A Thesis Submitted for the Degree of PhD at the University of Warwick

Permanent WRAP URL:

http://wrap.warwick.ac.uk/161735

warwick.ac.uk/lib-publications

Foetal Origins of Adult Neurocognitive Performance

by

Robert John Eves
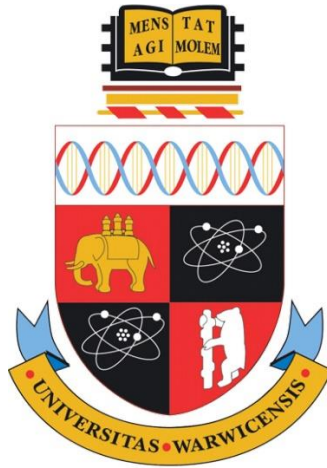
Thesis submitted for the degree of

Doctor of Philosophy in Psychology

Department of Psychology

University of Warwick

April 2021

Table of Contents

List of Tables

List of Figures

Abbreviations

ADD Attention-Deficit Disorder

ADHD Attention-Deficit/Hyperactivity Disorder

ADHD-H Attention-Deficit/Hyperactivity Disorder- Hyperactive/impulsivity

ADHD-I Attention-Deficit/Hyperactivity Disorder- Inattention

AGA Appropriate for Gestational Age

ANT Attention Network Task

APIC Adults Born Preterm International Collaboration

AYLS Arvo Ylppö Longitudinal Study

BLS Bavarian Longitudinal Study

BMI Body Mass Index

BPD Bronchopulmonary Dysplasia

CI Confidence Interval

DOHaD Developmental Origins of Health and Disease

DQ Developmental Quotient

DSM Diagnostic and Statistical Manual of Mental Disorders

EF Executive Functions

ELBW Extremely Low Birthweight

EP Extremely preterm

EU – European Union

HESVA Helsinki Study of Very Low Birth Weight Adults

IPD – Individual Participant Data

IQ Intelligence Quotient

ISCED International Standard Classification of Education

IUGR Intrauterine growth restriction

IVH Intraventricular Haemorrhage

K-ABC Kaufman Assessment Battery for Children

LBW Low Birthweight

MICE Multiple Imputation by Chained Equations

NICE National Institute for Health and Care excellence

NICU Neonatal Intensive Care Unit

NSI Neurosensory Impairment

NTNU Norges Teknisk-Naturvitenskaplige Universitet study

NZ VLBW New Zealand Very Low Birthweight study

PIRI Parent Infant Relationship Index

PVL Periventricular leukomalacia

RECAP Research on European Children and Adults Born Preterm

SD Standard Deviation

SES Socioeconomic Status

SGA Small for Gestational Age

SMD Standardised Mean Difference

TRAB-AS Tester Rating of Adult Behaviour – Attention Span

UCLH University College London Hospital study

UK United Kingdom

USA United States of America

VICS Victorian Infant Collaborative Study

VLBW Very Low Birthweight

VP Very preterm

WAIS Wechsler Adult Intelligence Scale

WASI Wechsler Abbreviated Scale of Intelligence

WHO World Health Organisation

Acknowledgements

There are many people I wish to thank for helping me complete this thesis. First and foremost, my primary supervisor Dieter Wolke. Dieter's passion for research is clearly limitless while his drive to produce only the highest quality science has been a continual source of inspiration. I feel I am a very different researcher than I was three years ago, for which Dieter has been a primary factor.

I have also had the pleasure of working with a number of other exceptional researchers as part of the RECAP preterm and APIC consortia, including my secondary supervisor Marina Mendonça. It has been rewarding from both a professional and personal level to work with world leading psychologists, neonatologists, epidemiologists and statisticians. It is an experience I will forever cherish.

I would also like to express my appreciation and gratitude to a number of friends and family members who have provided invaluable support and reassurance. In particular, I feel my mum's joy for learning and my dad's love of science strongly influenced me, something I am very grateful for. Last but by no means least, a humongous thank you to Sofia. No one has shared the highs and lows of this PhD with me as often as Sofia has.

Declaration/ Unique Contribution

I hereby confirm that I completed this thesis independently, that I have not heretofore presented this thesis to another department or university, and that I have listed all references used, and have given credit to all additional sources of assistance. The work presented was carried out by me except in the cases outlined below:

Data collection

The data presented in this thesis have been collected and made available by the BLS, EPICure, AYLS, HESVA, NTNU, VICS, NZ VLBW and UCLH cohorts.

Study 1 – Chapter 6: In print at JAMA Pediatrics

Contributions:

Marina Mendonça

Nicole Baumann

Yanyan Ni

Brian Darlow

John Horwood

Lianne Woodward

 Lex Doyle

Jeanie Cheong

Peter Anderson,

 Peter Bartmann

Neil Marlow

Samantha Johnson

Petteri Hovi

Eero Kajantie

Chiara Nosarti

Marit Indredavik

Kari-Anne Evensen

Katri Räikkönen

Kati Heinonen

Jennifer Zeitlin

Dieter Wolke

Study 2 – Chapter 7: Published in the Journal of Developmental and Behavioral Pediatrics

Contributions:
Marina Mendonça,

Adrian von Mühlenen,

Samantha Johnson

Helen O'Reilly

Peter Bartmann

Neil Marlow

Dieter Wolke

Reference:

Eves, R., von Mühlenen, A., Mendonça, M., Johnson, S., O'Reilly, H., Bartmann, P., Marlow, N., & Wolke, D. (2020). The Role of Executive and General Cognitive Functioning in the Attention Problems of Very and Extremely Preterm Adults. *Journal of Developmental and Behavioral Pediatrics: JDBP*, *41*(6), 461–469. https://doi.org/10.1097/DBP.0000000000000806

Study 3 – Chapter 8: Published in British Journal of Obstetrics and Gynaecology

Contributions:

Marina Mendonça

Peter Bartmann

Dieter Wolke

Reference:

Eves, R, Mendonça, M., Bartmann, P., & Wolke, D. (2020). Small for gestational age—cognitive performance from infancy to adulthood: An observational study. *BJOG: An International Journal of Obstetrics & Gynaecology*, *127*(13), 1598–1606. https://doi.org/10.1111/1471-0528.16341

One ancillary paper has also been published:

Spiegler, J., Eves, R., Mendonça, M., & Wolke, D. (2019). Association of physical activity and cardiorespiratory function or BMI and body composition in preterm-born individuals: A systematic review. Acta Paediatrica (Oslo, Norway: 1992), 108(7), 1205–1214. https://doi.org/10.1111/apa.14726

Funding

Abstract

Foetuses are affected by being born very preterm/very low birthweight (VP/VLBW) or being born at low birthweight for gestation (SGA). Both factors have been associated with lower IQ while VP/VLBW has been associated with executive functioning performance, such as working memory, inhibitory control, and attention problems. Several questions remain: 1. Are VP/VLBW's effects on adult IQ universal (found regardless of country or culture)? 2. Is it general cognitive functioning (IQ) or specific executive functions that explain associations of attention problems with VP/VLBW birth? 3. Are SGA's effects on IQ apparent in early childhood but grow out of them by adulthood?

Using individual level data from eight international cohorts, study 1 found that VP/VLBW adults had IQ scores 12 points lower than controls. Among VP/VLBW participants, the presence of intraventricular haemorrhage, lower birthweight for gestation, and lower maternal education were major risk factors for lower IQ. In study 2, attention differences between VP/VLBW adults and controls were investigated in two cohorts. Lower childhood IQ was consistently associated with adult attention problems. IQ explained more of the differences between VP/VLBW and controls in adulthood than any specific executive function. Study 3 investigated IQ development in the Bavarian Longitudinal Study, finding that the IQ of SGA individuals was consistently lower than those born at appropriate weight for gestation throughout the first 26 years of life. While SGA was associated with an 8 IQ point deficit, socioeconomic status and the quality of the parent infant relationship both had larger associations (14 and 10 points, respectively).

Overall, VP/VLBW and SGA birth are universally associated with lower adult IQ. Additionally, VP/VLBW's lower IQ is pervasive with further consequences for attention problems. Finally, low familial socioeconomic status has additional adverse effects on IQ and should be considered in future research and intervention for VP/VLBW or SGA children.

Structure of Thesis

Chapter 1 describes how two infant groups at risk, those born very preterm and those born small for gestational age, are at immediate risk of increased mortality and morbidity. It details how these two groups have similarities but also differ in both aetiology and outcomes. Chapter 1 also describes potential methodological issues with classifying these infants at risk. Finally how these measures of infant health may have long term implications for long term development is discussed.

Chapter 2 explores how cognitive performance is measured, from more general assessments deriving intelligence quotient (IQ) or developmental quotient (DQ) to measures of specific cognitive processes such as executive functioning. In addition, the theoretical relationship between these constructs and attention outcomes, is described.

Chapter 3 describes the current state of the literature on long term cognitive and behavioural outcomes for those born either very preterm or small for gestational age, how relative performance to controls may change across the lifespan and outlines the socioenvironmental factors that may also affect their cognitive outcomes.

Chapter 4 describes the research questions addressed in this thesis.

Chapter 5 describes the methodology, including the cohorts and research consortiums that agreed to contribute data in order to undertake the analyses in this thesis.

Chapter 6 to 8 are the three research studies completed for this thesis.

Chapter 9 summarises and discusses findings across the three studies. It integrates the combined results with critical implications for future research.

Chapter 1 **Introduction to the importance of early foetal factors**

Low birthweight (birthweight < 2500g) has been consistently linked to pathological foetal development and to long term health. Originating with the finding that regions in England and Wales with higher infant mortality were also found to have higher mortality rates due to adult cardiovascular disease a half century later (Barker et al., 1989). Subsequently the research into how early foetal factors influence long term outcomes area has expanded considerably. Further long term outcomes of interest have included psychopathology (Räikkönen & Pesonen, 2009), social outcomes (Kajantie et al., 2008; Mendonça et al., 2019), wealth (Bilgin et al., 2018; Strauss, 2000), behaviour (Breeman et al., 2016; Strang-Karlsson et al., 2008) and most importantly for this thesis, cognition (Pyhala et al., 2011).

While birthweight was the first and most commonly used proxy measure of pathological foetal development (Camerota & Bollen, 2016), evidence has indicated low birthweight can arise due to a number of factors and these underlying reasons may be responsible for variations in strength of association to later outcomes (Katz et al., 2013). Thus, research has looked to differentiate between the two main subgroups of infants born at low birthweight, those born immature and before term, i.e. preterm and those who are born at low birthweight relative to their gestational age, otherwise known as small for gestational age (SGA) (Hughes et al., 2017). In order to accurately determine links between early foetal health and later adult functioning, accurate determination and distinction of these risk groups is required. In the first chapter, the definitions, prevalence, and different causes for these risk groups are discussed as well as the challenges in accurately defining preterm and SGA birth.

1.1   Preterm Birth

While a healthy pregnancy lasts approximately 40 weeks, the World Health Organisation (WHO) defines preterm birth as any birth occurring before 37 weeks of gestation (see Table 1). Around 10.6% of children are born preterm worldwide (Chawanpaiboon et al., 2019). While the vast majority of preterm infants are born

between 32 and 36 weeks gestation (moderately and late preterm), approximately 11.3 % will be born very preterm (VP), from 28 weeks to 32 weeks gestation (Chawanpaiboon et al., 2019). Finally, there are those born extremely preterm (EP), at less than 28 weeks gestation, and make up approximately 4.1% of all preterm births.

In order to categorise these subgroups, it is necessary to accurately date the length of the pregnancy. There are alternative ways to date the length of a pregnancy, varying in accuracy and current use. The most accurate and most common method in high income countries today is to use ultrasounds of the foetal crown-rump during the first trimester to date the pregnancy (American College of Obstetricians and Gynecologists, 2017). However, this method is inaccessible in many low-income countries where ultrasounds scans are not readily available (Kim et al., 2018). A commonly used, inexpensive a1ternative to ultrasounds scans is using information on last menstrual period (Macaulay et al., 2019), where the time from the mother's first day of her last menstrual period until the day of birth is calculated (Deputy et al., 2017). This is thought to be somewhat inaccurate in determining gestational age due to uncertainty of the date of the last menstrual period or due to individual variation in menstrual cycle and has been found to systematically overstate gestational age in comparison to ultrasound (Macaulay et al., 2019; Savitz et al., 2002). Finally, there are neonatal estimates, such as the Dubowitz examination, where clinicians estimate gestational age based upon physical features of the infant (Dubowitz et al., 1970). It has been found that the Dubowitz method does not systematically underestimate or overestimate gestational age in comparison to ultrasound but does have a relatively large degree of inaccuracy, with over 40% of infants having gestational age estimates differing by more than 2 weeks (Lee et al., 2017). Overall, there are multiple ways to estimate gestational age, which when used in tandem reduces the degree of error (Blondel et al., 2002). Thus, preterm subgroups, and most importantly for this thesis very and extremely preterm subgroups, can be accurately defined. As preterm birth is the leading cause of infant mortality (Harrison & Goldenberg, 2016), the need to understand the causes and consequences of preterm birth remain pertinent.

Table 1: Definitions for commonly used categories for foetal groups

Foetal Categories	Definition

Gestational Age	length of time that a foetus grows inside the mother's uterus
Term	37 to 41 weeks+ 6 days
Post term	> 42 weeks
Preterm	< 37 weeks
Late preterm	34 to 36 weeks + 6 days
Moderate preterm	32 to 33 weeks+ 6 days
Very preterm	< 32 weeks or 28 to 31 weeks+ 6 days when further differentiating with extremely preterm birth
Extremely preterm	< 28 weeks
Birthweight	The body weight of an infant at its birth
Low birthweight	< 2500 g
Very Low birthweight	< 1500 g
Extremely low birthweight	< 1000 g
Birthweight for gestational age	The birthweight of an infant relative to their gestational, measured either as a percentile or Z score
Small for gestational age	Birthweight <10th Percentile for Gestational age
Appropriate for gestational age	Birthweight between 10th and 90th Percentile for Gestational age
Large for gestational age	Birthweight >90th Percentile for Gestational age
Intrauterine Growth Restricted (IUGR)	Growth during pregnancy being significantly lower than controls as measured by at least 2 ultrasound scans
Constitutionally Small	An SGA infants who displays no deviation from their individual growth potential as measured by at least 2 ultrasound scans

1.1.1 Causes and risk factors for preterm birth

Preterm birth occurs either spontaneously or due to labour being induced, with the causes and risk factors potentially differing upon this. For many cases of preterm birth, a precise mechanism or cause is difficult to establish. However, the two predominant causes are thought to be infection and inflammation (A. K. Boyle et al., 2017; Goldenberg et al., 2008). Risk factors that have been established for preterm birth are thought to result in either increased stimulation of an infection or increased inflammation (Goldenberg et al., 2008). Kramer (1987) suggested potential risk factors for preterm birth can be split into the following seven groups: genetic/constitutional, maternal demographics and psychosocial factors, obstetric factors, nutritional factors, maternal morbidities, substance intake factors or prenatal care. Subsequently much research has investigated more specific risk factors under this framework. In the case of spontaneous preterm labour, risk factors have been found to include ethnicity, adolescent pregnancy or advanced maternal age, parity (the number of pregnancies the mother has had that resulted in a surviving infant), gestational weight gain, low maternal body mass index during pregnancy, maternal smoking, and the timing and quality of antenatal care (Beeckman et al., 2013; Blencowe, Cousens, et al., 2013; Kramer, 1987; McKinnon et al., 2016). With regard to induced preterm labour, there is a large overlap of risk factors with spontaneous preterm labour (Prunet et al., 2017). However, a large risk factor is evidence of intrauterine growth restriction (IUGR), which without intervention might threaten the life of the infant (Ganzevoort et al., 2020; Goldenberg et al., 2008).

