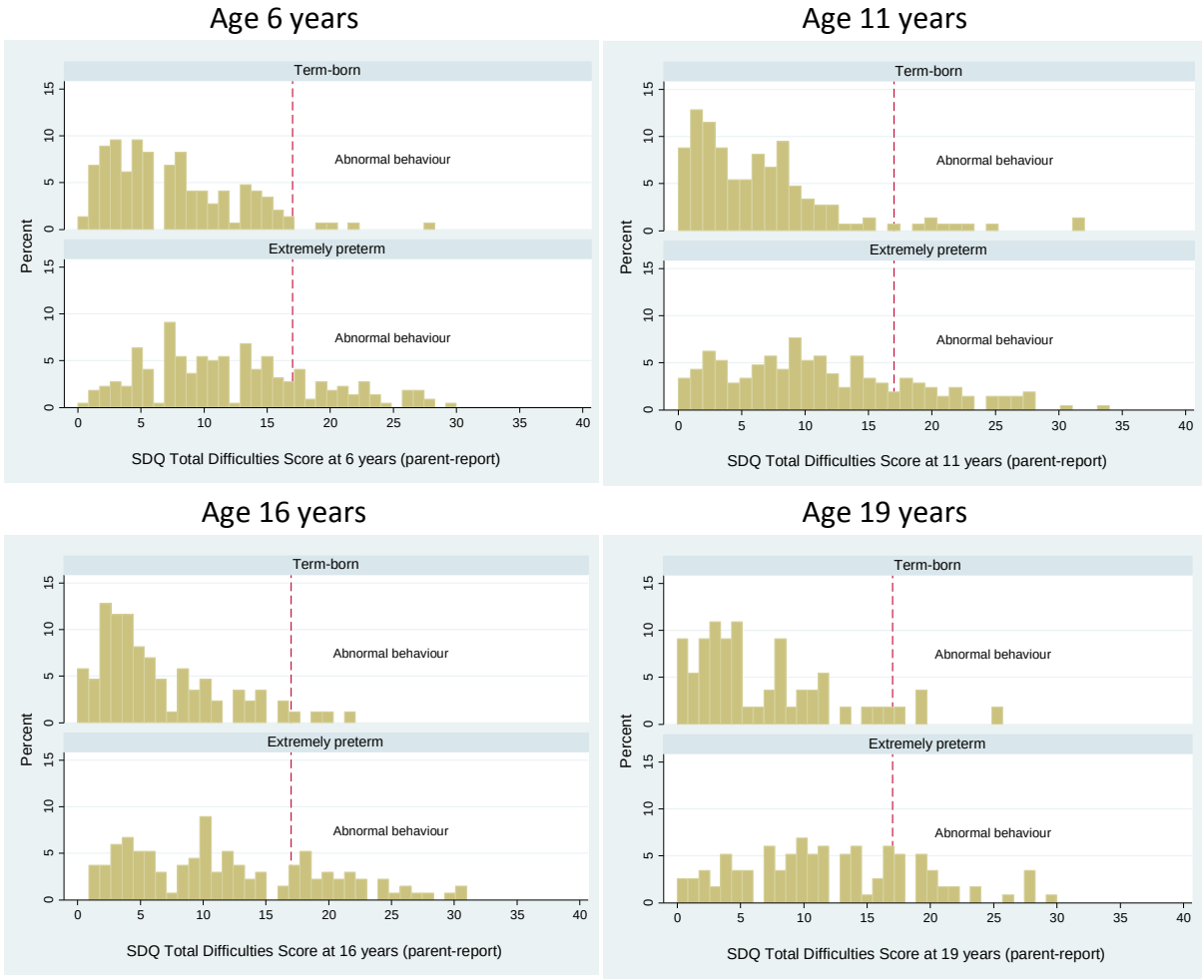
Summary statistics for the SDQ Total Difficulties and subscale scores are shown in Table 8.5 and the distributions of the scores are displayed in Figure 8.2 for each age of assessment. They were right-skewed in the term-born comparison group with more participants scoring at the low end of the scale indicating fewer difficulties. In comparison, the scores in the EPT group had a wider spread and a flatter distribution. The individual longitudinal trajectories are shown in Figure 8.3. They were not as flat as the cognitive trajectories and appear more variable over time, with a handful of individuals having large peaks. The mean SDQ scores and 95% confidence intervals at each age are presented in Table 8.6 and displayed graphically in Figure 8.4. The mean SDQ Total Difficulty scores in the EPT group were around five points higher than the comparison group at each assessment and remained stable in both groups over time. The difference appears to be mainly driven by the Hyperactivity/Inattention and Peer Problems subscales, though the gap between the two groups in the Emotional Symptoms score seems to widen over time. There was a slight increase in difficulties related to Conduct Problems and Prosocial Behaviour in the EPT group, but not as marked a difference as in the other domains. The Hyperactivity/Inattention score comprises of three items that relate to impulsivity and two items that relate to inattention; it is normally the latter behaviour that is affected by prematurity. However, both impulsivity and inattention accounted for an approximately equal proportion of the difference between the Hyperactivity/Inattention score between the two groups. The percentage of participants classified as being in the abnormal range for each subscale are also presented in Table 8.5 and displayed in Figure 8.5

Table 8.5: Parent-report SDQ Total Difficulties and subscale scores in the EPT and term-born group by age of assessment

	Age 6 years[a]		Age 11 years		Age 16 years		Age 19 years[b]	
	EPT (n=222)	Term (n=148)	EPT (n=209)	Term (n=148)	EPT (n=134)	Term (n=86)	EPT (n=117)	Term (n=55)
Age at assessment:								
Mean [SD]	6.3 [0.5]	6.1 [0.5]	10.9 [0.4]	11.0 [0.6]	17.1 [0.3]	17.0 [0.3]	19.3 [0.5]	19.2 [0.6]
Median {IQR}	6.3 {5.9, 6.7}	6.2 {5.8, 6.4}	10.9 {10.7, 11.2}	10.9 {10.6, 11.3}	17.0 {16.8, 17.3}	17.0 {16.7, 17.2}	19.3 {18.8, 19.7}	19.3 {18.8, 19.6}
Total Difficulties Score:								
Mean [SD]	12.3 [6.6]	7.4 [5.1]	11.1 [7.5]	6.2 [6.0]	11.9 [7.6]	6.4 [5.1]	12.2 [7.0]	6.8 [5.8]
Median {IQR}	12 {7, 17}	6 {3, 11}	10 {5, 16}	5 {2, 8}	10 {5.3, 18}	5 {3, 9}	12 {7, 17}	5 {3, 10}
Abnormal range (17-40), n (%)	56 (25.5)	6 (4.1)	48 (23.0)	10 (6.8)	40 (29.9)	4 (4.7)	35 (30.2)	5 (9.1)
Emotional Symptoms:								
Mean [SD]	2.4 [2.1]	1.9 [1.9]	2.7 [2.5]	1.6 [2.1]	3.2 [2.5]	1.5 [1.7]	3.9 [2.7]	2.1 [2.4]
Median {IQR}	2 {1, 4}	1 {0, 3}	2 {1, 4}	1 {0, 2}	3 {1, 5}	1 {0, 2}	4 {2, 6}	1 {0, 4}
Abnormal range (5-10), n (%)	34 (15.4)	12 (8.2)	51 (24.4)	13 (8.8)	41 (30.6)	7 (8.1)	43 (36.8)	12 (21.8)
Conduct Problems:								
Mean [SD]	2.2 [2.0]	1.5 [1.4]	1.5 [1.7]	1.1 [1.6]	1.3 [1.8]	1.3 [1.7]	1.6 [1.5]	1.4 [1.5]
Median {IQR}	2 {1, 3}	1 {0, 2}	1 {0, 3}	0 {0, 2}	1 {0, 2}	1 {0, 2}	1 {0, 2}	1 {0, 2}
Abnormal range (4-10), n (%)	51 (23.2)	13 (8.8)	27 (12.9)	11 (7.4)	18 (13.4)	8 (9.3)	12 (10.3)	4 (7.3)
Hyperactivity/Inattention:								
Mean [SD]	5.4 [2.7]	3.1 [2.4]	4.4 [2.8]	2.4 [2.3]	4.1 [2.9]	2.3 [2.3]	3.9 [2.7]	2.0 [2.3]
Median {IQR}	5 {3, 7.5}	3 {1, 5}	4 {2, 6}	2 {0.5, 3}	4 {2, 6}	1.6 {1, 3}	3 {2, 6}	1 {0, 3}
Abnormal range (7-10), n (%)	81 (36.5)	13 (8.8)	48 (23.0)	10 (6.8)	27 (20.2)	4 (4.7)	19 (16.2)	3 (5.5)
Peer Problems:								
Mean [SD]	2.2 [2.2]	1.0 [1.4]	2.5 [2.5]	1.0 [1.6]	3.2 [2.5]	1.4 [1.7]	2.9 [2.3]	1.3 [1.6]
Median {IQR}	2 {0, 3}	0 {0, 2}	2 {0, 4}	0 {0, 2}	3 {1, 5}	1 {0, 2}	2 {1, 5}	0 {1, 2}
Abnormal range (4-10), n (%)	49 (22.2)	10 (6.8)	68 (32.5)	11 (7.4)	58 (43.3)	8 (9.3)	42 (35.9)	5 (9.1)
Prosocial Behaviour:								
Mean [SD]	7.5 [2.3]	8.4 [1.5]	8.3 [2.3]	9.0 [1.3]	7.8 [2.5]	8.4 [1.7]	7.6 [2.6]	8.3 [1.9]
Median {IQR}	8 {6, 9}	9 {7, 10}	9 {7, 10}	10 {9, 10}	9 {6, 10}	9 {7, 10}	9 {6, 10}	9 {7, 10}
Abnormal range (0-4), n (%)	18 (8.3)	0 (0.0)	16 (7.7)	0 (0.0)	17 (12.7)	4 (4.7)	17 (14.7)	3 (5.6)

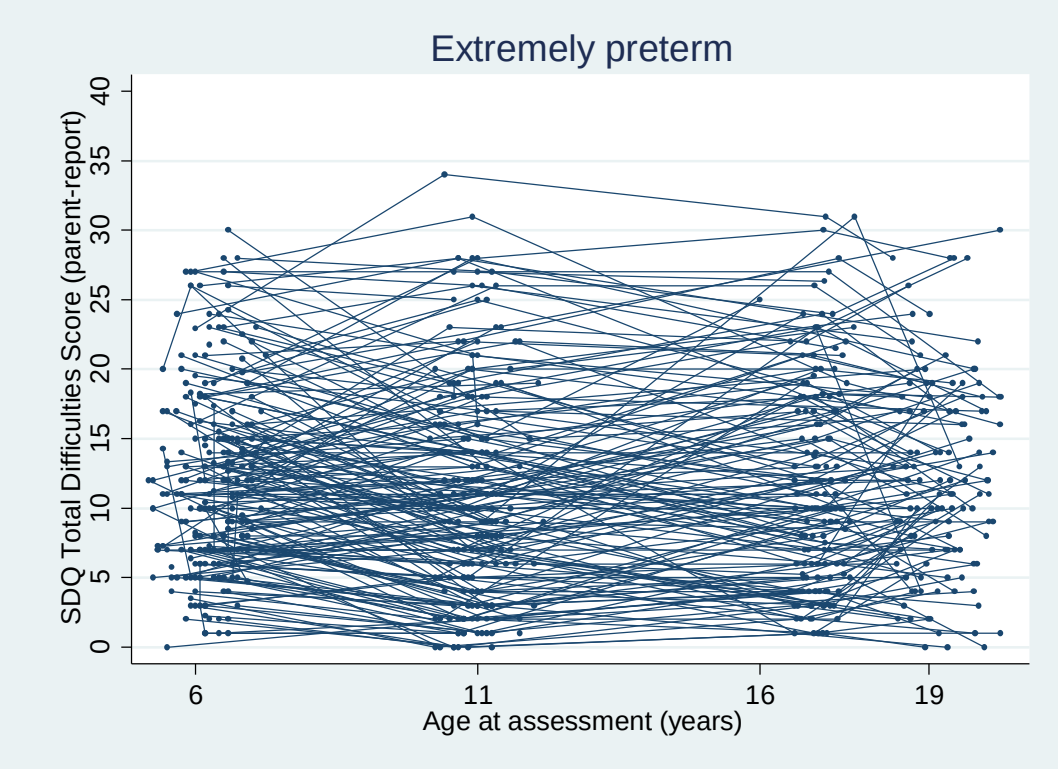
[a] 4 EPT and 2 term-born participants had at least one missing subscale score at age 6 years.
[b] 1 EPT and 1 term-born participant had at least one missing subscale score at age 19 years.
Higher scores indicate more behavioural problems (except for Prosocial Behaviour where a higher score indicates greater sociability).

Figure 8.2: Distribution of parent-report SDQ Total Difficulties Score in the EPT and term-born group by age of assessment



for each group. For the SDQ Total Difficulties score, around 25% of EPT participants fell into the abnormal range across all assessments compared to around 5% of the term-born comparison group. The percentage of EPT participants classified as abnormal was also greater for the Emotional Symptoms and Peer Problems scores and the gap appeared to widen over time, whereas the difference appeared to reduce over time for the Hyperactivity/Inattention subscale.

Figure 8.3: Longitudinal trajectories of parent-report SDQ Total Difficulties Scores in the EPT and term-born group from age 6 to 19 years



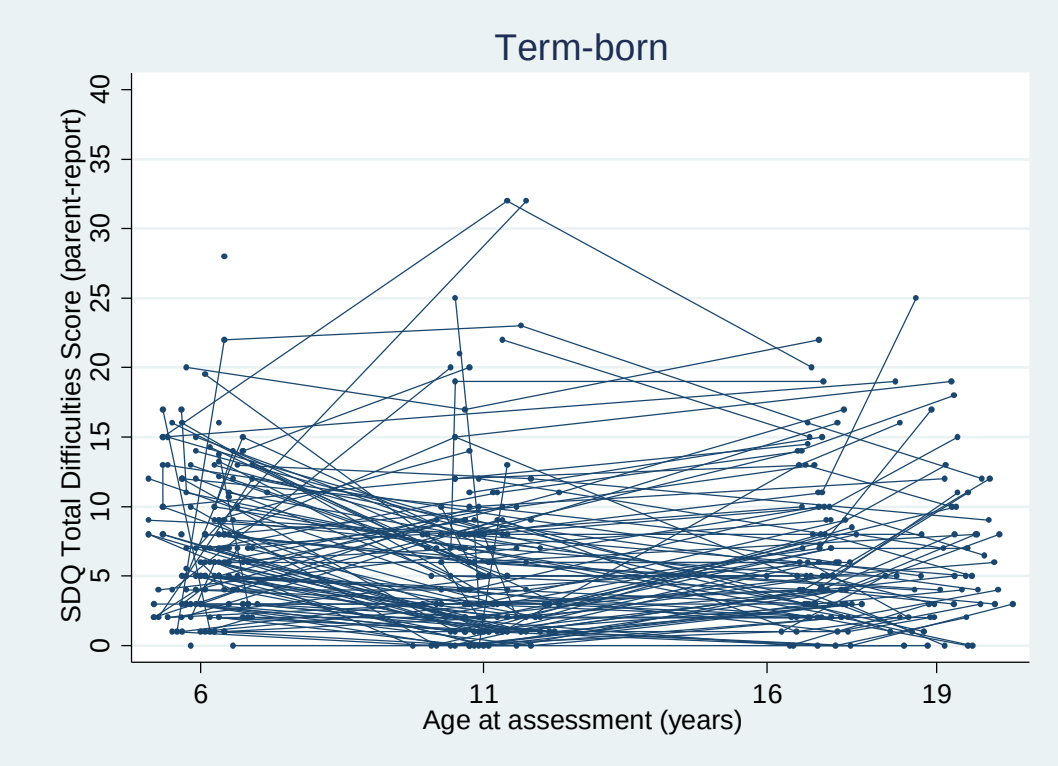


Table 8.6: Mean difference plus 95% confidence internals for parent-report SDQ Total Difficulties and subscale scores in the EPT and term-born group by age of assessment

	Age 6 years[a]		Age 11 years		Age 16 years		Age 19 years[b]	
	EPT (n=222)	Term (n=148)	EPT (n=209)	Term (n=148)	EPT (n=134)	Term (n=86)	EPT (n=117)	Term (n=55)
Total Difficulties Score:								
Mean	12.3	7.4	11.1	6.2	11.9	6.4	12.2	6.8
(95% confidence interval)	(11.4 to 13.1)	(6.6 to 8.3)	(10.1 to 12.1)	(5.2 to 7.1)	(10.6 to 13.2)	(5.3 to 7.5)	(10.9 to 13.5)	(5.3 to 8.4)
Mean difference (95% confidence interval)	4.8 (3.6 to 6.1)		4.9 (3.5 to 6.4)		5.5 (3.6 to 7.3)		5.3 (3.2 to 7.5)	
Emotional Symptoms:								
Mean	2.4	1.9	2.7	1.6	3.2	1.5	3.9	2.1
(95% confidence interval)	(2.2 to 2.7)	(1.6 to 2.2)	(2.4 to 3.1)	(1.3 to 2.0)	(2.8 to 3.6)	(1.1 to 1.9)	(3.4 to 4.4)	(1.5 to 2.8)
Mean difference (95% confidence interval)	0.6 (0.2 to 1.0)		1.1 (0.6 to 1.6)		1.7 (1.1 to 2.3)		1.7 (0.9 to 2.6)	
Conduct Problems:								
Mean	2.2	1.5	1.5	1.1	1.3	1.3	1.6	1.4
(95% confidence interval)	(1.9 to 2.4)	(1.2 to 1.7)	(1.3 to 1.7)	(0.8 to 1.3)	(1.0 to 1.6)	(0.9 to 1.6)	(1.3 to 1.8)	(1.0 to 1.8)
Mean difference (95% confidence interval)	0.7 (0.3 to 1.1)		0.5 (0.1 to 0.8)		0.1 (-0.4 to 0.5)		0.2 (-0.3 to 0.6)	
Hyperactivity/Inattention:								
Mean	5.4	3.1	4.4	2.4	4.1	2.3	3.9	2.0
(95% confidence interval)	(5.1 to 5.8)	(2.7 to 3.5)	(4.0 to 4.7)	(2.1 to 2.8)	(3.6 to 4.6)	(1.8 to 2.8)	(3.4 to 4.3)	(1.4 to 2.6)
Mean difference (95% confidence interval)	2.3 (1.8 to 2.9)		1.9 (1.4 to 2.5)		1.8 (1.1 to 2.6)		1.8 (1.0 to 2.7)	
Peer Problems:								
Mean	2.2	1.0	2.5	1.0	3.2	1.4	2.9	1.3
(95% confidence interval)	(1.9 to 2.5)	(0.8 to 1.2)	(2.1 to 2.8)	(0.8 to 1.3)	(2.8 to 3.7)	(1.0 to 1.7)	(2.5 to 3.3)	(0.8 to 1.7)
Mean difference (95% confidence interval)	1.2 (0.8 to 1.6)		1.4 (1.0 to 1.9)		1.9 (1.3 to 2.5)		1.6 (1.0 to 2.3)	
Prosocial Behaviour:								
Mean	7.5	8.4	8.3	9.0	7.8	8.4	7.6	8.3
(95% confidence interval)	(7.2 to 7.8)	(8.1 to 8.6)	(8.0 to 8.6)	(8.8 to 9.2)	(7.4 to 8.2)	(8.0 to 8.8)	(7.1 to 8.1)	(7.8 to 8.8)
Mean difference (95% confidence interval)	-0.9 (-1.3 to -0.4)		-0.8 (-1.2 to -0.3)		-0.6 (-1.2 to -0.02)		-0.7 (-1.5 to 0.07)	

[a] 4 EPT and 2 term-born participants had at least one missing subscale score at age 6 years. [b] 1 EPT and 1 term-born participant had at least one missing subscale score at age 19 years.
[c] Missing data at 6 years: 3 EPT, 11 years: 3 EPT and 2 term-born, 16 years: 8 EPT, 19 years: 2 EPT.
Higher scores indicate more behavioural problems (except for Prosocial Behaviour where a higher score indicates greater sociability).

Figure 8.4: Mean and 95% confidence intervals for SDQ Total Difficulties and subscale scores in the EPT and term-born group at age 6, 11, 16 and 19

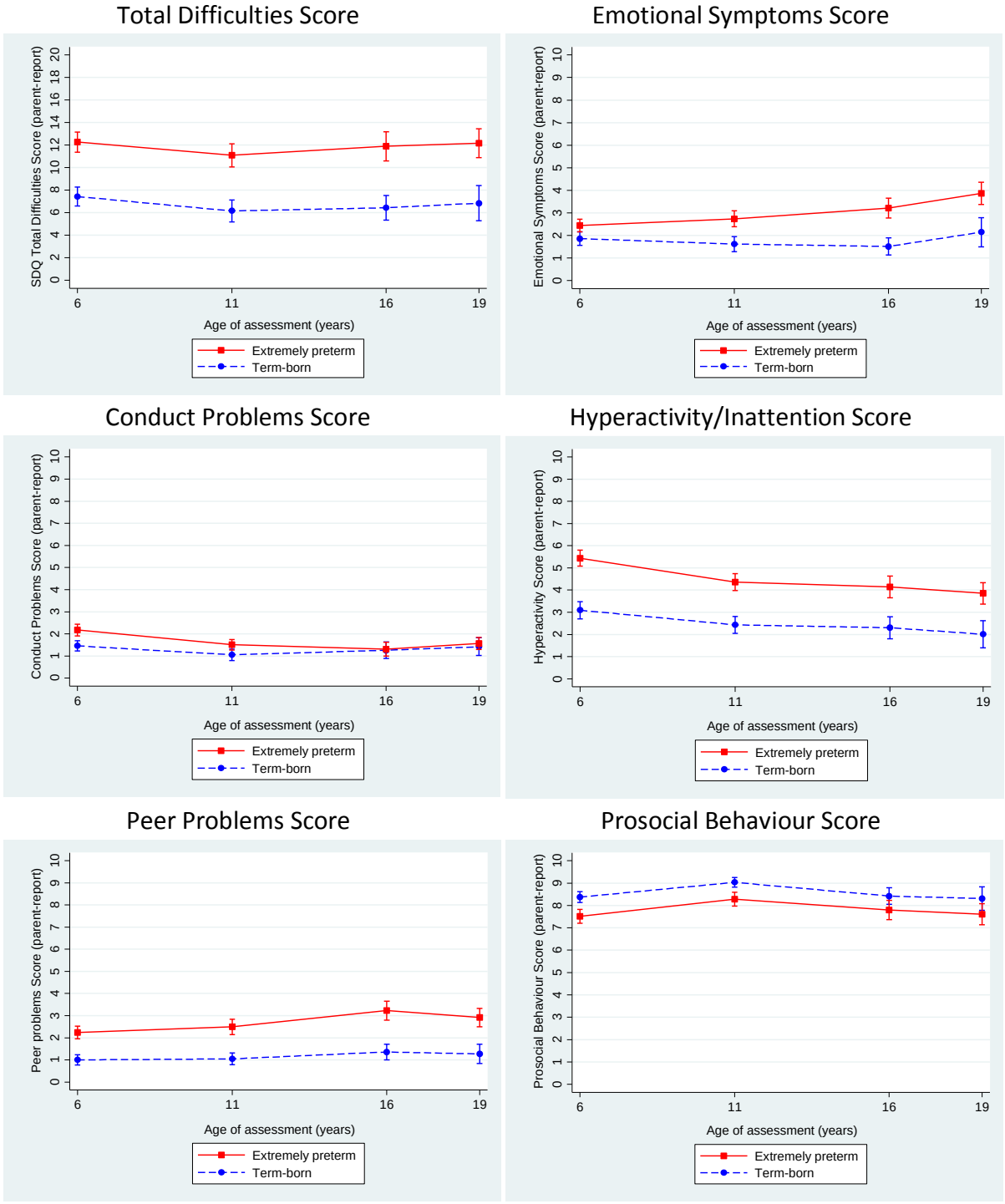
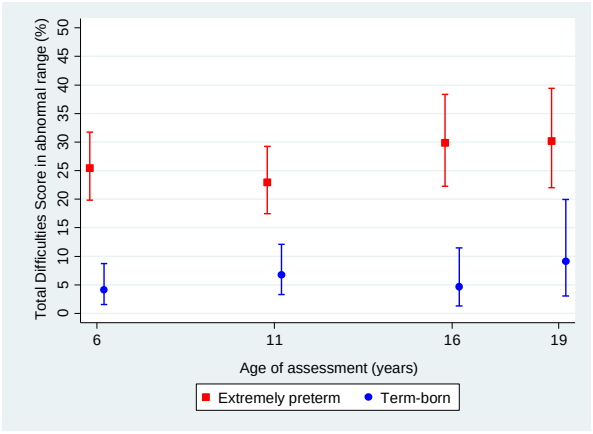


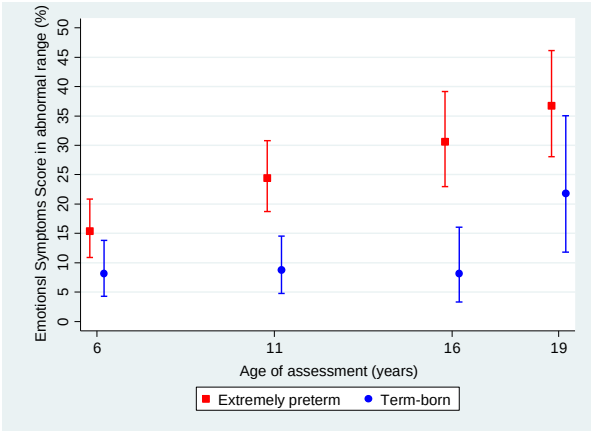
Figure 8.5: Percentage in abnormal range and 95% confidence intervals for SDQ Total Difficulties and subscale scores in the EPT and term-born group at age 6, 11, 16 and 19

Total Difficulties Score



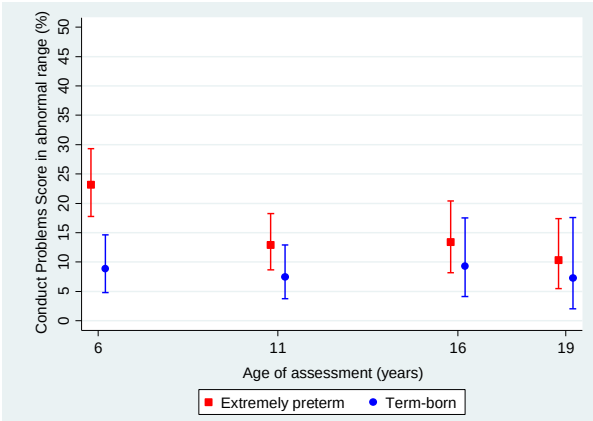
RR=4.48 (95% CI: 2.82 to 7.11)

Emotional Symptoms Score



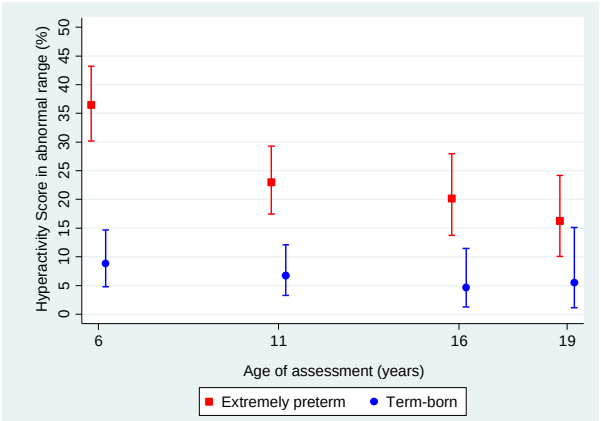
RR=2.31 (95% CI: 1.59 to 3.34)

Conduct Problems Score



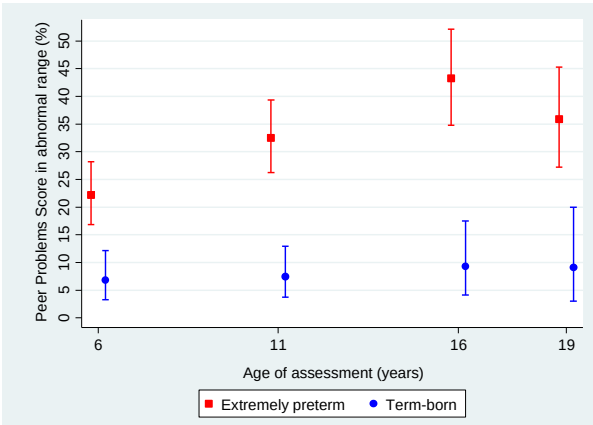
RR=1.98 (95% CI: 1.27 to 3.1)

Hyperactivity/Inattention Score



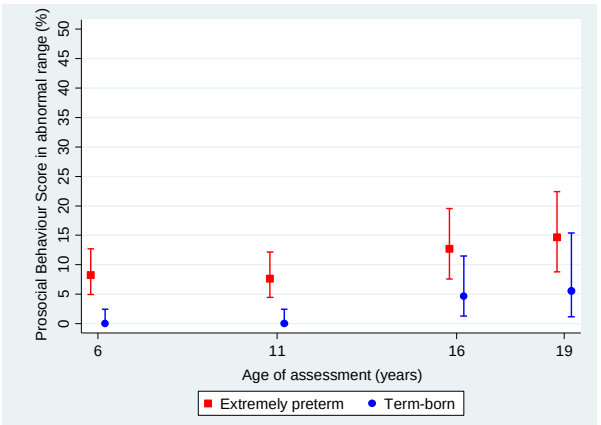
RR=3.80 (95% CI: 2.52 to 5.74)

Peer Problems Score



RR=3.93 (95% CI: 2.65 to 5.83)

Prosocial Behaviour Score



RR=5.92 (95% CI: 2.44 to 14.3)

8.4.2 Comparative analysis of SDQ scores in the EPT and term-born group

8.4.2.1 The SDQ Total Difficulties Score

The key stages of model development for the Total Difficulties Score are summarised in Table 8.7. This model was based on 443 participants (249 EPT and 194 term-born participants) and 1114 scores. Age was centred at age six years to make the intercept coefficient more meaningful; it represents the mean difference between the EPT group and comparison group at age six. The reference model (model one) fitted age as a random effect and various parameters were added in turn and retained if the fit of the model was significantly improved ($p \leq 0.05$ in the LRT). The equations, LRT statistics and corresponding p-values are displayed for each model. In model two, EPT exposure status was added as a fixed effect, and in model three, an EPT*Age interaction was included but was not statistically significant. This implies that the cognitive trajectories of the EPT and term-born participants have a different intercept and but a similar slope over time. Fitting age as a non-linear quadratic function (model four) significantly improved the fit of the model, but curvature did not vary by group (model five). Heteroscedasticity was assumed to be constant within and between groups in models one to five, meaning that the estimates for the between- and within-individual differences were assumed to have constant variation for all participants. In model six the within-individual variance was allowed to differ between the EPT and term-born groups, and in model seven the between-individual variance was allowed to differ. Both changes resulted in models with a significantly improved fit. The estimated coefficients, their standard errors and p-values (Wald test) for the fixed and random parts of the final models for the Total Difficulties Score (model seven) and component subscales are presented in Table 8.8. The predicted mean

Table 8.7: Log-likelihood statistics for key model-fitting stages in the mixed model analysis of the SDQ Total Difficulties Score in the EPT and term-born group

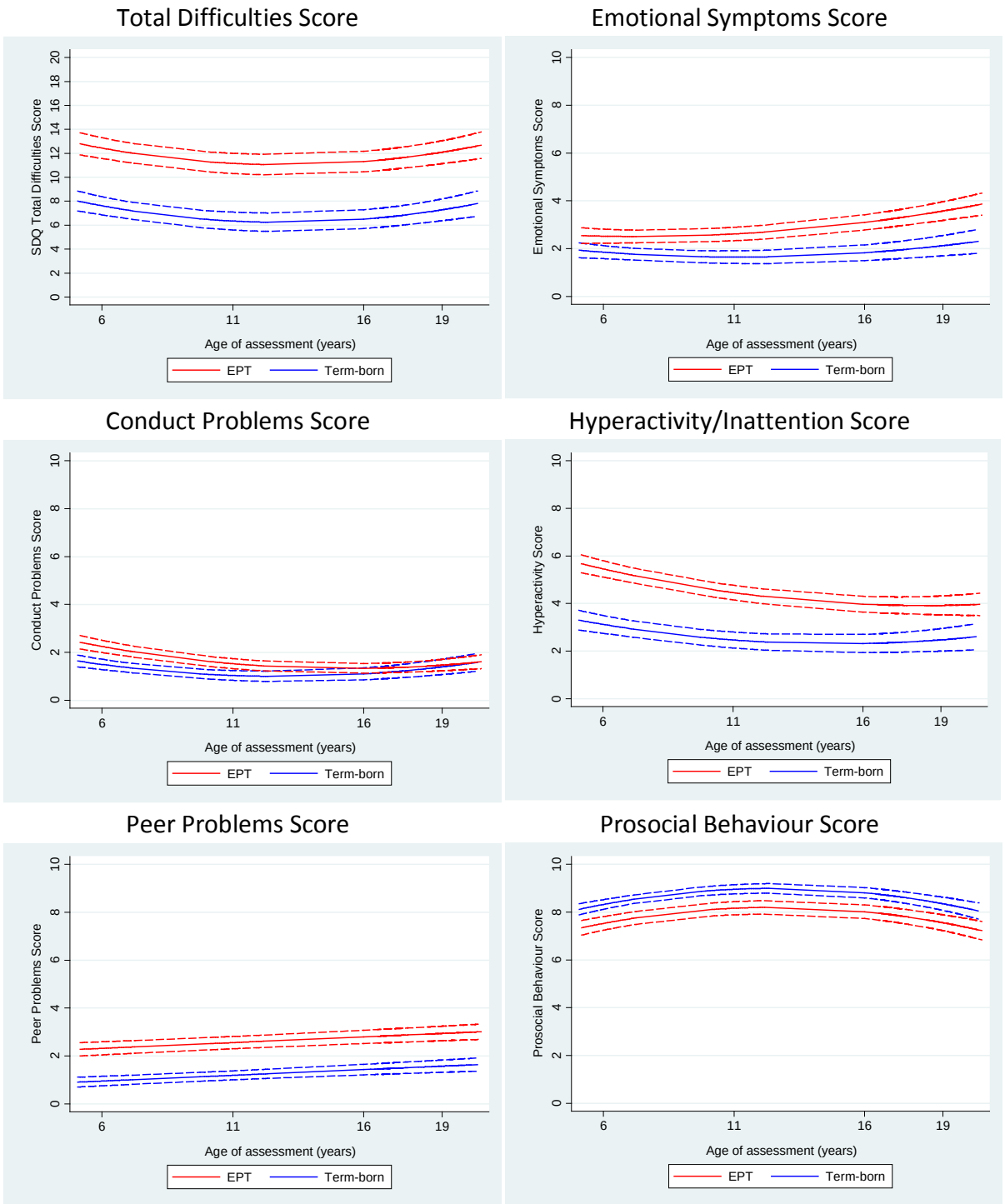
Model Number	Addition to model	Model	Likelihood ratio statistic (LRT)[a]	p-value[a]
1	Linear age	$y_{ij} = (\beta_0 + \mu_{0j}) + (\beta_1 + \mu_{1j})\, Age_{ij} + e_{ij}$	reference	
2	EPT	$y_{ij} = (\beta_0 + \mu_{0j}) + (\beta_1 + \mu_{1j})\, Age_{ij} + \beta_2 EPT + e_{ij}$	68.1 (1 d.f.)	<0.001
3	EPT*Age interaction	$y_{ij} = (\beta_0 + \mu_{0j}) + (\beta_1 + \mu_{1j})\, Age_{ij} + \beta_2 EPT + \beta_3 EPT*Age_{ij} + e_{ij}$	0.04 (1 d.f.)	0.85
4	Non-linear age	$y_{ij} = (\beta_0 + \mu_{0j}) + (\beta_1 + \mu_{1j})\, Age_{ij} + \beta_2 EPT + \beta_3 Age^2_{ij} + e_{ij}$	18.5 (1 d.f.)	<0.001
5	EPT*Non-linear age interaction	$y_{ij} = (\beta_0 + \mu_{0j}) + (\beta_1 + \mu_{1j})\, Age_{ij} + \beta_2 EPT + \beta_3 Age^2_{ij} + \beta_4 EPT* Age^2_{ij} + e_{ij}$	0.4 (1 d.f.)	0.52
6	Different level-1 (within-individual) variance for group	$y_{ij} = (\beta_0 + \mu_{0j}) + (\beta_1 + \mu_{1j})\, Age_{ij} + (\beta_2 + e_{1ij}) EPT + \beta_3 Age^2_{ij} + e_{0ij} Term$	4.0 (1 d.f.)	0.05
7	Different level-2 (between-individual) variance for group	Term: $y_{ij} = (\beta_0 + \mu_{00j}) + (\beta_1 + \mu_{01j}) Age_{ij} + \beta_3 Age^2_{ij} + e_{0ij}$ EPT: $y_{ij} = (\beta_0 + \mu_{10j}) + (\beta_1 + \mu_{11j}) Age_{ij} + \beta_2 + \beta_3 Age^2_{ij} + e_{1ij}$	14.4 (3 d.f.)	0.002

[a] The LRT statistic reported for model 2 is compared to the reference model. The LRT statistics for subsequent models are compared to the last model that improved the fit with p≤0.05.

Table 8.8: Mixed model analysis of the SDQ Total Difficulties Score and subscales in the EPT and term-born group

SDQ Total Difficulties Score (n=443)			Emotional Symptoms Score (n=443)			Conduct Problems Score (n=443)		
Parameter	**Estimate**	**95% CI**	**Parameter**	**Estimate**	**95% CI**	**Parameter**	**Estimate**	**95% CI**
Fixed effects			**Fixed effects**			**Fixed effects**		
Constant	7.62	(6.87 to 8.38)	Constant	1.85	(1.58 to 2.13)	Constant	1.50	(1.29 to 1.72)
EPT	4.81	(3.76 to 5.87)	EPT	0.67	(0.29 to 1.05)	EPT	0.75	(0.42 to 1.07)
Age	-0.40	(-0.57 to -0.23)	Age	-0.08	(-0.15 to -0.01)	Age	-0.14	(-0.20 to -0.09)
Age^2	0.03	(0.02 to 0.04)	EPT*Age	0.06	(0.01 to 0.11)	EPT*Age	-0.05	(-0.09 to -0.02)
Random effects			Age^2	0.008	(0.003 to 0.013)	Age^2	0.01	(0.007 to 0.014)
Within-individual			**Random effects**			**Random effects**		
EPT SD	3.70	(3.40 to 4.03)	Within-individual SD	1.52	(1.42 to 1.63)	Within-individual		
Term SD	3.47	(3.07 to 3.94)	Between-individual			EPT SD	1.14	(1.05 to 1.23)
Between-individual			EPT intercept SD	1.68	(1.42 to 1.99)	Term SD	1.04	(0.93 to 1.16)
EPT intercept SD	5.97	(5.28 to 6.76)	EPT slope SD	0.12	(0.09 to 0.17)	Between-individual		
EPT slope SD	0.29	(0.20 to 0.41)	EPT corr(intercept, slope)	-0.02	(-0.38 to 0.34)	EPT intercept SD	1.63	(1.42 to 1.86)
EPT corr(intercept, slope)	-0.20	(-0.46 to 0.09)	Term intercept SD	0.98	(0.65 to 1.48)	EPT slope SD	0.10	(0.08 to 0.14)
Term intercept SD	4.07	(3.32 to 4.98)	Term slope SD	0.06	(0.01 to 0.26)	EPT corr(intercept, slope)	-0.69	(-0.80 to -0.53)
Term slope SD	0.25	(0.14 to 0.46)	Term corr(intercept, slope)	0.88	(-1.00 to 1.00)	Term intercept SD	0.99	(0.78 to 1.27)
Term corr(intercept, slope)	-0.06	(-0.51 to 0.41)				Term slope SD	0.07	(0.04 to 0.13)
						Term corr(intercept, slope)	0.05	(-0.52 to 0.59)
Hyperactivity/Inattention Score (n=444)			**Peer Problems Score (n=443)**			**Prosocial Behaviour Score (n=443)**		
Parameter	**Estimate**	**95% CI**	**Parameter**	**Estimate**	**95% CI**	**Parameter**	**Estimate**	**95% CI**
Fixed effects			**Fixed effects**			**Fixed effects**		
Constant	3.13	(2.75 to 3.50)	Constant	0.95	(0.75 to 1.15)	Constant	8.32	(8.12 to 8.52)
EPT	2.33	(1.84 to 2.83)	EPT	1.37	(1.05 to 1.68)	EPT	-0.79	(-1.10 to -0.48)
Age	-0.18	(-0.26 to -0.11)	Age	0.05	(0.03 to 0.07)	Age	0.20	(0.15 to 0.26)
EPT*Age	-0.07	(-0.12 to -0.02)	**Random effects**			Age^2	-0.02	(-0.02 to -0.01)
Age^2	0.01	(0.01 to 0.02)	Within-individual			**Random effects**		
Random effects			EPT SD	1.38	(1.27 to 1.50)	Within-individual		
Within-individual			Term SD	1.17	(1.03 to 1.31)	EPT SD	1.24	(1.14 to 1.35)
EPT SD	1.61	(1.49 to 1.75)	Between-individual			Term SD	1.17	(1.05 to 1.31)
Term SD	1.32	(1.18 to 1.47)	EPT intercept SD	1.85	(1.61 to 2.12)	Between-individual		
Between-individual			EPT slope SD	0.09	(0.06 to 0.14)	EPT intercept SD	1.97	(1.74 to 2.24)
Intercept SD	2.14	(1.93 to 2.37)	EPT corr(intercept, slope)	-0.12	(-0.44 to 0.23)	EPT slope SD	0.14	(0.11 to 0.18)
Slope SD	0.14	(0.11 to 0.18)	Term intercept SD	0.91	(0.65 to 1.28)	EPT corr(intercept, slope)	-0.29	(-0.49 to -0.06)
Corr(intercept, slope)	-0.36	(-0.52 to -0.18)	Term slope SD	0.07	(0.04 to 0.15)	Term intercept SD	0.85	(0.61 to 1.20)
			Term corr(intercept, slope)	0.09	(-0.64 to 0.73)	Term slope SD	0.11	(0.08 to 0.17)
						Term corr(intercept, slope)	-0.38	(-0.71 to 0.09)

Figure 8.6: Predicted mean SDQ score and 95% confidence intervals for SDQ Total Difficulties and subscale scores in the EPT and term-born group



trajectories and 95% confidence intervals for the fixed portion of the model are displayed in Figure 8.6.

The fixed part of the model for the Total Difficulties Score is as follows:

Total Difficulties Score = 7.62 + 4.81EPT - 0.40(Age-6) + 0.03(Age-6)^2

On average, the scores of EPT participants were 4.81 points above their term-born peers at age 6 (95% CI: 3.76 to 5.87, p<0.001); almost one standard deviation above the comparison group (SD 5.1). Scores in both groups decreased slightly to age 11 when they started increasing slightly again. For the random part of the model, the estimated within-individual variation was 3.5 points for classmates and 3.7 points for EPT participants, so the scores of EPT participants at different time points varied slightly more around their average score than in term-born participants. The estimated function for between-individual variation in classmates is given by:[d]

$$\text{var}(\mu_{00j} + \mu_{01j}\text{Age}_{ij}) = 4.07^2 - 2 \times 0.07(\text{Age-6})_{ij} + 0.25^2(\text{Age-6})^2{}_{ij}$$

and for EPT participants,

$$\text{var}(\mu_{10j} + \mu_{11j}\text{Age}_{ij}) = 5.97^2 - 2 \times 0.34(\text{Age-6})_{ij} + 0.29^{2(}\text{Age-6})^2{}_{ij}$$

These functions are plotted in Figure 8.7 and indicate that the variance in the Total Difficulties Score between EPT participants was higher than among classmates and increased slightly in both groups over time. The SD around the mean intercept was lower in classmates (4.07

[d] Covariance(intercept, slope) = correlation(intercept, slope) x SD(intercept) x SD(slope)

versus 5.97 points), as was the SD around the mean slope (0.25 versus 0.29 points). EPT individuals were at increased risk of having clinically significant overall difficulties compared to their term-born peers (RR 4.48, 95% CI: 2.82 to 7.11) (Figure 8.5).

8.4.2.2 The SDQ Subscale Scores

Results from the final fitted models are presented in Table 8.8 and model predictions in Figure 8.6. There was a significant difference between the EPT and comparison group for every subscale at age six, although the difference seen in the Total Difficulties Score was mainly driven by hyperactivity/inattention (Mean difference: 2.33, 95% CI: 1.84 to 2.83) and peer relationship problems (Mean difference: 1.37, 95% CI: 1.05 to 1.68). Both groups displayed a similar trajectory for peer relationship problems but there was evidence of a slight decline in hyperactivity/inattention problems in the EPT group over time. There was evidence of a different trajectory of emotional symptoms between groups which, although increased in both groups with age, this was to a greater degree in the EPT individuals. The mean score for Conduct Problems was marginally elevated in the EPT group and showed a decline relative to the controls over time, and the mean score for Prosocial Behaviour was slightly depressed (worse performance) and followed a similar trajectory in both groups. The risk of being classified in the abnormal range was significantly higher for EPT participants compared to term-born classmates for every subscale (Figure 8.5).

8.4.2.3 The SDQ Impact Score

The risk of having behavioural problems that have a substantial impact on home life, friendships, school or work and/or leisure activities was significantly greater for EPT participants, increasing from 29% (63/222) at 2.5 years to 43% (49/117) at 19 years, compared to 5% at 2.5 years to 13% at 19 years among classmates (Figure 8.8 and Table 8.9). The risk

ratio of having an abnormal score of 2 or above in EPT group compared to the term-born group was 4.28 (95% CI: 2.89 to 6.35, p<0.001).

Figure 8.7: Predicted between-individual variance function for SDQ Total Difficulties Score in the EPT and term-born group

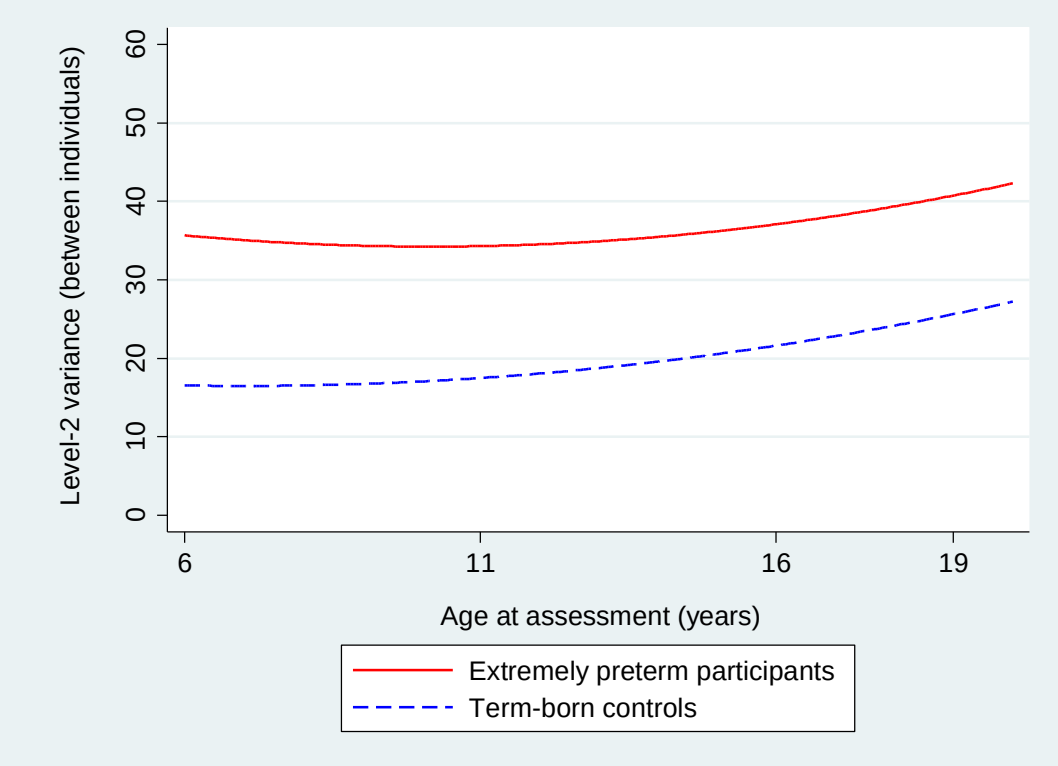
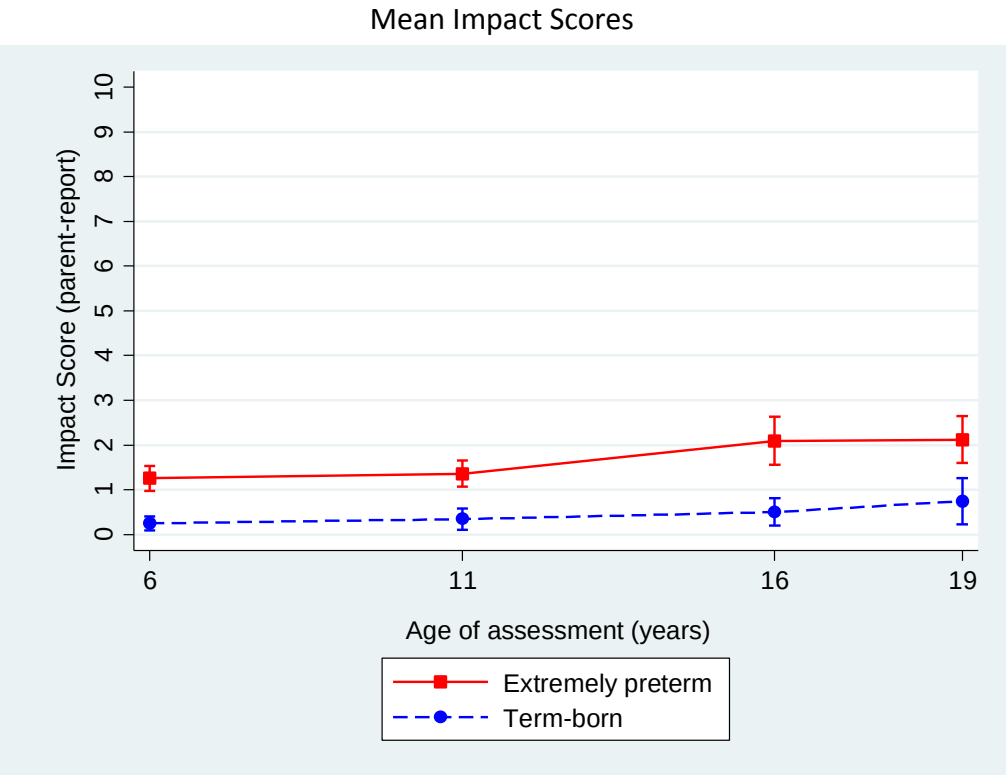


Figure 8.8: SDQ Impact Scores in the EPT and term-born group at age 6, 11, 16 and 19



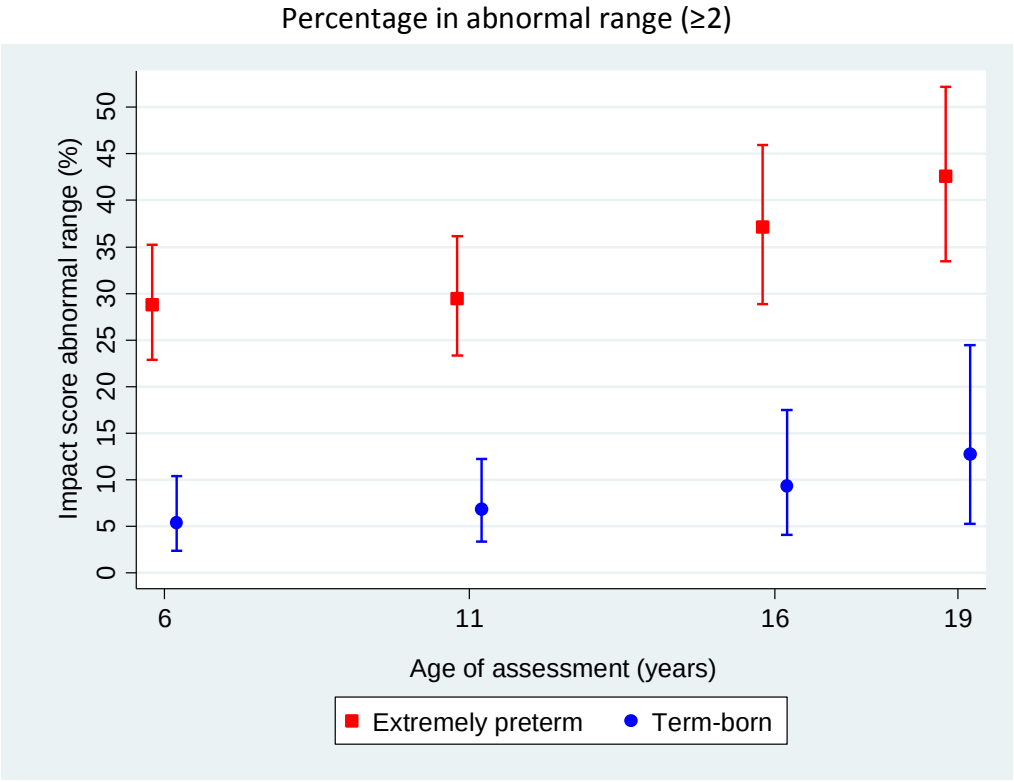


Table 8.9: Parent-report SDQ Impact Scores in the EPT and term-born group by age of assessment

	Age 6 years		Age 11 years		Age 16 years		Age 19 years	
	EPT (n=222)	**Term (n=148)**	**EPT (n=209)**	**Term (n=148)**	**EPT (n=134)**	**Term (n=86)**	**EPT (n=117)**	**Term (n=55)**
Impact Score[a]:								
Mean	1.3	0.3	1.4	0.3	2.1	0.5	2.1	0.8
(95% confidence interval)	(1.0 to 1.5)	(0.1 to 0.4)	(1.1 to 1.7)	(0.1 to 0.6)	(1.6 to 2.6)	(0.2 to 0.8)	(1.6 to 2.6)	(0.2 to 1.3)
Mean difference (95% confidence interval)	1.0 (0.6 to 1.4)		1.0 (0.6 to 1.4)		1.6 (0.9 to 2.3)		1.4 (0.6 to 2.2)	
Abnormal range (2-10), n (%)	63 (28.8)	8 (5.4)	61 (29.5)	10 (6.9)	49 (37.2)	8 (9.3)	49 (42.6)	7 (12.7)

[a] Missing data at 6 years: 3 EPT, 11 years: 3 EPT and 2 term-born, 16 years: 8 EPT, 19 years: 2 EPT.
Higher scores indicate more behavioural problems (except for Prosocial Behaviour where a higher score indicates greater sociability).

8.4.3 Diagnostics and model fit

The standardised level one residuals were approximately Normally distributed (Figure 8.9), and were fairly randomly scattered when plotted against the predicted scores and age at assessment. Most of the standardised level one residuals had values <|2| though there were a handful of outliers. The predicted random intercepts and slopes were also approximately Normally distributed (Figure 8.10), though the predicted random intercepts in both groups had a slight positive skew indicating that a greater proportion of participants had higher predicted intercepts than might be expected in a Normal distribution. The caterpillar plots in Figure 8.11 rank the level two residual for each child with 95% confidence intervals. Almost all of the confidence intervals overlapped zero in the predicted random slope plots, but in the predicted random intercept plots a small group of individuals at the lower and upper end of the plots (particularly in the EPT group) had confidence intervals for their residuals that did not overlap zero. Since these residuals represent individual departures from the overall average intercept predicted by the fixed parameters β_0 and β_1, this means that these are the individuals that differed significantly from the average at the 5% level.

Figure 8.9: Standardised level one residual diagnostic plots for SDQ Total Difficulties Score

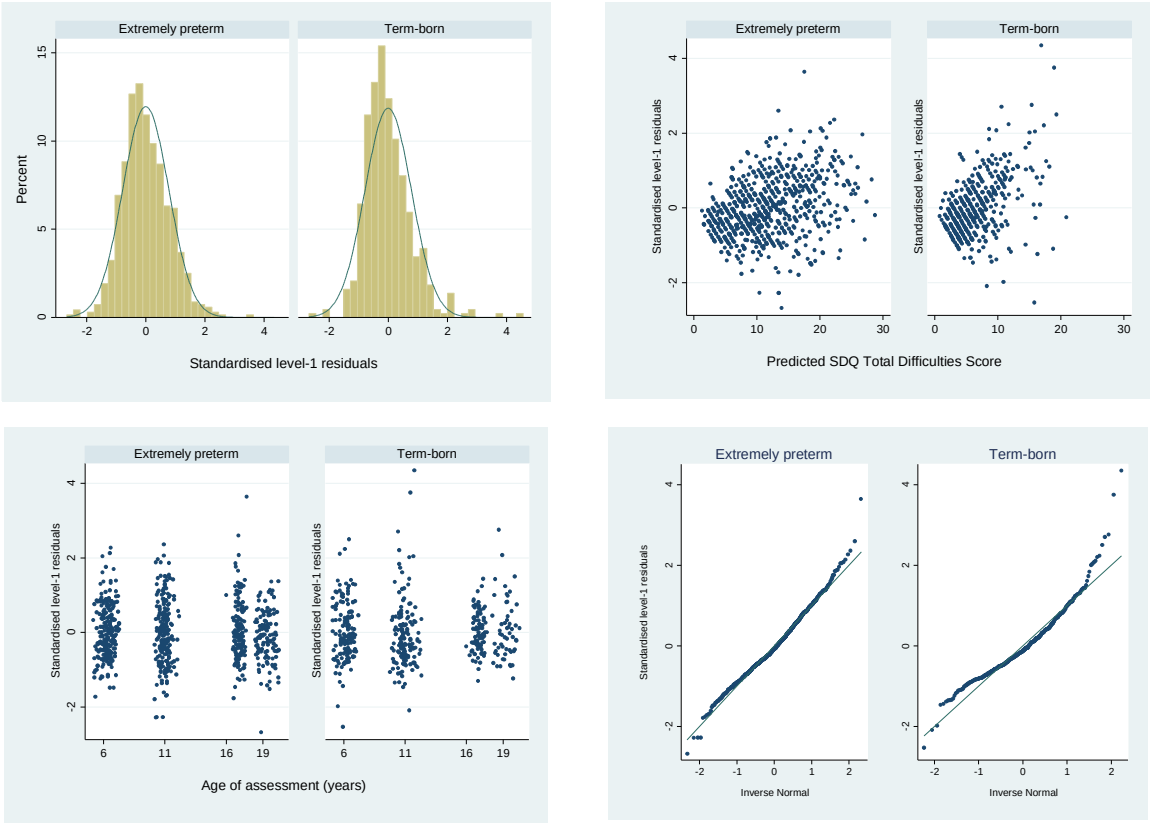


Figure 8.10: Predicted random intercepts and slopes for SDQ Total Difficulties Score

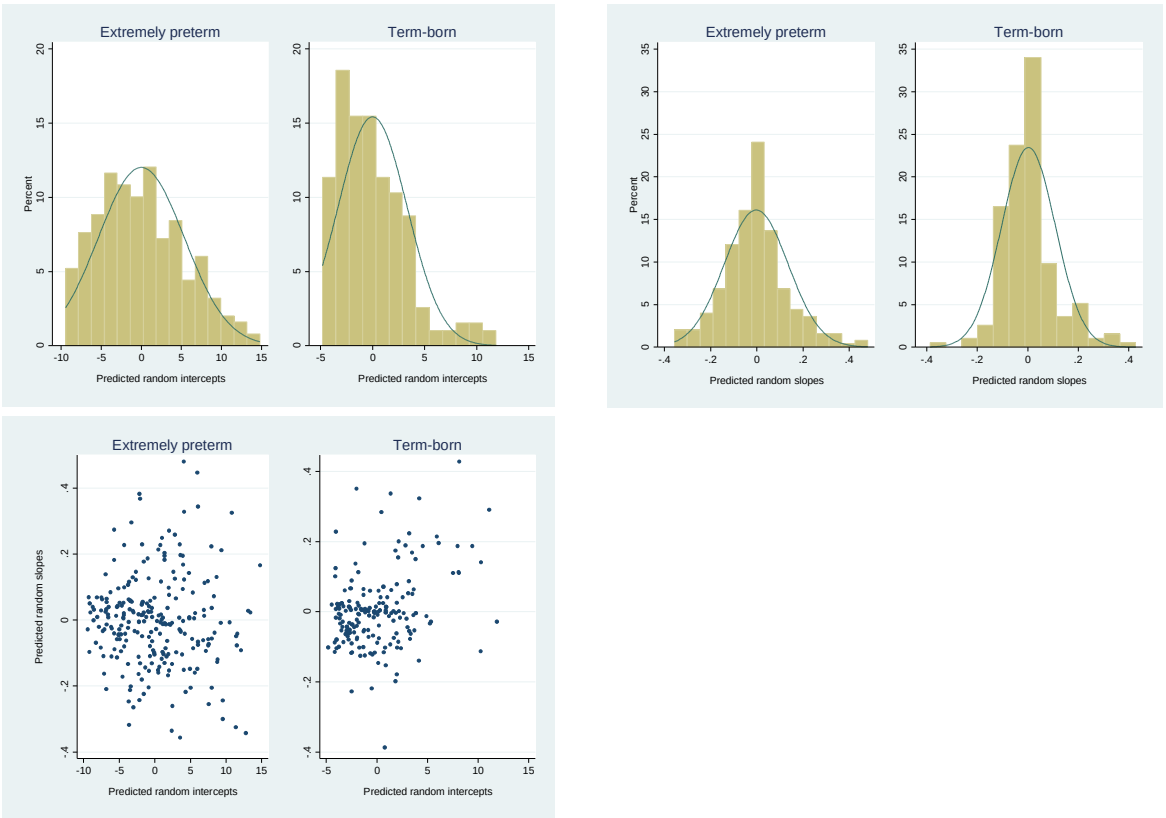
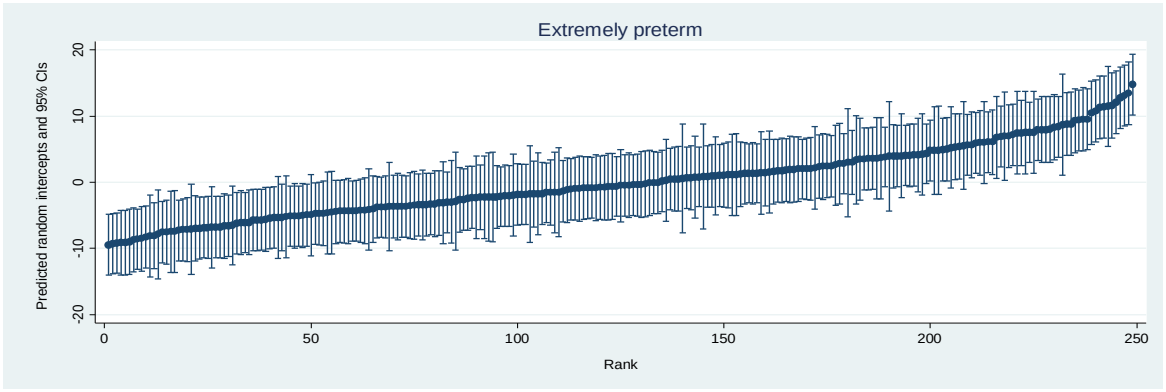
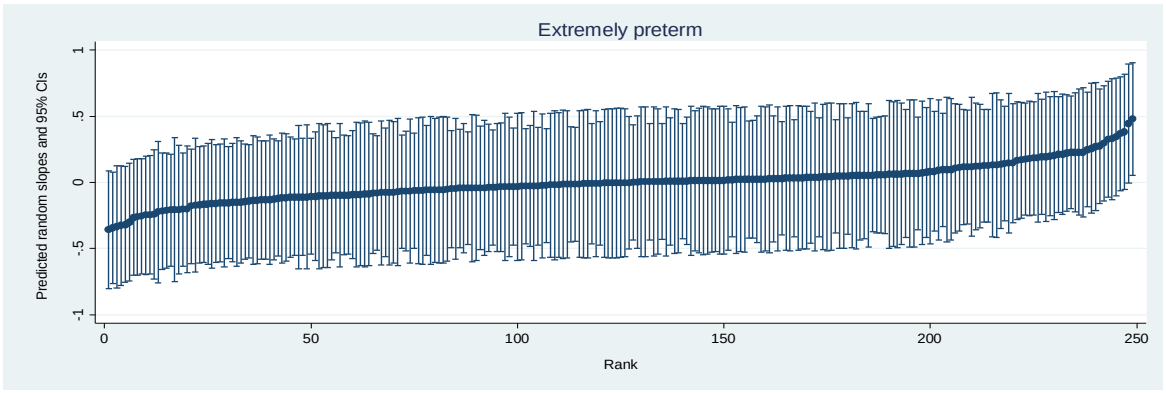
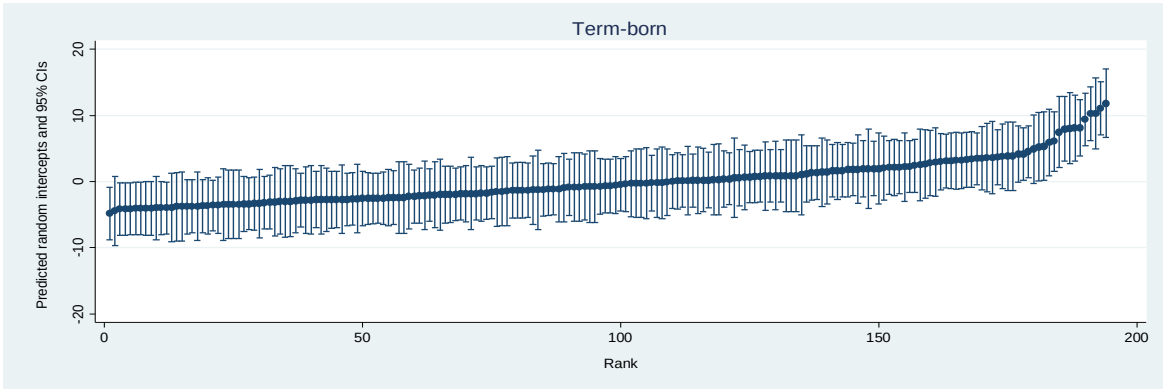
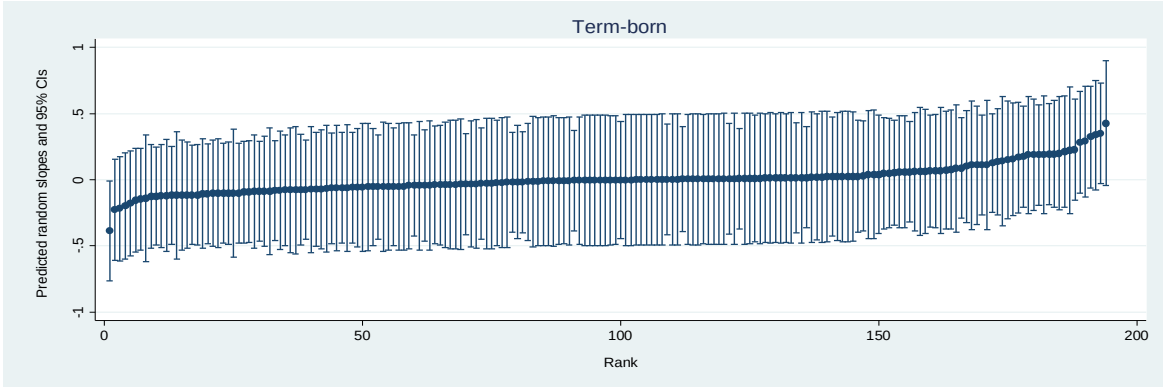


Figure 8.11: Caterpillar plots of random intercept and slope predictions and 95% confidence intervals versus ranking for SDQ Total Difficulties Score









8.4.4 Stratified analysis of the SDQ Total Difficulties Scores in the EPT and term-born group by participant sex and maternal education

The mean Total Difficulties Score and 95% confidence intervals at each age for EPT participants and term-born classmates, stratified by sex and maternal education, are presented in Table 8.10. Observed and predicted mean trajectories adjusted for sex and maternal education are displayed in Figure 8.12. The estimated Total Difficulties Scores of male participants were on average slightly higher in both the EPT and comparison group (mean difference 1.1, 95% CI: 0.05 to 2.15, p=0.04) (Table 8.10). The interaction between participant sex and group was not significant (LRT p=0.62) implying that the impact of sex on the Total Difficulties Score was similar in both groups. Higher maternal education did not have significant effect on scores in either group (mean difference -0.83, 95% CI: -1.92 to 0.26, p=0.14) (Table 8.11). Neither was there evidence of an interaction between maternal education and group (LRT p=0.27), suggesting this exerts the same effect on Total Difficulties Scores in both groups.