Overall, there are many potential risk factors for preterm birth, ranging from those that are impractical or impossible to change (constitutional factors), to those in which successful intervention is much more likely (prenatal care). Variations in these risk factors and prenatal care are likely to be related to the differing prevalence, survival rates and long-term outcomes of preterm individuals between regions and countries (Zeitlin et al., 2013) .

1.1.2   Prevalence and survival rates for preterm birth globally

Overall, 14.84 million infants are born preterm globally each year, of which 1.68 and 0.61 million will be born very and extremely preterm respectively (Blencowe, Cousens, et al., 2013; Chawanpaiboon et al., 2019). As a percentage, the rate of preterm births has increased globally from 9.8% in the year 2000 to 10.6% in 2014 (Chawanpaiboon et al., 2019). However, in Europe there is heterogeneity in trends with many countries showing little to no change in rates of preterm birth between 1996 and 2008 (Zeitlin et al., 2013). This is important to note as evidence suggests the incidence of preterm birth is higher in regions such as north Africa, at 13.4% of all live births, but lowest in Europe at 8.7% (Chawanpaiboon et al., 2019)

While there have been changes in the rates of preterm birth over time, survival rates have also changed dramatically (Saigal & Doyle, 2008). This has largely been a result of the improved medical care that started with mechanical ventilation in the 1980s followed by surfactant treatment, antenatal steroids and targeted oxygen therapy (Glass et al., 2015). Between 1983 and 1993 in the Netherlands, 28 day mortality decreased by 3.9% and 20.3% for VP and EP infants respectively(de Kleine et al., 2007). This trend has continued, with a 13% survival rate improvement being seen in EP infants in the UK between 1995 and 2006 (Costeloe et al., 2012) and a 7% improvement from 2004-2007 to 2014-2016 in Sweden (Norman et al., 2019). Another piece of evidence for increased survival is the changing threshold of viability in recent decades. The threshold of viability is commonly defined as the gestational age at which chance of survival is thought to be approximately 50%. In high income countries, the threshold of viability has improved from 25 or 26 weeks in the 1990's to 23 or 24 weeks by the 2000s (Glass et al., 2015). With increased survival, the focus has shifted to assess short and long-term morbidities associated with preterm birth. In particular, it has been posited that the result of decreased mortality in VP and EP infants may have resulted in an increase in infant morbidity (de Kleine et al., 2007). However, evidence from the USA has found rates of survival without major morbidity to be either stable or slightly decreasing between 1993 and 2012 for EP infants(Stoll et al., 2015).

1.1.3   Perinatal morbidities as a result of preterm birth

Those who are born preterm are born at a time when the organs are not fully developed, i.e immature. Two organs that are thought to be especially susceptible are the brain and the lungs (A. K. Boyle et al., 2017; Dammann et al., 2005; Saigal & Doyle, 2008). With regards to the brain, many significant developments occur between the second and third trimesters of pregnancy. This includes the proliferation of synapses, the maturation of cerebral pathways, changes to brain stem function, and brain gyrification (Peterson, 2003; White et al., 2010). When preterm birth occurs, this can disrupt these processes and result in a wide range of short and long-term alterations to the brain (Back & Miller, 2014). Neonatal outcomes such as grade 3 or 4 intraventricular haemorrhage (IVH), where bleeding occurs in the brain's ventricular system and extends to tissue, occurs in approximately 4% of VP infants (Bonamy et al., 2019). Similarly, 3.2% of VP infants suffer Periventricular leukomalacia (PVL), where cavities of white matter are seen surrounding the ventricles (Bonamy et al., 2019). In regard to lung problems, 26% of VP infants in Sweden were found to have died from or developed serious respiratory morbidity known as bronchopulmonary dysplasia (BPD) (Jensen et al., 2019).

While rates of PVL do appear to have dramatically fallen over the last 20 years (Norman et al., 2019), other key perinatal outcomes such as the rates of IVH and BPD have only decreased to a smaller extent (Norman et al., 2019) or remained stable (de Kleine et al., 2007). Overall there is strong evidence that preterm birth, and VP and EP birth especially, is a major public health problem for which understanding the long term outcomes are vital (Harrison & Goldenberg, 2016). However it would be overly simplistic to consider preterm birth as the sole factor that influences the health and later development of the infant. In the following section the other main cause of low birthweight, small for gestational age, will be discussed.

1.2 Small for Gestational Age

1.2.1 Differentiation of prematurity, low birthweight, and small for gestational age

Historically, low birthweight was used as a crude marker of preterm birth (Hughes et al., 2017). The strong relationship between gestational age and birthweight has been well established, for example in data from as early as the 1920s, it was found that 72% of LBW infants were also born preterm (Capper, 1928). This is in part why very preterm or very low birthweight infants are often merged into a single risk group of very preterm and/or very low birthweight (VP/VLBW). However, there is also an important differentiation between the two measure. In the above example, 28% of infants are born LBW but at term (37-41 weeks). Individuals who are born with relatively low birthweight for their gestational age are referred to as small for gestational age (SGA). SGA birth is thought to be a proxy indicator of Intrauterine Growth Restriction (IUGR), suggesting that growth within the womb has been disrupted (Schlaudecker et al., 2017). However, there is an important distinction between IUGR and SGA which will be further addressed in section 1.3.2.

While gestational age can be clearly measured in weeks and birthweight can be clearly measured in grams, there is far more scientific literature discussing how SGA should be best defined and calculated. The WHO expert committee defined SGA as a weight below the 10th percentile for gestational age as compared to a gender specific reference population (de Onis & Habicht, 1996). The specifics and difficulties of this definition will be further discussed in chapter 1.3.4. Overall, groups of LBW infants will consist of preterms who have grown adequately, term born infants who are SGA and infants who are both preterm and SGA. Understanding the shared and different causes and risk factors for these different groups is important, as it may allow for more specialised preventions (Katz et al., 2013). For example, an intervention that prevents preterm birth may not additionally benefit the other low birthweight group of SGA infants and vice versa (Katz et al., 2013). It is also important to differentiate these groups as both their short term and long-term outcomes, such as mortality and IQ, may also differ (Hutton et al., 1997; Katz et al., 2013).

1.2.2 **Overlap between preterm birth and SGA**

While preterm birth is thought to be the leading cause of perinatal mortality and morbidity, Intrauterine Growth Restriction (IUGR) is thought to be the second leading cause (Walker & Marlow, 2008). IUGR is a particularly important shared factor for both induced preterm birth and SGA (Goldenberg et al., 2008). This is because obstetricians may choose to induce labour due to evidence of IUGR , resulting in a preterm and often SGA birth. IUGR has a variety of different methods for diagnosis, however the ideal diagnosis method has been suggested as the use of serial ultrasounds (Albu et al., 2014; Nyberg et al., 2004). Serial ultrasounds can be used to identify foetuses whose growth in the womb differs to the growth of healthy foetuses (Løhaugen et al., 2013) or who show significant deviation in growth between an initial scan and a subsequent scan (Deter et al., 2018). While IUGR often leads to a smaller infant at birth it is different from SGA, a measurement solely made at birth (Sharma et al., 2016). For example, there are cases where a foetus shows no disruption of growth in the womb but is a small foetus throughout pregnancy (Sharma et al., 2016). This is often referred to as an infant being "constitutionally small" and results in an SGA birth without evidence of IUGR (Gardosi et al., 2018). It has been speculated that the percentage of SGA infants who are constitutionally small could be as high as 70% of all SGA infants (Alberry & Soothill, 2007). Conversely, there are also cases where an initially large developing foetus develops IUGR and does not grow to its fullest potential. In this case, despite the IUGR, the birthweight may be relatively normal for gestation and as such referred to as an appropriate weight for gestational age (AGA) infant (Sharma et al., 2016). Overall, due to the overlap between preterm birth and SGA, partially because of IUGR, unravelling the consequences of both factors is complicated. Many studies have failed to consider both factors concordantly in analyses, which means potentially important confounding is ignored. In this thesis, the importance of these two factors will be considered for long term neurocognitive outcomes which will allow for more accurate and specific risk factors to be determined.

1.2.3   Causes and risk factors for SGA

SGA and preterm birth overlap in regard to other risk factors, in addition to IUGR. Factors that have been associated with SGA birth at term include low maternal age, maternal smoking, primiparous birth (the first birth for the mother) and short maternal stature (McCowan & Horgan, 2009; Muhihi et al., 2016). However, maternal stature and primiparous birth were also found to be risk factor for AGA preterm birth, suggesting they are not unique risk factors for either subgroup (Muhihi et al., 2016). The risk of shorter maternal stature for term SGA, preterm AGA, and preterm SGA births has also been confirmed by individual participant data meta-analysis (Kozuki et al., 2015). In prior research, Lang, Lieberman and Cohen (1996) tested 23 separate risk factors for spontaneous preterm birth and SGA. 21 from the 23 were crudely associated with either preterm birth or SGA, with 16 factors associated with both. While this again suggested a large, shared overlap, once confounding was considered their results indicated the specific importance of genetic and constitutional factors that were solely significant for SGA birth. Thus, there are both shared and unique risk factors for both preterm birth and SGA. The importance of genetic and constitutional factors are somewhat controversial within the SGA literature (Iliodromiti et al., 2017), and as a result has led to divergence in how best to calculate SGA.

1.2.4   Differences in SGA classification and references

As previously noted, the WHO expert committee defined SGA as a weight below the 10th percentile for gestational age as compared to a gender specific reference population (de Onis & Habicht, 1996). However, many aspects of the WHO's definition has been debated. Firstly, rather than using 10% as the cut off, some researchers have instead used less than 2 Standard Deviations (approximately <2.3%). The stricter the cut-off used, the more pathological the SGA group should theoretically become (Charkaluk et al., 2012). While the stricter definition may increase the specificity associated with SGA for a certain outcome, it also potentially reduces the sensitivity with a number of SGA infants being misclassified as AGA. Secondly, the term "reference population" is not a simple matter. This debate originated from Goldenberg et al. (1989), where the wide range of reference populations being used

was shown. For example, some references excluded infants with congenital anomalies or those from diabetic mothers while other charts had no such exclusion criteria. Additionally, there is debate as to whether a global reference is ideal or whether references should be customised based upon nationality or ethnic group. Recently, both the WHO (Kiserud et al., 2017) and the INTERGROWTH Project (Villar et al., 2014) independently looked to determine whether if given the same recruitment criteria (healthy mothers, at the same altitude and living in areas with low levels of pollution) do infants grow to similar height and weight regardless of nationality or ethnicity. Unfortunately, the two studies differed in their conclusions, with INTERGROWTH suggesting a global reference is appropriate while WHO indicating a local reference is optimal. Thus the choice of reference, either global or local, with or without certain exclusion criteria, will likely have a subsequent impact on the number of infants classified as SGA.

For preterm infants, as well as the choice of a local or global reference, the decision to define SGA using a neonatal or a foetal reference is also important. Neonatal references use large databases of birthweights for infants born at different gestational ages, assuming an equal rate of SGA across the gestational age range. This appears to be flawed, due to the fact that preterm infants are also more likely to be growth restricted (Yanney & Marlow, 2004). In particular, infants with IUGR are usually given more obstetric intervention such as induced labour, resulting in a preterm birth (Yanney & Marlow, 2004). Thus by comparing the birthweight of preterms to other preterms, only the most extremely small infants will be diagnosed as SGA, with the more mildly small preterms being potentially erroneously classified as AGA. To address this, foetal references may be used instead. Foetal based references compare the birthweight of preterms to the estimated weight of healthy developing foetuses still in the womb at equivalent gestational ages. This increases the number of preterms classified as SGA, in some cases doubling the rate (Pritchard et al., 2019), see figure 1 below. However, this also results in measures of SGA and gestational age becoming more correlated with one another, an important fact to consider when considering both variables in an analysis of long-term outcomes.
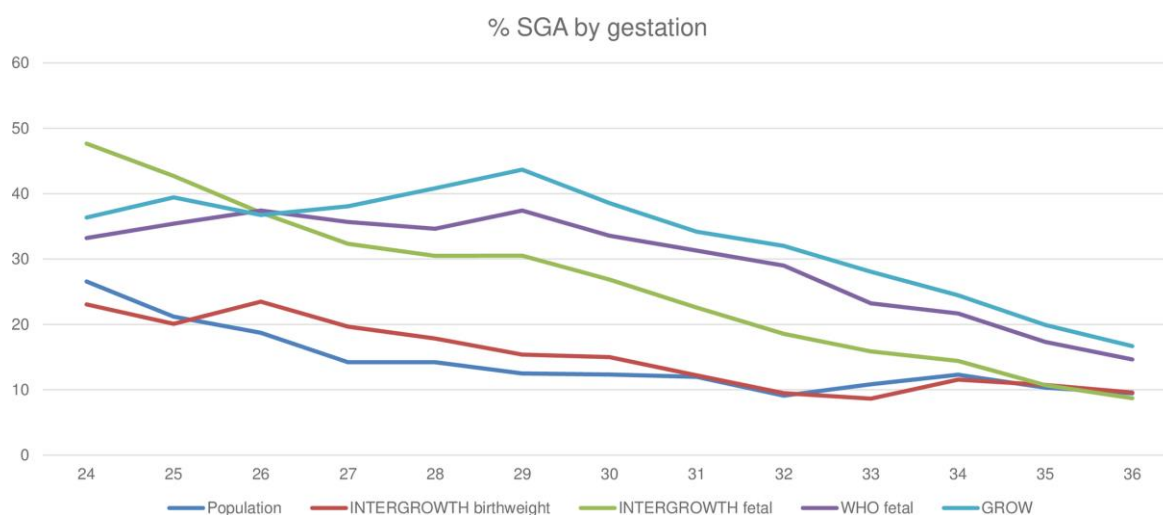
Figure 1: SGA by gestational age, adapted from Pritchard et al. (2019)

Finally, some researchers have suggested customisation of SGA references based on constitutional factors such as maternal height. This is based on the idea that some infants will be small, but not pathologically small, due to their mother's smaller stature. Evidence suggests that SGA infants who are "constitutionally small" do have better outcomes than SGA infants with IUGR (von Beckerath et al., 2013). Thus, if customised SGA references can reliably differentiate pathologically small from constitutionally small, it would improve prediction accuracy for a range of perinatal and long-term outcomes. Unfortunately, results for such outcomes are somewhat contrasting. In comparison to non-customised references, some research suggests customised references better identify those with metabolic disturbances (Verkauskiene et al., 2008) while others find no differences for predicting stillbirth, infant death, or neonatal morbidity (Ding et al., 2013; Iliodromiti et al., 2017). Additionally, customising based on maternal stature may be inappropriate if smaller maternal stature is a risk factor itself for infant mortality, as has been previously demonstrated (Monden & Smits, 2009)

Overall, there are currently many references for those investigating SGA in preterm infants: local or global, foetal or neonatal, customised or not. As well as this, the simpler decision of which cut-off to use (10% or 2 SD) has also differed between research groups. These complexities make determining the effect of SGA on both child and adult outcomes difficult, while further complicating comparisons across different cohorts that have utilised different methods. To conclude, it is important to determine

whether the choice of SGA reference influences the effect of SGA on both short- and long-term outcomes. This may explain potential differences in findings across cohorts, especially when vastly different references and thresholds are utilised.