Table 8.10: SDQ Total Difficulties Scores in the EPT and term-born group by age of assessment stratified by sex of participant and maternal education

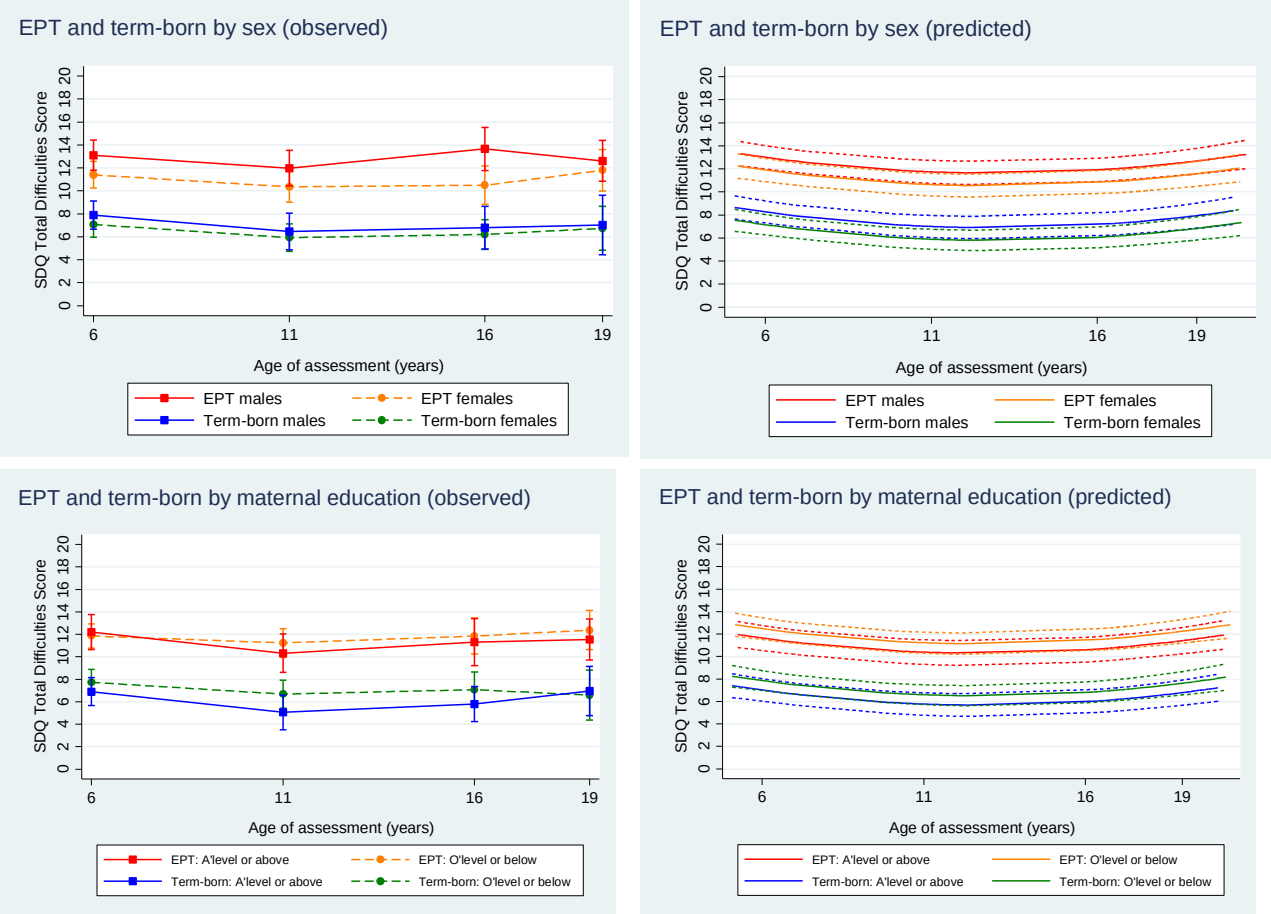
	Age 6 years		Age 11 years		Age 16 years		Age 19 years	
	EPT (n=220)	Term (n=146)	EPT (n=209)	Term (n=148)	EPT (n=134)	Term (n=86)	EPT (n=116)	Term (n=55)
Infant sex								
Female								
No. (%)	109 (49.5)	81 (55.5)	113 (54.1)	86 (58.1)	75 (56.0)	52 (60.5)	62 (53.4)	35 (63.6)
Mean	11.4	7.1	10.4	5.9	10.5	6.2	11.8	6.7
(95% confidence interval)	(10.2 to 12.6)	(6.0 to 8.2)	(9.0 to 11.7)	(4.7 to 7.1)	(8.8 to 12.2)	(4.9 to 7.5)	(10.0 to 13.6)	(4.8 to 8.7)
Male								
No. (%)	111 (50.5)	65 (44.5)	96 (45.9)	62 (41.9)	59 (44.0)	34 (39.5)	54 (46.6)	20 (36.4)
Mean	13.1	7.9	12.0	6.5	13.7	6.8	12.6	7.0
(95% confidence interval)	(11.8 to 14.4)	(6.6 to 9.1)	(10.4 to 13.5)	(4.9 to 8.1)	(11.8 to 15.5)	(4.9 to 8.7)	(10.8 to 14.4)	(4.4 to 9.6)
Mean difference (95% CI)	1.7 (-0.04 to 3.5)	0.8 (-0.9 to 2.5)	1.6 (-0.4 to 3.7)	0.5 (-1.4 to 2.5)	3.2 (0.6 to 5.7)	0.6 (-1.6 to 2.8)	0.8 (-1.8 to 3.4)	0.3 (-3.0 to 3.5)
Maternal education								
Up to O'level (or equivalent)								
No. (%)	133 (63.6)	81 (57.9)	127 (63.2)	85 (60.0)	80 (61.1)	46 (56.1)	58 (53.2)	28 (51.9)
Mean	11.9	7.7	11.2	6.7	11.9	7.1	12.4	6.6
(95% confidence interval)	(10.8 to 12.9)	(6.6 to 8.9)	(10.0 to 12.5)	(5.4 to 7.9)	(10.2 to 13.5)	(5.5 to 8.7)	(10.6 to 14.1)	(4.4 to 8.8)
A'level (or equivalent) and above								
No. (%)	76 (36.4)	59 (42.1)	74 (36.8)	57 (40.0)	51 (38.9)	36 (43.9)	56 (46.8)	26 (48.1)
Mean	12.2	6.9	10.3	5.1	11.3	5.8	11.5	7.0
(95% confidence interval)	(10.7 to 13.8)	(5.7 to 8.1)	(8.6 to 12.0)	(3.5 to 6.6)	(9.2 to 13.4)	(4.3 to 7.3)	(9.7 to 13.4)	(4.8 to 9.1)
Mean difference (95% CI)	0.3 (-1.5 to 2.2)	-0.8 (-2.6 to 0.9)	-0.9 (-3.1 to 1.2)	-1.6 (-3.6 to 0.4)	-0.5 (-3.2 to 2.1)	-1.3 (-3.6 to 1.0)	-0.8 (-3.4 to 1.7)	0.4 (-2.8 to 3.6)

Higher scores indicate more behavioural problems.

Table 8.11: SDQ Total Difficulties Scores in the EPT and term-born group by age of assessment adjusted for sex of participant and maternal education

	Adjusted for sex (n=443)		Adjusted for maternal education (n=421)	
	Estimate	**95% CI**	**Estimate**	**95% CI**
Fixed effects				
Constant	7.14	(6.26 to 8.03)	7.88	(6.99 to 8.77)
EPT	4.76	(3.70 to 5.81)	4.61	(3.54 to 5.69)
Age	-0.40	(-0.57 to -0.22)	-0.40	(-0.57 to -0.22)
Age^2	0.03	(0.02 to 0.04)	0.03	(0.02 to 0.04)
Male	1.10	(0.05 to 2.15)	-	-
Higher maternal education	-	-	-0.83	(-1.92 to 0.26)
Random effects				
Level 1 (within-individual)				
Term	3.48	(3.07 to 3.94)	3.41	(3.00 to 3.87)
EPT	3.70	(3.41 to 4.03)	3.68	(3.38 to 4.01)
Level 2 (between-individual)				
Term intercept SD	4.05	(3.30 to 4.96)	4.05	(3.30 to 4.96)
Term slope SD	0.25	(0.14 to 0.46)	0.28	(0.16 to 0.47)
Term correlation(intercept, slope)	-0.06	(-0.51 to 0.43)	-0.07	(-0.49 to 0.38)
EPT intercept SD	5.92	(5.23 to 6.70)	5.93	(5.22 to 6.73)
EPT slope SD	0.29	(0.20 to 0.41)	0.29	(0.21 to 0.41)
EPT correlation(intercept, slope)	-0.20	(-0.45 to -0.09)	-0.23	(-0.48 to 0.06)

Figure 8.12: Observed and predicted mean SDQ Total Difficulties Scores plus 95% confidence intervals in the EPT and term-born classmates stratified by sex and maternal education



8.4.5 Stratified analysis of SDQ Total Difficulties Scores of EPT participants by gestational age, neonatal brain injury, behavioural screening score at age 2.5, cognitive and functional impairment

The mean Total Difficulties Score and 95% confidence intervals at each age for EPT participants, stratified by GA, brain injury diagnosed in the neonatal period, CBCL score at 2.5 years, moderate/severe cognitive and functional impairment are presented in Table 8.12. The observed and predicted mean trajectories adjusted for these factors are displayed Figure 8.13. In the predicted models (Table 8.13), there was no significant difference in the Total Difficulties Score between EPT participants with moderate/severe neonatal brain injury and those with no/mild brain injury diagnosed in the neonatal period (Mean difference 1.44, 95% CI: -0.47 to 3.35, p=0.14). The interaction term between neonatal brain injury and age was also not significant (LRT p=0.33), indicating that the severity of overall behavioural symptoms between those with none/mild and moderate/severe brain injury was similar over time.

On average, the Total Difficulties Scores of participants born less than 25 weeks of gestation were 1.75 points higher than participants born at 25 weeks (95% CI: 0.15 to 3.35, p=0.03) but the interaction between GA and age was not significant (LRT p=0.10) (Table 8.13). A CBCL classification of clinical behavioural problems (>90th centile) at age 2.5 years was strongly associated with a higher SDQ Total Difficulties scores from six years onwards, with average scores 6.9 points higher compared to participants with a normal or borderline classification (95% CI: 5.01 to 8.70, p<0.0.01). However, the interaction term between CBCL classification and age was not significant (LRT p=0.25), implying that despite scores being persistently different, trajectories were stable in the normal/borderline and clinical group.

Participants with moderate/severe cognitive impairment at their last assessment had higher Total Difficulties Scores compared to participants with mild or no cognitive impairment (Mean difference 4.27, 95% CI: 2.76 to 5.77, p<0.001), with trajectories also stable over time in both groups (LRT for cognitive impairment*age interaction term p=0.12). Results were very similar in the comparison of participants with moderate/severe functional disability (which is largely driven by cognitive impairment) at their last assessment compare to participants with mild or no functional disability (mean difference 4.12, 95% CI: 2.61 to 5.63, p<0.001), with a non-significant interaction with age (LRT p=0.15).

Table 8.12: SDQ Total Difficulties Scores in EPT participants by age of assessment stratified by neonatal brain injury, gestational age, behavioural screening score at age 2.5, cognitive impairment and functional disability

	Age 6 years (n=241)	Age 11 years (n=219)	Age 16 years (n=138)	Age 19 years (n=127)
Neonatal brain injury				
No/mild brain injury				
No. (%)	176 (80.0)	162 (77.9)	108 (80.6)	96 (83.5)
Mean (95% CI)	12.0 (11.1 to 13.0)	10.9 (9.7 to 12.1)	11.4 (10.0 to 12.8)	12.1 (10.7 to 13.4)
Moderate/severe brain injury				
No. (%)	44 (20.0)	46 (22.1)	26 (19.4)	19 (16.5)
Mean (95% CI)	13.2 (11.0 to 15.4)	11.9 (9.8 to 14.0)	13.9 (10.9 to 17.0)	13.3 (9.7 to 16.9)
Mean difference (95% CI)	1.2 (-1.0 to 3.4)	1.0 (-1.5 to 3.5)	2.5 (-0.7 to 5.8)	1.3 (-2.2 to 4.7)
Gestational age at delivery				
25 weeks				
No. (%)	131 (59.5)	122 (58.4)	76 (56.7)	69 (59.5)
Mean (95% CI)	11.6 (10.4 to 12.7)	10.5 (9.1 to 11.8)	11.0 (9.3 to 12.7)	11.6 (10.0 to 13.2)
<25 weeks				
No. (%)	89 (40.5)	87 (41.6)	58 (43.3)	47 (40.5)
Mean (95% CI)	13.3 (11.9 to 14.7)	12.0 (10.5 to 13.6)	13.1 (11.2 to 15.0)	13.0 (10.9 to 15.1)
Mean difference (95% CI)	1.7 (-0.04 to 3.5)	1.6 (-0.5 to 3.6)	2.1 (-0.5 to 4.7)	1.4 (-1.2 to 4.0)
CBCL at age 2.5 years				
Normal/borderline				
No. (%)	167 (79.1)	160 (79.6)	107 (80.5)	93 (82.3)
Mean (95% CI)	10.7 (9.8 to 11.6)	9.3 (8.3 to 10.3)	10.6 (9.3 to 11.9)	11.1 (9.7 to 12.4)
Clinical				
No. (%)	44 (20.9)	41 (20.4)	26 (19.5)	20 (17.7)
Mean (95% CI)	17.9 (16.0 to 19.9)	17.0 (14.6 to 19.4)	17.4 (14.1 to 20.8)	17.0 (14.0 to 20.0)
Mean difference (95% CI)	7.2 (5.2 to 9.2)	7.7 (5.4 to 10.0)	6.8 (3.8 to 9.9)	5.9 (2.7 to 9.1)

Higher scores indicate more behavioural problems. CBCL Child Behaviour checklist.

[a] Functional disability was defined as cerebral palsy, cognitive, visual or hearing impairment at last assessment.

Table 8.12: SDQ Total Difficulties Scores in EPT participants by age of assessment stratified by neonatal brain injury, gestational age, behavioural screening score at age 2.5, cognitive impairment and functional disability (continued)

	Age 6 years (n=241)	Age 11 years (n=219)	Age 16 years (n=138)	Age 19 years (n=127)
Cognitive impairment				
No/mild impairment				
No. (%)	117 (53.2)	118 (56.5)	76 (56.7)	63 (54.3)
Mean (95% CI)	10.9 (9.7 to 12.1)	9.0 (7.8 to 10.3)	9.7 (8.3 to 11.2)	10.4 (8.7 to 12.1)
Moderate/severe impairment				
No. (%)	103 (46.8)	91 (43.5)	58 (43.3)	53 (45.7)
Mean (95% CI)	13.8 (12.5 to 15.0)	13.8 (12.3 to 15.3)	14.7 (12.7 to 16.8)	14.3 (12.5 to 16.1)
Mean difference (95% CI)	2.9 (1.1 to 4.6)	4.8 (2.8 to 6.8)	5.0 (2.6 to 7.5)	3.9 (1.4 to 6.4)
Functional disability[a]				
None/mild disability				
No. (%)	113 (51.4)	114 (54.5)	75 (56.0)	61 (52.6)
Mean (95% CI)	11.0 (9.8 to 12.2)	8.9 (7.6 to 10.2)	9.8 (8.3 to 11.2)	10.3 (8.6 to 12.1)
Moderate/severe disabilit13				
No. (%)	107 (48.6)	95 (45.5)	59 (44.0)	55 (47.4)
Mean (95% CI)	13.6 (12.4 to 14.9)	13.7 (12.2 to 15.2)	14.6 (12.6 to 16.6)	14.2 (12.5 to 16.0)
Mean difference (95% CI)	2.6 (0.9 to 4.4)	4.8 (2.9 to 6.8)	4.8 (2.4 to 7.3)	3.9 (1.4 to 6.4)

Higher scores indicate more behavioural problems. CBCL Child Behaviour checklist.

[a] Functional disability was defined as cerebral palsy, cognitive, visual or hearing impairment at last assessment.

Table 8.13: : SDQ Total Difficulties Scores in EPT participants by age of assessment adjusted for neonatal brain injury, gestational age, behavioural screening score at age 2.5, cognitive impairment and functional disability

Parameter	Adjusted for neonatal brain injury (n=248)		Adjusted for gestational age (n=249)		Adjusted for CBCL at 2.5 years (n=235)		Adjusted for cognitive impairment (n=249)		Adjusted for functional disability[a] (n=249)	
	Estimate	95% CI	Estimate	95% CI	Estimate	95% CI	Estimate	95% CI		
Fixed effects										
Constant	12.10	(11.09 to 13.12)	11.68	(10.55 to 12.80)	10.84	(8.99 to 11.81)	10.37	(9.22 to 11.51)	10.36	(9.19 to 11.52)
Age	-0.34	(-0.57 to -0.12)	-0.34	(-0.57 to -0.11)	-0.37	(-0.60 to -0.14)	-0.33	(-0.55 to -0.10)	-0.33	(-0.55 to -0.10)
Age^2	0.02	(0.01 to 0.04)	0.02	(0.01 to 0.04)	0.03	(0.01 to 0.04)	0.02	(0.01 to 0.04)	0.02	(0.01 to 0.04)
Moderate-severe brain Injury	1.44	(-0.47 to 3.35)	-	-	-	-	-	-	-	-
Gestational age <25 weeks	-	-	1.75	(0.15 to 3.35)	-	-	-	-	-	-
CBCL abnormal at 2.5 years	-	-	-	-	6.90	(5.10 to 8.70)	-	-	-	-
Moderate/severe cognitive impairment	-	-	-	-	-	-	4.27	(2.76 to 5.77)	-	-
Moderate/severe functional disability	-	-	-	-	-	-	-	-	4.12	(2.61 to 5.63)
Random effects										
Level 1 (within-individual)	3.70	(3.40 to 4.02)	3.69	(3.40 to 4.02)	3.71	(3.41 to 4.04)	3.69	(3.39 to 4.01)	3.69	(3.39 to 4.01)
Level 2 (between-individual)										
Intercept SD	5.95	(5.26 to 6.74)	5.91	(5.22 to 6.69)	5.16	(4.48 to 5.94)	5.71	(5.03 to 6.47)	5.72	(5.05 to 6.49)
Slope SD	0.29	(0.20 to 0.41)	0.29	(0.20 to 0.41)	0.29	(0.21 to 0.42)	0.29	(0.20 to 0.41)	0.29	(0.21 to 0.41)
Correlation(intercept, slope)	-0.21	(-0.46 to 0.08)	-0.20	(-0.45 to 0.08)	-0.17	(-0.45 to 0.15)	-0.27	(-0.51 to 0.002)	-0.27	(-0.51 to 0.007)

CBCL Child Behaviour checklist.

[a] Functional disability was defined as cerebral palsy, cognitive, visual or hearing impairment at last assessment.

Figure 8.13: Observed and predicted mean SDQ Total Difficulties Score plus 95% confidence intervals in EPT participants stratified by neonatal brain injury, gestational age, behavioural screening score at age 2.5, cognitive impairment and functional disability

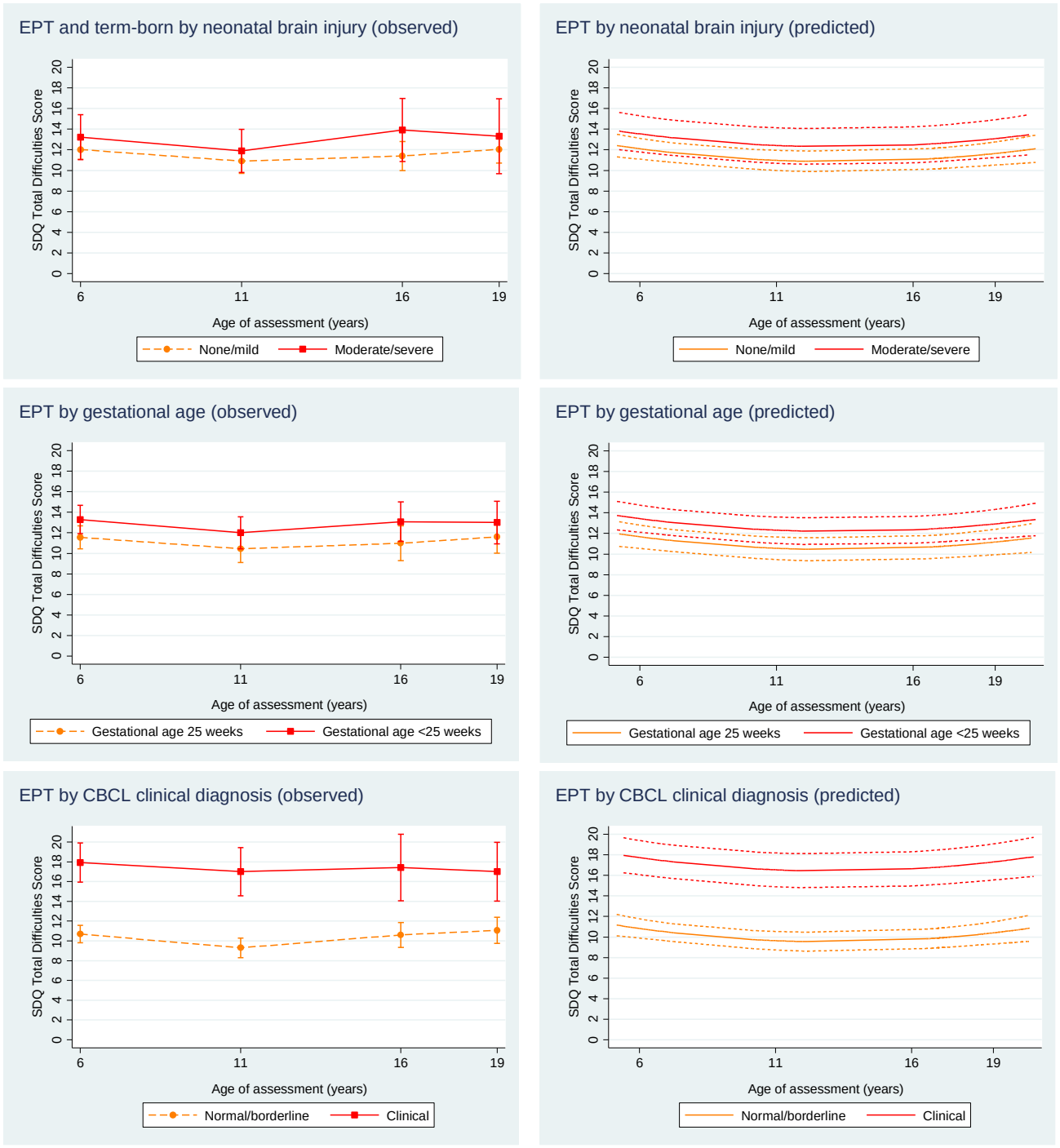
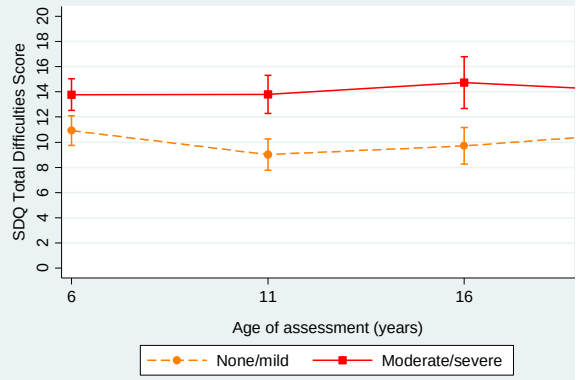
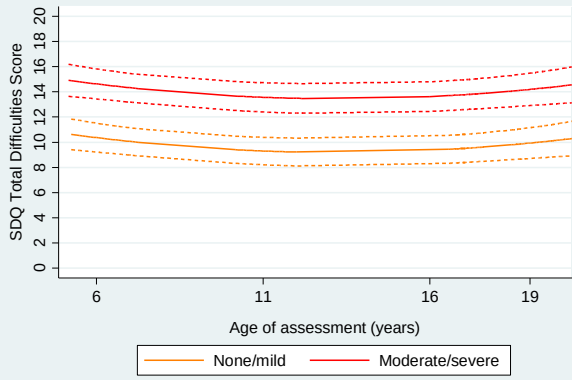


Figure 8.13: (continued)

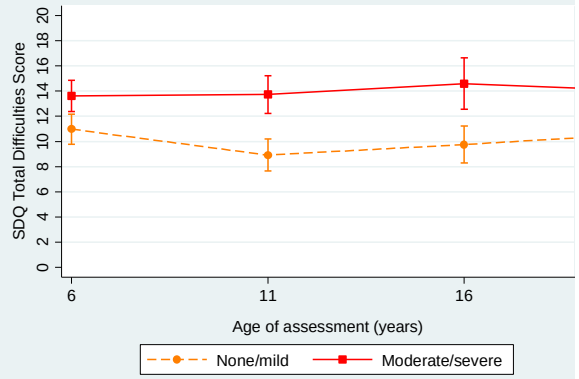
EPT by cognitive impairment (observed)



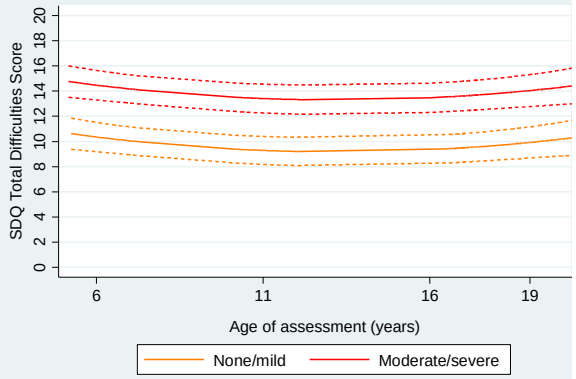
EPT by cognitive impairment (predicted)



EPT by functional disability (observed)



EPT by functional disability (predicted)



8.5 Complete case analysis of SDQ scores of the EPT and term-born group

All analyses were repeated for the 119 completers (82 EPT and 37 term-born) and the results are shown in Table 8.14 to Table 8.16 and Figure 8.14 to Figure 8.17. The SDQ Total Difficulties Score and the four component subscales were slightly lower in the complete case analysis compared to the analysis including all participants, in both groups, indicating fewer behavioural problems in those with complete follow-up. The difference in the Total Difficulties Score changed from 4.81 points (95% CI: 3.76 to 5.87) to 6.21 (95% CI: 4.32 to 8.10). All results remained statistically significant, except for the Conduct Problems Score; difference in means for completers: 0.3 points (95% CI: -0.21 to 0.80, p=0.25). Participant sex and maternal education were not significant when added as main effects or as interaction effects with group to the model for Total Difficulties Score (Table 8.15 and Figure 8.16).

The stratified analysis of the Total Difficulties Scores of EPT participants by GA, neonatal brain injury, behavioural screening score at age 2.5, cognitive and functional impairment and was also repeated for complete cases (Table 8.16 and Figure 8.17). Neonatal brain injury remained non-significant and the effect of GA was no longer significant. CBCL score at 2.5 years remained strongly associated with the trajectory of the Total Difficulties Score, and cognitive impairment and functional disability at last assessment remained significant.

Figure 8.14: Complete case: mean plus 95% confidence intervals for SDQ Total Difficulties and subscale scores the EPT and term-born group at age 6, 11, 16 and 19

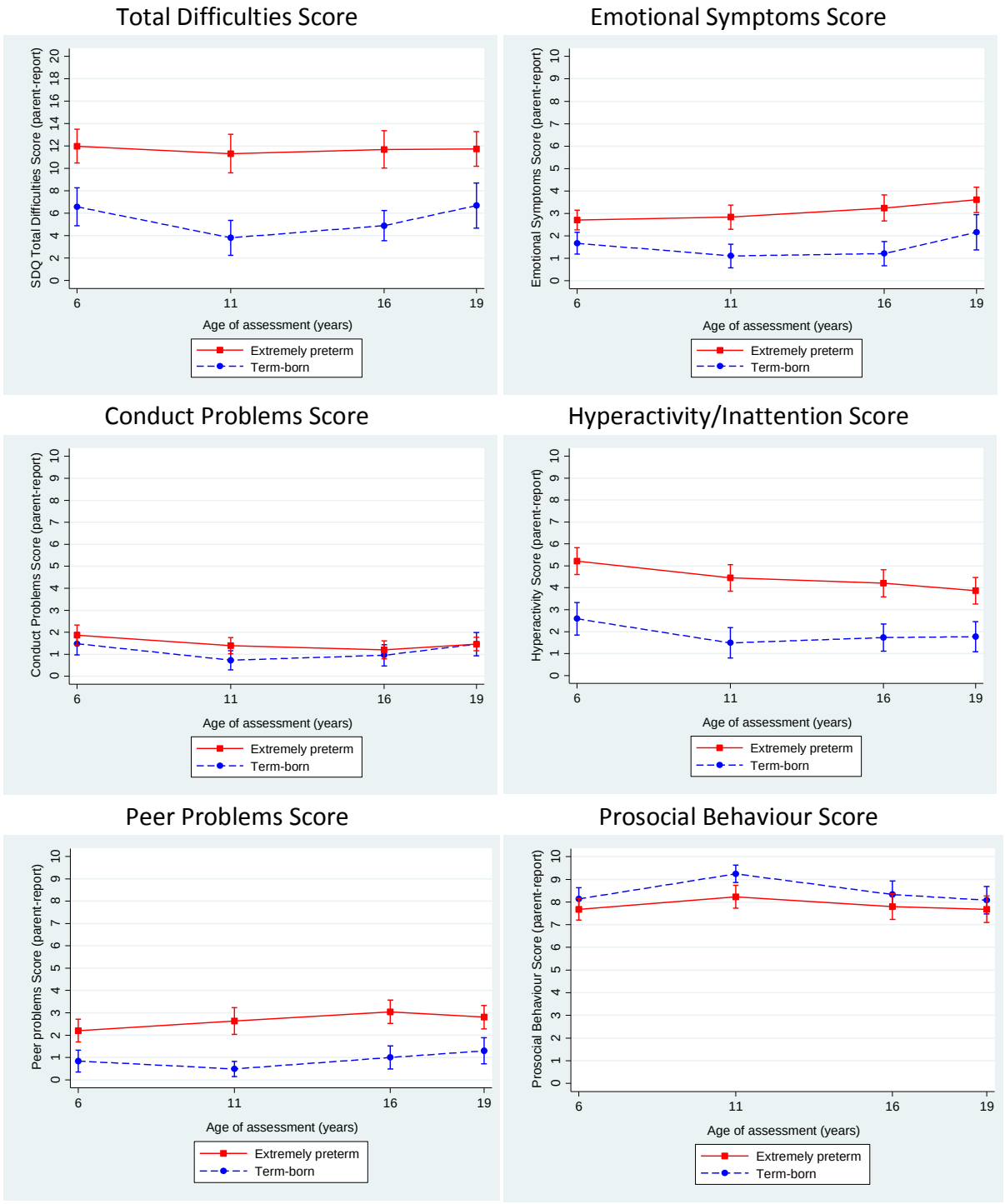


Table 8.14: Complete case: mixed model analysis of the SDQ Total Difficulties Score and subscales in the EPT and term-born group

SDQ Total Difficulties Score (n=119)			Emotional Symptoms Score (n=119)			Conduct Problems Score (n=119)		
Parameter	**Estimate**	**95% CI**	**Parameter**	**Estimate**	**95% CI**	**Parameter**	**Estimate**	**95% CI**
Fixed effects			**Fixed effects**			**Fixed effects**		
Constant	6.11	(4.69 to 7.53)	Constant	1.49	(0.97 to 2.02)	Constant	1.58	(1.10 to 2.06)
EPT	6.21	(4.32 to 8.10)	EPT	1.39	(0.77 to 2.00)	EPT	0.30	(-0.21 to 0.80)
Age	-0.45	(-0.69 to -0.20)	Age	-0.10	(-0.21 to 0.003)	Age	-0.19	(-0.26 to -0.11)
Age^2	0.03	(0.02 to 0.05)	Age^2	0.01	(0.004 to 0.02)	Age^2	0.01	(0.007 to 0.017)
Random effects			**Random effects**			**Random effects**		
Within-individual			Within-individual			Within-individual	1.05	(0.96 to 1.14)
EPT SD	3.54	(3.17 to 3.95)	EPT SD	1.58	(1.42 to 1.76)	Between-individual		
Term SD	2.91	(2.46 to 3.44)	Term SD	1.25	(1.08 to 1.45)	Intercept SD	1.51	(1.27 to 1.81)
Between-individual			Between-individual			Slope SD	0.10	(0.08 to 0.14)
EPT intercept SD	6.27	(5.17 to 7.59)	Intercept SD	1.18	(0.87 to 1.62)	Corr(intercept, slope)	-0.63	(-0.77 to -0.43)
EPT slope SD	0.29	(0.19 to 0.43)	Slope SD	0.10	(0.06 to 0.15)			
EPT corr(intercept, slope)	-0.18	(-0.51 to 0.19)	Corr(intercept, slope)	0.35	(-0.43 to 0.83)			
Term intercept SD	4.27	(3.10 to 5.86)						
Term slope SD	0.34	(0.23 to 0.51)						
Term corr(intercept, slope)	-0.51	(-0.77 to -0.10)						
Hyperactivity/Inattention Score (n=119)			**Peer Problems Score (n=119)**			**Prosocial Behaviour Score (n=119)**		
Parameter	**Estimate**	**95% CI**	**Parameter**	**Estimate**	**95% CI**	**Parameter**	**Estimate**	**95% CI**
Fixed effects			**Fixed effects**			**Fixed effects**		
Constant	2.64	(1.92 to 3.36)	Constant	0.57	(0.20 to 0.94)	Constant	8.28	(7.91 to 8.65)
EPT	2.54	(1.74 to 3.35)	EPT	1.75	(1.18 to 2.32)	EPT	-0.64	(-1.17 to -0.11)
Age	-0.21	(-0.32 to -0.11)	Age	0.05	(0.02 to 0.08)	Age	0.20	(0.11 to 0.29)
Age^2	0.01	(0.003 to 0.02)	**Random effects**			Age^2	-0.02	(-0.02 to -0.01)
Random effects			Within-individual			**Random effects**		
Within-individual			EPT SD	1.22	(1.09 to 1.36)	Within-individual	1.22	(1.11 to 1.33)
EPT SD	1.58	(1.42 to 1.77)	Term SD	0.96	(0.82 to 1.13)	Between-individual		
Term SD	1.12	(0.96 to 1.30)	Between-individual			EPT intercept SD	1.77	(1.43 to 2.18)
Between-individual			EPT intercept SD	2.15	(1.77 to 2.60)	EPT slope SD	0.13	(0.10 to 0.18)
Intercept SD	2.20	(1.85 to 2.62)	EPT slope SD	0.10	(0.07 to 0.15)	EPT corr(intercept, slope)	-0.05	(-0.38 to 0.29)
Slope SD	0.13	(0.10 to 0.18)	EPT corr(intercept, slope)	-0.29	(-0.58 to 0.06)	Term intercept SD	0.59	(0.22 to 1.60)
Corr(intercept, slope)	-0.45	(-0.65 to -0.20)	Term intercept SD	1.04	(0.70 to 1.55)	Term slope SD	0.09	(0.05 to 0.17)
			Term slope SD	0.11	(0.08 to 0.17)	Term corr(intercept, slope)	-0.06	(-0.86 to 0.82)
			Term corr(intercept, slope)	-0.51	(-0.79 to -0.05)			

Figure 8.15: Complete case: predicted mean SDQ score and 95% confidence intervals for SDQ Total Difficulties and subscale scores in the EPT and term-born group

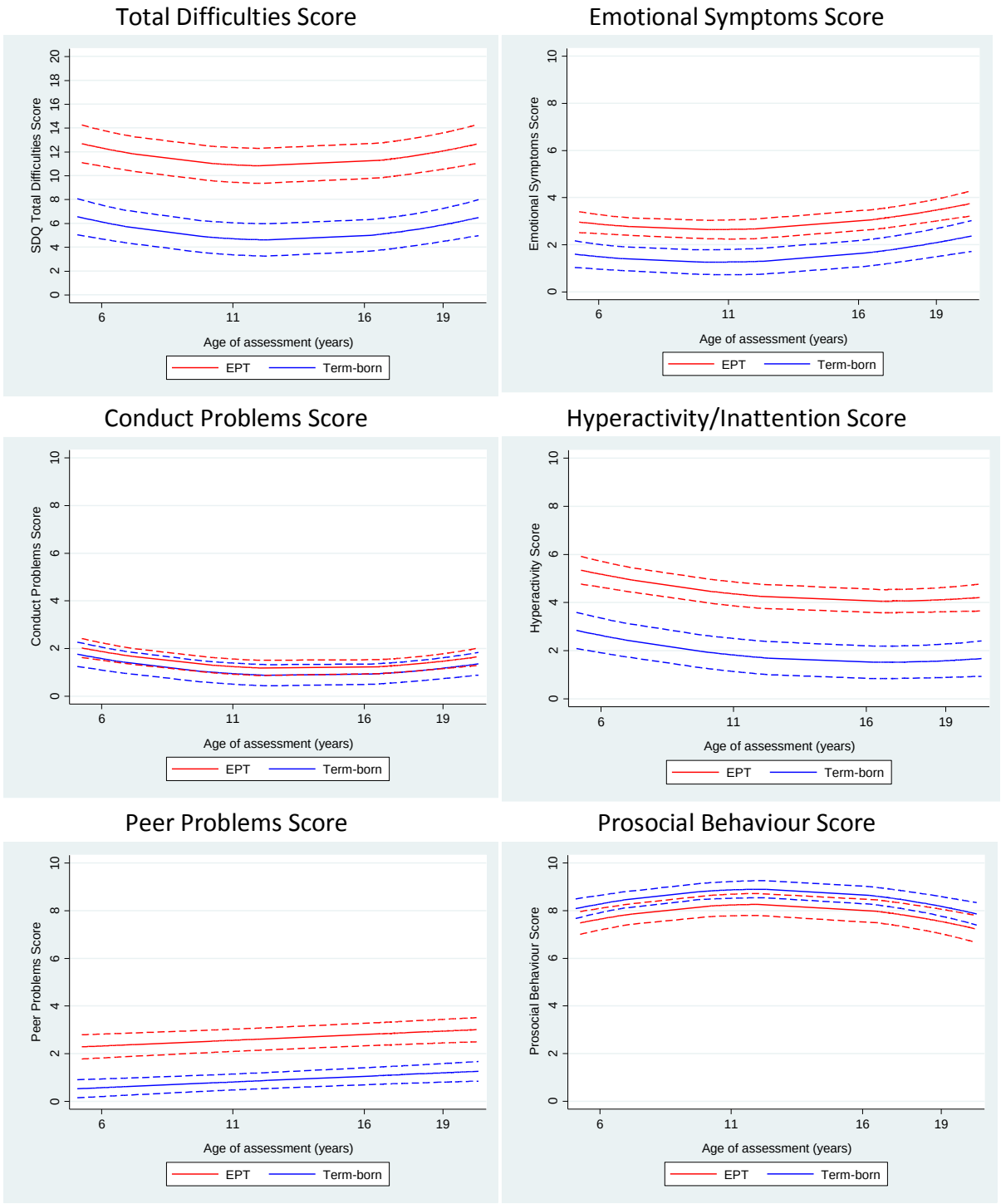


Table 8.15: Complete case: mixed model analysis of SDQ Total Difficulties Scores in the EPT and term-born group adjusted for sex and maternal education

	Adjusted for sex (n=119)		Adjusted for maternal education (n=116)	
	Estimate	**95% CI**	**Estimate**	**95% CI**
Fixed effects				
Constant	5.60	(3.98 to 7.23)	6.17	(4.46 to 7.87)
EPT	6.20	(4.31 to 8.09)	5.99	(4.12 to 7.85)
Age	-0.45	(-0.69 to -0.20)	-0.46	(-0.70 to -0.22)
Age^2	0.03	(0.02 to 0.05)	0.03	(0.02 to 0.05)
Male	1.25	(-0.66 to 3.16)	-	-
Higher maternal education	-	-	-0.35	(-2.21 to 1.51)
Random effects				
Level 1 (within-individual)				
Term	2.91	(2.46 to 3.45)	2.77	(2.33 to 3.30)
EPT	3.54	(3.17 to 3.95)	3.48	(3.11 to 3.89)
Level 2 (between-individual)				
Term intercept SD	4.33	(3.15 to 5.95)	4.33	(3.17 to 5.91)
Term slope SD	0.34	(0.23 to 0.51)	0.36	(0.25 to 0.52)
Term correlation(intercept, slope)	-0.52	(-0.78 to 0.11)	-0.53	(-0.78 to -0.14)
EPT intercept SD	6.17	(5.08 to 7.49)	5.93	(4.87 to 7.23)
EPT slope SD	0.29	(0.19 to 0.43)	0.29	(0.20 to 0.43)
EPT correlation(intercept, slope)	-0.18	(-0.51 to 0.19)	-0.18	(-0.51 to 0.19)

Figure 8.16: Complete case: observed and predicted mean SDQ Total Difficulties Scores plus 95% confidence intervals in the EPT and term-born group stratified by sex and maternal education

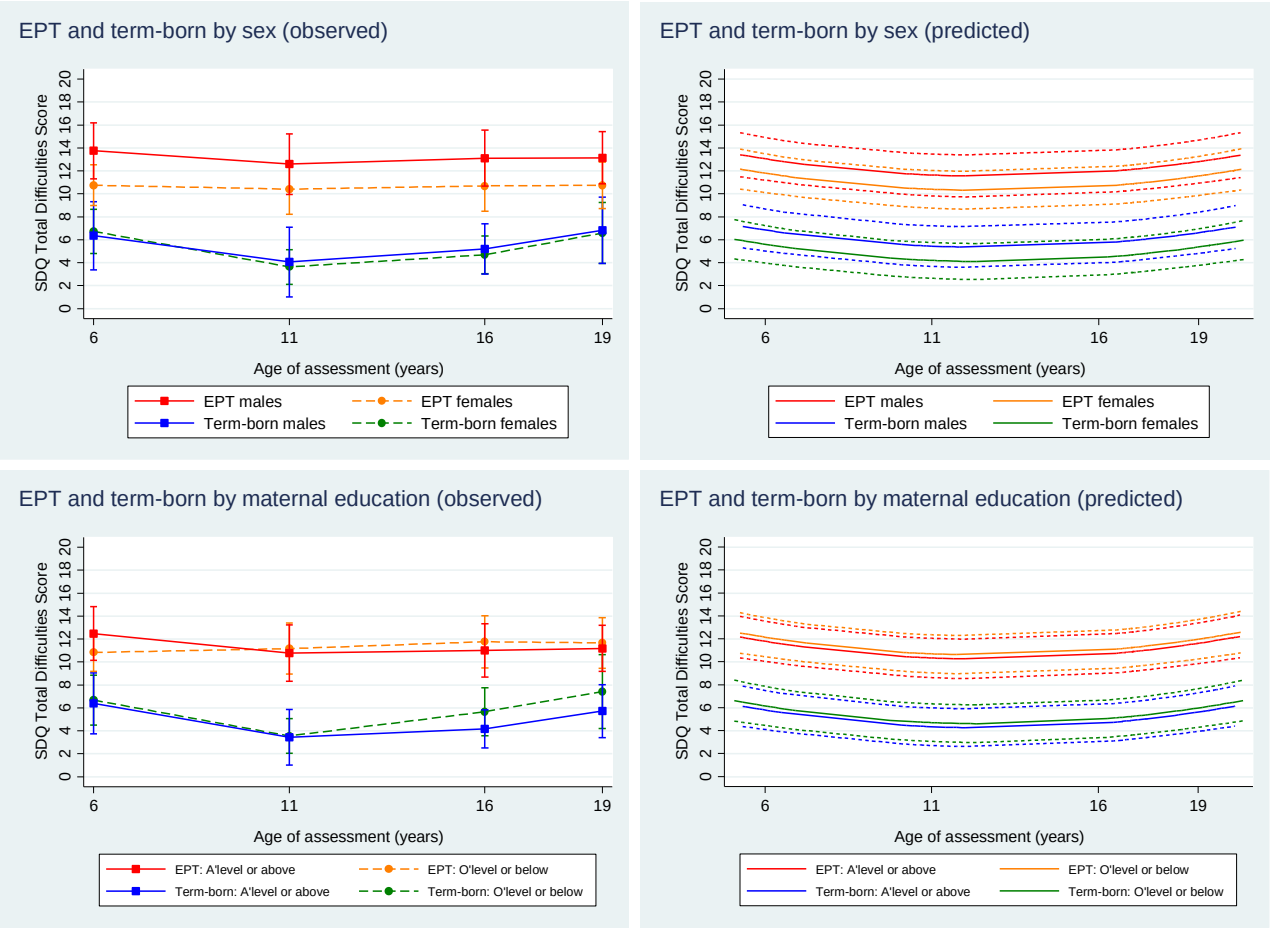


Table 8.16: Complete case: mixed model analysis of SDQ Total Difficulties Scores in EPT participants adjusted for neonatal brain injury, gestational age, behavioural screening score at age 2.5, cognitive impairment and functional disability

Parameter	Adjusted for neonatal brain injury (n=82)		Adjusted for gestational age (n=82)		Adjusted for CBCL at 2.5 years (n=82)		Adjusted for cognitive impairment (n=82)		Adjusted for functional disability[a] (n=82)	
	Estimate	95% CI	Estimate	95% CI	Estimate	95% CI	Estimate	95% CI		
Fixed effects										
Constant	11.74	(10.11 to 13.37)	10.77	(8.81 to 12.74)	10.40	(9.90 to 11.77)	10.22	(8.31 to 12.12)	10.17	(8.25 to 12.09)
Age	-0.007	(-0.10 to 0.09)	-0.007	(-0.10 to 0.09)	-0.007	(-0.10 to 0.09)	-0.007	(-0.10 to 0.09)	-0.007	(-0.10 to 0.09)
Moderate-severe brain Injury	0.06	(-3.65 to 3.77)	-	-	-	-	-	-	-	-
Gestational age <25 weeks	-	-	2.10	(-0.66 to 4.87)	-	-	-	-	-	-
CBCL abnormal at 2.5 years	-	-	-	-	8.53	(5.17 to 11.89)	-	-	-	-
Moderate/severe cognitive impairment	-	-	-	-	-	-	3.39	(0.69 to 6.10)	-	-
Moderate/severe functional disability	-	-	-	-	-	-	-	-	3.40	(0.70 to 6.10)
Random effects										
Level 1 (within-individual)	3.51	(3.15 to 3.91)	3.51	(3.15 to 3.91)	3.51	(3.15 to 3.90)	3.51	(3.15 to 3.91)	3.51	(3.15 to 3.91)
Level 2 (between-individual)										
Intercept SD	6.25	(5.16 to 7.57)	6.18	(5.09 to 7.49)	5.23	(4.25 to 6.45)	6.01	(4.95 to 7.30)	6.00	(4.94 to 7.29)
Slope SD	0.29	(0.19 to 0.42)	0.29	(0.19 to 0.42)	0.29	(0.19 to 0.43)	0.29	(0.19 to 0.42)	0.29	(0.19 to 0.42)
Correlation(intercept, slope)	-0.17	(-0.50 to 0.20)	-0.18	(-0.51 to 0.18)	-0.07	(-0.45 to 0.33)	-0.18	(-0.51 to 0.19)	-0.17	(-0.50 to 0.20)

CBCL Child Behaviour checklist.

[a] Functional disability was defined as cerebral palsy, cognitive, visual or hearing impairment at last assessment.

Figure 8.17: Complete case: observed and predicted SDQ Total Difficulties Scores plus 95% confidence intervals in EPT participants stratified by neonatal brain injury, gestational age, behavioural screening score at age 2.5, cognitive impairment and functional disability

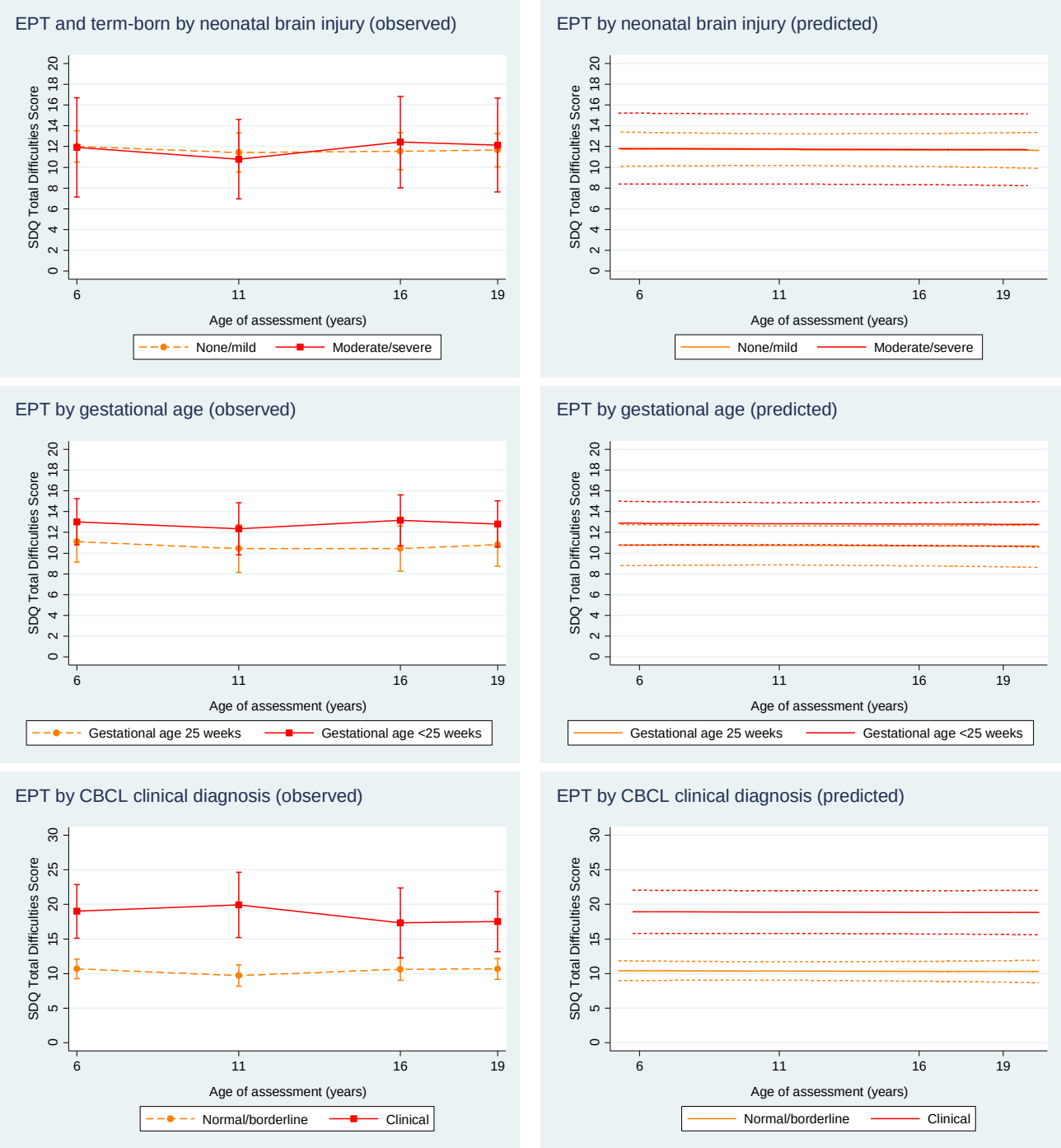
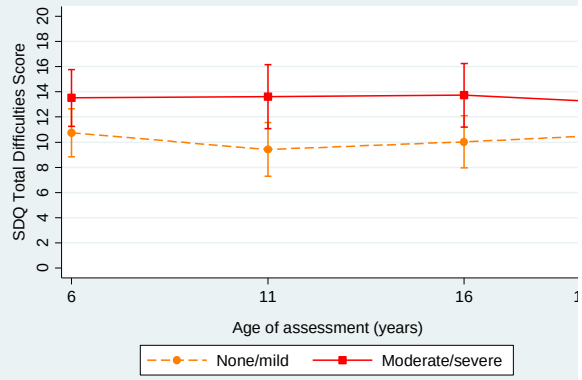
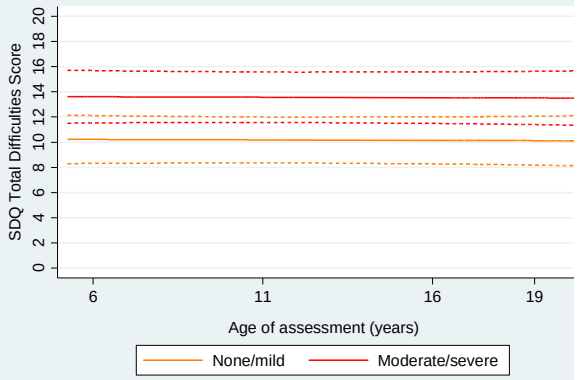


Figure 8.17: (continued)

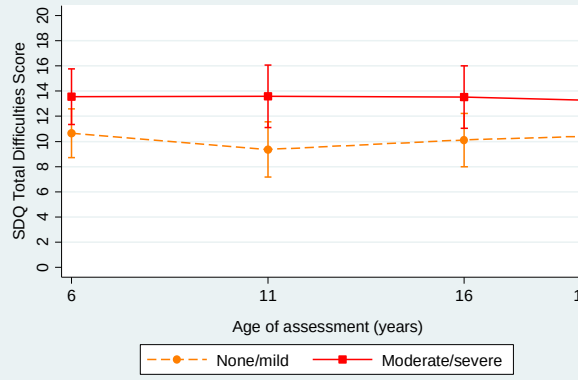
EPT by cognitive impairment (observed)



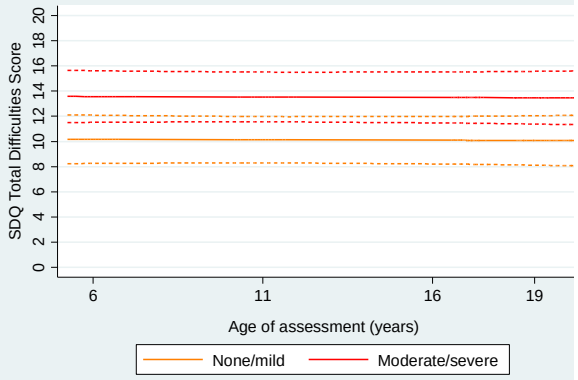
EPT by cognitive impairment (predicted)



EPT by functional disability (observed)



EPT by functional disability (predicted)



8.6 Discussion

8.6.1 Main findings

This is the first analysis investigating trajectories of neurobehavioural problems from childhood to young adulthood in EPT survivors and their term-born counterparts. Consistent with previous cross-sectional reports,[282,324,328,330] the results show that, overall, EPT individuals have significantly more neurobehavioural problems than term-born peers at six years of age and these differences remains stable throughout adolescence and into young adulthood. In addition the findings suggest that, although on a different level, the trajectories of overall neurobehavioural problems are similar in both EPT and term-born individuals. However, the assessment of overall neurobehavioural problems can mask between-group differences in separable symptom domains, as is demonstrated in the present study. The higher prevalence of overall problems was driven by the significant excess of emotional problems, hyperactivity/inattention and peer relationships in the EPT group, which is consistent with the triad of problems that characterizes the preterm behavioural phenotype.[257] Moreover, differences in trajectories were found both between symptom domains and between groups.

8.6.2 Relation to other studies

For hyperactivity/inattention, scores were persistently higher among the EPT individuals but they decreased slightly over time. The decline in the signs of ADHD with age for both term born and VLBW individuals has been noted in previous cross-sectional studies which have also shown that the between-group difference in the prevalence of ADHD declines over time.[324,330] The findings show that the group difference in the signs of ADHD (SDQ scores) remains similar

over time but the risk for clinically significant problems in EPT individuals declines from childhood to young adulthood.

For peer relationship problems, trajectories in SDQ scores were similar between groups with persistent between-group differences across childhood and adolescence, but an increased risk of clinically significant problems among EPT individuals over time compared with those born at term, which peaked at age 16. For emotional problems trajectories were different with a greater increase in SDQ scores among the EPT participants and an increased risk of clinically significant problems over time. This may reflect the generally later onset of emotional problems in adolescence observed in the natural course of these disorders. Generally, the greater persistence of problems over time in this study compared with other VLBW/ELBW cohorts[282,324,330] may also be related to the greater neurodevelopmental immaturity at birth, higher risk of neonatal brain injury and reduced neurodevelopmental plasticity that is conferred by birth at extremely low gestations compared with birth at later gestations. Indeed, moderate/severe cognitive impairment, an index of such factors, accounted for some of the difference in the overall neurobehavioural problems with those having cognitive difficulties having persistently greater problems.

EPT participants were also over four times more likely to have neurobehavioural problems that had a significant detrimental impact on their home life, friendships, and school and leisure activities compared with their term-born counterparts. Attenuating the risk of neurobehavioural problems and providing interventions for those with symptoms should therefore be a key aim of clinical follow-up. A clinical classification of behavioural problems at 2.5 years on the CBCL was found to be strongly associated with an increased risk of behavioural problems in later childhood. An association between parent-reported early

childhood behaviour or emotional problems and later psychiatric problems and disorders has also been shown in this and other EPT cohorts.[41,152,281] Given the stability of neurobehavioural problems from childhood to young adulthood shown here, and the increasing risk of emotional and peer problems over time, screening for behavioural and emotional problems in early childhood would be a valuable strategy for identifying those at risk and for facilitating the provision of early intervention or support.[338,339]

8.6.3 Strengths and limitations

A major strength of this large dual nation study of EPT survivors is its population-based design; findings are based on a prospective cohort of all live births in all centres in a geographically defined region and participants were followed up longitudinally from age 6 to age 19. In addition, longitudinal statistical analysis techniques were employed to examine the trajectories of neurobehaviour from early childhood to adulthood, which has many advantages over cross-sectional analysis techniques that simply compare group averages at each time point. In common with other longitudinal studies, the number of participants lost to follow-up increased over time, and was related to markers of social disadvantage and disability. It is possible therefore that the true extent of neurobehavioural impairment, has been underestimated particularly at older ages, though there was no evidence of a difference in CBCL scores at age 2.5 between participants with complete and incomplete data. The findings were strengthened by the analysis of individuals with complete follow-up which corroborated the main results, although this subset of participants had a lesser degree of behavioural difficulties in general compared to all participants.

8.6.4 Conclusion

Attention and peer relationship problems in individuals born extremely preterm persist into early adulthood, and emotional problems increase as they enter adolescence. The increased risk of clinically significant behavioural problems continues to have a substantial impact on their everyday lives as young adults. A positive behavioural screen in infancy is strongly associated with early adult outcomes and the introduction of screening into routine clinical follow-up should be considered, to enable targeted early interventions.

Chapter 9: Discussion

9.1 Summary of results

Neonatal interventions at extremely low GA are becoming more frequent, leading to an increasing number of admissions into NICUs. EPT survivors take up a disproportionate amount of NICU funding and overall costs, and as they grow are more likely to place an additional burden on health care and education services as well as the human costs to the individuals themselves and their families. The early identification and management of factors that mediate long term outcome is necessary to develop, target and evaluate interventions to promote resilience and help predict developmental course. Many risk factor analyses of neurodevelopmental impairment in preterm populations have been published, but this vast literature has not been formally summarised. Furthermore, there is a dearth of studies reporting longitudinal analysis of neurodevelopmental trajectories from early childhood to adulthood. The first aim of this thesis was to perform a systematic review of the world literature over the last two decades, to consolidate the evidence on the prognosis of neurodevelopmental outcome in children born VPT or with VLBW. The second aim was to conduct a longitudinal analysis of a cohort of EPT participants followed up into early adulthood to investigate the trajectories of long term sequelae over time, and to examine the association of neurodevelopmental course in relation to the predictive factors identified in the systematic review.