1.2.5 Survival rates and perinatal morbidities as a result of SGA birth

In studies that have used SGA below the 10th percentile, SGA infants have been found to have significantly higher mortality rates than AGA infants in the neonatal period (Iliodromiti et al., 2017). For example a risk ratio of 1.83 was seen in low/middle income countries, a risk ratio of 2.50 in Scotland and an adjusted odds ratio of 3.18 in the USA (Iliodromiti et al., 2017; Katz et al., 2013; Mendez-Figueroa et al., 2017). However, in a similar fashion to preterm birth, SGA birth has seen a reduction in infant mortality over time presumed to be due to an improvement in medical care (Claas et al., 2010).

As well as increased mortality, neonatal morbidities are also found to be more common for SGA infants than AGA infants. For example, SGA birth has been linked to higher rates of BPD, seizures, lower Apgar scores (a summary measure of infant health including pulse rate) and higher chance of being admitted to a neonatal intensive care unit (Chauhan et al., 2017; Iliodromiti et al., 2017). This evidence suggests that SGA infants are at an increased risk during the neonatal period. However, difficulty arises in synthesizing research into perinatal morbidities and survival in the same way as preterm birth, due to the aforementioned wider variation in how SGA has been defined and calculated. This continues to cause difficulty when similarly investigating longer term outcomes for SGA individuals, such as major disabilities.

1.3 Major disability as a result of preterm birth or SGA

In the previous sections, the neonatal morbidities following preterm and SGA birth have been discussed. The consequences of these morbidities are thought to be higher risks of blindness, deafness and cerebral palsy (Bolisetty et al., 2014). To quantify the rates of long term major disability, VP infants were found to be 2.34 and 1.94 times more likely to suffer from visual and hearing impairments than those born at term

(Hirvonen et al., 2018). For cerebral palsy, while the prevalence is just 0.14% in term borns, it has been found for 4.3% of VP and 8.2% of EP individuals (Oskoui et al., 2013). With regard to changes over time, the prevalence of cerebral palsy in VP children has not declined (Spittle et al., 2018; Vincer et al., 2014). Thus, while mortality has decreased substantially in VP and EP infants as a result of improved medical care, a decrease in long term morbidity may not have followed.

For SGA infants there has been less research into major disability rates. Surprisingly, there was no significant differences in percentage of participants with a long term disability for term born SGA and AGA adults (Strauss, 2000). This differs from research on cerebral palsy in childhood where term born individuals with a weight percentile below the 1st percentile were found to be 6.6 times more likely to be diagnosed with cerebral palsy than AGA controls (Jacobsson et al., 2008). However, weight below either the 5th or 1st percentile did not predict cerebral palsy within preterm individuals, suggesting that the impact of SGA differs across the gestational age range. Within cohorts of EP infants, there is contrasting evidence for the effect of SGA on major disability. For example, it was found that in comparison to those born EP and AGA, EP infants who were SGA had higher rates of cerebral palsy and other major disabilities such as blindness and major cognitive impairment (De Jesus et al., 2013). However, this finding within EP groups has not been further validated for cerebral palsy (Guellec et al., 2011) or neurosensory impairment generally (Bickle Graz et al., 2015). For specific visual and hearing impairments, research is limited as these are generally rare events and as such are not commonly reported for SGA children (Vollmer & Edmonds, 2019). Overall, results indicate SGA status may be associated with major disability such as cerebral palsy but SGA's effects may differ depending on the individual's gestational age.

1.4   **Why does it matter?**

In this chapter, the complexity in classifying two groups of infants at high risk of neonatal mortality and morbidity has been discussed. Since survival for preterm and SGA groups have improved, the focus has moved to the quality of survival. Although long term major disability is high in VP and EP individuals especially (Blencowe, Lee,

et al., 2013; Hirvonen et al., 2018), there has been an increased interest in other areas of functioning such as cognition, attention and mental health, to name but a few (Wolke et al., 2019). In particular, to understand the long-term cognitive development for preterm and SGA infants, a critical review of general and specific cognitive functioning is required. Notably, understanding how specific cognitive factors and attention relate to general cognitive performance will be discussed in chapter 2. This is important subsequently for chapter 3, where literature on the specific and general cognitive outcomes for preterm or SGA individuals is reviewed.

Chapter 2 **Theoretical models of the relationship of general cognitive, specific cognitive and behavioural outcomes.**

This chapter will first discuss intelligence and general cognitive performance as a whole and how it is assessed. It will critically examine how performance on general cognitive tests requires skills in specific cognitive areas, such as components of executive functioning. Following this, it will review the potential relationship between specific cognitive skills, such as working memory, and behavioural dysfunction, such as attention problems, based on research from the most commonly used models in the literature.

2.1 Rationale and History of IQ testing

While intelligence is often discussed in both popular and scientific literature, a clear definition is somewhat challenging. Sternberg defined intelligence as 'the capacity to learn from experience, using metacognitive processes to enhance learning and the ability to adapt to the surrounding environment' (Sternberg, 2005). This modern definition is similar to past ideas as put forward by Alfred Binet, often credited with developing the first standardised intelligence test at the beginning of the 20th century. Binet considered intelligence to have three distinct components: direction, adaptation, and criticism (Binet & Simon, 1916). Binet used this framework to develop their original test, used in the French education system to differentiate "gifted" from "mentally retarded" children (Binet & Simon, 1916). It has set the foundations for today's most prominent theories for intelligence, including the Catell–Horn–Caroll theory of intelligence (Cattell, 1971; Flanagan & Dixon, 2014), and the most commonly used intelligence tests, the Wechsler intelligence scales (Boake, 2002; Wechsler, 2008). The Catell–Horn–Caroll theory of intelligence describes intelligence as having an overall 'G', indicating general intelligence that can be stratified into 2 subcomponents: fluid and crystallized intelligence. While fluid intelligence is thought to measure skills that are largely fixed, such as inductive reasoning ability, crystallized intelligence reflects acquired knowledge, such as vocabulary. In a similar fashion to Catell's G, the Wechsler scales produces a full-scale IQ score, scaled on a normative

sample which is representative of the population of a whole. This results in the average IQ score being 100 and a standard deviation of 15, with scores below 85 (-1SD) or 70 (-2SD) indicative of mild and major cognitive impairment, respectively. The full scale is based on the results from two sub-scores, performance IQ and verbal IQ, largely an operationalisation of Cattell's theory of fluid and crystallised intelligence. The performance and verbal IQ subscales themselves are based on performance on a range of subtests. For example, a common performance IQ subtest is the block design subtest, see figure 2. It tests an individual's ability in spatial reasoning as an individual must recreate a 2D pattern from a number of smaller, patterned 3D blocks (Lind et al., 2014). In contrast, verbal IQ assesses how individuals find similarities between words (e.g. Lion and Whale; where they are both mammals) or to ask participants to accurately define words as to test semantic knowledge. Overall, by having an individual perform multiple IQ subtests, it provides an overview of the individual's general cognitive functioning, allowing for accurate determination of who is performing above or below the expected level.

There are three important points to note regarding the utility of IQ tests. Firstly, the standardised norms quickly become outdated due to a phenomenon known as the Flynn effect (Flynn, 2007). This is the fact that IQ scores have demonstrated both a substantial and sustained increase over time, estimated at approximately 3 IQ points per decade (Trahan et al., 2014). While this effect is important for the entire population, it is especially critical for those scoring near to the 70 point cut-off for major cognitive impairment. For example, an individual who scores 72 on an IQ test with norms that are 10 years obsolete would result in no diagnosis of major cognitive impairment despite the fact their score is likely 2 SD below their peers (Flynn, 2007).

Secondly, the age at assessment is potentially an important factor. IQ tests are not appropriate before the age of 2 with developmental quotient (DQ) tasks needing to instead be used. However, there are potentially fundamental differences in the underlying constructs DQ and IQ measure, with DQ's ability to predict later IQ performance also questioned in the general population (Aylward, 2009). In addition, there is the question of how stable IQ is from childhood into adulthood (Schneider et al., 2014). The predictive ability of DQ and early IQ for later IQ will be discussed in more depth in the following section, as to determine the stability of general cognitive performance over time.

Finally, as IQ performance is a combination of a number of specific tests, there is the argument that some of these "building blocks" may be the fundamental primary deficit for a number of cognitive and behavioural outcomes (Diamond, 2013). Following the debate surrounding the predictive ability of DQ and early IQ assessments for later IQ, the subsequent section will provide an overview of the models and hypotheses surrounding specific cognitive dysfunction and behavioural problems and how they relate to performance on IQ tests.
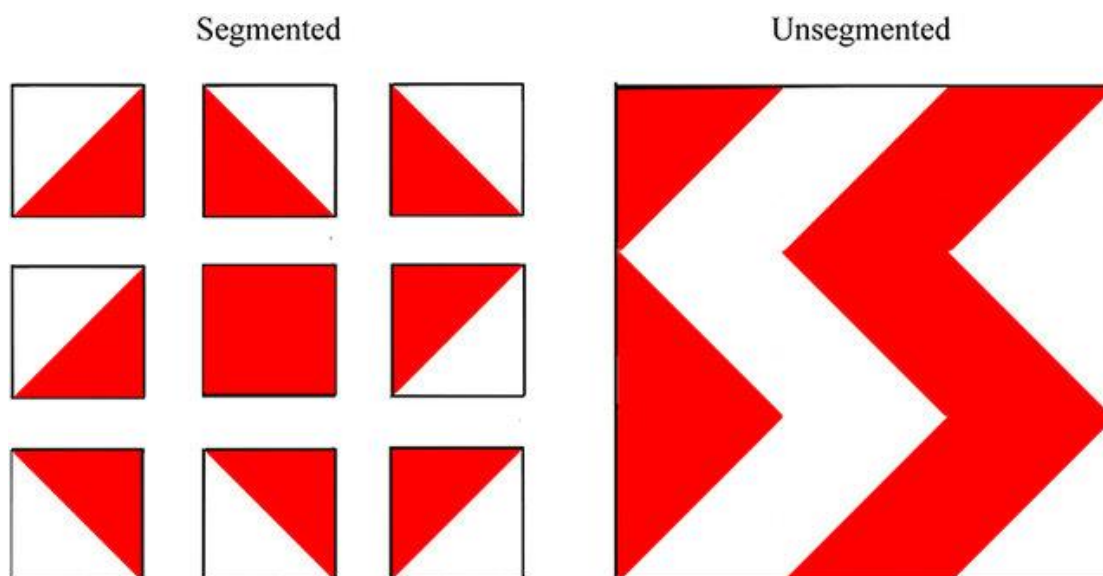
Figure 2 Examples of the block design subtest, adapted from Lind et al. (2014)

2.2   Overview of Developmental Quotient and early IQ assessments and their associations with adult IQ

Before the age of 2, DQ assessments are predominantly focused on neurologic, motor and sensorimotor skills rather than more advanced cognitive skills measured in a traditional intelligence quotient task (Aylward, 2009). For example, in the widely used Griffiths developmental task there are five subscales: locomotor scale; personal-social scale; hearing and speech development; hand and eye co-ordination; and performance tests scale (Griffiths, 1970). Similarly, the Bayley scale of infant development is another commonly used DQ measure, in which mental and motor sub scores are calculated in order to try and identify infants and toddlers at risk for

cognitive impairment (Werner & Bayley, 1966). Thus, while later IQ tests look to assess skills such as pattern recognition or working memory, DQ tests look to measure more basic functions such as walking, understanding of simple instructions, and ability to imitate other (Flensborg-Madsen & Mortensen, 2018).

IQ and DQ tests both produce an ordinal score. This allows for individuals to be placed on a scale, with higher scores indicating greater cognitive performance. When participants are followed longitudinally, there is the ability to test whether early DQ scores are predictive of later IQ. Results appear to mostly support the hypothesis that early DQ is predictive of later IQ performance. For example, DQ performance at 2 years and full scale IQ at 6 years were found to have a correlation of r=0.46 (Girault et al., 2018). Additionally, the simple age at which early developmental milestones are reached has been found to significantly correlate with adult full scale IQ, with correlations as high as r=0.19 (Flensborg-Madsen & Mortensen, 2015; Flensborg-Madsen & Mortensen, 2018). Therefore early DQ assessments, while measuring different skills to IQ tests, do appear to provide an important way to identify infants as risk for later lower IQ performance.

In addition, there has also been research into how highly correlated IQ is, from earlier assessments in childhood to IQ assessed in adolescence and adulthood. Largely, it has been found that that the later the assessment in childhood, the more predictive it is of IQ in late adolescence or adulthood. For example, it was found that the correlation between IQ at 17 years was more strongly correlated with IQ at 8 years (r=.77) than earlier assessments at 6 years (r=.67) or 3.5 years (r =.44) (Gottfried et al., 2006). Other IQ research has demonstrated remarkable stability into later life, with a correlation of r=.88 for scores between 12 years and 23 years of age (Schneider et al., 2014) or r=.66 between age 11 and 80 (Deary et al., 2004). Importantly, the stability of IQ may differ depending on relative performance, with those with particularly low scores in childhood showing greater stability than those with initially average or high scores (Schneider et al., 2014). Thus, while the later the child IQ tests is performed the better it will predict adult performance, it still appears that child IQ tests are especially important tools for identifying those at risk for long term, lower cognitive performance.

2.3   **Executive Functioning models and its links to IQ**

As previously mentioned, IQ tests are built upon multiple subtests. These subtests include measurement of underlying functions such as working memory, which are traditionally included under the umbrella of "executive functions". Executive functions (EF) can be defined as a set of higher order neurocognitive processes, required for goal orienting and decision making (Diamond, 2013). However, there are a number of executive functioning models, differing in the components included, how the components link, their stability over time, and their implications for general cognitive functioning and behaviour. In the following section, these differences will be critically analysed and what consequences this has for research looking at general and specific cognitive outcomes.

According to Diamond, executive functions can be categorised into 3 main components: working memory, inhibitory control, and cognitive flexibility (Diamond, 2013). Working memory can be defined as "holding information in mind and mentally working with it" (Diamond, 2013). Inhibitory control can be defined as being able to "control one's behaviour to override a strong internal predisposition and instead do what is more appropriate or needed" (Diamond, 2013). Finally, cognitive flexibility refers to "the ability to switch between different mental sets, tasks, or strategies" (Archambeau & Gevers, 2018). Diamond's model has many similarities to other popular EF models, such as Miyake (2000) and Anderson (2002). Miyake uses different terminology but similarly defined EF as having three predominant components: updating, inhibiting, and shifting that build to form a unitary EF ability (Miyake et al., 2000). Anderson (2002) diverges slightly further, with four discrete but inter related executive domains of attentional control, cognitive flexibility, goal setting, and information processing (Anderson, 2002).

While models differ in how they categorise "core" EFs, they also differ in their implications. For example, as seen in figure 3, Diamond's model proposes that working memory performance drives performance on inhibitory control, cognitive flexibility and onto further advanced behaviours such as fluid intelligence. However, this contrasts to Anderson's model, which instead believes inhibition drives cognitive flexibility, of which working memory is merely a sub-component, see figure 4. Thus, these differences in models have important implications for interventions and

especially for determining the potentially specific deficits that may cause other cognitive and behavioural problems.