The systematic review of studies that have published multivariable prediction models for neurodevelopmental outcome in children born VPT or VLBW aged over 18 months identified 78 eligible articles in the following outcome domains; motor (n=28 articles), cognitive (n=31 studies), behaviour (n=15 studies), vision (n=3 studies), hearing (n=0 studies) and composite outcome (n=27 studies). The data from these studies were synthesised and resulted in three separate reviews of risk factors for poor cognitive and language development, CP and motor impairment, and behavioural problems and psychiatric disorders. The evidence for most risk factors was mixed or unclear, however some factors emerged as strongly prognostic, and the influence of these on neurodevelopmental trajectories in the EPICure cohort were investigated. Firstly, the main findings from the comparative analysis of cognitive and behavioural trajectories between the EPT and term-born control participants in the EPICure cohort are summarised. Then the risk factors identified in the systematic review and incorporated into the longitudinal analyses of the EPICure cohort are discussed. The difficulties in predicting neurodevelopmental outcome in preterm populations is then considered, followed by recommendations for future research.

9.1.1 Cognitive and behavioural trajectories in children born EPT compared to term-born children

In the longitudinal analysis of a dual nation, population-based study, impaired cognitive function in EPT children persisted into early adulthood with no evidence of substantial 'catch-up' with term-born peers, or further decline. IQ scores were on average 25 points lower in EPT individuals, and scores were more variable both within individuals and between individuals in the EPT group. Factors having a detrimental effect on cognitive function among EPT participants included male sex, brain injury diagnosed in the neonatal period, lower maternal education and earlier GA at birth. Overall behavioural difficulties were greater in individuals

born EPT versus those born at term at every age and trajectories in both groups remain stable over time. The main differences in behaviour were driven by hyperactivity, inattention and peer problems, and emotional symptoms in EPT individuals increased as they entered adolescence. The risk of behavioural problems having a significant impact on home life, friendships, school or work and/or leisure activities was over four times greater in individuals born EPT compared to those born at term. A clinical classification of behavioural problems at age 2.5 using the CBCL was strongly associated with increased risk of behavioural problems in later childhood and early adulthood, and moderate/severe cognitive impairment in the EPT group accounted for some but not all of the difference seen in overall behavioural difficulties.

9.1.2 Brain injury diagnosed in the neonatal period and neurodevelopmental outcome

The systematic review of risk factors for CP and motor impairment found strong evidence that IVH grade 3-4 alone or in combination with PVL was predictive of CP. This is supported by many large well-conducted studies that have focussed on the association of IVH and PVL with CP which have reported strong linear trends with grade.[207-210] However, there was no evidence in the studies reviewed that brain injury diagnosed in the neonatal period was associated with general motor impairment in EPT children, with or without major disability. It was not possible to examine the impact of neonatal brain injury on the trajectories of motor development in the EPICure study due to the lack of consistency in the outcome measures used to assess motor function over time. Many studies that have focussed exclusively on the relationship between neonatal brain injury and later cognitive development have also reported strong linear trends with grade.[207,208,210,255] This relationship was also apparent in the longitudinal analysis of cognitive trajectories in the EPICure study. Having moderate to severe neonatal brain injury was found to have an adverse effect on outcome for EPT individuals, who had IQ

scores 10 points lower on average compared to EPT individuals with no/mild brain injury. However, the prognostic value of neonatal brain injury in the multivariable models reported in the review of risk factors for poor cognitive and language development was mixed. This is possibly because cognitive and language development are multifactorial, unlike a diagnosis of CP which is more directly related to focal brain injury, so that when other factors are entered into a model, the influence of perinatal factors become less pronounced. The unclear findings may also reflect the heterogeneous selection criteria and methods of dealing with missing data related to severe disability across the studies. The systematic review of risk factors for behavioural problems revealed a paucity of good quality studies with long-term follow-up, so there was a lack of evidence on the prognostic value of neonatal brain injury for behavioural problems. In the longitudinal analysis of behavioural difficulties in the EPICure cohort, participants with moderate/severe neonatal brain injury were more likely to experience difficulties throughout childhood, compared to and those with no/mild brain injury, although this difference was not statistically significant.

9.1.3 Participant sex and neurodevelopmental outcome

In the systematic review of motor outcomes, male sex was found to be of limited use as a prognostic factor for CP, but there was fairly strong evidence that it was predictive of later motor impairment in children free of major disability. Other studies that have focussed solely on the association of infant sex with motor function in VPT children have reported mixed findings;[227-230] however these were not restricted to children free of major disability. In the review of cognitive outcomes, there was evidence that male sex was a significant predictor of global cognitive dysfunction and language impairment in VPT/VLBW infants, but after the age of five years, the impact of male sex diminished. This was not the case in the longitudinal

analysis of the EPICure cohort where the IQ scores of EPT males were almost 9 points below EPT females, persisting across childhood and adolescence; with no significant difference between males and females in the term-born comparison group. The systematic review of behavioural problems provided no evidence on whether infant sex was prognostic or not, but in the longitudinal analysis of the EPICure cohort there was evidence of a small but increased Total Difficulties Score among both males born EPT and males born at term.

9.1.4 Social disadvantage and neurodevelopmental outcome

The association between level of maternal education and SES and cognitive outcome in both term and preterm populations has been frequently reported,[295,299,317] There was evidence that a lower level of parental education was good predictor of global cognitive dysfunction, both in early infancy and childhood in the systematic review of cognitive outcomes. Whereas the impact of perinatal risk factors for cognitive impairment appeared to diminish over time, the impact of parental education remained influential. Non-white ethnicity, which is a reasonable proxy for poverty and social disadvantage,[251,252] was also found to be predictor in infancy. While the review of behavioural outcomes was inconclusive overall, markers of socio-economic deprivation were the only factors that appeared to be consistent predictors of general behavioural problems. In the longitudinal analysis of the EPICure cohort, there was weak evidence that higher maternal education was marginally beneficial for cognitive development in both the EPT and term-born comparison group and no evidence that it improved behavioural outcomes. This may be due to the higher drop-out rate in younger, less educated mothers, which restricted the educational range included in the analysis, making it less sensitive to detect a difference. It is also possible that maternal education alone is not a sufficient marker of social disadvantage.

9.1.5 Gestational age and neurodevelopmental outcome

GA was found to be of limited use as prognostic factor for either CP or other motor impairments and for cognitive or language development in the systematic reviews. The lack of association in these reviews was surprising as the relationship between older GA and lower prevalence of CP[184,225,226] and higher IQ scores[40] is well established across the whole spectrum of GA from 25 to 40 weeks. One explanation is the restriction to cohorts <33 weeks; the prevalence of CP declines quite steeply after 32 weeks, hence discrimination below this threshold may become less apparent.[226] It also appeared to be a less powerful predictor of cognitive impairment when restricted to the more preterm subgroups, however this was not the case in the longitudinal analysis of the EPICure cohort in which the IQ scores of participants born less than 25 weeks of gestation were more than 4 points lower than participants born at 25 weeks throughout childhood and adolescence. Another possible explanation is that GA is associated with other critical risk factors so its power as an independent predictive factor may be reduced due to high correlation with factors that are more strongly causally related to CP and cognitive delay. So while the impact of GA is clearly evident when its association with neurodevelopmental outcome is examined in isolation, its predictive power disappears once other factors are introduced into a multivariable model.

9.1.6 Early behavioural screening and neurodevelopmental outcome

A dearth of studies investigating prognostic factors for psychiatric disorders were identified in the systematic review, however the few that were reported that poor performance in behavioural screening questionnaires in infancy or early childhood were predictive of later disorders. [41,119,152] This finding is supported by other studies that have examined the predictive validity of screening tests and the stability of diagnoses over time.[281,282] [119] In the longitudinal

analysis of the EPICure cohort, a CBCL classification of clinical behavioural problems at age 2.5 years was strongly associated with a high SDQ Total Difficulties score from six years onwards compared to participants with a normal or borderline classification. Thus early screening of behavioural problems may have some predictive validity and clinical utility in preterm populations and would enable early interventions for those most likely to benefit.

9.2 Recommendations for future research

The systematic review of 78 published articles reporting multivariable risk factor models for neurodevelopmental outcome in surviving VPT/VLBW children, revealed some shortcomings in methodology and reporting that could be improved in future studies, and confirmed that that there is a paucity of properly designed and well-conducted prognostic modelling studies in this field. The findings and recommendations of this critical review should be used as a basis for the design, analysis and reporting of future studies seeking to develop multivariate risk factor or prognostic models in this population. The use of meta-analysis to determine the strength of association of the prognostic factors identified in the review; type of neonatal brain injury, participant sex, parental education and social disadvantage, would also be a valuable future direction.

There is an urgent need for larger population cohorts with comparison groups followed up routinely beyond two years as the more subtle outcomes such as impairment of executive function and fine motor skills cannot be reliably assessed at this age, and the natural course of some disorders, such as ADHD or ASD, have their onset later in childhood. Future studies should be adequately powered to study less common conditions in this population and there should be greater standardisation of outcome and risk factor measurements, particularly the

use of standard diagnostic evaluations to assess psychiatric disorders. Recruiting controls at delivery would provide more robust comparative data, and improving methods of follow-up to reduce the extent of dropout would also merit attention. The use routinely collected follow-up data, for example data linkage with official statistics on educational attainment collected by the Department of Education may be a way to overcome this problem.

Future studies should adopt appropriate statistical analytical techniques to analyse longitudinal outcome data and the impact of risk factors on these trajectories. Further evaluation of different patterns of neurodevelopmental over time is needed before reliable prognostic models can be developed in VPT and EPT populations. Additional research is required to improve the prediction of individual differences, and to identify the neuropathological differences underlying different developmental trajectories and their interaction with environmental factors over time.

9.3 Conclusion

The comprehensive systematic review of the world literature over the last two decades to identify predictive factors for neurodevelopmental outcome in children born VPT or with VLBW, identified a handful of risk factors that appear to be robust. These include brain injury diagnosed during the neonatal period, specifically moderate to severe IVH or PVL, male sex, and markers of social disadvantage such as a lower level of maternal or parental education. While GA is strongly associated with long-term outcome when modelled alone and across the entire GA spectrum, its predictive power in cohorts with GA below 32 weeks is limited in multivariable models which include other important risk factors. There were generally mixed findings for other major clinical events in the neonatal period such as NEC, BPD, ROP, PH and

sepsis. The unclear findings for many risk factors could reflect differences in study design, study population, methodological quality and lack of standardization of measures. Or it could simply reflect the fact that prognostic modelling in such a heterogeneous population is challenging and complex, with multiple risk factors acting sequentially over time, and often the existence of multiple impairments within the same individual. It is possible that different developmental phenotypes exist, for example in children with multiple impairments, or children who acquired a neonatal brain injury, and that it is necessary to develop different prognostic models within particular preterm subgroups.

The main conclusions from the longitudinal analysis of individuals born EPT is that being born so prematurely appears to place limits on brain plasticity and function which does not recover over time; with the most vulnerable being males and those who experienced brain injury early in life. Acute forms of brain injury such as IVH can cause severe damage to the central nervous system which can lead to CP, but widespread diffuse white matter damage can also have long-term effects. Loss of tissue and consequent reduction in number of neurons and synapses is likely to affect how the brain learns and functions. These structural abnormalities may disturb neurodevelopmental processes and impede the brain from maintaining a normal developmental trajectory.[321,322] If EPT participants fail to achieve optimum levels of cognitive function and are still experiencing behavioural problems once they have reached maturity, then this will have potentially profound implications for their health and well-being in later adulthood and old age.[323]

References

1. Tucker J, McGuire W. Epidemiology of preterm birth. *British Medical Journal* 2004; **329**: 675-8.
2. Goldenberg RL, Culhane JF, Iams JD, Romero R. Epidemiology and causes of preterm birth. *The Lancet* 2008; **371**: 75-84.
3. Baron IS, Rey-Casserly C. Extremely preterm birth outcome: a review of four decades of cognitive research. *Neuropsychological Review* 2010; **20**: 430-52.
4. Mangham LJ, Petrou S, Doyle LW, Deraper ES, Marlow N. The cost of preterm birth throughout childhood in England and Wales. *Pediatrics* 2009; **123**: 312-27.
5. Hankins GDV, Pacheco LD, Soo-Tung Wen T. Fetal assessment near the limits of viability. *Seminars in Perinatology* 2013; **37**: 401-3.
6. Wilkinson DJC. Gestational ageism. *Archives of Pediatrics and Adolescent Medicine* 2012; **166**(6): 567-72.
7. Guillen U, Weiss EM, Munson D, et al. Guidelines for the management of extremely premature deliveries: a systematic review. *Pediatrics* 2015; **136**(2): 343-50.
8. Lorenz JM, Wooliever DE, Jetton JR, Paneth N. A quantitative review of mortality and developmental disability in extremely premature newborns. *Archives of Pediatrics and Adolescent Medicine* 1998; **152**(5): 425-35.
9. Engle WA, Newborn AAoPCoFa. Age terminology during the perinatal period. *Pediatrics* 2004; **114**(5): 1362-4.
10. Blencowe H, Cousens S, Oestergaard MZ, et al. National, regional, and worldwide estimates of preterm birth rates in the year 2010 with time trends since 1990 for selected countries: a systematic analyis and implications. *The Lancet* 2010; **379**: 2162-72.
11. Slattery MM, Morrison JJ. Preterm delivery. *The Lancet* 2002; **360**: 1489-97.
12. Mercer BM, Goldenberg RL, Meis PJ, et al. The preterm prediction study: prediction of preterm rupture of membranes through clinical findings and ancillary testing. *American Journal of Obstetrics and Gynecology* 2000; **183**: 738-45.
13. England MC, Benjamin A, Abenhaim HA. Increased Risk of Preterm Premature Rupture of Membranes at Early Gestational Ages among Maternal Cigarette Smokers. *American Journal of Perinatology* 2013; **30**(10): 821-6.
14. Rennie & Robertson's textbook of neonatology. 5th ed. London: Elsevier Ltd; 2012.
15. Jackson RA, Gibson KA, Wu YW, Croughan MS. Perinatal outcomes in singletons following in vitro fertilization: a meta-analysis. *Obstetrics and Gynecology* 2004; **103**: 551-63.
16. Born too soon: the global action report on preterm birth: The World Health Organisation (WHO); 2012.
17. Stewart AL, Reynolds EO, Lipscomb AP. Outcome for infants of very low birth weight: a survey of the world literature. *The Lancet* 1981; **317**(8228): 1038-40.
18. Costeloe KL, Hennessy EM, Haider S, Stacey F, Marlow N, Draper ES. Short term outcomes after extreme preterm birth in England: comparison of two birth cohorts in 1995 and 2006 (the EPICure studies) *British Medical Journal* 2012; **345**: DOI: 10.1136/bmj.e7976.
19. Moore T, Hennessy EM, Myles J, et al. Neurological and developmental outcome in extremely preterm children born in England in 1995 and 2006: The EPICure studies. *British Medical Journal* 2012; **345**(7886): e7961.

20. Ancel PY, Goffinet F. Survival and morbidity of preterm children born at 22 through 34 weeks' gestation in France in 2011: results of the EPIPAGE-2 cohort study. *Jama, Pediatr* 2015; **169**(3): 230-38.
21. Hintz SR, Kendrick DE, Wilson-Costello DE, et al. Early-childhood neurodevelopmental outcomes are not improving for infants born at <25 weeks' gestational age. *Pediatrics* 2011; **127**(1): 62-70.
22. Fanaroff AA, Stoll BJ, Wright LL, et al. Trends in neonatal morbidity and mortality for very low birthweight infants. *American Journal of Obstetrics & Gynecology* 2007; **196**(2): 147.e1-8.
23. Stoll BJ, Hansen BM, Bell EF, et al. Trends in Care Practices, Morbidity, and Mortality of Extremely Preterm Neonates, 1993-2012. *JAMA* 2015; **314**(10): 1039-51.
24. Doyle LW, Roberts G, Anderson PJ, Victorian Infant Collaborative Study G. Changing long-term outcomes for infants 500-999 g birth weight in Victoria, 1979-2005. *Arch Dis Child Fetal Neonatal Ed* 2011; **96**(6): F443-7.
25. Gestation specific mortality 2012: England and Wales: Office for national statistics
26. Medlock S, Ravelli ACJ, Tamminga P, Moi BWM, Abu-Hanna A. Prediction of mortality in very premature infants: a systematic review of prediction models. *PLoS ONE* 2011; **6**(9).
27. Hakansson S, Farooqi A, Holmgren PA, Serenius F, Hogberg U. Proactive management promotes outcome in extremely preterm infants: a population-based comparison of two perinatal management strategies. *Pediatrics* 2004; **114**: 58-64.
28. Lorenz JM, Paneth N, Jetton JR, Ouden LD, Tyson JE. Comparison of management strategies for extreme prematurity in New Jersey and the Netherlands: Outcomes and resource expenditure. *Pediatrics* 2001; **108**(6): 1269-74.
29. Saigal S, Doyle LW. An overview of mortality and sequelae of preterm birth from infancy to adulthood. *Lancet* 2008; **371**(9608): 261-9.
30. Perlman. Neurobehavioral deficits in premature graduates of intensive care - potential medical and neonatal environmental risk factors. *pediatrics* 2001; **108**: 1339-48.
31. Papile LA, Burstein J, Burstein R, Koffler H. Incidence and evolution of subependymal and intraventricular haemorrhage: a study of infants with birthweights less than 1500g. *Journal of pediatrics* 1978; **92**: 529-34.
32. de Vries LS, Eken P, Dubowitz LMS. The spectrum of leukomalcia using cranial ultrasound. *Behavioural Brain Research* 1992; **49**: 1-6.
33. Groenendaal F, Termote JU, Heide-Jalving M. Complications affecting preterm neonates from 1991 to 2006: what have we gained? *Acta Paediatrica* 2010; **99**: 354-8.
34. International Committee for the Classification of Retinopathy of Prematurity (ICCROP). The International Classification of Retinopathy of Prematurity revisited. *Archives of Opthalmology* 2005; **123**(7): 991-9.
35. Ehrenkranz RA, Walsh MC, Vohr B, et al. Validation of the National Institutes of Health Consensus Definition of Bronchopulmonary Dysplasia. *Pediatrics* 2005; **116**: 1353-60.
36. Bell MJ, Ternberg JL, Feigin RD, et al. Neonatal necrotizing enterocolitis: theraupeutic decisions based on clinical staging. *Annals of Surgery* 1978; **187**: 1-7.
37. Vergnano S, Menson E, Kennea N, et al. Neonatal infections in England: the NeonIN surveillance network. *Arch Dis Child Fetal Neonatal Ed* 2011; **96**: F9-F14.
38. Johnson S, Fawke J, Hennessy E, et al. Neurodevelopmental disability through 11 years of age in children born before 26 weeks of gestation. *Pediatrics* 2009; **124**(2): e249-e57.

39. Larroque B, Ancel P-Y, Marret S, et al. Neurodevelopmental disabilities and special care of 5-year-old children born before 33 weeks of gestation (the EPIPAGE study): A longitudinal cohort study. *Lancet* 2008; **371**(9615): 813-20.
40. Bhutta AT, Cleves MA, Casey PH, Cradock MM, Anand KJS. Cognitive and behavioral outcomes of school-aged children who were born preterm: A meta-analysis. *JAMA* 2002; **288**(6): 728-37.
41. Johnson S, Hollis C, Kochhar P, Hennessy E, Wolke D, Marlow N. Psychiatric Disorders in Extremely Preterm Children: Longitudinal Finding at Age 11 Years in the EPICure Study. *Journal of the American Academy of Child and Adolescent Psychiatry* 2010; **49**(5): 453-63.e1.
42. Farooqi A, Hagglof B, Sedin G, Gothefors L, Serenius F. Mental health and social competencies of 10- to 12-year-old children born at 23 to 25 weeks of gestation in the 1990s: A Swedish national prospective follow-up study. *Pediatrics* 2007; **120**(1): 118-33.
43. Delobel-Ayoub M, Arnaud C, White-Koning M, et al. Behavioral problems and cognitive performance at 5 years of age after very preterm birth: The EPIPAGE study. *Pediatrics* 2009; **123**(6): 1485-92.
44. Johnson S, Hennessy E, Smith R, Trikic R, Wolke D, Marlow N. Academic attainment and special educational needs in extremely preterm children at 11 years of age: The EPICure study. *Archives of Disease in Childhood: Fetal and Neonatal Edition* 2009; **94**(4): F283-F9.
45. Marlow N, Hennessy EM, Bracewell MA, Wolke D. Motor and executive function at 6 years of age after extremely preterm birth. *Pediatrics* 2007; **120**(4): 793-804.
46. Johnson S, Wolke D, Hennessy E, Marlow N. Educational outcomes in extremely preterm children: Neuropsychological correlates and predictors of attainment. *Developmental Neuropsychology* 2011; **36**(1): 74-95.
47. Moore T, Hennessy EM, Myles J, et al. Neurological and developmental outcome in extremely preterm children born in England in 1995 and 2006: the EPICure studies. *British Medical Journal* 2012; **345**: DOI: 10.1136/bmj.e7961.
48. Ment LR, Vohr B, Allan W, et al. Change in cognitive function over time in very low-birth-weight infants. *JAMA* 2003; **289**(6): 705-11.
49. Taylor HG, Minich NM, Klein N, Hack M. Longitudinal outcomes of very low birth weight: neuropsychological findings. *Journal of the international neuropsychological society* 2004; **10**: 149-63.
50. Allin M, Walshe M, Fern A, et al. Cognitive maturation in preterm and term born adolescents. *Journal of neurology, neurosurgery and psychiatry* 2008; **79**: 381-6.
51. Sansavini A, Pentimonti J, Justice L, et al. Language, motor and cognitive development of extremely preterm children: modeling individual growth trajectories over the first three years of life. *Journal of communication disorders* 2014; **49**: 55-68.
52. Royston P, Moons KGM, Altman DG, Vergouwe Y. Prognosis and prognostic research: developing a prognostic model. *British Medical Journal* 2009; **338**: 1373-7.
53. Altman DG, Vergouwe Y, Royston P, Moons KGM. Prognosis and prognostic research: validating a prognostic model. *British Medical Journal* 2009; **338**: 1432-5.
54. Tyson JE, Parikh NA, Langer J, Green C, Higgins RD. Intensive care for extreme prematurity - Moving beyond gestational age. *NEJM* 2008; **358**(16): 1672-81.
55. Ambalavanan N, Carlo WA, Tyson JE, et al. Outcome trajectories in extremely preterm infants. *Pediatrics* 2012; **130**(1): e115-e25.
56. Boland RA, Davis PG, Dawson JA, Doyle LW. Predicting death or major neurodevelopmental disability in extremely preterm infants born in Australia. *Arch Dis Child Fetal Neonatal Ed* 2013; **98**(3): F201-4.

57. Costeloe K, Hennessy E, Gibson AT, Marlow N, Wilkinson AR. The EPICure study: Outcomes to discharge from hospital for infants born at the threshold of viability. *Pediatrics* 2000; **106**(4 I): 659-71.
58. Hack M, Fanaroff AA. Outcomes of children of extremely low birthweight and gestational age in the 1990's. *Early Human Development* 1999; **53**: 193-218.
59. Arnold CC, Kramar MS, Hobbs CA, Mclean FH, Usher RH. Very low birth weight: a problematic cohort for epidemiologic studies of very small or immature neonates. *American Journal of Epidemiology* 1991; **134**(6): 604-13.
60. Van den Berg T, Heymans MW, Leone SS, et al. Overview of data-synthesis in systematic reviews of studies on outcome prediction models. *BMC Medical research methodology* 2013; **13**(42).
61. Riley RD, Hayden JA, Steyerberg EW, et al. Prognosis Research Strategy (PROGRESS) 2: Prognostic Factor Research. *PLOS Medicine* 2013; **10**(2): e1001380. .
62. Grimmer I, Metze B, Walch E, Scholz T, Buhrer C. Predicting neurodevelopmental impairment in preterm infants by standardized neurological assessments at 6 and 12 months corrected age. *Acta Paediatrica, International Journal of Paediatrics* 2010; **99**(4): 526-30.
63. Harijan P, Beer C, Glazebrook C, et al. Predicting developmental outcomes in very preterm infants: Validity of a neonatal neurobehavioral assessment. *Acta Paediatrica, International Journal of Paediatrics* 2012; **101**(7): e275-e81.
64. Roze E, Meijer L, Van Braeckel KNJA, Ruiter SAJ, Bruggink JLM, Bos AF. Developmental trajectories from birth to school age in healthy term-born children. *Pediatrics* 2010; **126**(5): e1134-42.
65. Marlow N, Wolke D, Bracewell MA, Samara M. Neurologic and Developmental disability at six years of age after extremely preterm birth. *New England Journal of Medicine* 2005; **352**: 9-19.
66. Roberts G, Anderson PJ, Doyle LW. The stability of the diagnosis of developmental disability between ages 2 and 8 in a geographic cohort of very preterm children born in 1997. *Archives of Disease in Childhood* 2012; **95**: 786-90.
67. Hack M, Taylor HG, Drotar D, et al. Poor predictive validity of the Bayley Scales of Infant Development for cognitive function of extremely low birth weight children at school age. *Pediatrics* 2005; **116**(2): 333-41.
68. Collins G, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *Journal of clinical epidemiology* 2015; **68**(2): 112-21.
69. Hayden JA, Van der Windt DA, Cartwright JL, Cote P, Bombardier C. Asessing bias in studies of prognostic factors. *Annals of Internal Medicine* 2013; **158**: 280-6.
70. Hoogendoom WE, Van Poppel MNM, Bongers PM, Koes BW, Bouter LM. Systematic review of psychosocial factors at work and private life as risk factors for back pain. *Spine* 2000; **25**(16): 2114-25.
71. Tangri N, Kitsios GD, Inker LA, et al. Risk prediction models for patients with chronic kidney disease. *Annals of internal medicine* 2013; **158**.
72. Jackson R, Ameratunga S, Broad J, et al. The GATE frame: critical appraisal with pictures. *Evidence Based Medicine* 2006; **11**: 35-8.
73. Collaboration TC. Cochrane handbook for systematic reviews of interventions version 5.1.0. 2011.
74. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *British Medical Journal* 2009; **339**: b2535.

75. Aarnoudse-Moens CSH, Weisglas-Kuperus N, Duivenvoorden HJ, Oosterlaan J, Van Goudoever JB. Neonatal and parental predictors of executive function in very preterm children. *Acta Paediatrica, International Journal of Paediatrics* 2013; **102**(3): 282-6.
76. Ambalavanan N, Nelson KG, Alexander G, Johnson SE, Biasini F, Carlo WA. Prediction of neurologic morbidity in extremely low birth weight infants. *Journal of perinatology : official journal of the California Perinatal Association* 2000; **20**(8 Pt 1): 496-503.
77. Brevaut-Malaty V, Busuttil M, Einaudi MA, et al. Evaluation of a tool to screen at preschool age for minor cognitive disorders liable to affect schooling among children born premature. *Child's Nervous System* 2009; **25**(11): 1437-45.
78. Broitman E, Ambalavanan N, Higgins RD, et al. Clinical data predict neurodevelopmental outcome better than head ultrasound in extremely low birth weight infants. *The Journal of pediatrics* 2007; **151**(5): 500-5, 5.e1-5052.
79. Costantine MM, How HY, Coppage K, Maxwell RA, Sibai BM. Does peripartum infection increase the incidence of cerebral palsy in extremely low birthweight infants? *American Journal of Obstetrics and Gynecology* 2007; **196**(5): e6-e8.
80. Gargus RA, Vohr BR, Tyson JE, et al. Unimpaired outcomes for extremely low birth weight infants at 18 to 22 months. *Pediatrics* 2009; **124**(1): 112-21.
81. 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. *Archives of Pediatrics and Adolescent Medicine* 2000; **154**(7): 725-31.
82. Kumar P, Shankaran S, Ambalavanan N, et al. Characteristics of extremely low-birth-weight infant survivors with unimpaired outcomes at 30 months of age. *Journal of Perinatology* 2013; **33**(10): 800-5.
83. Laptook AR, O'Shea TM, Shankaran S, Bhaskar B. Adverse neurodevelopmental outcomes among extremely low birth weight infants with a normal head ultrasound: prevalence and antecedents. *Pediatrics* 2005; **115**(3): 673-80.
84. Vohr BR, Wright LL, Dusick AM, 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.
85. Vohr BR, Wright LL, Dusick AM, et al. Center Differences and Outcomes of Extremely Low Birth Weight Infants. *Pediatrics* 2004; **113**(4 I): 781-9.
86. Voss W, Neubauer AP. Outcome of extremely low birth weight infants at school age. The influence of perinatal factors on neurodevelopment. *Padiatrische Praxis* 2009; **73**(3): 373-82.
87. Voss W, Neubauer AP. Outcome of extremely low birth weight infants at school age. Perinatal risk factors on neurodevelopment? *Gynakologische Praxis* 2010; **34**(2): 319-28.
88. Adams-Chapman I, Bann CM, Vaucher YE, Stoll BJ, Eunice Kennedy Shriver National Institute of Child H, Human Development Neonatal Research N. Association between feeding difficulties and language delay in preterm infants using Bayley Scales of Infant Development-Third Edition. *Journal of Pediatrics* 2013; **163**(3): 680-5.e1-3.
89. Andrews WW, Cliver SP, Biasini F, et al. Early preterm birth: association between in utero exposure to acute inflammation and severe neurodevelopmental disability at 6 years of age. *American Journal of Obstetrics and Gynecology* 2008; **198**(4): 466.e1-.e11.
90. Arnaud C, Daubisse-Marliac L, White-Koning M, et al. Prevalence and associated factors of minor neuromotor dysfunctions at age 5 years in prematurely born children: The EPIPAGE study. *Archives of Pediatrics and Adolescent Medicine* 2007; **161**(11): 1053-61.

91. Ayoubi JM, Audibert F, Boithias C, et al. Perinatal factors affecting survival and survival without disability of extreme premature infants at two years of age. *European journal of obstetrics, gynecology, and reproductive biology* 2002; **105**(2): 124-31.
92. Beaino G, Khoshnood B, Kaminski M, et al. Predictors of the risk of cognitive deficiency in very preterm infants: The EPIPAGE prospective cohort. *Acta Paediatrica, International Journal of Paediatrics* 2011; **100**(3): 370-8.
93. Beaino G, Khoshnood B, Kaminski M, et al. Predictors of cerebral palsy in very preterm infants: The EPIPAGE prospective population-based cohort study. *Developmental Medicine and Child Neurology* 2010; **52**(6): e119-e25.
94. Bolisetty S, Dhawan A, Abdel-Latif M, et al. Intraventricular hemorrhage and neurodevelopmental outcomes in extreme preterm infants. *Pediatrics* 2014; **133**(1): 55-62.
95. Brevaut-Malaty V, Busuttil M, Einaudi MA, Monnier AS, D'Ercole C, Gire C. Longitudinal follow-up of a cohort of 350 singleton infants born at less than 32 weeks of amenorrhea: neurocognitive screening, academic outcome, and perinatal factors. *European Journal of Obstetrics Gynecology and Reproductive Biology* 2010; **150**(1): 13-8.
96. Charkaluk ML, Truffert P, Fily A, Ancel PY, Pierrat V. Neurodevelopment of children born very preterm and free of severe disabilities: The Nord-Pas de Calais Epipage cohort study. *Acta Paediatrica, International Journal of Paediatrics* 2010; **99**(5): 684-9.
97. Cooke RWI. Perinatal and postnatal factors in very preterm infants and sequent cognitive and motor abilities. *Archives of Disease in Childhood: Fetal and Neonatal Edition* 2005; **90**(1): F60-F3.
98. Davis N, Ford G, Anderson P, Doyle L. Developmental coordination disorder at 8 years of age in a regional cohort of extremely-lowbirthweight or very preterm infants. *Developmental Medicine & Child Neurology* 2007; **49**(5): 325-30.
99. de Waal CG, Weisglas-Kuperus N, van Goudoever JB, Walther FJ. Mortality, neonatal morbidity and two year follow-up of extremely preterm infants born in the netherlands in 2007. *PLoS ONE* 2012; **7**(7): e41302.
100. Delobel-Ayoub M, Kaminski M, Marret S, et al. Behavioral outcome at 3 years of age in very preterm infants: the EPIPAGE study. *Pediatrics* 2006; **117**(6): 1996-2005.
101. Dewey D, Creighton DE, Heath JA, et al. Assessment of developmental coordination disorder in children born with extremely low birth weights. *Developmental Neuropsychology* 2011; **36**(1): 42-56.
102. Doyle LW. Outcome at 5 years of age of children 23 to 27 weeks' gestation: Refining the prognosis. *Pediatrics* 2001; **108**(1): 134-41.
103. Emsley HCA, Wardle SP, Sims DG, Chiswick ML, D'Souza SW. Increased survival and deteriorating developmental outcome in 23 to 25 week old gestation infants, 1990-4 compared with 1984-9. *Archives of Disease in Childhood: Fetal and Neonatal Edition* 1998; **78**(2): F99-F104.
104. Farooqi A, Hagglof B, Sedin G, Serenius F. Impact at age 11 years of major neonatal morbidities in children born extremely preterm. *Pediatrics* 2011; **127**(5): e1247-e57.
105. Fily A, Pierrat V, Delporte V, Breart G, Truffert P. Factors associated with neurodevelopmental outcome at 2 years after very preterm birth: the population-based Nord-Pas-de-Calais EPIPAGE cohort. *Pediatrics* 2006; **117**(2): 357-66.
106. Ford RM, Neulinger K, O'Callaghan M, Mohay H, Gray P, Shum D. Executive function in 7-9-year-old children born extremely preterm or with extremely low birth weight: Effects of biomedical history, age at assessment, and socioeconomic status. *Archives of Clinical Neuropsychology* 2011; **26**(7): 632-44.

107. Franz AR, Pohlandt F, Bode H, et al. Intrauterine, early neonatal, and postdischarge growth and neurodevelopmental outcome at 5.4 years in extremely preterm infants after intensive neonatal nutritional support. *Pediatrics* 2009; **123**(1): e101-9.
108. Gianni ML, Picciolini O, Vegni C, Gardon L, Fumagalli M, Mosca F. Twelve-month neurofunctional assessment and cognitive performance at 36 months of age in extremely low birth weight infants. *Pediatrics* 2007; **120**(5): 1012-9.
109. Goyen TA, Lui K. Developmental coordination disorder in "apparently normal" schoolchildren born extremely preterm. *Archives of Disease in Childhood* 2009; **94**(4): 298-302.
110. Hack M, Taylor HG, Schluchter M, Andreias L, Drotar D, Klein N. Behavioral outcomes of extremely low birth weight children at age 8 years. *Journal of developmental and behavioral pediatrics : JDBP* 2009; **30**(2): 122-30.
111. Hansen BM, Hoff B, Uldall P, Greisen G. Perinatal risk factors of adverse outcome in very preterm children: A role of initial treatment of respiratory insufficiency? *Acta Paediatrica, International Journal of Paediatrics* 2004; **93**(2): 185-9.
112. Helderman JB, O'Shea TM, Kuban KC, et al. Antenatal antecedents of cognitive impairment at 24 months in extremely low gestational age newborns. *Pediatrics* 2012; **129**(3): 494-502.
113. Hok-Wikstrand M, Hard AL, Niklasson A, Hellstrom A. Early postnatal growth variables are related to morphologic and functional ophthalmologic outcome in children born preterm. *Acta Paediatrica, International Journal of Paediatrics* 2010; **99**(5): 658-64.
114. Holmstrom GE, Kallen K, Hellstrom A, et al. Ophthalmologic outcome at 30 months' corrected age of a prospective Swedish cohort of children born before 27 weeks of gestation: the extremely preterm infants in sweden study. *JAMA Ophthalmol* 2014; **132**(2): 182-9.
115. Howard K, Roberts G, Lim J, et al. Biological and environmental factors as predictors of language skills in very preterm children at 5 years of age. *Journal of Developmental and Behavioral Pediatrics* 2011; **32**(3): 239-49.
116. Iijima S, Arai H, Ozawa Y, Kawase Y, Uga N. Clinical patterns in extremely preterm (22 to 24 weeks of gestation) infants in relation to survival time and prognosis. *American Journal of Perinatology* 2009; **26**(6): 399-406.
117. Janssen AJWM, der Sanden MWGNv, Akkermans RP, Tissingh J, Oostendorp RAB, Kollee LAA. A model to predict motor performance in preterm infants at 5 years. *Early Human Development* 2009; **85**(9): 599-604.
118. Janssen AJWM, Nijhuis-van Der Sanden MWG, Akkermans RP, Oostendorp RAB, Kollee LAA. Influence of behaviour and risk factors on motor performance in preterm infants at age 2 to 3 years. *Developmental Medicine and Child Neurology* 2008; **50**(12): 926-31.
119. Johnson S, Hollis C, Kochhar P, Hennessy E, Wolke D, Marlow N. Autism Spectrum Disorders in Extremely Preterm Children. *Journal of Pediatrics* 2010; **156**(4): 525-31.e2.
120. Jones KM, Champion PR, Woodward LJ. Social competence of preschool children born very preterm. *Early Human Development* 2013; **89**(10): 795-802.
121. Kiechl-Kohlendorfer U, Ralser E, Pupp Peglow U, Pehboeck-Walser N, Fussenegger B. Early risk predictors for impaired numerical skills in 5-year-old children born before 32weeks of gestation. *Acta Paediatrica, International Journal of Paediatrics* 2013; **102**(1): 66-71.
122. Leversen KT, Sommerfelt K, Ronnestad A, et al. Predicting neurosensory disabilities at two years of age in a national cohort of extremely premature infants. *Early Human Development* 2010; **86**(9): 581-6.

123. Leversen KT, Sommerfelt K, Ronnestad A, et al. Prediction of neurodevelopmental and sensory outcome at 5 years in norwegian children born extremely preterm. *Pediatrics* 2011; **127**(3): e630-e8.
124. Lowe J, MacLean PC, Shaffer ML, Watterberg K. Early working memory in children born with extremely low birth weight: Assessed by object permanence. *Journal of Child Neurology* 2009; **24**(4): 410-5.
125. Marston L, Peacock JL, Calvert SA, Greenough A, Marlow N. Factors affecting vocabulary acquisition at age 2 in children born between 23 and 28 weeks' gestation. *Developmental Medicine and Child Neurology* 2007; **49**(8): 591-6.
126. Matsuda Y, Kouno S, Hiroyama Y, et al. Intrauterine infection, magnesium sulfate exposure and cerebral palsy in infants born between 26 and 30 weeks of gestation. *European Journal of Obstetrics, Gynecology, & Reproductive Biology* 2000; **91**(2): 159-64.
127. Mercier CE, Dunn MS, Ferrelli KR, Howard DB, Soll RF. Neurodevelopmental outcome of extremely low birth weight infants from the vermont oxford network: 1998-2003 and the Vermont Oxford Network ELBW infant follow-up study group. *Neonatology* 2010; **97**(4): 329-38.
128. Messinger D, Lambert B, Bauer CR, Bann CM, Hamlin-Smith K, Das A. The relationship between behavior ratings and concurrent and subsequent mental and motor performance in toddlers born at extremely low birth weight. *Journal of Early Intervention* 2010; **32**(3): 214-33.
129. Mikkola K, Ritari N, Tommiska V, et al. Neurodevelopmental outcome at 5 years of age of a national cohort of extremely low birth weight infants who were born in 1996-1997. *Pediatrics* 2005; **116**(6): 1391-400.
130. Moore T, Johnson S, Hennessy E, Marlow N. Screening for autism in extremely preterm infants: Problems in interpretation. *Developmental Medicine and Child Neurology* 2012; **54**(6): 514-20.
131. Neubauer AP, Voss W, Kattner E. Outcome of extremely low birth weight survivors at school age: The influence of perinatal parameters on neurodevelopment. *European Journal of Pediatrics* 2008; **167**(1): 87-95.
132. Orchinik LJ, Taylor HG, Espy KA, et al. Cognitive outcomes for extremely preterm/extremely low birth weight children in kindergarten. *Journal of the International Neuropsychological Society* 2011; **17**(6): 1067-79.
133. Peralta-Carcelen M, Bailey K, Rector R, Gantz M. Behavioral and socioemotional competence problems of extremely low birth weight children. *Journal of Perinatology* 2013; **33**(11): 887-92.
134. Pin TW, Eldridge B, Galea MP. Motor trajectories from 4 to 18months corrected age in infants born at less than 30weeks of gestation. *Early Human Development* 2010; **86**(9): 573-80.
135. Potharst ES, Houtzager BA, Van Sonderen L, et al. Prediction of cognitive abilities at the age of 5years using developmental follow-up assessments at the age of 2 and 3 years in very preterm children. *Developmental Medicine and Child Neurology* 2012; **54**(3): 240-6.
136. Potharst ES, Van Wassenaer-Leemhuis AG, Houtzager BA, et al. Perinatal risk factors for neurocognitive impairments in preschool children born very preterm. *Developmental Medicine and Child Neurology* 2013; **55**(2): 178-84.
137. Sann L, Bourgeois J, Stephant A, Putet G. A study of 490 premature babies born before 31 weeks' gestation age. Mortality and development before 2 years. *Revue Francaise de Gynecologie et d'Obstetrique* 1999; **94**(6): 500-10.

138. Sann L, Bourgeois J, Stephant A, Putet G. Outcome of 249 infants of less than 29 weeks gestational age. *Archives de Pediatrie* 2001; **8**(3): 250-8.
139. Sansavini A, Guarini A, Savini S. Linguistic and cognitive delays in very preterm infants at 2 years: General or specific delays? *Revista de Logopedia, Foniatria y Audiologia* 2011; **31**(3): 133-47.
140. Schlapbach LJ, Aebischer M, Adams M, et al. Impact of sepsis on neurodevelopmental outcome in a swiss national cohort of extremely premature infants. *Pediatrics* 2011; **128**(2): e348-e57.
141. Schmidhauser J, Caflisch J, Rousson V, Bucher HU, Largo RH, Latal B. Impaired motor performance and movement quality in very-low-birthweight children at 6 years of age. *Developmental Medicine and Child Neurology* 2006; **48**(9): 718-22.
142. Schmidt B, Asztalos EV, Roberts RS, Robertson CM, Sauve RS, Whitfield MF. Impact of bronchopulmonary dysplasia, brain injury, and severe retinopathy on the outcome of extremely low-birthweight infants at 18 months: Results from the trial of indomethacin prophylaxis in preterms. *JAMA: Journal of the American Medical Association* 2003; **289**(9): 1124-9.
143. Spittle AJ, Treyvaud K, Doyle LW, et al. Early Emergence of Behavior and Social-Emotional Problems in Very Preterm Infants. *Journal of the American Academy of Child and Adolescent Psychiatry* 2009; **48**(9): 909-18.
144. Stahlmann N, Rapp M, Herting E, Thyen U. Outcome of extremely premature infants at early school age: health-related quality of life and neurosensory, cognitive, and behavioral outcomes in a population-based sample in northern Germany. *Neuropediatrics* 2009; **40**(3): 112-9.
145. Stephens BE, Bann CM, Watson VE, et al. Screening for Autism spectrum disorders in extremely preterm infants. *Journal of Developmental and Behavioral Pediatrics* 2012; **33**(7): 535-41.
146. Stoelhorst GM, Martens SE, Rijken M, et al. Behaviour at 2 years of age in very preterm infants (gestational age < 32 weeks). *Acta Paediatrica* 2003; **92**(5): 595-601.
147. Stoelhorst GMSJ, Rijken M, Martens SE, et al. Developmental outcome at 18 and 24 months of age in very preterm children: A cohort study from 1996 to 1997. *Early Human Development* 2003; **72**(2): 83-95.
148. Taylor HG, Klein N, Anselmo MG, Minich N, Espy KA, Hack M. Learning problems in kindergarten students with extremely preterm birth. *Archives of Pediatrics and Adolescent Medicine* 2011; **165**(9): 819-25.
149. Taylor HG, Klein N, Drotar D, Schluchter M, Hack M. Consequences and risks of <1000-g birth weight for neuropsychological skills, achievement, and adaptive functioning. *Journal of Developmental and Behavioral Pediatrics* 2006; **27**(6): 459-69.
150. Tommiska V, Heinonen K, Kero P, et al. A national two year follow up study of extremely low birthweight infants born in 1996-1997. *Archives of Disease in Childhood: Fetal and Neonatal Edition* 2003; **88**(1): F29-F34.
151. Toome L, Varendi H, Mannamaa M, Vals MA, Tanavsuu T, Kolk A. Follow-up study of 2-year-olds born at very low gestational age in Estonia. *Acta Paediatrica, International Journal of Paediatrics* 2013; **102**(3): 300-7.
152. Treyvaud K, Ure A, Doyle LW, et al. Psychiatric outcomes at age seven for very preterm children: rates and predictors. *J Child Psychol Psychiatry* 2013; **54**(7): 772-9.

153. Trittmann JK, Nelin LD, Klebanoff MA. Bronchopulmonary dysplasia and neurodevelopmental outcome in extremely preterm neonates. *European Journal of Pediatrics* 2013; **172**(9): 1173-80.
154. Vincer MJ, Allen AC, Joseph KS, Stinson DA, Scott H, Wood E. Increasing prevalence of cerebral palsy among very preterm infants: a population-based study. *Pediatrics* 2006; **118**(6): e1621-6.
155. Vohr BR, Wright LL, Poole WK, McDonald SA. Neurodevelopmental outcomes of extremely low birth weight infants <32 weeks' gestation between 1993 and 1998. *Pediatrics* 2005; **116**(3): 635-43.
156. Voss W, Jungmann T, Wachtendorf M, Neubauer AP. Long-term cognitive outcomes of extremely low-birth-weight infants: The influence of the maternal educational background. *Acta Paediatrica, International Journal of Paediatrics* 2012; **101**(6): 569-73.
157. Wong HS, Huertas-Ceballos A, Cowan FM, Modi N. Evaluation of early childhood social-communication difficulties in children born preterm using the quantitative checklist for autism in toddlers. *The Journal of Pediatrics* 2014; **164**(1): 26-33.
158. Wood NS, Costeloe K, Gibson AT, Hennessy EM, Marlow N, Wilkinson AR. The EPICure study: Associations and entecedents of neurological and developmental disability at the 30 months of age following extremely preterm birth. *Archives of Disease in Childhood: Fetal and Neonatal Edition* 2005; **90**(2): F134-F40.
159. Zanudin A, Burns Y, Gray PH, Danks M, Poulsen L, Watter P. Perinatal events and motor performance of children born with ELBW and nondisabled. *Pediatr* 2013; **25**(1): 30-5.
160. Zanudin A, Gray PH, Burns Y, Danks M, Watter P, Poulsen L. Perinatal factors in non-disabled ELBW school children and later performance. *Journal of Paediatrics and Child Health* 2013; **49**(1): E62-E7.
161. Larroque B, Breart G, Kaminski M, et al. Survival of very preterm infants: Epipage, a population based cohort study. *Arch Dis Child Fetal Neonatal Ed* 2004; **89**: F139-F44.
162. Lee HC, Green C, Hintz SR, et al. Prediction of death for extremely premature infants in a population-based cohort. *Pediatrics* 2010; **126**: e644.
163. Amiel-Tison C, Stewart A. Follow-up studies during the first five years of life: a pervasive assessment of a neurological function. *Archives of Disease in Childhood* 1989; **64**: 496-502.
164. Josse D. Brunet-Lezine revise: Echelle de developpement psychomoteur de la premiere enfance. Paris: Etablissements d'applications psychotechniques; 1997.
165. Griffiths R. The Griffiths Mental Development Scales from birth to two years. Henley. Henley: The Test Agency Limited; 1996.
166. Bayley N. Bayley Scales of Infant Development-II. 2nd ed: Psychological Corporation; San Antonio, TX; 1993.
167. Bayley N. Bayley Scales of Infant Development-III. 3rd ed: Harcourt Assessment; San Antonio, TX; 2006.
168. Moons KG, Royston P, Vergouwe Y, Grobbee DE, Altman DG. Prognosis and prognostic research: what, why, and how? *British Medical Journal* 2009; **338**: b606.
169. Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. *Journal of clinical epidemiology* 1996; **49**(12): 1373-9.
170. Harrell FE, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in medicine* 1996; **15**(4): 361-87.

171. Steyerberg EW, Eijkemans MJC, Harrell FE. Prognostic modeling with logistic regression analysis: in search of a sensible strategy in small data sets. *Medical Decision Making* 2001; **21**(1): 45-56.
172. Royston P, Sauerbrei W. Multivariable model-building: a pragmatic approach to regression analysis based on fractional polynomials for modelling continuous variables. Chichester UK: John Wiley and Sons Ltd; 2008.
173. Royston P, Altman DG, Sauerbrei W. Dichotomising continuous predictors in multiple regression: a bad idea. *Statistics in medicine* 2006; **25**(1): 127-41.
174. Altman DG, Royston P. The cost of dichotomising continuous variables [statistical notes]. *British Medical Journal* 2006; **332**: 1080.
175. Sun GW, Shook TL, Kay GL. Innapropriate use of bivariable analysis to screen risk factors for use in multivariable analysis. *Journal of clinical epidemiology* 1996; **49**(8): 907-16.
176. Steyerberg EW, Eijkemans MJ, Harrell FE, Habbema JD. Prognostic modelling with logistic regression analysis: a comparison of selection and estimation methods in small data sets. *Statistics in medicine* 2000; **19**(8): 1059-79.
177. Austin PC, Tu JV. Automated variable selection methods for logistic regression produced unstable models for predicting acute myocardial infarction mortality. *journal of clinical epidemiology* 2004; **57**(11): 1138-46.
178. Harrell FE. Regression modelling strategies: with application to linear models, logistic regression and survival analysis. New York: Springer-Verlag; 2001.
179. Wolke D, Sohne B, Ohrt B, Riegel K. Follow-up of preterm children: important to document drop-outs [letter]. *Lancet* 1995; **345**(8947): 447.
180. 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; **79**(2): F83-F7.
181. Steyerberg EW. Clinical prediction models. New York: Springer; 2008.
182. Geersing G-J, Bouwmeester W, Zuithoff P, Spijker R, Leeflang M, Moons K. Search filters for finding prognostic and diagnostic prediction studies in Medline to enhance systematic reviews. *PLoS ONE* 2012; **7**(2).
183. Platt MJ, Cans C, Johnson A, et al. Trends in cerebral palsy among infants of very low birthweight (<1500 g) or born prematurely (<32 weeks) in 16 European centres: a database study. *Lancet* 2007; **369**: 43-50.
184. Himmelmann K, Uvebrant p. The panorama of cerebral palsy in Sweden. XI. Changing patterns in the birth-year period 2003–2006. *Acta Paediatrica* 2014; **103**(6): 618-24.
185. Association AP. Diagnostic and Statistical Manual of Mental Disorders. 4th edition ed. Washington, DC: American Psychiatric Association; 1994.
186. Foulder-Hughes LA, Cooke RW. Motor, cognitive, and behavioural disorders in children born very preterm. *Developmental medicine and child neurology* 2003; **45**(2): 97-103.
187. Gilbert C. Changing challenges in the control of blindness in children. *Eye* 2007; **21**(10): 1338-43.
188. Farooqi A, Hagglof B, Sedin G, Gothefors L, Serenius F. Chronic conditions, functional limitations, and special health care needs in 10- to 12-year-old children born at 23 to 25 weeks' gestation in the 1990s: a Swedish national prospective follow-up study. *Pediatrics* 2006; **118**(5): e1466-77.
189. O'Connor AR, Stephenson T, Johnson A, et al. Long-term ophthalmic outcome of low birth weight children with and without retinopathy of prematurity. *Pediatrics* 2002; **109**(1): 12-8.

190. Saigal S, Stoskopf B, Boyle M, et al. Comparison of current health, functional limitations, and health care use of young adults who were born with extremely low birth weight and normal birth weight. *Pediatrics* 2007; **119**(3): e562-e73.
191. van Dommelen P, Verkerk PH, van Straaten HLM. Hearing loss by week of gestation and birth weight in very preterm neonates. *The Journal of pediatrics* 2015; **166**(4): 840-3.
192. Doyle LW, Keir E, Kitchen WH, Ford GW, Rickards AL, Kelly EA. Audiologic assessment of extremely low birth weight infants: A preliminary report. *Pediatrics* 1992; **90**(5 I): 744-9.
193. Amiel-Tison C. Neuromotor status. In: Taeusch H, Yogman M, eds. Follow-up management of the high-risk infant. Boston, MA: Little, Brown and Company; 1987: 115-26.
194. Bax M, Goldstein M, Rosenbaum Pea. Proposed definition and classification of cerebral palsy. *Developmental Medicine and Child Neurology* 2005; **47**: 571-6.
195. Bax M. Terminology and classification of cerebral palsy. *Developmental Medicine and Child Neurology* 1964; **6**: 295-7.
196. Surveillance of cerebral palsy in Europe (SCPE): a collaboration of cerebral palsy surveys and registers. *Developmental Medicine and Child Neurology* 2000; **42**(12): 816-24.
197. Touwen BCL. Examination of the child with minor neurological dysfunction. 2nd ed. Suffolk, England: The Lavenham Press LTD; 1979.
198. Hadders-Algra M, Mavinkurve-Groothuis AM, Groen SE, Stremmelaar EF, Martijn A, Butcher PR. Quality of general movements and the development of minor neurological dysfunction at toddler and school age. *Clinical Rehabilitation* 2004; **18**: 287-99.
199. Palisano R, Rosenbaum P, Walter S, Russell D, Wood E, Galuppi B. Development and reliability of a system to classify gross motor function in children with cerebral palsy. *Developmental Medicine and Child Neurology* 1997; **39**(4): 214-23.
200. Henderson SE, Sugden DA. Movement Assessment Battery for Children. London, UK: Psychological Corporation; 1992.
201. Piper MC, Darrah J. Motor assessment of the developing infant. USA: W.B. Saunders Company; 1994.
202. Bayley N. Bayley Scales of Infant Development. New York: Psychological Corporation; 1969.
203. Bruininks RH, Bruininks BD. BOT-2: Bruininks-Oseretsky test of motor proficiency (2nd ed.) Circle Pines, MN: American Guidance Service; 2005.
204. Bruininks RH. Bruininks-Oseretsky Test of Motor Proficiency: examiners manual. Circle Pines, MN: American Guidance Service; 1978.
205. Largo RH, Fischer JE, Caflisch JA. Manual for the Zurich Neuromotor Assessment. Zurich: AWE-Verlag; 2002.
206. Burns Y, O'Callaghan M, McDonell B, Rogers Y. Movement and motor development in ELBW infants at 1 year is related to cognitive and motor abilities at 4 years. *Early Human Development* 2004; **80**: 19-29.
207. Adams-Chapman I, Hansen NI, Stoll BJ, Higgins R. Neurodevelopmental outcome of extremely low birth weight infants with posthemorrhagic hydrocephalus requiring shunt insertion. *Pediatrics* 2008; **121**(5): e1167-77.
208. Sherlock RL, Anderson PJ, Doyle LW, et al. Neurodevelopmental sequelae of intraventricular haemorrhage at 8 years of age in a regional cohort of ELBW/very preterm infants. *Early Human Development* 2005; **81**(11): 909-16.
209. Kuban KCK, Allred EN, O'Shea MT, et al. Cranial ultrasound lesions in the NICU predict cerebral palsy at age 2 years in children born at extremely low gestational age. *Journal of Child Neurology* 2009; **24**(1): 63-72.

210. Marret S, Marchand-Martin L, Picaud JC, et al. Brain injury in very preterm children and neurosensory and cognitive disabilities during childhood: the EPIPAGE cohort study. *PLoS ONE* 2013; **8**(5): e62683.
211. Klebermass K, Olischar M, Waldhoer T, Fuiko R, Pollak A, Weninger M. Amplitude-integrated EEG pattern predicts further outcome in preterm infants. *Pediatric Research* 2011; **70**(1): 102-8.
212. Patra K, Wilson-Costello D, Taylor HG, Mercuri-Minich N, Hack M. Grades I-II intraventricular hemorrhage in extremely low birth weight infants: Effects on neurodevelopment. *Journal of Pediatrics* 2006; **149**(2): 169-73.
213. Payne AH, Hintz SR, Hibbs AM, et al. Neurodevelopmental outcomes of extremely low-gestational-age neonates with low-grade periventricular-intraventricular hemorrhage. *Jama, Pediatr* 2013; **167**(5): 451-9.
214. Doyle LW, Ehrenkranz RA, Halliday HL. Early (<8 days) postnatal corticosteroids for chronic lung disease in preterm infants. *Cochrane Database of Systematic Reviews* 2014; (5).
215. Doyle LW, Ehrenkranz RA, Halliday HL. Late (>7 days) postnatal corticosteroids for chronic lung disease in preterm infants. *Cochrane Database of Systematic Reviews* 2014; (5).
216. Doyle LW, Bowman E, Callanan C, et al. Postnatal corticosteroids and sensorineural outcome at 5 years of age. *Journal of Paediatrics and Child Health* 2000; **36**(3): 256-61.
217. Wilson-Costello D, Walsh MC, Langer JC, et al. Impact of postnatal corticosteroid use on neurodevelopment at 18 to 22 months' adjusted age: effects of dose, timing, and risk of bronchopulmonary dysplasia in extremely low birth weight infants. *Pediatrics* 2009; **123**(3): e430-7.
218. Roberts D, Dalziel SR. Antenatal corticosteroids for accelerating fetal lung maturation for women at risk of preterm birth. *Cochrane Database of Systematic Reviews* 2006; (3).
219. Carlo WA, McDonald SA, Fanaroff AA, et al. Association of antenatal corticosteroids with mortality and neurodevelopmental outcomes among infants born at 22 to 25 weeks' gestation. *JAMA* 2011; **306**(21): 2348-58.
220. Wong D, Abdel-Latif M, Kent A, Network N. Antenatal steroid exposure and outcomes of very premature infants: a regional cohort study. *Arch Dis Child Fetal Neonatal Ed* 2014; **99**(1): F12-20.
221. Lee BH, Stoll BJ, McDonald SA, Higgins RD. Neurodevelopmental outcomes of extremely low birth weight infants exposed prenatally to dexamethasone versus betamethasone. *Pediatrics* 2008; **121**(2): 289-96.
222. Amin SB, Kamaluddeen M, Sangem M. Neurodevelopmental outcome of premature infants after exposure to antenatal indomethacin. *American Journal of Obstetrics and Gynecology* 2008; **199**(1): 41.e1-.e8.
223. Doyle LW, Crowther CA, Middleton P, Marret S. Antenatal magnesium sulfate and neurologic outcome in preterm infants: a systematic review. *Obstetrics and Gynecology* 2009; **113**: 1327-33.
224. Conde-Agudelo A, Romero R. Antenatal magnesium sulfate for the prevention of cerebral palsy in preterm infants less than 34 weeks' gestation: a systematic review and metaanalysis *American Journal of Obstetrics and Gynecology* 2009; **200**(6): 596-606.
225. Himpens E, Van den Broeck C, Oostra A, Vanhaesebrouck P. Prevalence, type, distribution, and severity of cerebral palsy in relation to gestational age: a meta-analytic review. *Developmental Medicine and Child Neurology* 2008; **50**: 334-40.

226. Oskoui M, Coutinho F, Dykeman J, Jette N, Pringsheim T. An update on the prevalence of cerebral palsy: a systematic review and meta-analysis. *Developmental Medicine and Child Neurology* 2013; **55**: 509-19.
227. Hintz SR, Kendrick DE, Vohr BR, Poole WK, Higgins RD. Gender differences in neurodevelopmental outcomes among extremely preterm, extremely-low-birthweight infants. *Acta Paediatrica, International Journal of Paediatrics* 2006; **95**(10): 1239-48.
228. Kent AL, Wright IMR, Abdel-Latif ME, Bowen J, Bajuk B, Vincent T. Mortality and adverse neurologic outcomes are greater in preterm male infants. *Pediatrics* 2012; **129**(1): 124-31.
229. Neubauer V, Griesmaier E, Ralser E, Kiechl-Kohlendorfer U. The effect of sex on outcome of preterm infants-A population-based survey. *Acta Paediatrica* 2012; **101**(9): 906-11.
230. Skiold B, Alexandrou G, Padilla N, Blennow M, Vollmer B, Aden U. Sex differences in outcome and associations with neonatal brain morphology in extremely preterm children. *Journal of Pediatrics* 2014; **164**(5): 1012-8.
231. Kesler SR, Reiss AL, Vohr B, et al. Brain volume reductions within multiple cognitive systems in male preterm children at age 12. *Journal of Pediatrics* 2008; **152**: 513-20.
232. Anderson P, Doyle LW. Neurobehavioral outcomes of school-age children born extremely low birth weight or very preterm in the 1990s. *JAMA* 2003; **289**(24): 3264-72.
233. Mulder H, Pitchford NJ, Hagger MS, Marlow N. Development of executive function and attention in preterm children: A systematic review. *Developmental Neuropsychology* 2009; **34**(4): 393-421.
234. Taylor HG, Klein N, Minich NM, Hack M. Middle-school-age outcomes in children with very low birthweight. *Child development* 2000; **71**(6): 1495-511.
235. Aarnoudse-Moens CSH, Smidts DP, Oosterlaan J, Duivenvoorden HJ, Weisglas-Kuperus N. Executive function in very preterm children at early school age. *Journal of Abnormal Child Psychology* 2009; **37**(7): 981-93.
236. Barre N, Morgan A, Doyle LW, Anderson PJ. Language abilities in children who were very preterm and/or very low birth weight: a meta-analysis. *Journal of Pediatrics* 2011; **158**(5): 766-74.e1.
237. Kaufman AS, Kaufman NL. Kaufman Assessment Battery for Children. Circle Pines, MN: American Guidance Service; 1983.
238. Wechsler D. Wechsler's Preschool and Primary Scale of Intelligence-Revised. New York: The Psychological Corporation Harcourt Brace, Inc.; 1989.
239. Eliot CD. Differential Ability Scales. San Antonio, TX: The Psychological Corporation; 1990.
240. Tewes U, Rossmann R, Schallberger U. Der Hamburg-Wechsler-Intelligenztest fur Kinder (HAWIK-III). Bern: Huber-Verlag; 1999.
241. Wechsler D. WISC-IV administration and scoring manual. San Antonio, TX: The Psychological Corporation; 2003.
242. Woodcock RC, McGrew KS, Mather S. Woodcock-Johnson III Tests of Cognitive Abilities. Itasca, IL: Riverside Publishing; 2001.
243. Fenson L, Dale PS, Reznick JS, et al. McArthur Communicative Development Inventories: Users guide and technical manual. San Diego, CA: Singular Publishing; 1993.
244. Kaufman AS, Kaufman NL. Kaufman Survey of Early Academic and Language Skills (K-SEALS). Circle Pines, MN: American Guidance Service; 1993.

245. Woodcock RC, McGrew KS, Mather S. Woodcock-Johnson III Tests of Achievement. Itasca, IL: Riverside Publishing; 2001.
246. Korkman M, Kirk U, Kemp S. NEPSY - A Developmental NEuroPSYchological Assessment: Psychological Corporation: San Antonio, TX; 1998.
247. Kaufman L, Neurk H-C, Graf M, Krinzinger H, Delazer M, Willmes K. TEDI-MATH-Test zur Erfassung numerisch-rechnerischer Fertigkeiten vom Kindergarten bis zur 3. Klasse. Bern Verlag Hans Huber; 2009.
248. Wechsler D. Wechsler Individual Achievemnet Test, 2nd edition. London: Harcourt Assessment Inc.; 2004.
249. Peacock JL, Marston L, Marlow N, Calvert SA, Greenough A. Neonatal and infant outcome in boys and girls born very prematurely. *Pediatric Research* 2012; **71**(3): 305-10.
250. Ko G, Shah P, Lee SK, Asztalos E. Impact of maternal education on cognitive and language scores at 18 to 24 months among extemely preterm neonates. *American Journal of Perinatology* 2013; **30**(9): 723-30.
251. Kelly Y, Panico L, Bartley M, Marmot M, Nazroo J, Sacker A. Why does birthweight vary among ethnic groups in the UK? Findings from the Millennium Cohort Study. *Journal of public health* 2008; **31**(1): 131-7.
252. Nepomnyaschy L. Socioeconomic gradients in infant health across race and ethnicity. *Maternal child health journal* 2009; **13**: 720-31.
253. Duncan AF, Watterberg KL, Nolen TL, et al. Effect of ethnicity and race on cognitive and language testing at age 18-22 months in extremely preterm infants. *The Journal of Pediatrics* 2012; **160**(6): 966-71.
254. Brooks-Gunn J, Klebanov PK, Duncan GJ. Ethnic differences in children's intelligence test scores: role of economic deprivation, home environment, and maternal characterisitcs. *Child development* 1996; **67**(2): 396-408.
255. Woodward LJ, Clark CAC, Bora S, Inder TE. Neonatal white matter abnormalities an important predictor of neurocognitive outcome for very preterm children. *PLoS ONE* 2012; **7**(12).
256. Lowe JR, Erickson SJ, Schrader R, Duncan AF. Comparison of the Bayley II Mental Developmental Index and the Bayley III Cognitive Scale: Are we measuring the same thing? *Acta Paediatrica* 2012; **101**(2): e55-e8.
257. Johnson S, Marlow N. Preterm birth and childhood psychiatric disorders. *Pediatric Research* 2011; **69**(5 PART 2): 11R-8R.
258. Johnson S, Hollis C, Hennessy E, Kochhar P, Wolke D, Marlow N. Screening for autism in preterm children: diagnostic utility of the Social Communication Questionnaire. *Archives of Disease in Childhood* 2011; **96**(1): 73-7.
259. Glascoe FP. Screening for developmental and behavioral problems. *Mental Retardation and Developmental Disabilities Research Reviews* 2005; **11**: 173-9.
260. Limperopoulos C, Bassan H, Sullivan NR, et al. Positive screening for autism in ex-preterm infants: prevalence and risk factors. *Pediatrics* 2008; **121**(4): 758-65.
261. Indredavik MS, Vik T, Heyerdahl S, Kulseng S, Fayers P, Brubakk A-M. Psychiatric symptoms and disorders in adolescents with low birth weight. *Arch Dis Child Fetal Neonatal Ed* 2004; **89**: F445-F50.
262. Saigal S, Pinelli J, Hoult L, Kim MM, Boyle M. Psychopathology and social competencies of adolescents who were extremely low birth weight. *Pediatrics* 2003; **111**(5): 969-75.
263. Moster D, Terje R, Markestad T. Long-term medical and social consequences of preterm birth. *NEJM* 2008; **359**: 262-73.