Different EF components can be further sub-categorised and measured using a number of different tasks. Working memory can be assessed using either visual or auditory working memory tasks, however both measures require the storage and crucially manipulation of information. For example, the backwards digit span task requires the participant to repeat numbers in the reverse order of that presented by an examiner (Weiss et al., 2015). This is in contrast to simpler measures, such as short-term memory tests (e.g. the forward digit span task), which require participants to simply recall information as previously presented (Cowan, 2008). Inhibitory control can be measured with a number of different tasks including the stroop task, the Eriksen flanker task or the attention network task (ANT) (Eriksen & Eriksen, 1974; Fan et al., 2002; Stroop, 1935). In the stroop test for example, participants override a strong internal predisposition to say a specific colour word, and instead do what's more appropriate or needed, by saying the ink colour of the word (Stroop, 1935). In regard to cognitive flexibility, the key element is being able to change perspectives spatially, and can be measured using tasks such as the Wisconsin card sorting task where participants are required to flexibly sort cards by a specific criterion (e.g. colour, shape or number) (Grant & Berg, 1948).
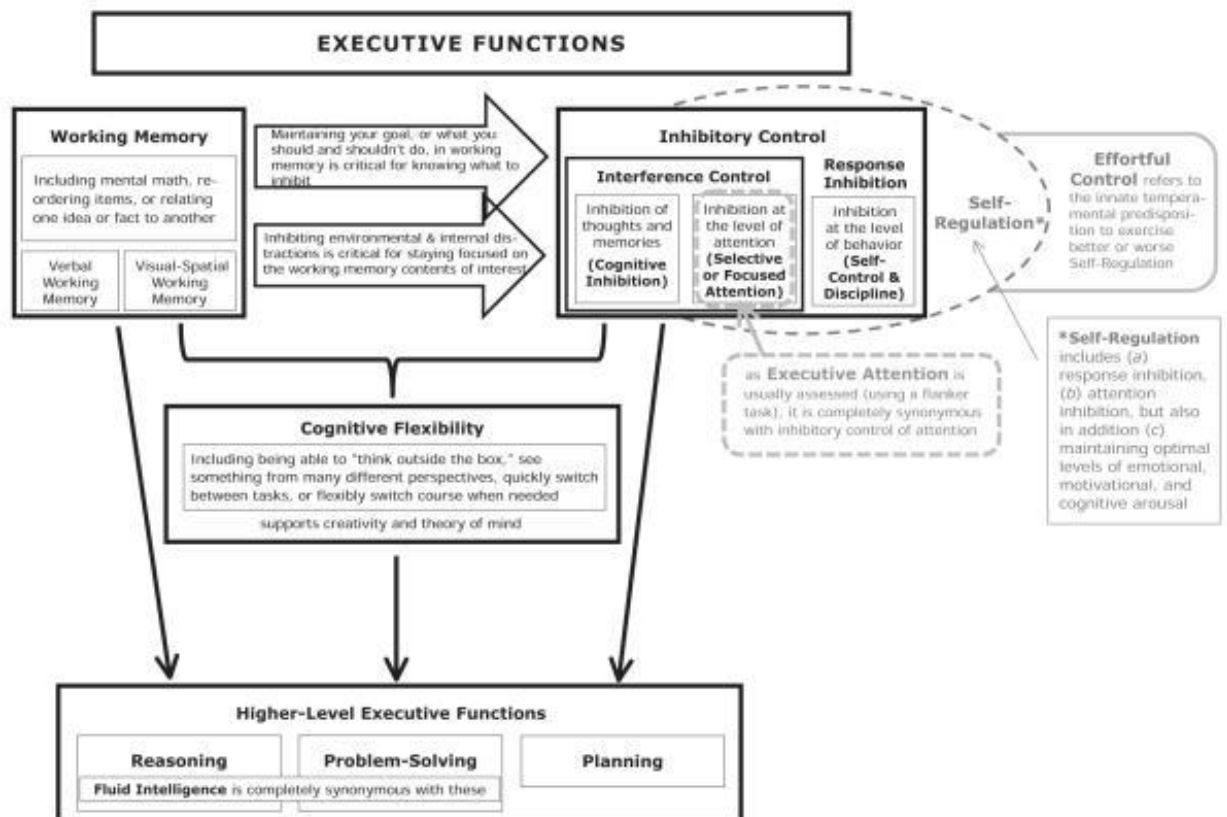
Figure 3: Diamond's model of executive functioning, adapted from Diamond (2013)
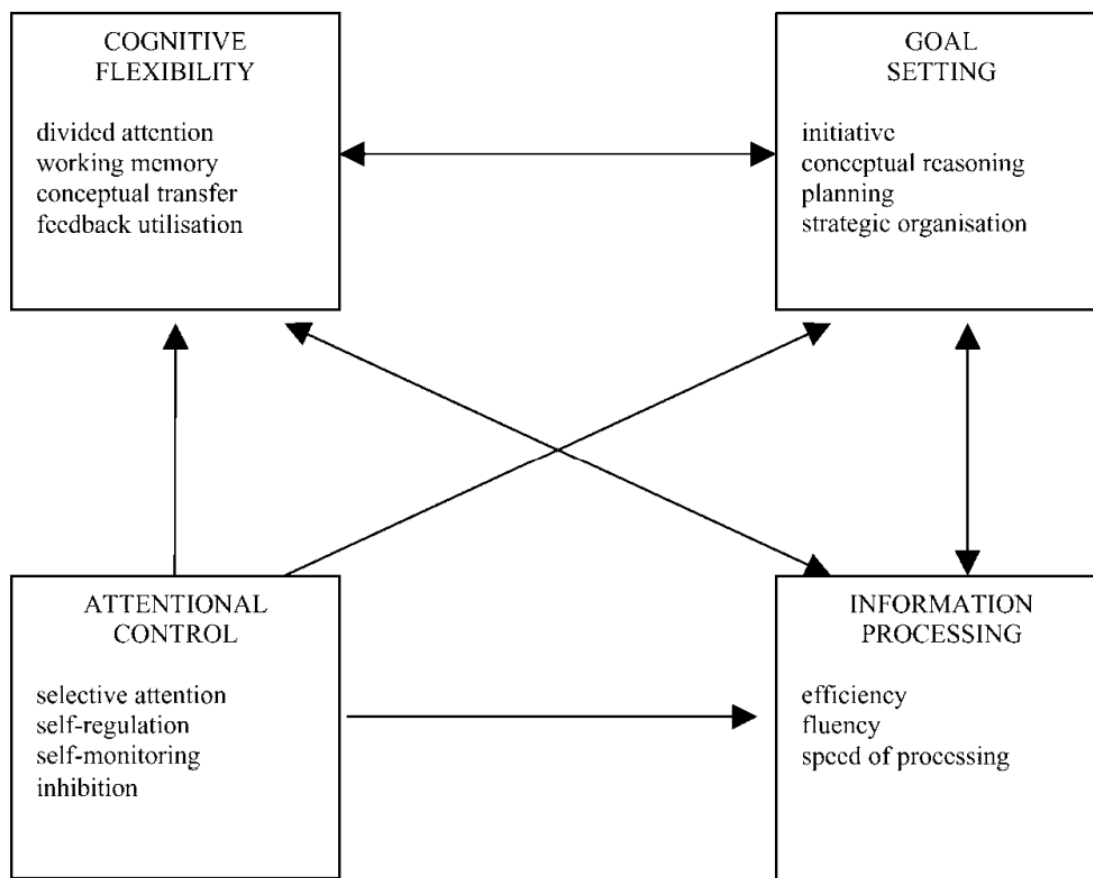
Figure 4: Anderson's model of executive functioning, adapted from Anderson (2002)

In regard to how EF components link to other cognitive functions and real-world outcomes, working memory performance has been proposed as a particularly important sub-component. For example, working memory performance at age 5 has been found to be a better predictor than IQ for school outcomes at 6 years (Alloway & Alloway, 2010). Additionally, working memory is often a sub-test on IQ tests and as such clearly forms a basis for an individual's full-scale IQ score. Therefore, there is discussion on the degree to which EF measures are associated with, responsible for, or predict full scale IQ. For example, combined visuo-spatial working memory and cognitive flexibility performance was found to correctly determine whether an individual scored above or below 85 IQ points 96.2% of the time (Alloway, 2010). Similarly, a latent construct of updating, of which working memory was a key aspect, was found to explain up to 45% of the variance in full scale IQ (Friedman et al., 2006). While these results show the considerable overlap between EF and IQ, there is also conflicting evidence. For example, from 10 different measures of

executive functioning ability, it was found that only 3 were significantly correlated with full scale IQ in healthy children, with a measure of phonological verbal fluency the largest correlate at r=0.30 (Ardila et al., 2000).

Another important aspect of the overlap between EF and IQ is the stability of EF over time. While the strong evidence for IQ stability has already been discussed, for example correlations of r=.66 between IQ at age 11 and 80 (Deary et al., 2004), evidence does not appear to show such high levels of stability for EF. For example, inhibition performance in kindergarten was found to not significantly predict inhibition in the first grade (Vandenbroucke et al., 2017). Alternatively, while there are significant correlations between EF performance at ages 8 and 12, the correlation is only moderate at r=.38 (Harms et al., 2014). Thus, while IQ appears largely stable throughout development, EF measures do not show such a similarly stable pattern. Overall, while there is clearly overlap between EF measures and full-scale IQ, the degree to which likely depends on the EF measures and tests used. In addition, if EF performance varies more substantially throughout development, the degree to which it overlaps with IQ may also vary.

2.4 Relationship between specific cognitive skills and attention.

What is particularly important in regard to EF measures, are their potential links, associations or causes of other outcomes of interest. While EF dysfunction may partially explain poorer IQ scores, they have been particularly associated with attention problems (Willcutt et al., 2005). The clinical diagnosis for those with severe attention problems is attention deficit hyperactivity disorder (ADHD), characterised by behaviours such as an inability to focus (inattention) or by fidgeting behaviour (hyperactivity). A particular prominent theory is that ADHD, and attention problems in general, are a result of a primary deficit in executive functioning (Pennington & Ozonoff, 1996). In addition, there has been much research into how specifically working memory, inhibitory control and to a lesser extent cognitive flexibility are key factors for ADHD (Coutinho et al., 2018; Farrant et al., 2014; Martel et al., 2007; Rommelse et al., 2007; Woltering et al., 2013).

For ADHD, there has also been discussion of whether individuals who solely display inattentive symptoms without hyperactivity, sometimes also referred to as attention-deficit disorder (ADD) (Diamond, 2005), should be considered to have a separate disorder compared to those with combined ADHD including hyperactivity. The underlying cognitive dysfunction for both disorders are thought to differ, with those solely displaying inattention symptoms thought to have a specific, pure deficit in working memory (Diamond, 2005). Alternatively, it may be that ADHD and attention problems generally are linked to more general cognitive problems, of which EF dysfunction is merely a sub-component. Meta-analysis has demonstrated that ADHD individuals have lower IQ than healthy controls, but that it is a relatively small effect at approximately 3 IQ points (Bridgett & Walker, 2006). However the decision to control for IQ when investigating EF and ADHD has been debated, in part due to the argument that ADHD may cause lower performance on a wide range of abilities, including IQ (Dennis et al., 2009).

2.5   Chapter Summary

In the current chapter the links between general cognitive functioning, executive components, and attention have been reviewed. The debate surrounding which specific skills are responsible for other problems causes difficulty in identifying the general and specific problems for individuals. For example, while there is substantial evidence that executive dysfunction and attention problems are linked, there is debate about whether one should control for general cognitive functioning (e.g. IQ) to differentiate a general from a specific deficit. In the next chapter, the general and specific cognitive outcomes of VP/VLBW individuals will be reviewed. How VP/VLBW's performance on measures of general and specific cognitive outcomes relate to specific attention problems will then be subsequently reviewed. Following this, the cognitive outcomes of SGA individuals, born either at term or preterm, will be discussed.

Chapter 3   **The association of VP/VLBW and SGA with later**
cognitive and behavioural outcomes

The result of disruption to foetal development has been noted as having immediate consequences for the development of the brain. One indicator of the effect on functional development is that more VP/VLBW and SGA infants suffer from major neurocognitive impairment compared to those born healthily at term. As well as major cognitive impairment, the utility of DQ and IQ tests is the ability to identify reliably and consistently those with lower but more subtle cognitive deficits. In this chapter, research into the general cognitive, specific cognitive and behavioural outcomes of those born VP/VLBW will first be critically assessed. Next, the general cognitive outcomes of SGA participants will be addressed. As VP/VLBW or SGA birth are two of many potentially important factors for cognitive performance, this chapter will also address other important factors that may moderate, mediate or otherwise influence the effect of foetal factors on cognitive performance. The current limitations of past research will also be raised, as these formed the basis for the research questions in this thesis.

3.1   **General cognitive outcomes of VP/VLBW in infancy, childhood, and adulthood**

Major cognitive and developmental impairment is an area of particular concern following preterm birth. Evidence suggests that even before 2 years of age, developmental differences between VP/VLBW and term born infants are apparent. However, the degree to the difference between VP/VLBW infants and term born controls may depend on whether the age of the VP/VLBW participant is corrected; calculated by reducing the chronological age of the infant by the number of weeks

they were born preterm. For example, when assessed on the Bayley scale at 4, 8, 12 and 16 months, VLBW infants were found to score lower on all developmental domains than full term infants when using uncorrected scores but did not significantly differ after correction (Barrera et al., 1987). Other studies of VP/VLBW infants generally have found lower performance on DQ measures. In another study at 24 months, even after age correction, it was found that 40% of VP infants scored below 1 SD on either the Bayley mental or motor subscales, with 7% of those scoring 2 SD below the norm on both sub-scales (Stoelhorst et al., 2003). In studies of EP, the EPICure cohort and the Victorian Infant Collaborative Study Group (VICS) both found EP toddlers to have lower overall DQ scores, with 30% and 18.3% scoring 2 SD below the norm in each study respectively (Victorian Infant Collaborative Study Group, 1997; Wood et al., 2000). In a meta-analysis of VP/VLBW individuals up to 5 years of age, it was estimated that 24.6% score below 2 SD while 32.4 % score below 1 SD points on standardised measures of cognitive impairment (Blencowe, Lee, et al., 2013). However, when comparing VP/VLBW to standardised norms there is a need to proceed with caution. As previously noted, standardised norms quickly become outdated due to the Flynn effect, with the scores of VP/VLBW individuals needing to be instead compared to control groups, born in the same region and time. This has been found to change the rates of impairment dramatically for VP/VLBW groups, with 2.4 times more VP/VLBW individuals being defined as impaired when compared to controls rather than obsolete norms (Wolke et al., 1994). Thus, all studies looking at cognitive performance of VP/VLBW individuals must have a valid control group to accurately compare to in order to accurately determine levels of impairment.

Results therefore appear to show discernible differences between VP/VLBW and term born controls on early DQ measures. While a disproportionate number of infants score 2 SD below the norm, there is also a large number of infants demonstrating a minor cognitive impairment. This lower cognitive performance appears to show stability into childhood. A meta-analysis of 14 VP/VLBW cohorts suggests DQ scores are predictive of later childhood IQ scores with a correlation of 0.61 (Luttikhuizen dos Santos et al., 2013). Thus, information of VP/VLBW's DQ performance is likely to provide accurate information on the IQ scores of VP/VLBW children.