264. Lindstrom K, Lindblad F, Hjern A. Psychiatric morbidity in adolescents and young adults born preterm: a Swedish national cohort study. *pediatrics* 2009; **123**: e47-e53.
265. Hille ETM, den Ouden AL, Saigal S, et al. Behavioural problems in children who weigh 1000g or less at birth in four countries. *The Lancet* 2001; **357**: 1641-43.
266. Sapolsky RM. Why stress is bad for your brain. *Science* 1996; **273**: 749-50.
267. Burnett AC. Prevalence of psychiatric diagnoses in preterm and full-term children, adolescents and young adults: a meta-analysis. *Psychological Medicine* 2011; **41**: 2463-74.
268. Goodman R. The Strengths and Difficulties Questionnaire: a research note. *Journal of Child Psychology & Psychiatry* 1997; **1997**(38): 581-6.
269. Achenbach TM. Manual for the Child Behavior Checklist/2-3 and 1992 profile. Burlington: University of Vermont, Department of Psychiatry; 1992.
270. Briggs-Gowan MJ, Carter AS, Irwin JR, Wachtel K, Cicchetti DV. The Brief Infant Toddler Social and Emotional Assessment: screening for social-emotional problems and delays in competence. *Journal of Pediatric Psychology* 2004; **29**: 143-55.
271. Briggs-Gowan MJ, Carter AS. The Infant Toddler Social and Emotional Assessment (ITSEA). Pre-publication version. Boston, MA: Yale University; 2000.
272. Sparrow SS, Carter AS, Cicchetti DV. Vineland Screener: Overview, reliability, validity, administration and scoring. New Haven, CT: Yale University; 2000.
273. Goodman R, Scott S. Comparing the Strengths and Difficulties Questionnaire and the Child Behavior Checklist: Is small beautiful? *Journal of Abnormal Child Psychology* 1999; **27**(1): 17-24.
274. Goodman R, Ford T, Richards H, Gatward R, Meltzer H. The Development and Well-Being Assessment: description and initial valuation of an integrated assessment of child and adolescent psychopathology. *Journl of child psychology and psychiatry* 2000; **41**: 645-55.
275. Siegel B. Pervasive Developmental Disorders Screening Test-II. San Antonio, TX: PsychCorp; 2004.
276. Lord C, Rutter M, DiLavore PCea. Autism Diagnostic Observation Schedule Manual. Los Angeles, CA: Western Psychological Services; 2001.
277. Robins DL, Dumont-Mathieu TM. Early screening for autism spectrum disorders: update on Modified Checklist for Autism in Toddlers and other measures. *Journal of Developmental and Behavioral Pediatrics* 2006; **27 (Suppl 2)**: S111-9.
278. Gadow KD, Sprafkin J. Child Symptom Inventory-4 Norms Manual. Stony Brook, NY: Checkmate Plus Ltd.; 1997.
279. Allison C, Baron-Cohen S, Wheelwright S, Charman T, Richler J, Pasco Gea. The Q-CHAT (Quantitative CHecklist for Autism in Toddlers): a normally distributed quantitative measure of autistic traits at 18-24 months of age: preliminary report *Journal of Autism and Developmental Disorders* 2008; **38**: 1414-25.
280. Rutter M, Bailey A, Lord C. The Social Communication Questionnaire. Los Angeles, CA: Western Psychological Services; 2003.
281. Treyvaud K, Doyle LW, Lee KJ, et al. Social-emotional difficulties in very preterm and term 2 year olds predict specific social-emotional problems at the age of 5 years. *Journal of Pediatric Psychology* 2012; **37**(7): 779-85.
282. Gray RF, Indurkhya A, McCormick MC. Prevalence, stability, and predictors of clinically significant behavior problems in low birth weight children at 3, 5, and 8 years of age. *Pediatrics* 2004; **114**(3): 736-43.
283. Johnson S, Marlow N. Growing up after extremely preterm birth: Lifespan mental health outcomes. *Seminars in Fetal and Neonatal Medicine* 2014; **19**: 97-104.

284. Dennis M, Francis DJ, Cirino PT, Schachar R, Barnes MA, Fletcher JM. Why IQ is not a covariate in cognitive studies of neurodevelopmental disorders. *Journal of the International Neuropsychological Society* 2009; **15**: 331-43.
285. Kuban KC, O'Shea MT, Allred EN, Tager-Flusberg H, Goldstein DJ, Leviton A. Positive screening on the Modified Checklist for Autism in Toddlers (M-CHAT) in extremely low gestational age newborns. *Journal of Pediatrics* 2009; **154**: 535-40.
286. Guy A, Seaton SE, Boyle EM, et al. Infants born late/moderately preterm are at increased risk for a positive autism screen at 2 years of age. *Journal of pediatrics* 2015; **166**(2): 269-75.
287. Rubin DB. Multiple imputation for nonresponse in surveys. New York: Wiley; 1987.
288. Wood AM, White IR, Royston P. How should variable selection be performed with multiply imputed data? *Statistics in Medicine* 2008; **27**: 3227-46.
289. Hedeker D, Gibbons RD. Application of random-effects pattern-mixture models for missing data in longitudinal studies. *Psychological Methods* 1997; **2**(1): 64-78.
290. McGurn B, Deary IJ, Starr JM. Childhood cognitive ability and risk of late-onset Alzheimer and vascular dementia. *Neurology* 2008; **71**(14): 1051-56.
291. Deary IJ, Whiteman MC, Starr JM, Whalley LJ, Fox HC. The impact of childhood intelligence on later life: following up the Scottish mental surveys of 1932 and 1947. *Journal of Personality and Social Psychology* 2004; **86**(1): 130-47.
292. Calvin CM, Batty GD, Der G, et al. Childhood intelligence in relation to major causes of death in 68 year follow-up: prospective population study. *British Medical Journal* 2017; **357**.
293. Schneider W, Niklas F, Schmiedeler S. Interlectual development from early childhood to early adulthood: the impact of early IQ differences on stability and change over time. *Learning and Individual Differences* 2014; **32**(0): 156-62.
294. Hoekstra RA, Bartels M, Boomsma DI. Longitudinal genetic study of verbal and nonverbal IQ from early childhood to young adulthood. *Learning and Individual Differences* 2007; **17**(2): 97-114.
295. Eryigit Madzwamuse S, Baumann N, Jaekel J, Bartmann P, Wolke D. Neuro-cognitive performance of very preterm or very low birth weight adults at 26 years. *Journal of Child Psychology & Psychiatry* 2015; **56**(8): 857-64.
296. Lohaugen GC, Gramstad A, Evensen KA, et al. Cognitive profile in young adults born preterm at very low birthweight. *Developmental Medicine & Child Neurology* 2010; **52**(12): 1133-8.
297. Hack M, Flannery DJ, Schluchter M, Cartar L, Borawski E, Klein N. Outcomes in young adulthood for very-low-birth-weight infants. *NEJM* 2002; **346**(3): 149-57.
298. Breeman LD, Jaekel J, Baumann N, Bartmann P, Wolke D. Preterm cognitive function into adulthood. *Pediatrics* 2015; **136**(3): 415-23.
299. Mangin KS, Horwood LJ, Woodward LJ. Cognitive development trajectories of very preterm and typically developing children. *Child development* 2017; **88**(1): 282-98.
300. Botting N, Powls A, Cooke RWI, Marlow N. Cognitive and educational outcome of very-low-birthweight children in early adolescence. *Developmental Medicine and Child Neurology* 1998; **40**(10): 652-60.
301. Koller H, Lawson K, Rose SA, Wallace I, McCarton C. Patterns of cognitive development in very low birth weight children during the first six years of life. *Pediatrics* 1997; **99**: 383-9.
302. Isaacs EB, Edmonds CJ, Chong WK, Lucas A, Morley R, Gadian DG. Brain morphometry and IQ measurements in preterm children. *Brain: a journal of neurology* 2004; **127**(12): 2595-607.

303. O'Brien F, Roth S, Stewart A, Rifkin L, Rushe T, Wyatt J. The neurodevelopmental progress of infants less than 33 weeks into adolescence. *Archives of Disease in Childhood* 2004; **89**(3): 207-11.
304. Stalnacke J, Lundequist A, Bohm B, Forssberg H, Smedler A. Individual cognitive patterns and developmental trajectories after preterm birth. *Child neuropsychology* 2015; **21**: 648-67.
305. Saigal S, Hoult LA, Streiner DL, Stoskopf BL, Rosenbaum PL. School difficulties at adolescence in a regional cohort of children who were extremely low birth weight. *Pediatrics* 2000; **105**(2): 325-31.
306. Luu TM, Vohr BR, Allan W, Schneider KC, Ment LR. Evidence for catch-up in cognition and receptive vocabulary among adolsecents born very preterm. *Pediatrics* 2011; **128**: 313-22.
307. Luu TM, Vohr BR, Sneider KC, et al. Trajectories of receptive language development from 3 to 12 years of age for very preterm children. *Pediatrics* 2009; **124**: 333-41.
308. Wood NS, Marlow N, Costeloe K, Gibson AT, Wilkinson AR. Neurologic and developmental disability after extremely preterm birth. *NEJM* 2000; **343**(6): 378-84.
309. Linsell L, Malouf R, Morris J, Kurinczuk JJ, Marlow N. Prognostic Factors for Poor Cognitive Development in Children Born Very Preterm or With Very Low Birth Weight: A Systematic Review. *Jama, Pediatr* 2015; **169**(12): 1162-72.
310. Robinson BF, Mervis CB. Extrapolated raw scores for the second edition of the Bayley Scales of Infant Development. *American Journal on Mental Retardation* 1996; **100**(6): 666-70.
311. Wechsler D. Wechsler Abbreviated Scale of Intelligence - Second Edition (WASI-II) manual. San Antonio, TX: NCS Pearson; 2011.
312. Wechsler D. Wechsler Adult Intelligence Scale – Fourth Edition (WAIS-IV). San Antonio, TX: NCS Pearson; 2008.
313. Flynn J. Searching for justice: the discovery of IQ gains over time. *American Psychologist* 1999; **54**(1): 5-20.
314. Wolke D, Ratschinski G, Ohrt B, Riegel K. The cognitve outcome of very preterm infants may be poorer than often reported: an empirical investigation of how methodological issues make a big difference. *European Journal of Pediatrics* 1994; **153**(12): 906-15.
315. Pyhala R, Lahti J, Heinonen K, et al. Neurocognitive abilities in young adults with very low birth weight. *Neurology* 2011; **77**(23): 2052-60.
316. Wolke D, Strauss VY-C, Johnson S, Gilmore C, Marlow N, Jaekel J. Universal Gestational Age Effects on Cognitive and Basic Mathematic Processing: 2 Cohorts in 2 Countries. *The Journal of Pediatrics* 2015; **166**(6): 1410-6.
317. Wong HS, Edwards P. Nature or nurture: a systematic review of the effect of socio-economic status on the developmental and cognitive outcomes of children born preterm. *Maternal child health journal* 2013; **17**: 1689-700.
318. Meng C, Bäuml JG, Daamen M, et al. Extensive and interrelated subcortical white and gray matter alterations in preterm-born adults. *Brain Structure and Function* 2015: 1-13.
319. Bornstein MH, Hahn CS, Wolke D. Systems and cascades in cognitive development and academic achievement. *Child development* 2013; **84**(1): 154-62.
320. Schools and pupils in England, January 2007 (final). London, United Kingdom: Department for Children Schools and Families, 2007.
321. Allin M, Henderson M, Suckling J, et al. Effects of very low birthweight on brain structure in adulthood. *Developmental Medicine & Child Neurology* 2004; **46**(1): 46-53.

322. Bäuml JG, Daamen M, Meng C, et al. Correspondence Between Aberrant Intrinsic Network Connectivity and Gray-Matter Volume in the Ventral Brain of Preterm Born Adults. *Cerebral Cortex* 2015; **25**(11): 4135-45.
323. Breeman LD, van der Pal S, Verrips GHW, Baumann N, Bartmann P, Wolke D. Neonatal treatment philosophy in Dutch and German NICUs: health-related quality of life in adulthood of VP/VLBW infants. *Quality of Life Research* 2016: 1-9.
324. Taylor HG, Margevicius S, Schluchter M, Andreias L, Hack M. Persisting behavior problems in extremely low birth weight adolscents. *Journal of Developmental and Behavioral Pediatrics* 2015; **36**(3): 178-87.
325. Hutchinson EA, De Luca CR, Doyle Lw, Roberts G, Anderson PJ. School-age outcomes of extremely preterm or extremely low birth weight children. *Pediatrics* 2013; **131**: e1053-e61.
326. Ritchie K, Bora S, Woodward LJ. Social development of children born very preterm: a systematic review. *Developmental Medicine and Child Neurology* 2015; **57**: 899-918.
327. Johnson S. Understanding social development following very preterm birth. *Developmental Medicine and Child Neurology* 2015; **57**: 890.
328. Elgen IB, Holsten F, Odberg MD. Psychiatric disorders in low birthweight young adults. Prevalence and association with assessments at 11 years. *European Psychiatry* 2013; **28**: 393-6.
329. Stevenson CJ, Blackburn P, Pharoah POD. Longitudinal study of behaviour disorders in low birthweight infants. *Arch Dis Child Fetal Neonatal Ed* 1999; **81**: F5-F9.
330. Breeman LD, Jaekel J, Baumann N, Bartmann P, Wolke D. Attention problems in very preterm children from childhood to adulthood: the Bavarian Longitudinal Study. *Journal of Child Psychology & Psychiatry* 2016; **57**(2): 132-40.
331. Samuelsson M, Holsti A, Adamson M, Serenius F, Hagglof B, Farooqi A. Behavioral patterns in adolescents born at 23 to 25 weeks of gestation. *Pediatrics* 2017; **140**(1).
332. Goodman A, Goodman R. Strengths and difficulties questionnaire as a dimensional measure of child mental health. *Journal of the American Academy of Child & Adolescent Psychiatry* 2009; **48**(4): 400-3.
333. Goodman A, Goodman R. Population mean scores predict child mental disorder rates: validating SDQ prevalence estimators in Britain. *Journal of Child Psychology and Psychiatry* 2011; **52**(1): 100-8.
334. Goodman R. Psychometric properties of the strengths and difficulties questionnaire. *Journal of the American Academy of Child & Adolescent Psychiatry* 2001; **40**: 1337-45.
335. Stone LL, Otten R, Engels RC, Vermulst AA, Janssens JM. Psychometric properties of the parent and teacher versions of the strengths and difficulties questionnaire for 4- to 12-year-olds: a review. . *Clinical Child and Family Psychology Review* 2010; **13**: 254-74.
336. Johnson S, Hollis C, Marlow N, Simms V, Wolke D. Screening for childhood mental health disorders using the Strengths and Difficulties Questionnaire (SDQ): The validity of multi-informant reports. *Developmental Medicine and Child Neurology* 2014; **56**(5): 453-9.
337. Goodman R. SDQ: Information for Researchers and Professionals about the Strengths & Difficulties Questionnaires. 1997. http://www.sdqinfo.com/.
338. Spittle A, Orton J, Anderson PJ, Doyle LW. Early developmental intervention programmes provided post hospital discharge to prevent motor and cognitive impairment in preterm infants. *Cochrane Database of Systematic Reviews* 2015; **11**: CD005495.
339. van Wassenaer-Leemhuis AG, Jeukens-Visser M, van Hus JWP, et al. Rethinking preventive post-discharge invention programmes for very preterm infants and their parents. *Developmental Medicine and Child Neurology* 2016; **58**(Supp. 4): 67-73.

Appendices

Appendix 1: Published protocol for systematic review of risk factors for neurodevelopmental outcome in children born very preterm

UNIVERSITY of York
Centre for Reviews and Dissemination

NHS
National Institute for Health Research

PROSPERO International prospective register of systematic reviews

Review title and timescale

1 Review title
Give the working title of the review. This must be in English. Ideally it should state succinctly the interventions or exposures being reviewed and the associated health or social problem being addressed in the review.
What factors predict the severity of neurodevelopmental outcome in infants born very preterm

2 Original language title
For reviews in languages other than English, this field should be used to enter the title in the language of the review. This will be displayed together with the English language title.

3 Anticipated or actual start date
Give the date when the systematic review commenced, or is expected to commence.
01/01/2013

4 Anticipated completion date
Give the date by which the review is expected to be completed.
31/12/2014

5 Stage of review at time of this submission
Indicate the stage of progress of the review by ticking the relevant boxes. Reviews that have progressed beyond the point of completing data extraction at the time of initial registration are not eligible for inclusion in PROSPERO. This field should be updated when any amendments are made to a published record.

The review has not yet started ×

Review stage	Started	Completed
Preliminary searches	Yes	Yes
Piloting of the study selection process	Yes	Yes
Formal screening of search results against eligibility criteria	Yes	Yes
Data extraction	No	No
Risk of bias (quality) assessment	No	No
Data analysis	No	No

Provide any other relevant information about the stage of the review here.

Review team details

6 Named contact
The named contact acts as the guarantor for the accuracy of the information presented in the register record.
Louise Linsell

7 Named contact email
Enter the electronic mail address of the named contact.
louise.linsell@npeu.ox.ac.uk

8 Named contact address
Enter the full postal address for the named contact.
National Perinatal Epidemiology Unit (NPEU), Richard Doll Building, Old Road Campus, Headington, Oxford, OX3 7LF

9 Named contact phone number
Enter the telephone number for the named contact, including international dialing code.
01865 617922

10 Organisational affiliation of the review
Full title of the organisational affiliations for this review, and website address if available. This field may be completed as 'None' if the review is not affiliated to any organisation.

National Perinatal Epidemiology Unit (NPEU)

Website address:
https://www.npeu.ox.ac.uk/

11 Review team members and their organisational affiliations
Give the title, first name and last name of all members of the team working directly on the review. Give the organisational affiliations of each member of the review team.

Title	First name	Last name	Affiliation
Ms	Louise	Linsell	National Perinatal Epidemiology Unit (NPEU)
Professor	Jenny	Kurinczuk	National Perinatal Epidemiology Unit (NPEU)
Professor	Neil	Marlow	Institute of Women's Health, University College London
Professor	Joan	Morris	Queen Mary University of London
Dr	Reem	Malouf	National Perinatal Epidemiology Unit (NPEU)

12 Funding sources/sponsors
Give details of the individuals, organizations, groups or other legal entities who take responsibility for initiating, managing, sponsoring and/or financing the review. Any unique identification numbers assigned to the review by the individuals or bodies listed should be included.
National Institute of Health Research (NIHR)

13 Conflicts of interest
List any conditions that could lead to actual or perceived undue influence on judgements concerning the main topic investigated in the review.
Are there any actual or potential conflicts of interest?
None known

14 Collaborators
Give the name, affiliation and role of any individuals or organisations who are working on the review but who are not listed as review team members.

Title	First name	Last name	Organisation details

Review methods

15 Review question(s)
State the question(s) to be addressed / review objectives. Please complete a separate box for each question.
What factors predict the severity of neurodevelopmental outcome in infants born very preterm?

16 Searches
Give details of the sources to be searched, and any restrictions (e.g. language or publication period). The full search strategy is not required, but may be supplied as a link or attachment.
MEDLINE, EMBASE and PSYCHINFO, all languages and publication date between 1st January 1990 and 26th March 2013.

17 URL to search strategy
If you have one, give the link to your search strategy here. Alternatively you can e-mail this to PROSPERO and we will store and link to it.
http://www.crd.york.ac.uk/PROSPEROFILES/6943_STRATEGY_20140020.pdf

I give permission for this file to be made publicly available
No

18 Condition or domain being studied
Give a short description of the disease, condition or healthcare domain being studied. This could include health and wellbeing outcomes.
Surviving infants born very preterm are at high risk of long term developmental problems, including cerebral palsy,

visual and auditory deficits, impairments in global and executive cognitive function, learning disabilities and behavioural problems. This risk is inversely related to birth weight and gestational age.

19 Participants/population
Give summary criteria for the participants or populations being studied by the review. The preferred format includes details of both inclusion and exclusion criteria.
Study population were born before 33 weeks gestation and/or had birth weight below 1250g.

20 Intervention(s), exposure(s)
Give full and clear descriptions of the nature of the interventions or the exposures to be reviewed
Multivariable risk prediction of at least one neurodevelopmental outcome assessed after the age of 18 months.

21 Comparator(s)/control
Where relevant, give details of the alternatives against which the main subject/topic of the review will be compared (e.g. another intervention or a non-exposed control group).
Not applicable.

22 Types of study to be included initially
Give details of the study designs to be included in the review. If there are no restrictions on the types of study design eligible for inclusion, this should be stated.
No restriction.

23 Context
Give summary details of the setting and other relevant characteristics which help define the inclusion or exclusion criteria.

24 Primary outcome(s)
Give the most important outcomes.
To identify risk factors that predict the severity of neurodevelopmental outcome and to assess the strength of evidence that supports this association.

Give information on timing and effect measures, as appropriate.

25 Secondary outcomes
List any additional outcomes that will be addressed. If there are no secondary outcomes enter None.
None.

Give information on timing and effect measures, as appropriate.

26 Data extraction, (selection and coding)
Give the procedure for selecting studies for the review and extracting data, including the number of researchers involved and how discrepancies will be resolved. List the data to be extracted.
Articles identified by the search strategies will be screened on title and abstract by one reviewer for definite exclusions and duplicates. The full text will be retrieved for remaining articles and will be screened by the same reviewer for the following eligibility criteria: 1. Contains original data 2. Study population born after 1st January 1990 3. Study population born before 33 weeks gestation or birth weight below 1250g 4. Study population is not a highly select group, and results can be applied to the general population of very preterm infants 5. An objective of the paper is to develop a multivariable risk prediction model (more than 2 variables) for a neurodevelopmental outcome assessed after the age of 18 months. If the reviewer is unsure of the eligibility of an article, it will be screened independently by a second reviewer and if a decision cannot be reached it will be referred to the rest of the review team. For all articles eligible for inclusion, both reviewers will complete a full data extraction form and risk of bias assessment. The following data will be extracted from each included article: study design, participant setting, centre selection, study location, year of birth, gestational age, birth weight, age at assessment, selection criteria of study population, control group recruited, sample size, completeness of data at follow-up, details of outcomes assessed, number of candidate risk factors assessed, variable selection, treatment of continuous variables, adjustment for confounders, method of analysis, model assumptions checked, missing data analysis, presentation of multivariable model, details of risk factors included in final model, strength of association, statistical validation and clinical validation. The data extraction forms will be cross-validated for discrepancies, and referred to the review team if the two reviewers cannot reach a resolution.

27 Risk of bias (quality) assessment

State whether and how risk of bias will be assessed, how the quality of individual studies will be assessed, and whether and how this will influence the planned synthesis.

The quality of studies will be assessed according to a standardised set of criteria recommended for use in reviews of prognosis. These criteria were modified slightly using suggested criteria from other sources. These guidelines focus on six areas of potential bias pertinent to studies of prognosis: study participation, study attrition, prognostic factor measurement, outcome measurement, confounding measurement and account, and analysis. The Cochrane Handbook for Systematic Reviews of interventions recognises four sources of potential systematic bias in randomised and non-randomised comparisons of healthcare interventions, including: selection bias, performance bias, attrition bias and detection bias. The six areas that will be evaluated in this review encompass the usual four sources of bias assessed, but have been adapted for use in studies of risk factor analysis or prognosis. Studies will be classified as acceptable quality if they at least partly satisfy all six areas of potential bias.

28 Strategy for data synthesis

Give the planned general approach to be used, for example whether the data to be used will be aggregate or at the level of individual participants, and whether a quantitative or narrative (descriptive) synthesis is planned. Where appropriate a brief outline of analytic approach should be given.

A narrative synthesis is planned, reporting the number and type of risk factors studied for each outcome domain: motor function, cognitive function, hearing, vision and behavioural/psychiatric. These domains will be further stratified by age of assessment: 1.5-2 years, 3-5 years, 6-12 years and above 13 years. The number and quality of studies examining each risk factor will be reported along with the strength and direction of association found. The level of evidence for each risk factor will be graded as high (multiple studies of acceptable quality), moderate (1 acceptable quality study and multiple lower quality studies), or low (1 study only or multiple lower quality studies).

29

Appendix 2: Search strategies for MEDLINE, EMBASE and PsycINFO databases

2a. MEDLINE search

1. exp infant, premature, diseases/ or exp gestational age/ or exp infant, low birth weight/ or exp infant, premature/ or exp premature birth/
2. exp infant/ or exp child/ or adolescent/ or young adult/
3. 1 and 2
4. exp animals/ not human/
5. 3 not 4
6. exp psychomotor performance/ or brain damage, chronic/ or brain injury, chronic/ or cerebral palsy/ or nervous system/ or central nervous system/ or neural pathways/ or neurologic manifestations/ or dyskinesias/ or dystonia/ or gait disorders, neurologic/ or neurobehavioural manifestations/ or memory disorders/ or intellectual disability/ or psychomotor disorders/ or seizures/ or neuromuscular diseases/ or muscular disorders, atrophic/ or movement disorders/ or dystonic disorders/ or brain/ab, gd, ph [abnormalities, growth & development, Physiology]
7. (((motor or psycho?motor or neur* or mobil* or movement* or physical* or brain) adj3 (develop* or impair* or disabilit* or disorder? or defect* or disease? or dysfunction* or function* or retard* or sever* or abilit* or assessment? or skill? or performance? or limitation? or co?ordination* or problem? or difficult* or manifestation? or abnormalit* or competenc* or capacit*)) or neuro?development* or cerebral palsy or gait or dyspraxia).ti,ab.
8. exp auditory diseases, central/ or hearing disorders/ or hearing loss/ or deafness/ or hearing loss, sensorineural/ or refractive errors/ or vision disorders/ or blindness/ or retinopathy of prematurity/
9. (((auditory or hearing or vision or visual or sensorineural or sensory or refract*) adj3 (impair* or disabilit* or disorder? or defect* or dysfunction* or function* or abilit* or assessment? or limitation? or process* or problem? or difficult* or abnormalit* or capacit*)) or deaf* or blindness or "retinopathy of prematurity").ti,ab.
10. mental processes/ or cognition/ or executive function/ or attention/ or learning/ or memory/ or cognition disorders/ or mild cognitive impairment/ or language/ or language development/ or verbal behaviour/ or language/ or mental competency/ or intelligence/ or intelligence tests/ or aptitude tests/ or language tests/ or cognition/ or executive function/
11. (((mental* or cogniti* or verbal* or language or intelligen* or learning or memory or spatial*) adj3 (outcome? or develop* or impair* or disabilit* or disorder? or defect* or dysfunction* or function* or retard* or sever* or abilit* or aptitude* or quotient* or assessment? or status or skill* or performance? or proficien* or limitation? or process* or problem* or difficult* or manifestation* or abnormalit* or competenc*)) or phonolog* or recall* or "executive function" or "word recognition").ti,ab.
12. neurobehavioural manifestations/ or mental disorders diagnosed in childhood/ or exp "attention deficit and disruptive behaviour disorders"/ or behaviour/ or child behaviour disorders/ or exp child development disorders, pervasive/ or developmental disabilities/ or learning disorders/ or intellectual disability/ or motor skills disorders/ or anxiety disorders/ or depressive disorder/ or exp child behaviour/ or mental disorders/ or psychological tests/ or neuropsychological tests/
13. (((behavio?r* or neurobehavio?r* or neuropsycholog* or psycholog* or psychiatr* or emotion* or internali* or externali* or adaptive*) adj3 (outcome? or morbidit* or symptom? or diagnos* or develop* or disorder* or defect* or dysfunction* or function* or sequelae or sever* or assessment* or status or behavio?r* or skill* or problem* or difficult* or manifestation* or abnormalit* or health or condit*)) or attention deficit* or depress* or anxiety or autis*).ti,ab.
14. exp education, special/ or achievement/ or educational status/
15. (((education* or academic or read* or writ* or math? or mathematic?) adj3 (outcome? or abilit* or aptitude* or assess* or skill* or perform* or proficien* or problem? or difficult* or achieve* or attain* or competenc* or status or develop* or level? or impact* or progress* or capacit*)) or numeracy or literacy or "special educational needs" or "special needs" or "special education*").ti,ab.
16. health status/ or quality of life/ or "outcome assessment (health care)"/ or child health services/ or ("heath status" or "quality of life").ti,ab.
17. child development/ or adolescent development/ or disabled persons/ or disabled children/ or hearing impaired persons/ or mentally disabled persons/ or visually impaired persons/ or morbidity/ or severity of illness index/ or survivors/
18. (development* adj (impair* or disabilit* or disorder? or defect* or retard* or sever* or assessment? or quotient or co?ordination* or problem? or difficult* or manifestation? or abnormalit*)).ti,ab.
19. or/6-18
20. 5 and 19
21. limit 20 to yr="1990-2014"
22. (journal article or review or meta analysis or systematic reviews or technical report).pt. or (meta?analys* or "systematic review").ti,ab.
23. (addresses or autobiography or bibliography or biography or case reports or clinical conference or comment or congresses or dictionary or directory or editorial or festschrift or in vitro or interactive tutorial or interview or lectures or legal cases or legislation or letter or news or newspaper article or overall or patient education handout or periodical index or portraits or published erratum or "scientific integrity review" or video-audio media).pt.
24. (21 and 22) not 23