Once into childhood, where childhood IQ tests can be used, there is substantive evidence for cognitive deficits for VP/VLBW individuals in comparison to term born controls. In a recent meta-analysis of 71 VP/VLBW cohorts born after the year 1990, an IQ deficit of 12.9 IQ points was found in comparison to term born control groups (Twilhaar et al., 2018). This is in concordance with past meta analyses comparing all preterm children to term born controls, where deficits of 10.9 and 11.9 IQ points have been reported (Bhutta et al., 2002; Kerr-Wilson et al., 2012). As well as determining the difference between preterms and controls on IQ tests, a secondary objective of these meta-analyses was to identify whether preterm cohorts born more recently have better cognitive outcomes than those born further in the past. To test this, meta-regression has been used where year of birth has been included as a covariate. However, none of the aforementioned meta-analyses found that differences between preterms and controls was reducing as year of birth got closer to the modern day. This would therefore suggest that preterms born today continue to demonstrate the cognitive problems that preterms of earlier generations have shown. However, these analyses are confounded due to the fact that different cohorts born in different eras are also born in different countries, assessed at different ages and use different cognitive tests. In a comparison of cohorts from the same region, VICS compared cognitive outcomes at 8 years of age in their 3 separate EP cohorts born in 1991, 1997 and 2005. They also found no evidence that cognitive outcomes were improving with time (Cheong et al., 2017). Overall, these results suggest that research into VP/VLBW individuals born 25-40 years ago is likely to be applicable to current generations of VP/VLBW individuals. Therefore, it is necessary to understand the difficulties that these VP/VLBW adults now demonstrate as well as understanding what can be done to improve outcomes.

The research so far into adulthood has also found cognitive deficits remaining for VP/VLBW individuals. For example, cohorts from Cleveland (USA), Trondheim (Norway), Helsinki (Finland), Victoria (Australia), and Bavaria (Germany) have reported differences in IQ scores to controls from as low as 5 points to as high as 15 points (Doyle et al., 2015; Eryigit Madzwamuse et al., 2015; Hack et al., 2002; Løhaugen et al., 2010; Pyhala et al., 2011). While no meta-analysis has specifically been performed to combine these results, there are a number of potential reasons for

the apparent variation in scores. These include different recruitment criteria, differences in selective dropout rates into adulthood, or how cases of participants too impaired to undergo IQ testing are handled. On the subject of VP/VLBW too impaired to take part, some cohorts may have decided to give these participants a proxy IQ score of 40, the lowest possible score on a traditional IQ test (Eryigit Madzwamuse et al., 2015). However, not all cohorts have done this or have not clearly reported this in their analyses. This makes comparing results across cohorts difficult, as those including proxy scores will almost certainly show a greater difference between VP/VLBW and term born controls than those who have not. Thus, future studies looking to assess the universality of cognitive outcomes of VP/VLBW adults must ensure that comparisons across cohorts are as valid and as fair as possible. As well as these systematic differences between cohorts, differences in IQ scores could be the result of biological, medical, or socioenvironmental risk and resiliency factors that occur at differing rates in different cohorts. For example, male sex, small for gestational age and low socio-economic status (SES) have all been proposed as factors that interact with VP/VLBW status to further reduce cognitive performance (Benavente-Fernández et al., 2019; Linsell et al., 2015; P. Shah & Kingdom, 2011; Wolke, 2019). While past meta-regressions in childhood have been utilised to investigate these potential covariates, they are restricted due to their use of aggregated data. The use of aggregated data increases the risk of ecological bias as the variation of a covariate across studies does not reflect the individual variation of the covariate within studies (Lambert et al., 2002). Instead, Individual Participant Data (IPD) meta-analyses involves obtaining the raw data from each cohort so that important covariates can be tested on the individual level. This results in IPD meta analyses having greater ability to detect important covariates than traditional meta-regression (Lambert et al., 2002). As a result, IPD meta-analyses can be utilised to identify specific risk factors for VP/VLBW individuals that would otherwise be missed using meta-regression.

Finally, there is currently little research looking at trajectories of VP/VLBW cognitive performance from infancy to adulthood. This research can investigate whether preterm individuals display cognitive deficits or delays in childhood that have either been sustained or diminished into adulthood. If a VP/VLBW individual suffers from a cognitive delay, they would be expected to perform poorly on early cognitive

assessments in infancy or childhood but display cognitive catch up once into adulthood. Evidence so far has demonstrated little cognitive catch up into adulthood but rather a cognitive deficit that sustains throughout development. Relative to term born controls, EPICure found a minimal increase of 0.5 IQ points per year between age 6 and 19 in their EP sample (Linsell et al., 2018). Similarly, the Stockholm Neonatal Project found performance IQ Z scores improved from -0.46 at age 5 to -0.38 at age 18, an increase equating to just 1.2 IQ points over the 13 years in their VLBW sample (Stålnacke et al., 2015). In the Bavarian longitudinal study, differences between VP/VLBW and controls remained remarkably stable, from 17 IQ points at age 4 to 16.4 IQ points at 26 years (Breeman et al., 2015). Finally, In the VICS cohort, EP/ELBW showed a slightly larger difference in IQ to controls at 18 years (11.1 IQ points) than they did at 5 years (9 IQ points) (Doyle et al., 2015). Therefore, the results from repeated measurement and trajectory analyses are in concordance with results that have reported solely on IQ for VP/VLBW children and adults at single timepoints, suggesting sustained lower cognitive performance.

Overall, there is evidence that preterm individuals demonstrate lower cognitive performance than term born controls. In particular, VP/VLBW and EP show the greatest deficits, in line with a dose-response between gestational age and cognitive performance. While evidence is most prominent in childhood, data from infancy suggests lower DQ performance is already measurable before the age of 2. Evidence into adulthood indicates that lower IQ scores remain and are stable, with research showing very minimal cognitive catch up as demonstrated by trajectory analyses. On the question of whether improvement of medical care are matched by improvements in IQ outcome, the results from childhood meta-analyses suggests that more recent cohorts suffer from similar levels of general cognitive impairment than older cohorts. What is less known is whether the IQ outcomes of VP/VLBW are universal in adulthood and what can be done to universally improve outcomes. In order to accurately compare VP/VLBW cohorts across countries, a systematic, methodological approach is required. This means making comparisons that treat participants too impaired to take part in the same way in each cohort; that use the same harmonised risk factors; and that consider factors such as selective dropout rate. To do this, IPD meta-analyses of IQ for VP/VLBW adults can be performed, as it is considered the

"gold standard" of systematic review (Tierney et al., 2015). However, as previously noted, IQ performance is a general measure of cognition and is based upon being able to perform in a number of specific tasks, such as executive functioning.

3.2 Specific neurocognitive and behavioural outcomes of preterms in both childhood and adulthood.

In order to fully understand the cognitive differences between VP/VLBW and term born individuals, the potential for specific cognitive dysfunction in VP/VLBW individuals must also be investigated. For the purpose of this review, differences between VP/VLBW and term born controls on executive functioning (EF) subcomponents will be based on differences for the three core EFs (working memory, inhibitory control, cognitive flexibility), for which there is general agreement upon (Best & Miller, 2010; Diamond, 2013; Miyake et al., 2000). See Diamond's model in Figure 3, chapter 2.3.

3.2.1 Executive Functions

VP/VLBW individuals are thought to suffer from lower working memory performance. In individual studies during childhood, adolescence and adulthood, effect sizes between VP/VLBW individuals and controls have been found to be 0.30, 0.63 and 0.50 respectively (Aanes et al., 2015; Fitzpatrick et al., 2016; Twilhaar et al., 2020). A meta-analysis of childhood EF measures has reported an effect size indicating that VP/VLBW individuals score on average 0.36 SD lower on working memory than controls (Aarnoudse-Moens et al., 2009). However, this study relied solely on studies assessing verbal working memory, an area that has been suggested to be less impaired in VP/VLBW individuals than auditory working memory (Martinussen et al., 2005; Mulder et al., 2011; Retzler et al., 2018). Nevertheless, in a more recent meta-analysis of preterm children of any gestation born since the 1990s, both visual or verbal working memory performance were equally impaired in comparison to controls (Houdt et al., 2019). In regard to inhibitory control, a relatively

small deficit for all preterm children has been found, with effect sizes of 0.25 and 0.40 as demonstrated by meta-analyses (Houdt et al., 2019; Mulder et al., 2011). For VP/VLBW individuals especially, slightly larger effect sizes of 0.30 and 0.48 have been reported in individual child and adult studies respectively (Nosarti et al., 2007; Ritter et al., 2013). On the final component of EF according to Diamond's model, meta-analyses of all preterm children or just VP/VLBW have both found a 0.5 standardised mean difference (SMD) on measures of cognitive flexibility in comparison to term born controls (Aarnoudse-Moens et al., 2009; Houdt et al., 2019). SMD, often used interchangeably with effect size or Cohen's D, compares the differences between groups as a proportion of their shared standard deviation, with SMD's of 0.2, 0.5 and 0.8 thought of as small, medium and large effect sizes respectively (Cohen, 1992).

There is strong evidence that VP/VLBW and preterms in general demonstrate poorer performance on measures of EF. However, the effect sizes or differences between VP/VLBW and term controls do appear to be smaller than those reported for IQ (Eryigit Madzwamuse et al., 2015). Additionally, while the dose response between lower gestational age and lower child IQ has been well supported, a dose response has not been found for lower gestational age and poorer EF performance (Houdt et al., 2019). Similarly, while male VP/VLBW children are thought to have lower IQs than female VP/VLBW children, a similar sex difference has not been found for EF (Houdt et al., 2019). Thus, there may be different reasons or causes for poorer IQ and EF outcomes among VP/VLBW individuals. These difference may also demonstrate the utility in analysing both specific EF measures and IQ as potential predictors of other behavioural outcomes. While EF and IQ are clearly overlapping, they may differ in their ability to predict other outcomes of interest, such as attention, for VP/VLBW individuals.

3.2.2 **Attention**

An area of particular concern after preterm birth are attention problems, which has been suggested to be one of the most affected behaviours after preterm birth (Aarnoudse-Moens et al., 2009). It has been found that VP/VLBW children are especially at a higher risk of ADHD diagnosis than controls with an odds ratio of 3.04 (Franz et al., 2018). When looking at symptomology in particular, it appears VP/VLBW children suffer from particular problems with inattention (SMD to controls of 1.31), with lesser but still apparent problems with hyperactivity and impulsivity (SMD = 0.74) (Franz et al., 2018).

Results therefore indicate that VP/VLBW have both attention problems and executive dysfunction. According to Willcutt, these two issues are closely linked, with the latter at least partially responsible for the former. As noted in the previous chapter, evidence from the general population has indeed linked executive functioning and attention problems (Willcutt et al., 2005). For VP/VLBW individuals, evidence within childhood suggests a relationship between lower executive functioning and greater attention problems. For example, working memory performance has been found to mediate the relationship between VP/VLBW birth and teacher or parent rated attention (de Kieviet et al., 2012; Nadeau et al., 2001). Similarly, attention problems in VP/VLBW children has also been associated with lower inhibitory control (Aarnoudse-Moens et al., 2009). In addition, as VP/VLBW are thought to display more issues with inattention than hyperactivity, it would be logical to assume that they may have a primary working memory deficit, as has been proposed for those in the general population with specific inattention problems (Diamond, 2005).

What is important to emphasise, as previously noted, executive functioning and general IQ are linked. The majority of prior research has failed to account for general cognitive ability when investigating the links between executive functioning and attention for VP/VLBW individuals. This is important as when trying to identify a specific deficit, it must differentiate from the lower general cognitive performance that is more common for VP/VLBW individuals. Another important caveat to VP/VLBW research in childhood is that attention problems have mainly been found when

assessed by parent, teacher, or experimenter rating. In adulthood, the opportunity to assess self-reported ADHD symptomology can instead be used. However, the sparse evidence from VP/VLBW adults suggest they do not self-report more problems with attention, with an Ontario cohort of ELBW adults finding no significant differences to controls (Lahat et al., 2014). This would potentially indicate that once into adulthood, the attention problems of VP/VLBW individuals have mostly subsided. However, this is in contrast to the Bavarian longitudinal study that has used parent reported or experimenter ratings of VP/VLBW adults and found attention problems persist (Breeman et al., 2016). Comparing these results is somewhat complicated due to the fact that as well as differing in how they assessed attention problems, cohorts have also differed in a number of other ways (age of assessment, initial medical care received, education system, gestation/birthweight criteria used). Thus, research is needed to identify whether this discrepancy is due to the importance of who is the informant or whether there are other important cohort factors also at play.

Overall, further research is required to identify the factors associated with the attention problems of VP/VLBW adults demonstrate and the importance of informant. Research in childhood has found links between executive functioning and attention but whether this link is seen in adulthood, depends on who is rating the individual, or after controlling for general cognitive ability has not been thoroughly tested. Resolving these issues would especially provide greater confidence for those potentially looking to investigate how executive functioning-based interventions may help VP/VLBW children and adults.

3.2.3   The Preterm Phenotype

There has been much research to investigate the cognitive and behavioural outcomes for VP/VLBW children and preterms in general. This has resulted in the 'preterm phenotype' hypothesis, the attempt to bring together the results from the differing domains into one coherent understanding of the outcomes of preterms. Evidence suggests that when one domain is affected, such as intelligence, there is an increased likelihood that other domains, such attention, will be also affected (Wolke et al., 2019).

In a systematic review of the literature, Arpi and Ferrari (2013) determined that between birth and 2 years, VP/VLBW demonstrate greater motor problems, emotional dysregulation, and attention problems. They also determined that these problems persist into early childhood, with ADHD symptoms, anxiety, depression, somatic symptoms, and relational difficulties more common in those born VP/VLBW than term borns between 3 and 5 years of age. Further evidence for the multitude of problems arising from preterm birth is seen in a cohort of 5-year-old preterm children born before 30 weeks of gestation. It was found that 44% had more than one disability in domains of intelligence, neurological assessment and motor performance while just 17% demonstrated a single disability (van Baar et al., 2005). Later assessments into childhood and adolescence have continued to demonstrate the preterm phenotype, consisting of comorbid attention problems, emotional problems and atypical social development (Johnson & Marlow, 2011). Overall, there is substantial evidence that preterms, and VP/VLBW individuals especially, rarely show a single dysfunction but often show a number of comorbidities. Whether lower general or specific cognitive functioning is the primary reason for other issues, such as attention problems, needs further investigation.

3.2.4 **Summary of the links between preterm birth with cognitive and behavioural outcomes.**

To conclude, there is a growing appreciation that the cognitive problems displayed by VP/VLBW do not happen in a vacuum but often co-occur with other problems. Results from DQ and child IQ have found substantial differences between VP/VLBW and term born controls, that do not appear to subside into adulthood. As well as differences on general measures of cognition, there is also findings that VP/VLBW have lower performance on a range of specific cognitive measures, including EF measures such as working memory and inhibitory control. It has been hypothesised that specific cognitive problems, such as executive dysfunction, are responsible for the attention problems often seen in VP/VLBW individuals. However, research investigating this

specific association is limited for VP/VLBW adults, while potentially being part of a more general cognitive problem has been generally overlooked for VP/VLBW individuals.

3.3 General cognitive outcomes of SGA individuals in infancy, childhood, and adulthood

Small for Gestational Age has also been consistently associated with lower cognitive performance in infancy and childhood, as demonstrated by lower scores on DQ and IQ tests relative to those born appropriate weight for gestational age (AGA) (Bie et al., 2010; Sacchi et al., 2020; Silva et al., 1984). It has been hypothesised that the effect of SGA on cognition differs depending on whether the individual is born at term or preterm, with SGA's effects being disproportionately larger for those born very preterm (P. Shah & Kingdom, 2011). This is based on evidence that SGA is more closely related to mortality in preterms than it is in term borns, suggesting SGA in preterms is greater evidence for pathology (Ananth & Vintzileos, 2009). Similarly, it has been hypothesised that infants born both preterm and SGA suffer from a "double jeopardy" due to the co-occurrence of the two risk factors (Gutbrod et al., 2000)

Two recent meta-analyses into the effects of early growth on child cognitive outcomes in preterms and term borns provide slightly conflicting results. On the one hand, Sacchi (2020) found evidence that being born SGA is associated with lower cognitive scores of -0.23 SMD in preterms and -0.34 SMD for term borns. Additionally, for IUGR specifically, Sacchi (2020) found differences of -0.36 and -0.39 for preterm and term borns respectively. Thus, Sacchi found the effects of poor early growth to have a marginally greater effects on term borns than preterms. This however differs to Murray (2015), who found not just a bigger effect of IUGR on neurodevelopment but also a greater effect for preterms, at -0.7 SMD for preterms and -0.5 SMD for term borns (Murray et al., 2015). Therefore, there appears to be strong evidence that poor early growth as measured by SGA or IUGR is associated with worse cognitive outcomes in both preterm and term borns. However, it is currently unclear whether the effects are more detrimental for term born or preterm individuals. Additionally, while these two

meta-analyses and the majority of the research have focused on childhood outcomes, little research has investigated whether these finding are also found into adulthood.

Currently, there are contradictory findings of long-term effects of SGA on IQ, potentially suggesting diminishing cognitive differences between SGA and AGA adults. In term born adults, those born SGA have been found to have lower IQ, academic performance and professional achievement than those born AGA (Østgård et al., 2014; Strauss, 2000). In contrast, VP/VLBW adult cohorts have found no differences (Eryigit Madzwamuse et al., 2015; Pyhala et al., 2011), or that it depends on whether post-natal catch up growth was achieved (Brandt et al., 2003; Lundgren et al., 2001). An issue with the majority of these adult studies is that they only look at IQ scores at one timepoint in adulthood. Drop out over time may result in less statistical power than child studies with greater number of participants. Additionally, there is also the potential of systematic attrition bias, where participants of lower socioeconomic status or with more health problems may be less likely to participate in follow up observations (Howe et al., 2013; Nohr & Liew, 2018). If the cohort has sustained large amounts of selective dropout from childhood to adulthood, then the remaining SGA participants may be, for example, higher functioning and are therefore less likely to show large IQ differences with AGA participants. To address these issues, longitudinal analyses using repeated measurements of DQ and IQ from infancy to adulthood can be used. Firstly, by including more cognitive assessments the statistical power to detect a difference between SGA and AGA participants naturally increases. Secondly, by only including those with IQ scores in adulthood in the longitudinal analysis, systematic attrition cannot influence the cognitive trajectories of the remaining SGA participants. If cognitive catch up is shown throughout development by the included SGA group, this cannot be due to the fact that lower functioning participants were systematically lost to follow up, but instead suggests that these remaining SGA participants are demonstrating cognitive catch up relative to AGA participants once into adulthood.

Finally, as discussed in section 1.3.4, there are many ways to calculate SGA. While the majority of cohort studies that looked at cognitive outcomes used neonatal references to determine SGA status, there has been a growing discussion on the use of

foetal references. As these references differ in whom they classify as SGA, it is logical to assume the use of different reference will also change the relationship between SGA and IQ. Theoretically, as neonatal references are more stringent (they classify less participants as SGA), it should only class the very smallest participants as SGA. Thus, under the assumption of a dose response relationship of SGA severity with IQ, it should result in SGA being associated with a larger difference in cognitive scores than when a foetal reference has been used. Very few studies have compared the two different types of reference and their association with cognitive outcomes. Neta (2011) found that as expected, when predicting cognitive impairment (categorised as IQ scores above or below the 10th percentile) the neonatal reference had a lower sensitivity but higher specificity than the foetal reference (Neta et al., 2011). However, this was not replicated by Charkaluk et al. (2012) who found that despite classifying many more participants as SGA, the foetal reference had a similar odds ratio for cognitive impairment (categorised as IQ scores above or below 1SD) as the neonatal reference (Charkaluk et al., 2012).

To conclude, there is evidence that being born SGA is associated with lower cognitive performance, at least within childhood. Evidence into adulthood is limited but would appear to suggest that SGA participants may demonstrate cognitive catch up in comparison to AGA participants. However, as this line of research has not investigated the cognitive trajectories of SGA and AGA participants, this effect could be partially due to selective dropout or reduced statistical power due to dropout generally. Other methodological differences complicate the matter of SGA and IQ. Especially, whether to use neonatal or foetal references to classify SGA. While in term born participants, the difference between foetal and neonatal references will be relatively small, there is a larger disparity between references for VP/VLBW individuals, where the foetal reference will classify many more infants as SGA. How identifying these additional VP/VLBW infants as SGA affects the association between SGA and IQ is largely unknown and requires investigation.

3.4   **Factors that must be accounted for in cognitive outcomes of VP/VLBW and SGA individuals**

In this chapter so far it has been reviewed how early risk factors of VP/VLBW birth and SGA are associated with cognitive dysfunction in childhood and adulthood. However, it may be too simplistic to ignore other factors that influence cognitive outcomes across childhood and adolescence. Thus, there is a need to highlight other factors across the lifespan that may additionally influence cognitive performance. In particular those that may moderate or mediate the effect of VP/VLBW or SGA birth on cognitive trajectories of these individuals. In the following section the other factors that have been found to influence cognitive outcomes of VP/VLBW and SGA individuals will be discussed, with a specific focus on social factors.

As there is clear heterogeneity for cognitive outcomes both within and across preterm cohorts (Twilhaar et al., 2018) there are likely factors that can help or further hinder the cognitive performance of preterms. However, the identification of risk and especially resiliency factors, factors that help in "beating the odds" despite high risk for adverse developmental outcomes, have currently been under investigated (Wolke, 2019). Research in childhood suggests that there are potentially genetic, perinatal, and socio-environmental factors responsible for this variability (Anderson & Doyle, 2008). Past research has predominantly focused on medical complications at birth factors or sex differences. For example, BPD and IVH have both been found to be important covariates for IQ in meta-analyses or in individual studies of preterm children (Benavente-Fernández et al., 2019; Twilhaar et al., 2018). Additionally, it has been found that VP/VLBW males largely have poorer IQ performance than VP/VLBW females (Benavides et al., 2019; Linsell et al., 2015).

In contrast, there has been far less focus on social determinants of cognitive outcomes. Socioeconomic status, often measured indirectly through maternal or paternal education, has also been found to be predictive of VP/VLBW children's cognitive performance (Linsell et al., 2015; Wolke & Meyer, 1999). The effect sizes of these social factors can be substantial, with research indicating low maternal SES has a similar effect as brain injuries, in the form of white matter injury volume and

intraventricular haemorrhage (Benavente-Fernández et al., 2019). More specific measures of parental behaviour have also been found to be predictive of long-term IQ performance for VP/VLBW individuals. Independent of SES, VP/VLBW infants with a poor parent-infant relationship were found to have IQ scores nearly 9 points lower than those with a good relationship 26 years later as adults (Breeman et al., 2017). Finally, as well as IQ, socio-environmental factors are also important for educational outcomes. For example, VP/VLBW children whose mothers had high maternal sensitivity at age 6 performed at equivalent levels as term born controls on reading, spelling and writing at age 8 (Jaekel et al., 2015) and academic performance at 13 years (Wolke et al., 2013).

Regarding SGA birth, there has been less research investigating the comparative or interactive effects with medical or social factors. From DQ at 5 months to IQ at 56 months, neonatal complications (Apgar score, IVH etc.) were found to have a larger effect on cognitive development than the difference between SGA and AGA participants (Gutbrod et al., 2000). For social factors, at five years of age, it was found that maternal IQ, maternal smoking and factors relating to child rearing were all more important to IQ performance than the 5 IQ point difference between SGA and AGA children (Sommerfelt, 2000). Similarly, at 19 years of age, it has been found that a 1SD increase in relative birthweight is associated with an increase of IQ of 2.6 IQ points but this is a relatively small effect in comparison to the difference between high and low educated parents at 14.2 points (Weisglas-Kuperus et al., 2008)

Thus, in order to better understand and improve the cognitive outcomes of VP/VLBW and SGA adults, the study of both neonatal risks and social influences are required. VP/VLBW and SGA neurocognitive outcomes are by no means fixed and may be altered by a number of risk and resiliency factors across the lifespan. Some of these factors, such as parental behaviour, may be more modifiable than others, such as sex or brain injury. However, before any intervention is implemented, the reliability and universality of these outcomes must be established.

Chapter 4   Outstanding issues and research questions

In order to understand the long-term cognitive and behavioural outcomes of VP/VLBW and SGA adults, a holistic approach may be required. In particular, this thesis considers how factors are measured and how their importance changes across the life course. An over focusing on the early complications at birth, and in particular VP/VLBW birth, may result in a lack of consideration of SGA or socio-environmental factors that may have similar effects on neurocognitive outcomes. As these are potential areas for successful interventions, the universality of risk and resiliency factors should be considered using harmonised data from multiple international cohorts. This does not only increase statistical power due to the increased sample size but also a greater ability to understand what factors affect VP/VLBW adults universally and which factors are specific to one region or cohort. In addition, understanding whether specific EF deficits in VP/VLBW adults explain behavioural difficulties such as attention is important for both determining the nature of the preterm phenotype and for potential interventions. Finally, investigating developmental changes from infancy to adulthood for SGA individuals, either born preterm or at term, will provide evidence of the long term effects on cognitive development and how risk factors work independently or moderate one another. In addition, determining potential cognitive improvement in SGA individuals must also consider the importance of socio-environmental factors. These areas of research require further investigation and as such form the basis for the three research studies. In the next subsections, the research questions will be outlined and how they relate to current gaps in the literature, with a summary in Table 2.

Table 2: Summary of key research areas for this thesis

Table 2: Summary of key research areas for this thesis

	Very Preterm/ Very low Birthweight Research			Small for Gestational Age Research
Area of Research	IQ	Executive Functioning	Attention Problems	IQ
What is currently well known	VP/VLBW perform poorer than controls on early measures of developmental quotient and child IQ (Twilhaar et al., 2018)	VP/VLBW children perform poorer than controls on all key measures of executive functioning components of working memory, inhibitory control and cognitive flexibility (Houdt et al., 2019)	VP/VLBW children have higher levels of parent reported ADHD symptoms and are rated with shorter attention spans than term born controls (Johnson et al., 2016)	SGA children perform poorer than controls on early measures of developmental quotient and child IQ (Bie et al., 2010)

| Where there is controversy | Is there heterogeneity in the effect of VP/VLBW birth on adult IQ across cohorts (Eryigit Madzwamuse et al., 2015; Hack et al., 2002)? | Is executive functioning responsible for behavioural problems seen in VP/VLBW adults, such as attention problems (Lahat et al., 2014)? | Does the effect of SGA on IQ disproportionately affect preterm or term born individuals? Are findings dependent on using a neonatal or foetal reference to calculate SGA(Charkaluk et al., 2012; Sacchi et al., 2020; P. Shah & Kingdom, 2011) ? |

Where there is sparse or no evidence	What are the key neonatal or social variables that are universal risk and resiliency factors for adult IQ among VP/VLBW adults?	In adulthood, does executive functioning performance explain the difference between VP/VLBW adults and controls on measures of attention?	For VP/VLBW adults, are self-reported measures of attention problems in concordance with parent or experimenter ratings? Do VP/VLBW adults specifically show greater problems with inattention than hyperactivity?	Do SGA participants demonstrate "cognitive catch up" with smaller cognitive differences to controls in adulthood than in childhood?
Important for which research study in this thesis	Study 1	Study 2	Study 2	Study 3

4.1   Study 1

As previously discussed, there is heterogeneity in cognitive outcomes for VP/VLBW individuals in both childhood and adulthood. Comparisons across studies are complicated due to differing methodologies, follow up rates and inclusion criteria. In study 1, the first ever Individual Participant Data Meta-Analysis of IQ outcomes for VP/VLBW adults is presented. A range of antecedent factors are considered as to explain the IQ difference both within and across VP/VLBW cohorts.

Research Questions:

- **Firstly, what are the mean differences in IQ in adulthood for VP/VLBW individuals in comparison to same aged term born controls?**
- **Secondly, what are the individual level neonatal and social factors that must be investigated when considering the relationship between VP/VLBW birth and IQ?**
- **Thirdly, are cohorts that provide individual level data representative of all VP/VLBW cohorts that have published on adult IQ?**

4.2   Study 2

While the investigation of general cognitive ability for VP/VLBW adults has great utility, there is the argument it does not provide the specificity to identify key deficits. One key set of skills associated with both general cognitive performance and attention problems is executive functioning. Executive functioning has been proposed as a building block to IQ but also pivotal for attention problems in the general population and VP/VLBW children. Study 2, investigates whether specific cognitive (executive functioning) or general cognitive performance (IQ) are associated with attention problems for VP/VLBW or EP adults in comparison to controls.

Research Questions:

- **Are the greater attention problems seen in VP/VLBW or EP adults as compared to term-born adults best explained by specific executive functioning deficits, general cognitive abilities, or sex.**
- **Are attention problems and the factors associated with them replicated across cohorts? Are findings independent of the informant used (parent, self or experimenter rating)**

4.3   Study 3

In chapter 1, the difficulty in calculating SGA, especially for VP/VLBW infants, was discussed. In chapter 3, the limited research that has looked into how using foetal and neonatal SGA references impact findings on long term cognitive outcomes was reviewed. In study 3, as well as considering the importance of using foetal or neonatal SGA references on cognitive development, temporal and socio-environmental factors are also considered, in both VP/VLBW and term born samples. As research from childhood appears to have found larger differences in IQ than in adulthood, the longitudinal cognitive performance of SGA and AGA individuals is investigated while also considering the environment the infant was born into.

Research Questions :
- **Does the IQ of SGA and AGA participants differ over time, classified using either foetal or neonatal references?**
- **Are the effects of SGA on IQ disproportionately larger for participants born VP/VLBW than at term?**
- **Do the effects of SGA on IQ persist once socio-environmental risk factors are controlled for?**

Chapter 5   **Methodology**

This chapter provides an overview of the cohort studies utilised for the three empirical research chapters to follow. In the first study, data from eight VP/VLBW cohorts were used as part of the RECAP-preterm project, including the Bavarian Longitudinal Study (BLS) and the EPICure study. These two cohorts were then used in order to conduct the research for study 2. Finally, the BLS was solely used in order to complete study 3. As the BLS was used in all 3 studies, it will be the first to have its design, sample and measures described. This will largely focus on the measures and participant assessments relevant for the 3 studies and thus is not exhaustive for all measures assessed in the BLS. The EPICure study will next be described in a similar manner to the BLS. Following this, this section provides a summative overview of the 6 other cohorts that formed study 1 with a particular focus on the aspects in which cohorts differed from one another. Finally, as each research study has its own methods section, a summary of the measures and statistical analyses used will be provided.