2b. EMBASE search

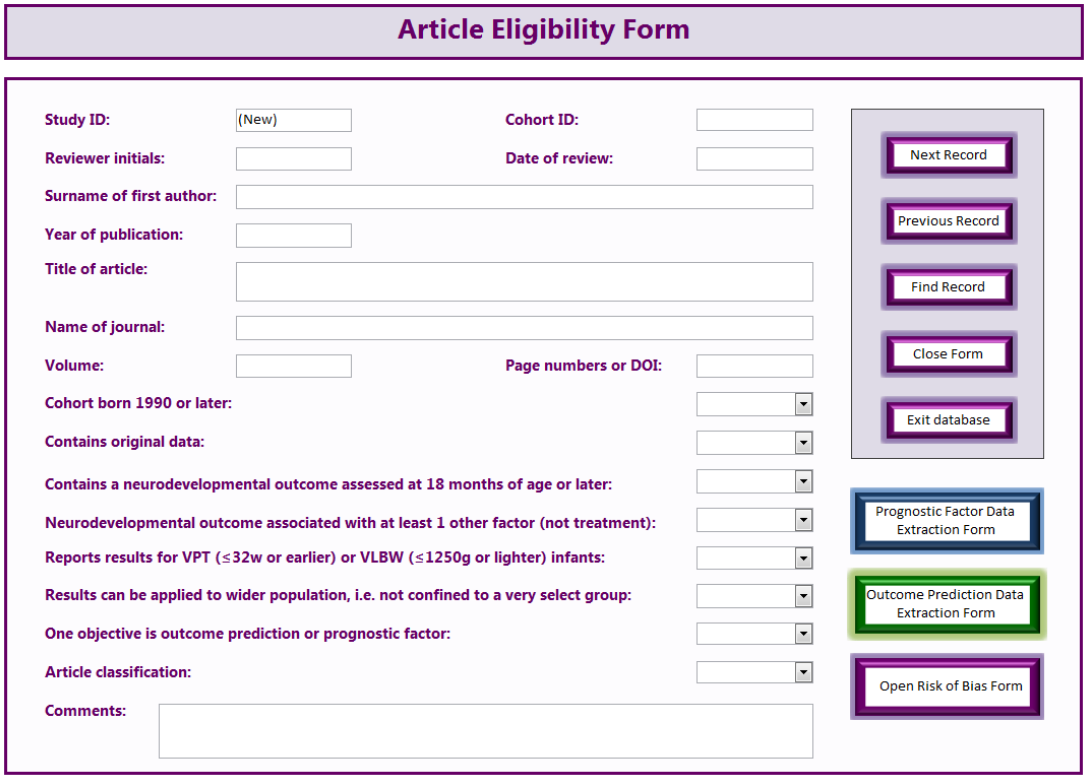
1. exp prematurity/ or exp immaturity/ or exp premature labor/ or exp low birth weight/ or exp gestational age/
2. exp child/ or exp infant/ or newborn/ or child development/ or adolescent development/ or adolescent/ or adult$.mp.
3. 1 and 2
4. exp animals/ not human/
5. 3 not 4
6. nervous system development/ or developmental coordination disorder/ or mental disease/ or neuromuscular function/ or chronic brain disease/ or cerebral palsy/ or functional disease/ or neurologic disease/ or central nervous system disease/ or physical disability/ or walking difficulty/ or motor dysfunction/ or gait disorder/ or motor retardation/ or motor performance/ or psychomotor retardation/ or psychomotor disorder/ or psychomotor development/
7. (((motor or psycho?motor or neur* or mobil* or movement* or physical* or brain) adj3 (develop* or impair* or disabilit* or disorder? or defect* or disease? or dysfunction* or function* or retard* or sever* or abilit* or assessment? or skill? or performance? or limitation? or co?ordination* or problem? or difficult* or manifestation? or abnormalit* or competenc* or capacit*)) or neuro?development* or cerebral palsy or gait or dyspraxia).ti,ab.
8. hearing impairment/ or hearing disorder/ or visual disorder/ or visual impairment/ or retinopathy of prematurity/
9. (((auditory or hearing or vision or visual or sensorineural or sensory or refract*) adj3 (impair* or disabilit* or disorder? or defect* or dysfunction* or function* or abilit* or assessment? or limitation? or process* or problem? or difficult* or abnormalit* or capacit*)) or deaf* or blindness or "retinopathy of prematurity").ti,ab.
10. language/ or language development/ or language disability/ or intelligence/ or intelligence quotient/ or intelligence test/ or cognitive defect/ or learning disorder/ or mental function/ or mental deficiency/ or intellectual impairment/ or cognition/ or attention/ or executive function/ or learning/ or memory/ or mental capacity/ or mental development/
11. (((mental* or cogniti* or verbal* or language or intelligen* or learning or memory or spatial*) adj3 (outcome? or develop* or impair* or disabilit* or disorder? or defect* or dysfunction* or function* or retard* or sever* or abilit* or aptitude* or quotient or assessment? or status or skill* or performance? or proficien* or limitation? or process* or problem* or difficult* or manifestation* or abnormalit* or competenc*)) or phonolog* or recall* or "executive function" or "word recognition").ti,ab.
12. psychiatric diagnosis/ or psychological aspect/ or psychologic test/ or psychosocial development/ or attention deficit disorder/ or child psychiatry/ or child behaviour/ or behaviour/ or anxiety disorder/ or autism/ or behaviour disorder/ or emotional disorder/ or psychosocial environment/ or hyperactivity/
13. (((behavio?r* or neurobehavio?r* or neuropsycholog* or psycholog* or psychiatr* or emotion* or internali* or externali* or adaptive*) adj3 (outcome? or morbidit* or symptom? or diagnos* or develop* or disorder* or defect* or dysfunction* or function* or sequelae or sever* or assessment* or status or behavio?r* or skill* or problem* or difficult* or manifestation* or abnormalit* or health or condit*)) or attention deficit* or depress* or anxiety or autis*).ti,ab.
14. special education/ or reading/ or educational status/ or ability/ or achievement/ or academic achievement/
15. (((education* or academic or read* or writ* or math? or mathematic?) adj3 (outcome? or abilit* or aptitude* or assess* or skill* or perform* or proficien* or problem? or difficult* or achieve* or attain* or competenc* or status or develop* or level? or impact* or progress* or capacit*)) or numeracy or literacy or "special educational needs" or "special needs" or "special education*").ti,ab.
16. quality of life/ or health status/ or ("heath status" or "quality of life").ti,ab.
17. child development/ or developmental disorder/ or development/ or handicapped child/ or disability/ or disability severity/ or morbidity/ or newborn morbidity/ or perinatal morbidity/ or survivor/
18. (development* adj (impair* or disabilit* or disorder? or defect* or retard* or sever* or assessment? or quotient or co?ordination* or problem? or difficult* or manifestation? or abnormalit*)).ti,ab.
19. or/6-18
20. 5 and 19
21. limit 20 to yr="1990-2014"
22. limit 21 to (meta analysis or "systematic review")
23. (article or report or review).pt. or (meta?analys* or "systematic review").ti,ab.
24. (book or book series or conference abstract or conference paper or conference proceeding or "conference review" or editorial or erratum or letter or note).pt.
25. (21 and 23) not 24
26. 22 or 25

2c. PsycINFO search

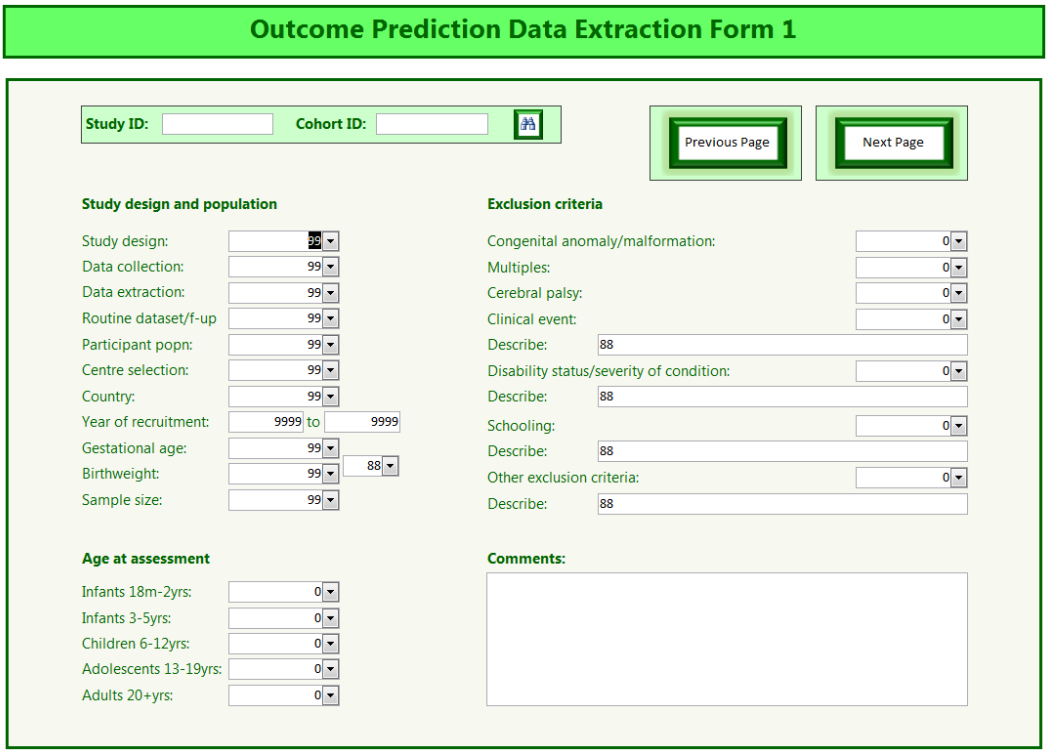
1. premature birth/ or low birth weight/ or gestation/ or gestational outcomes/ or low birth weight.mp. or prematur*.mp. or preterm*.mp. or gestation*.mp.
2. exp animals/ not human/
3. 1 not 2
4. limit 3 to (100 childhood or 120 neonatal or 140 infancy or 160 preschool age or 180 school age or 200 adolescence or 320 young adulthood)
5. motor development/ or nervous system disorders/ or central nervous system disorders/ or dyspraxia/ or movement disorders/ or motor performance/ or motor coordination/ or motor skills/ or physical mobility/ or physically handicapped/ or gait/ or cerebral palsy/ or neurodevelopmental disorders/ or neuromuscular disorders/ or brain development/ or physical disabilities/ or neural development/
6. (((motor or psycho?motor or neur* or mobil* or movement* or physical* or brain) adj3 (develop* or impair* or disabilit* or disorder? or defect* or disease? or dysfunction* or function* or retard* or sever* or abilit* or assessment? or skill? or performance? or limitation? or co?ordination* or problem? or difficult* or manifestation? or abnormalit* or competenc* or capacit*)) or neuro?development* or cerebral palsy or gait or dyspraxia).ti,ab,id.
7. sensory system disorders/ or hearing disorders/ or vision disorders/ or refraction errors/
8. (((auditory or hearing or vision or visual or sensorineural or sensory or refract*) adj3 (impair* or disabilit* or disorder? or defect* or dysfunction* or function* or abilit* or assessment? or limitation? or process* or problem? or difficult* or abnormalit* or capacit*)) or deaf* or blindness or "retinopathy of prematurity").ti,ab,id.
9. cognitive ability/ or cognitive impairment/ or cognition/ or intelligence/ or intelligence quotient/ or intellectual development disorder/ or attention/ or language/ or language delay/ or language development/ or language disorders/ or intellectual development disorder/ or memory/ or memory disorders/ or learning disabilities/ or learning disorders/ or executive function/ or spatial memory/ or "recall (learning)"/
10. (((mental* or cogniti* or verbal* or language or intelligen* or learning or memory or spatial*) adj3 (outcome? or develop* or impair* or disabilit* or disorder? or defect* or dysfunction* or function* or retard* or sever* or abilit* or aptitude* or quotient or assessment? or status or skill* or performance? or proficien* or limitation? or process* or problem* or difficult* or manifestation* or abnormalit* or competenc*)) or phonolog* or recall* or "executive function" or "word recognition").ti,ab,id.
11. emotional development/ or emotional disturbances/ or attention deficit disorder/ or attention deficit disorder with hyperactivity/ or autism/ or anxiety/ or anxiety disorders/ or behaviour disorders/ or behaviour problems/ or behaviour/ or depression/ or emotional development/ or psychosocial development/ or psychological development/ or externalizing problems/ or internalizing problems/ or "resilience (psychological)"/ or psychiatric symptoms/ or psychiatric disorders/ or child psychiatry/ or adaptive behaviour/
12. (((behavio?r* or neurobehavio?r* or neuropsycholog* or psycholog* or psychiatr* or emotion* or internali* or externali* or adaptive*) adj3 (outcome? or morbidit* or symptom? or diagnos* or develop* or disorder* or defect* or dysfunction* or function* or sequelae or sever* or assessment* or status or behavio?r* or skill* or problem* or difficult* or manifestation* or abnormalit* or health or condit*)) or attention deficit* or depress* or anxiety or autis*).ti,ab,id.
13. academic achievement/ or achievement/ or academic aptitude/ or performance/ or ability/ or special education/ or special needs/ or reading skills/ or reading ability/ or reading development/ or mathematical ability/ or verbal ability/ or verbal fluency/ or language proficiency/ or school readiness/
14. (((education* or academic or read* or writ* or math? or mathematic?) adj3 (outcome? or abilit* or aptitude* or assess* or skill* or perform* or proficien* or problem? or difficult* or achieve* or attain* or competenc* or status or develop* or level? or impact* or progress* or capacit*)) or numeracy or literacy or "special educational needs" or "special needs" or "special education*").ti,ab,id.
15. quality of life/ or ("heath status" or "quality of life").ti,ab,id.
16. development/ or neonatal development/ or infant development/ or early childhood development/ or childhood development/ or adolescent development/ or delayed development/ or developmental disabilities/ or morbidity/ or disorders/ or neonatal disorders/ or disabilities/ or survivors/
17. (development* adj (impair* or disabilit* or disorder? or defect* or retard* or sever* or assessment? or quotient or co?ordination* or problem? or difficult* or manifestation? or abnormalit*)).ti,ab.
18. or/5-17
19. 4 and 18
20. limit 19 to yr="1990-2014"
21. limit 20 to (journal article or reviews)
22. (abstract collection or bibliography or chapter or "column/opinion" or "comment/reply" or dissertation or editorial or encyclopedia entry or "erratum/correction" or letter or obituary or poetry or publication information or reprint or review-book or review-media or review-software & other).dt.
23. 21 not 22

Appendix 3: Data entry forms for data screening and data extraction of articles identified in the systematic review

3a. Article Eligibility Form completed for the full text screen of n=2,284 articles



3b. Study data extraction forms completed for the n=78 articles included in the review



Outcome Prediction Data Extraction Form 2

Study ID: **Cohort ID:** Previous Page Next Page

Size of sample

Total number of infants born/admitted in study period:	9999
Number of infants alive at discharge (or PMA):	9999
Number excluded due to exclusion criteria:	9999
Number of late deaths (after discharge):	9999
Number lost to follow-up (excluding late deaths):	9999

Age at assessment:

From: 99 yrs 99 mths **To:** 99 yrs 99 mths

Response rates

Number of infants/children assessed:	9999
Percentage assessed of those alive at assessment (or discharge if NK):	9999
Percentage assessed of those alive at assessment minus exclusions:	9999

Comments:

Outcome Prediction Data Extraction Form 3

Study ID: **Cohort ID:** Previous Page Enter models

Outcomes

Number of outcomes studied for risk: 99 Number of multivariate risk factor models presented: 99

Any composite outcomes: 0
Please tick the components:
- Motor: 0
- Cognitive: 0
- Behavior: 0
- Hearing: 0
- Vision: 0
- Death after discharge: 0

Any motor outcomes: 0
Please tick all that apply:
- General motor: 0
- Cerebral palsy: 0
- Dev coord dis: 0
- Other: 0
- Describe other: 88

Any cognitive outcomes: 0
Please tick all that apply:
- General cognition: 0
- IQ: 0
- Executive function: 0
- Language: 0
- Academic abilities: 0
- Schooling/SEN: 0
- Other: 0
- Describe other: 88

Any hearing outcomes: 0
Please tick all that apply:
- Deafness: 0
- Other: 0
- Describe other: 88

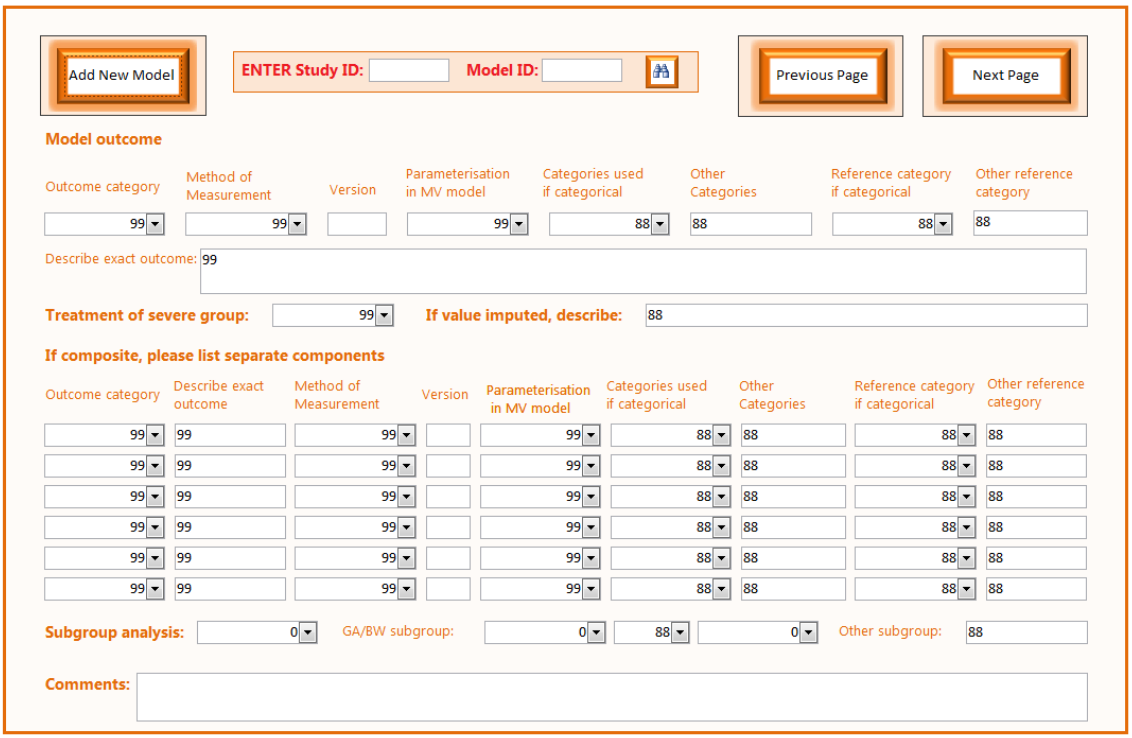
Any vision outcomes: 0
Please tick all that apply:
- Blindness: 0
- Other: 0
- Describe other: 88

Any behavioral outcomes: 0
Please tick all that apply:
- General: 0
- Psychiatric: 0
- ADHD: 0
- Autism: 0
- Other: 0
- Describe other: 88

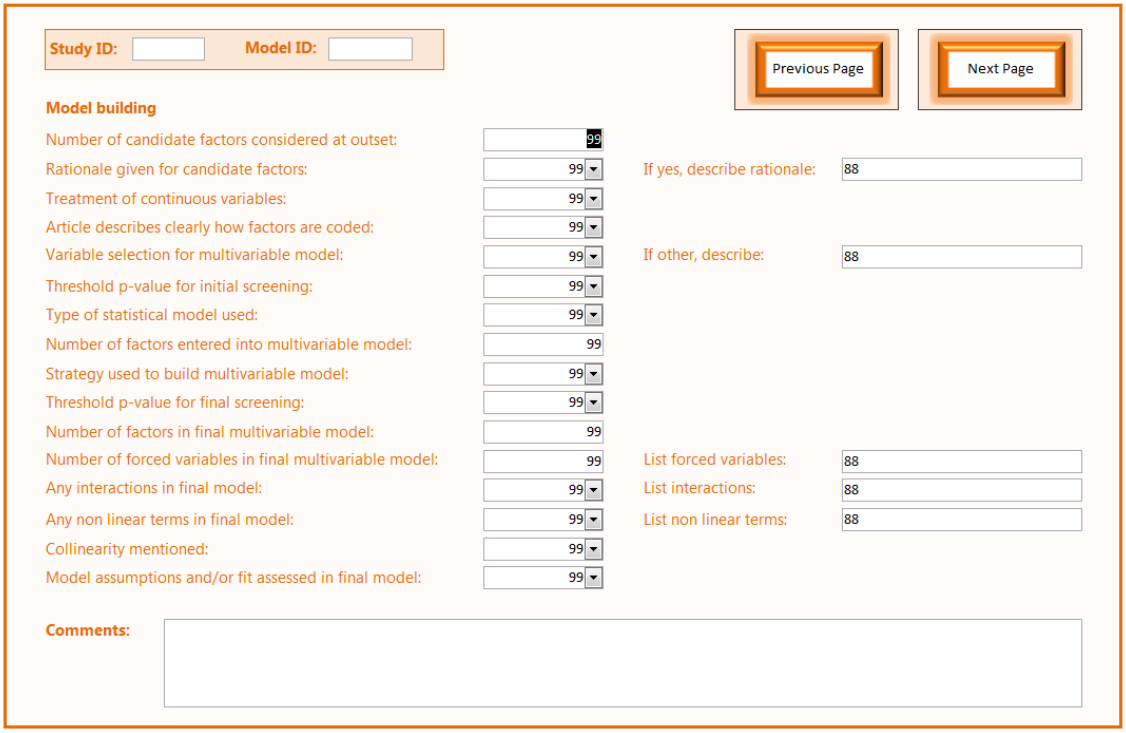
Comments:

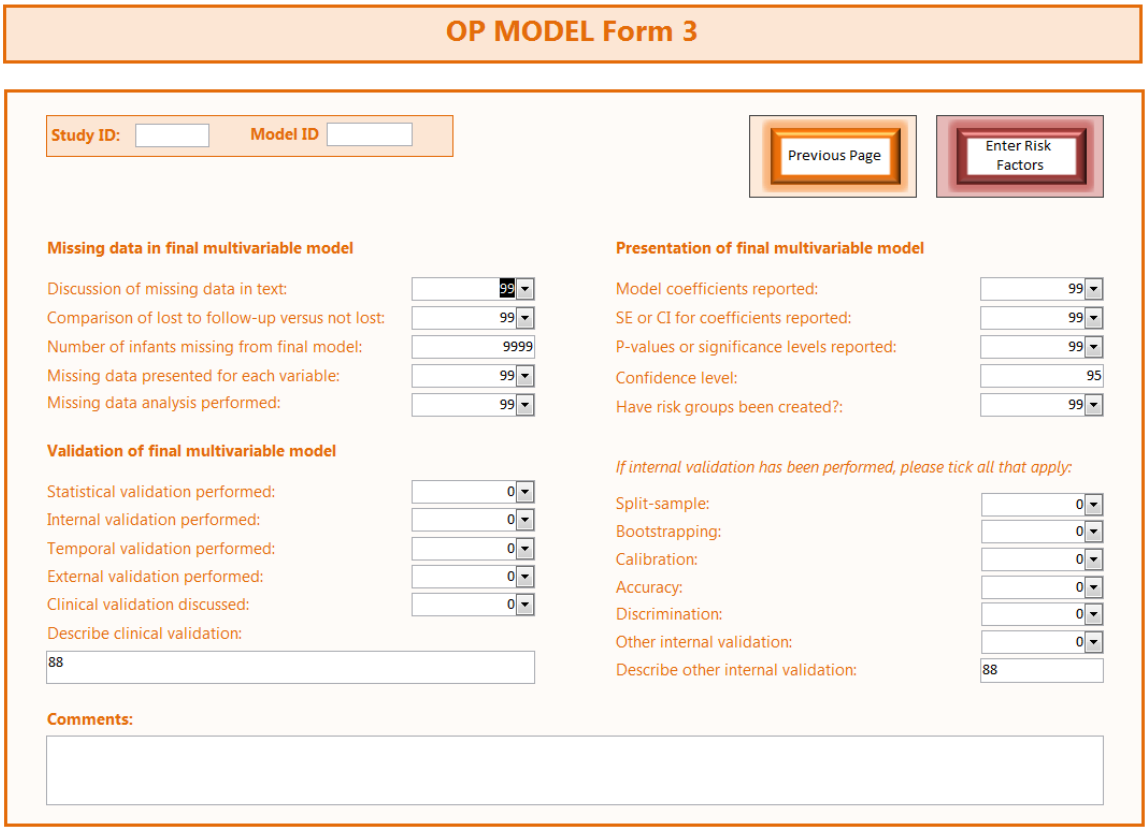
3c. Model data extraction forms completed for the n=78 articles included in the review

OP MODEL Form 1

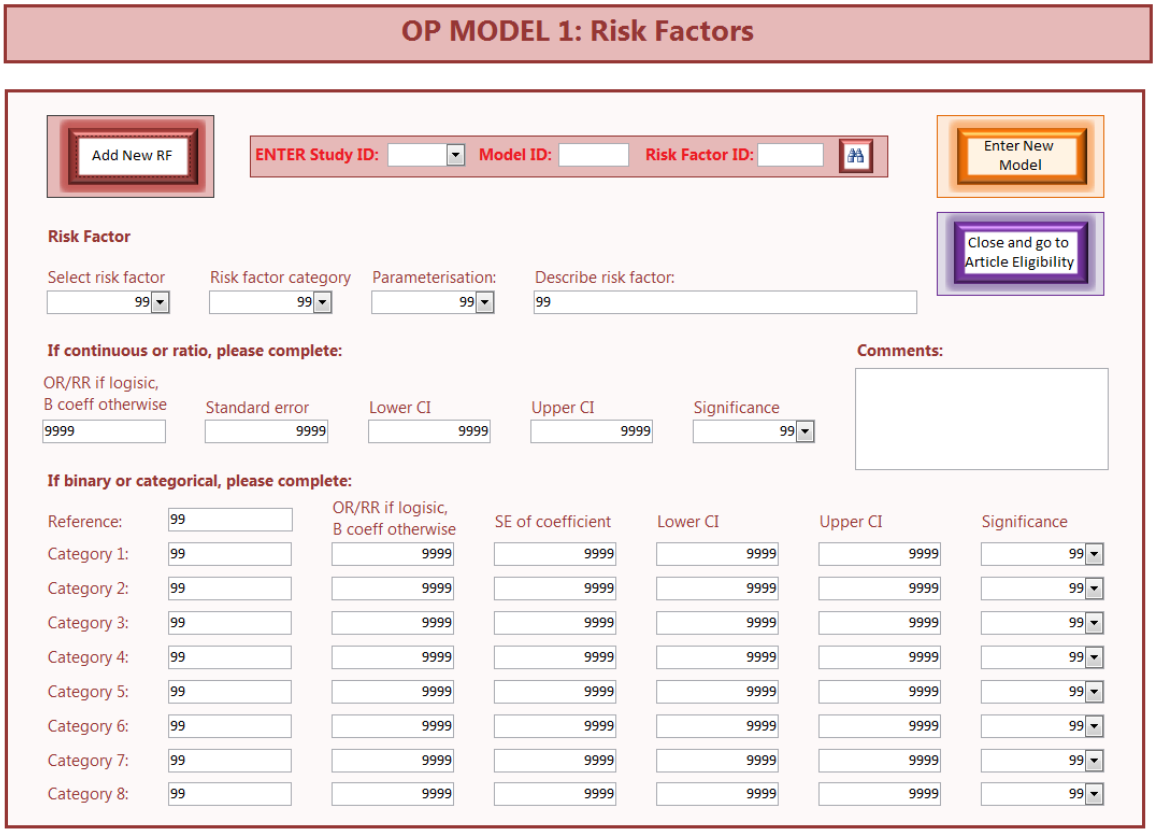


OP MODEL Form 2





3d. Risk factor data extraction form completed for the n=78 articles included in the review



Appendix 4: Risk of bias assessment guidelines based on the QUIPs tool

Risk of Bias Form (part I)

Potential bias (circle one)	**Items considered for assessment of potential opportunity for bias**	**Yes response**	**No response**
Study population The study sample represents the population of interest on key characteristics sufficient to limit potential bias to the results. **Yes** **Partly** **No** **Unsure**	▪ The source population or population of interest is adequately described for key characteristics and the study setting supports the applicability of results. ▪ Eligibility criteria and recruitment are adequately described and the inclusion/exclusion criteria applied uniformly to all screened for eligibility. ▪ There is adequate participation in the study by eligible participants, and was sufficient information given about eligibles who did not participate. ▪ The baseline characteristics of participants included in the study sample is adequately described for key characteristics and are representative of the population of interest. ▪ Cases and controls drawn from same population and participation rate similar in both groups.	Prospective cohort of all live births AND multicentre (3 or more) AND sample size alive at discharge >200 AND no major exclusions (unless study focus is infants free of major disability) AND sufficient information provided about flow of participants from recruitment to discharge.	Single centre NICU. OR sample size alive at discharge <50 OR major exclusions (unless focus of study is infants free of major disability).
Study attrition Loss to follow-up (from sample to study population) is not associated with key characteristics (i.e. the study data adequately represent the sample), sufficient to limit potential bias. **Yes** **Partly** **No** **Unsure**	▪ Study design was prospective. ▪ The completeness of follow-up was sufficiently high. ▪ Attempts to collect information on participants lost to follow-up are described. ▪ Reasons on lost to follow-up are provided. ▪ Participants lost to follow-up are adequately described for key characteristics. ▪ There are no important differences between key characteristics and outcomes in participants who completed follow-up and those who did not. ▪ If the risk of bias due to study attrition was moderate or high, measures taken to address this in the analysis, e.g. multiple imputation.	Number seen at assessment: >80% (up to 5 years) or >70% (over 5 years) AND reasons lost to follow-up reported with numbers AND comparison of lost versus not lost to follow-up with no important differences if response rate <90%, OR if important differences found addressed in the analysis.	Number seen at assessment: <65% (up to 5 years) or <50% (over 5 years). OR attrition/denominators not reported.
Prognostic factor measurement The prognostic factors of interest are adequately measured in study participants to sufficiently limit potential bias. **Yes** **Partly** **No** **Unsure**	▪ Clear definitions of the prognostics factors were provided and measurements described in sufficient detail to allow replication. ▪ Prognostic factors measured prior to outcomes occurring. ▪ Continuous variables are treated appropriately and rationale provided for cut-off values if analysed as categorical. ▪ Methods of measurement were accurate, valid, consistent and reliable, e.g. blinded assessment, validated scales used, not prone to recall bias. ▪ Adequate proportion of the study sample has complete data for prognostic factors. ▪ Appropriate methods were used to account for missing prognostic data in the analysis.	Data collection is prospective and risk factors recorded prior to outcome AND clear definition of risk factors given AND clear rationale for candidate risk factors, or a very wide coverage AND method of measurement a validated scale or strict diagnostic criteria AND continuous variables left as cont-inuous or rational provided for cut-offs AND number in final model with complete data on risk factors reported.	Definition of risk factors not clear OR use of non-validated scales OR diagnostic criteria not well-defined OR Inadequate proportion of those assessed included in final model.

Appendix 4: Risk of bias assessment guidelines based on the QUIPs tool (continued)

Risk of Bias Form (part II)

Potential bias (circle one)	**Items considered for assessment of potential opportunity for bias**	**Yes response**	**No response**
Outcome measurement The outcomes of interest are adequately measured in study participants to sufficiently limit potential bias **Yes** **Partly** **No** **Unsure**	▪ Clear definitions of the outcomes of interest were provided, including duration of follow-up. ▪ Methods of measurement were accurate, valid, consistent and reliable, e.g. blinded or objective assessment, validated scales used, strict diagnostic criteria. ▪ The method and setting of measurement was the same for all participants.	Evaluated prospectively AND comprehensive well-validated test with suitable norms/reference group or strict diagnostic/standard published criteria used AND performed by a small number of qualified study paediatricians/neurologists (or if assessed in local centres, all trained according to central protocol) AND blinded to previous medical history.	Mark down if: ▪ General or routine clinical exam with no protocol/strict diagnostic criteria with potential for misclassification. ▪ Short form or brief version of a more comprehensive test. ▪ Parent report or assessors not blinded. ▪ Reliability across assessors questionable. NO if 2 or more of the above.
Confounding measurement and account Important potential confounders are appropriately accounted for, limiting potential bias with respect to the prognostic factor of interest. **Yes** **Partly** **No** **Unsure**	▪ Important potential confounders are accounted for in the study design or in the analysis. ▪ All important confounders (including treatments) are measured. ▪ Clear definitions of the important confounders measured are provided. ▪ Methods of measurement were accurate, valid, consistent and reliable, e.g. blinded assessment, validated scales used, not prone to recall bias. ▪ The method and setting of measurement was the same for all participants. ▪ Appropriate methods were used to account for missing confounder data in the analysis.	List of candidate factors includes: 1) Gestational age or birth weight, 2) Gender, 3) Multiple pregnancy, 4) Socio-economic status or education. Mark down if: ▪ One or more of these factors is not considered (unless multiples are excluded or if the study cohort very restricted GA/BW population). ▪ Population is from an RCT and trial arm is not adjusted for.	None of these factors are considered as candidate factors, or if they are considered, are eliminated without statistical testing.
Statistical analysis and reporting The statistical analysis is appropriate for the design of the study, limiting potential for presentation of invalid results. **Yes** **Partly** **No** **Unsure**	▪ There is sufficient presentation of the data to assess the adequacy of the analysis. ▪ The strategy for model building was reported and acceptable. ▪ The analysis is appropriate for the design of the study. ▪ There is no selective reporting of results. ▪ Confidence intervals were provided for estimates of association.	Statistical model used appropriate for the study design and type of data AND a model building strategy used, e.g. stepwise, forward, backward selection AND strategy and results clearly reported AND completeness of reporting of results in final multivariable model with point estimates and measures of variance.	Mark down if: ▪ Statistical model not appropriate. ▪ No model building strategy, e.g. all candidates included without screening. ▪ Unclear reporting of strategy or results. ▪ Selective reporting of results. NO if 2 or more of the above.