5.1   Overview of the Bavarian Longitudinal Study

The Bavarian Longitudinal Study (BLS) started as a whole population sample of neonatal at-risk children born between January 1985 and March 1986 in Southern Bavaria, Germany. The target sample comprised all children who required admission to a children's hospital within the first 10 days after birth (N=7,505), of which 682 were born VP/VLBW with 510 surviving to discharge. Healthy infants who were cared for on normal postnatal wards in the same obstetric hospitals were recruited as controls (N=916;) (Riegel et al., 1995).

The BLS has had follow up assessments at 5 months, 20 months, 56 months, 6 years, 8 years, 13 years and 26 years (Eryigit Madzwamuse et al., 2015). For the 3 research studies, data from the 13 year assessment has not been utilised and as such will not be described further. While the assessments at 5 and 20 months were age corrected for

prematurity, all other assessments have been conducted according to chronological age.

Up to 56 months, the BLS remained a full population birth cohort. Assessments at 5, 20 and 56 months included neurological assessments, parent interviews, cognitive assessments, and observations of behaviour (Riegel et al., 1995).

Subsequent assessments reduced the number of participants in the BLS. From the initial sample, 1,543 children were examined at 6 and 8 years that included all very preterm (<32 weeks gestation, VP) and/or very low birth weight (<1500g, VLBW) and a sub-sample of children born ≥32 weeks gestation (randomly drawn within the stratification factors child sex, family socioeconomic status and degree of neonatal risk), and 344 healthy term control children.

The assessments at 26 years comprised a sample consisting of all reached VP/VLBW and term born control individuals. In each of the 3 research chapters, the number of participants differs depending on data availability and inclusion criteria. The eligible sample at 26 years was 411 VP/VLBW and 308 term born controls (Eryigit Madzwamuse et al., 2015). At most, 260 VP/VLBW participants and 229 control participants took part in some form of follow up at 26 years. However the specific numbers included for each research chapter are presented in their respective method's section. Participants were assessed for one whole day with assessments focused on neuronal, neurocognitive and behavioural development, health utility, and quality of life.

The outcomes of interest for the BLS in this thesis are DQ and IQ assessments measured throughout development and measures of adult attention. DQ assessments were the Griffiths Mental Development Scale at 5 and 20 months of age (Griffiths & Brandt, 1983). IQ was measured at 4 years of age using a composite of the Columbia Mental Maturity Scale (Eggert, 1972), the Active Vocabulary Test (Kiese & Kozielski, 1979), and the Beery-Buktenica Developmental Test of Visual-Motor Integration (Beery & Buktenica, 1982). At 6 years and 8 years, child IQ was again measured using German version of the Kaufmann Assessment Battery for Children (Kaufman &

Kaufman, 1983; Melchers & Preuß, 1991). At 26 years, the Wechsler Adult Intelligence Scale (WAIS III) was used (von Aster et al., 2006). Attention outcomes were measured at 26 years and consisted of self or parent reported ADHD symptoms as determined by Kooij's DSM-IV based ADHD adult rating scale (Kooij et al., 2005) and the tester rating of adult behaviour- attention span (TRAB- AS) (Wolke, 2012).

5.2   Overview of the EPICure study

The EPICure study identified all extremely preterm children (<26 weeks gestation, EP) who were born in the United Kingdom and Ireland between March and December 1995 (Wood et al., 2000). 315 survived to discharge and were followed at ages 2.5, 6, 11 and 19 years (Linsell et al., 2018).

An important differentiation between EPICure and the BLS is the recruitment of controls. While in the BLS they were recruited at birth in the same hospital, EPICure did not recruit controls until age 6. For each EP child in a mainstream school, a classmate matched for age and sex was found to serve as a control, matched on sex and ethnic group as the EP child. At age 6, this resulted 241 EP children and 160 classmates taking part (Marlow et al., 2005). At age 11, 219 EP children and 110 of the controls were reassessed with a further 43 new controls additionally recruited for a total of 153 controls. At age 19, the eligible sample consisted of 306 EP and 153 term born controls. At most, 129 EP adults and 65 controls took part in some form of follow up at 19 years (Linsell et al., 2018). However the specific numbers included for each research chapter are presented in their respective method's section.

The outcomes of interest for EPICure in this thesis were all measured in adulthood (19 years) and were the WAIS II as to test adult IQ, self-reported ADHD symptoms as determined by Kooij's DSM-IV based ADHD adult rating scale (Kooij et al., 2005) and the TRAB- AS (Wolke, 2012).

5.3   Overview of RECAP-preterm and APIC cohorts

This thesis was completed as part of the Research On European Children and Adults Born Preterm (RECAP-preterm) project. RECAP's aim is to improve the health, development, and quality of life of children and adults born very preterm. RECAP looks to combine data from European VP/VLBW cohorts in order identify universal outcomes following VP/VLBW birth. As well as RECAP, cohorts from the Adults Born Preterm International Collaboration (APIC) consortium were also invited to take part. In study 1, IQ data from 8 international VP/VLBW cohorts (including the BLS and EPICure) were brought together in order to undertake an IPD meta-analysis of IQ.

As well as region/country, cohorts differed on a number of other factors including year of birth, inclusion according to prematurity/birthweight, whether the cohorts were regional or national studies, the recruitment of controls, the percentage lost to follow up and the age at assessments in adulthood. A short description of each cohort can be found below. While general cohort differences are shown in Table 3, differences specifically related to the analysis of adult IQ are shown in Table 5 and described in chapter 6.

Arvo-Ylppö Longitudinal Study (AYLS)

The AYLS cohort started in 1985, originally as part of the Bavarian-Finnish Longitudinal Study. Thus the BLS and AYLS started with the same research objective and recruitment strategy. Within the region of Uusimaa, Finalnd, with 7 maternity units and perinatal centres the AYLS recruited an index group of all children who required admission to a children's hospital within the first 10 days after birth from March 1985 to March 1986 (N=1536), of which 108 VP/VLBW infants survived to discharge. In the same way as the BLS, healthy infants who were cared for on normal postnatal wards in the same obstetric hospitals were recruited as controls (N=658). The AYLS also has had follow up assessments at 5 months, 20 months, 56 months (Riegel et al., 1995) and again at 26 years of age (Heinonen et al., 2018). While the BLS later restricted the cohort to VP/VLBW and controls, the AYLS remained a full

birth cohort, including the full gestation range. At age 26 years, 28 VP/VLBW and 303 controls were assessed using the WAIS III (Heinonen et al., 2018)

Helsinki Study of Very Low Birth Weight Adults (HESVA)

The HESVA cohort comprised 335 VLBW infants who were born between 1978 and 1985 and treated in the NICU at the Helsinki University Central Hospital (Hovi et al., 2007). To recruit a control group, for each VLBW infant in the original study cohort the next consecutive and available singleton infant born in the same maternity hospital, of the same sex, gestational age of 37 weeks or above and who was not SGA were recruited retrospectively for the adult assessment. When these VLBW participants were 25 years old, 109 undertook the WAIS III along with 98 controls (Pyhala et al., 2011).

Norwegian University of Science and Technology (NTNU) Low Birth Weight in a Lifetime Perspective Study

The NTNU cohort comprised 121 VLBW infants who were admitted to the neonatal intensive care unit at the University Hospital in Trondheim, Norway, from 1986 to 1988. The term-born control participants were born to women residing in the Trondheim area. Of 5722 eligible participants, a random 10% sample were selected for participation. After removing those born SGA, who died or had congenital malformations there were an initial 459 control infants. At 26 years, 51 VLBW adults completed the WASI along with 75 controls (Laerum et al., 2019).

New Zealand Very Low Birthweight Cohort (NZ_VLBW)

The NZ VLBW cohort comprised all 413 VLBW infants admitted to a neonatal intensive care unit across the whole of New Zealand in 1986. At 28 years, 250 agreed to take part with 225 completing the WASI-II IQ test. At 28 years, a comparison sample of 100 age-matched term-born controls were recruited through a process of peer nomination by a cohort member, or via random sampling from the electoral rolls (n=24), aiming to ensure balance with respect to sex, ethnicity, and regional distribution (Darlow et al., 2015, 2020)

Victorian Infant Collaborative Study (VICS)

The VICS cohort comprised of 568 infants born below 28 weeks of gestation or less than 1000g (EP/ELBW) in the state of Victoria in 1991 and 1992. At birth, healthy infants with a birthweight greater than 2499g were recruited from each of the three tertiary perinatal hospitals in the state as to act as controls. 262 controls were matched to an EP/ELBW participant on expected date of birth, mother's country of birth (English-speaking versus other) and health insurance status (private or public) (Anderson, 2003). At 18 years of age, 224 EP/ELBW participants completed the WAIS II along with 146 controls.

University College London Hospitals (UCLH) VP cohort

A total of 406 infants born before 33 weeks' gestation (range, 24–32 weeks) between 1979 and 1984 were admitted to the neonatal unit of University College Hospital London within 5 days of birth (Kroll et al., 2017). A total of 302 survived, were discharged, and recruited. In adulthood, term-born controls were selected from the local area using community advertisements. Inclusion criteria were full-term birth (38–42 weeks) and birth weight > 2500 grams. At 30 years of age, 104 VP/VLBW adults along with 89 controls were assessed with the WAIS II.

Table 3: Summary of the VP/VLBW cohorts included in Study 1

Cohort	Birth Year	Inclusion Criteria	Regional/National	Age Assessed in Adulthood	Control Group Information
BLS	1985-86	VP/VLBW(<32 wk,1500g)	Regional	26	Recruited in Infancy
AYLS	1985-86	<37 Weeks (~28 VP/VLBW participants)	Regional	26	Recruited in Infancy
EPICure	1995	EP(<26 weeks)	National	19	Recruited at 6 or 11 years
HESVA	1978-85	VLBW(<1500g)	Regional	25	Recruited in Adulthood
NTNU	1986-88	VLBW(<1500g)	Regional	26	Recruited in Infancy
NZ_VLBW	1986	VLBW(<1500g)	National	28	Recruited In Adulthood
VICS	1991-92	EP/ELBW(<28 wk/<1000g)	Regional	18	Recruited in Infancy
UCLH	1979-84	VPT(<33 weeks with oversampling of EP)	Regional	30	Recruited in Adulthood

5.4   Measures

The measures used differed both in regard to outcomes of interest and predictor variables in study 1, 2 and 3. While each study has its own respective method's section that describe the measures used in detail, a summative overview is provided in Table 4.

Table 4: Key measures of interest for this thesis

Measure	Age the measure was collected	Description of Measure	Used as a Predictor or as an Outcome	Used in study 1,2 or 3 and in which cohorts
Gestational Age	Birth	Weeks the foetus has been carried for as estimated by ultrasound estimate, last menstrual period or neonatal examination:	Predictor	1,2,3 - All cohorts
Birth Weight	Birth	Measured in grams as taken from the neonatal records	Predictor	1,2,3 - All cohorts
SGA- Foetal	Birth	Birthweight percentile below the 10th centile in comparison to the	Predictor	3 - BLS

		estimated weight of foetuses at equivalent gestation		
SGA- Neonatal	Birth	Birthweight percentile below the 10th centile in comparison to other neonates of equivalent gestation	Predictor	3 - BLS
Birthweight Z score	Birth	A continuous, relative measure of the weight of infant for their gestation	Predictor	1 - All cohorts
Intraventricular Haemorrhage (IVH)	Birth	Categorical measure of whether bleeding occurred inside or around the ventricles	Predictor	1 - All cohorts
Bronchopulmonary Dysplasia (BPD)	Birth	Categorical measure of long-term breathing problems. Defined by the use of oxygen after a certain time period (post menstrual age or following birth)	Predictor	1 - All cohorts
Parental Socio-economic Status/ Maternal Education	Birth	a weighted composite score of maternal/paternal	Predictor	1,2,3 - All cohorts

		highest educational qualification and occupation of the head of family; grouped as low, middle or high		
Parent-Infant Relationship	Birth and 5 months	Scale derived from concerns regarding the mother and infant or father and infant relationship	Predictor	3 - BLS
Developmental Quotient	5 months and 20 months	Griffiths ordinal measure of early developmental abilities	Outcome	3 - BLS
Childhood IQ	56 months, 6 years, 8 years	IQ tests used to measure full scale child intellectual abilities (test used depends on age of assessment)	Predictor Study 2, Outcome Study 3	2,3 – BLS and EPICure
Adult IQ	18-30 years (depending on cohort)	WAIS/WASI used to measure full scale adult intellectual abilities (test used depends on cohort)	Outcome	1, 3 All cohorts

Adult ADHD Symptoms- Self Reported	19/26 years (EPICure and BLS)	Self-reported scale with two sub scores assessing inattention symptoms and hyperactivity/ impulsivity symptoms. Combined ADHD symptoms is calculated by totalling the two sub scores.	Outcome	2 - BLS and EPICure
Adult ADHD Symptoms- Parent Reported	26 years (BLS)	Parent reported scale with two sub scores assessing inattention symptoms and hyperactivity/ impulsivity symptoms. Combined ADHD symptoms is calculated by totalling the two sub scores.	Outcome	2 - BLS
Tester Rating of Adult Behaviour – Attention span (TRAB -AS)	19/26 years (EPICure and BLS)	Psychologists rated the participant's attention on a scale from very short span to very long	Outcomes	2 - BLS and EPICure

Working Memory	19/26 years (EPICure and BLS)	Letter Number Sequencing task (BLS) or backwards digit recall task (EPICure)	Predictor	2 - BLS and EPICure
Inhibitory Control	19/26 years (EPICure and BLS)	Attention Network Task – Inhibitory Control	Predictor	2 - BLS and EPICure

5.5   Statistical Analyses

Statistical analyses for all three studies were conducted using R (R Core Team, 2013) or SPSS version 24 (IBM Corp., Armonk, NY). For the three studies described in this thesis the following statistical methods were used for the main analyses:

- Study 1: 1 stage IPD meta-analysis using linear mixed models
- Study 2: Hierarchical linear regressions ran separately in each cohort
- Study 3: Longitudinal analysis using linear mixed models

Full details of the analytical approach are given in the respective statistical analyses sections in each research study.

Chapter 6   Intelligence of VP/VLBW individuals in adulthood and its antecedents: an IPD meta-analysis

Abstract

Importance: Birth before 32 weeks of gestation (very preterm; VP) or below 1500g (very low birthweight; VLBW) is associated with lower cognitive performance in childhood. There are few investigations of the associations of neonatal complications and maternal education with VP/VLBW cognitive performance in adulthood.

Objective: To determine differences in intelligence quotient (IQ) between VP/VLBW and term born adults and the association of adult IQ with cohort factors, neonatal complications, and maternal education among VP/VLBW participants.

Data sources: Systematic review of published data from PubMed and an individual participant data (IPD) meta-analysis from cohorts included in two pre-established consortia..

Study selection: Prospective longitudinal cohort studies of adults (mean age >17 years) born VP/VLBW with respective term-born control groups assessing full scale IQ.

Data extraction and synthesis for IPD: Following PRISMA-IPD guidelines, eight of 13 identified cohorts provided data from 1068 VP/VLBW adults and 1067 controls born between 1978 and 1995. IPD meta analyses were performed using a one-stage approach, treating the effects of VP/VLBW and cohort as random effects.

Main Outcomes and Measures: Full scale IQ scores were converted to Z scores within each cohort using the combined standard deviation of both VP/VLBW participants and controls with scores centered on the mean of the controls.

Results: In unadjusted analyses, VP/VLBW adults had mean IQ scores 0.78 standard deviations lower than controls (95% confidence interval -0.90, -0.66),

equivalent to 12 IQ points. Other than birth year, cohort level factors were not related to IQ scores. Among VP/VLBW participants, lower gestational age, lower birthweight Z scores, any intraventricular hemorrhage, bronchopulmonary dysplasia, and lower maternal education were all significantly associated with lower IQ scores.

Conclusions and Relevance:

Being born VP/VLBW is associated with lower IQ in young adulthood. Lower gestational age, lower weight for gestation, neonatal complications, and lower maternal education were all important risk factors for lower IQ among VP/VLBW adults.

Introduction

A key life outcome after being born very preterm (<32 weeks of gestation; VP) or very low birthweight (<1500g; VLBW) is intelligence, defined as 'the capacity to learn from experience, using metacognitive processes to enhance learning and adapt to the surrounding environment' (Sternberg, 2005). Standardized intelligence tests in the general population provide an intelligence quotient (IQ) score with a normative mean of 100 and standard deviation (SD) of 15. IQ is associated with a range of life course outcomes including physical health, premature death, educational attainment and socio-economic success (Basten et al., 2015; Deary et al., 2004; Hegelund et al., 2018; Strenze, 2007). Thus adult IQ is a global indicator of the long-term outcomes of VP/VLBW individuals.

While individual studies typically show that VP/VLBW adults have lower IQ than term-born adults (Eryigit Madzwamuse et al., 2015; Hack, 2006; Linsell et al., 2018), no specific meta-analysis of adult IQ has been published. Meta-analyses of childhood IQ show that VP/VLBW children score on average 11 to 13 IQ points lower than term-born children (0.73 to 0.86 SD) (Allotey et al., 2018; Brydges et al., 2018; Sentenac et al., 2020; Twilhaar et al., 2018). Sustained differences into adulthood cannot be assumed, as evidenced by smaller IQ differences between normal and low birthweight individuals into adulthood (Kormos et al., 2014). Furthermore, considerable variation across VP/VLBW cohorts has been found, potentially explained by factors such as gestation or birthweight inclusion criteria or later selective attrition (Sentenac et al., 2020). Furthermore, variation in IQ may be due to individual level neonatal or demographic factors such as sex, low birthweight for gestational age, neonatal morbidities or maternal education (Benavente-Fernández et al., 2019; Linsell et al., 2015; P. Shah & Kingdom, 2011; Wolke, 2019).

To investigate cohort and individual level factors, individual participant data (IPD) meta-analyses have been proposed as superior to traditional meta-analyses using aggregated data (Lambert et al., 2002). Importantly, this allows for accurate harmonization of data across cohorts and increased statistical power for detecting

individual level risk factors (Tierney et al., 2015). Furthermore, the influence of cohort specific factors, such as rates of attrition, can be investigated.

We performed an IPD meta-analysis on IQ scores in adulthood with three objectives. First, to compare the difference in adult IQ for VP/VLBW individuals with term-born controls. Second, to determine the cohort or individual level factors associated with IQ scores among VP/VLBW adults. Third, to conduct a sensitivity analysis to assess if IQ scores from the IPD cohorts are representative of all VP/VLBW adult cohorts.

Methods

Protocol and registration

This study is part of the Research on European Children and Adults Born Preterm (RECAP preterm) consortium (https://recap-preterm.eu/). As well as seven adult RECAP cohorts, six non-European cohorts participating in the Adults Born Preterm International Collaboration (APIC) consortium (https://www.apic-preterm.org) were invited to take part. All cohorts had received country specific ethical review, with participants providing written, informed consent. All cohorts adhered to the Helsinki Declaration on ethical principles for research involving human subjects.

This IPD meta-analysis is registered with PROSPERO (CRD42020162043), and reporting is consistent with the Preferred Reporting Items for Systematic Review and Meta-analyses of Individual Participant Data.

Eligibility criteria

Prospective longitudinal cohorts of VP/VLBW adults who had undertaken a standardized IQ test at a mean age at assessment greater than 17 years were eligible to participate. Additionally, all cohorts required a term-born control group to compute harmonized and comparable IQ Z-scores, minimizing bias due to test or secular trends (Flynn effect) (Trahan et al., 2014).

Identifying additional studies: Information sources and search strategy

To assess whether the cohorts from RECAP/APIC were representative of all VP/VLBW adult cohorts included, we undertook a PubMed search. This sensitivity analysis used the same eligibility criteria as previously described with the search terms (Intelligence OR IQ OR Cognition OR Cognitive) AND (Adult OR Adulthood OR Late Adolescence) AND (Preterm OR Gestation OR Birthweight OR Birth Weight). The date of the last search was 07/09/2020.

Study selection processes

Eligibility for inclusion was assessed by RE and YN. Any disagreements regarding eligibility were resolved by discussion. Of seven RECAP preterm adult cohorts, two were excluded: the ESTER study (Suikkanen et al., 2020), which did not perform a full scale IQ test and POPS (Weisglas-Kuperus et al., 2008), which had no control group. Of six potential APIC cohorts, the McMaster (Saigal et al., 2016) and Melbourne (Doyle et al., 2003) studies had not assessed adult IQ while the Cleveland study (Hack, 2006) only had summary data available.

Data collection processes and data items

Following initial data scoping, encrypted data from each cohort was transferred to the University of Warwick. Data including IQ scores, neonatal variables, maternal education, neurosensory impairment (NSI) in childhood and attrition rates were collected for all cohorts. All data were only accessible to authorized personnel from RECAP preterm.

Data Harmonization

To harmonize IQ, scores were converted to Z scores within each cohort using the combined SD of both VP/VLBW participants and controls with scores centered on the mean of the controls. Neonatal data included gestational age, sex, birthweight, bronchopulmonary dysplasia (BPD), intraventricular hemorrhage (IVH), and multiple birth. BPD definitions varied with cohorts using the criteria of oxygen dependency at 36 weeks of postmenstrual age or for more than 28 days after birth. For each definition, separate sub-analyses determined each criteria's independent association with IQ. IVH was classified according to Papile et al (Papile et al., 1978),

but some cohorts provided either Grades 3 and 4 or Grades 2 and 3 combined. Thus, IVH was harmonized into 'no IVH' versus 'any IVH' (Grades 1-4), and a sub-analysis compared 'No IVH or IVH Grades 1-2' versus 'IVH Grades 3-4' in cohorts where this was possible. Multiple birth was classified as a binary variable (0 = singleton, 1= multiple). Birthweight Z scores were determined using Fenton's international reference (Fenton & Kim, 2013).

Maternal education was harmonized according to the International Standard Classification of Education (ISCED) into low (ISCED level 0-2), medium (3-5), and high (6-8) (UNESCO Institute for Statistics, 2012). Evidence of childhood neurosensory impairment (NSI) was collated from evidence of severe visual impairment (blind in both eyes), hearing impairment (uncorrected by hearing aids), non-ambulatory cerebral palsy, or child cognitive impairment (childhood IQ <70). If data were missing for a participant on a certain NSI variable, it was treated as no evidence for that impairment. Evidence of NSI was combined into a binary childhood NSI variable (any evidence of impairment versus no evidence of impairment) (Appendix 1). Additionally, four cohort level factors were determined. For each cohort, the percentage of eligible VP/VLBW adult participants who did not have adult IQ scores was calculated (VP/VLBW attrition %). To focus on selective attrition, the percentage of VP/VLBW participants with prior diagnoses of childhood NSI who did not have adult IQ scores was also calculated (VP/VLBW with NSI attrition %). Data on the mean age at assessment for the VP/VLBW participants in each cohort and their year of birth were also collated.

IPD integrity and risk of bias assessment in individual studies:

Data were checked for consistency with prior publications from the included cohorts, with any discrepancies resolved by communication with investigators from the cohorts. RE and YN assessed cohort quality and comparability using the Newcastle-Ottawa scale (see Appendix 2) (Wells et al., 2012).

Specification of outcomes and effect measures:

The primary outcome of interest was the full-scale IQ Z score of VP/VLBW adults compared with term-born controls.

Synthesis methods

All participants with adult IQ scores were included. Missing neonatal data were imputed solely for VP/VLBW participants and missing maternal education data imputed for VP/VLBW and controls using multiple imputation by chained equations (MICE) (Buuren & Groothuis-Oudshoorn, 2011), resulting in <5% of the data being imputed (Table 5). In the first analysis, a simple comparison of IQ scores between VP/VLBW and controls was conducted using a one-stage linear mixed model. We analyzed the effects of VP/VLBW on IQ using a random intercepts model for each cohort and a random slope for the effect of VP/VLBW on IQ by cohort, estimated using maximum likelihood in the R package lme4 (Bates et al., 2015). Subsequently, the association between VP/VLBW and IQ was considered after adjusting for sex and maternal education, after removing VP/VLBW participants with childhood NSI, or differentiating between VP/VLBW with or without neonatal morbidities (IVH or BPD). All analyses again used a one-stage approach with random intercepts and slopes.

Additional analyses

To explore antecedents of IQ scores among VP/VLBW participants, a one-stage IPD analysis was performed. Cohort factors (age at assessment, birth year, VP/VLBW attrition (%), VP/VLBW with NSI attrition (%)) were added as fixed effects. Individual level neonatal factors and maternal education were then added as fixed effects. Beta estimates from all factors are reported from both univariable and multivariable analyses to determine their independent and combined associations. Statistically significant associations were determined by P values <.05.

Finally, a sensitivity meta-analysis using aggregate data was performed combining all cohorts used in the IPD analyses and summary data from the Cleveland study and additional cohorts identified through the PubMed search for whom IPD were not

requested. The standardized mean differences (SMD) in IQ between VP/VLBW adults and controls in each cohort were pooled using a random effects meta-analysis using the R package meta (Balduzzi et al., 2019). Heterogeneity across cohorts was assessed using Cochran's Q value and I^2 while a subgroup analysis differentiated between IPD and non-IPD cohorts as to test for selection bias.

Results

<u>Study Selection and IPD Obtained</u>

Eight (Darlow et al., 2020; Doyle et al., 2015; Eryigit Madzwamuse et al., 2015; Heinonen et al., 2018; Kroll et al., 2017; Lærum et al., 2019; Linsell et al., 2018; R. Pyhälä et al., 2011) of nine RECAP/APIC cohorts with adult IQ data contributed 1068 VP/VLBW participants and 1067 controls to the IPD meta-analysis. Summary data for each cohort included in the IPD analyses are detailed in Table 5.

Table 5: Summary of cohorts included in the IPD analysis

Cohort	Birth Year	Country	Mean age Assessed	IQ test	Initial Eligibility Criteria	Initial VP/VLBW Surviving to Discharge (N)	Eligible Adult VP/VLBW (N)	VP/VLBW Attrition %	VP/VLBW+ NSI Attrition %	VP/VLBW with IQ scores (N)	Control Group (N) + Information	Harmonization Issues	VP/VLBW Neonatal / Maternal Education Data Imputed %
AYLS (Heinonen et al., 2018)	1985-1986	Finland	26	WAIS III (1997)	Preterm <37 Weeks (reduced to VP/VLBW for this analysis)	108	68	59	80	28	303 - Recruited Infancy	None	0%
BLS (Eryigit Madzwa	1985-1986	Germany	26	WAIS III (1997)	VP/VLBW (<32	510	411	51	76	203	192- Recruited Infancy	None	<1%

muse et al., 2015)					weeks/ <1500 g)							
EPICure (Linsell et al., 2018)	1995	UK & Ireland	19	WAIS II (1981)	EP (<26 weeks)	315	306	59	69	124	64- Recruited at ages 6 or 11	None	1%
HESVA (R. Pyhälä et al., 2011)	1978-1985	Finland	25	WAIS III (1997)	VLBW (<1500 g)	334	254	57	69	109	98-Recruited in Adulthood	Maternal education measured in adulthood. NSI did not include IQ <70 and could not differentiate ambulatory/non-

											ambulatory cerebral palsy .	
NTNU (Lærum et al., 2019)	1986-1988	Norway	26	WASI (1999)	VLBW (<1500 g)	86	82	37	60	51	75 -Recruited Infancy	Maternal education measured at 14 years	6%
NZ VLBW (Darlow et al., 2020)	1986	New Zealand	28	WASI-II (2011)	VLBW (<1500 g)	338	323	30	64	225	100- Recruited in Adulthood	None	1%

UCLH (Kroll et al., 2017)	1979-1984	UK	30	WAIS II (1981)	VP(<33 weeks, reduced to VP/VLBW for this analysis)	302[A]	220[A]	46	85	104	89- Recruited in Adulthood	BPD was not available so was fully imputed. Maternal education reported by the participant in adulthood. NSI solely based on IQ <70 at 8 years	21%
VICS (Doyle et al., 2015	1991-1992	Australia	18	WAIS II (1981)	EP/ELBW (<28 weeks/ <1000 g)	299	277	19	37	224	146- Recruited Infancy	None	6%

Footnotes: Very preterm/Very Low Birthweight (VP/VLBW), Extremely preterm/Extremely Low Birthweight (EP/ELBW), Wechsler Adult Intelligence Scale (WAIS), Intelligence Quotient (IQ), Neurosensory Impairment (NSI), bronchopulmonary dysplasia (BPD).

[A]Cohort information regarding attrition data, eligible adult sample and initial sample from UCLH is based on the criteria <33 weeks' gestation rather the VP/VLBW (<32 weeks or <1500 g) criteria imposed subsequently. While 122 preterm individuals (<33 weeks) took part in adulthood, only 104 were VP/VLBW and included in this analysis.

IPD Study Participant Characteristics

The eight IPD cohorts were from seven high-income countries, six were regional and two were national. The mean age at IQ assessment across cohorts ranged from 18 to 30 years and birth year ranged from 1978 to 1995. Increasing birth year (i.e. being from a more recent cohort) was associated with lower birthweight and earlier gestational age (Appendices 3 and 4). In total, 48% and 43% of VP/VLBW and control participants were male, respectively. See Appendices 5 and 6 for further information on VP/VLBW and control participants.

IPD integrity and risk of bias within studies

IPD cohorts were rated for quality according to the Newcastle-Ottawa scale (range 0-9). The mean was 7.9 with cohorts rated highly in regard to representativeness, ascertainment of exposure and assessment of outcome. However, studies differed in inclusion criteria and attrition rates among VP/VLBW participants, which were above 50% in 4/8 cohorts (Appendix 2).

Results of syntheses

One Stage IPD meta-analysis with all participants

Including all participants from IPD cohorts, VP/VLBW mean IQ scores were 0.78 SD lower than those of controls (95% CI -0.90, -0.66). When sex and maternal education were included, the estimate of VP/VLBW on IQ reduced minimally from -0.78 to -0.74 SD (95% CI -0.85, -0.63). Excluding participants with childhood NSI reduced the IQ difference from -0.78 to -0.65 SD (95% CI -0.76, -0.55). Regarding neonatal morbidities, being born VP/VLBW and with any IVH grade was associated with a larger difference in IQ than without (-0.99 SD, 95% CI -1.19, -0.79 vs -0.70 SD, 95% CI -0.84, -0.57). A similar difference for VP/VLBW individuals with or without BPD was also found (-0.93 SD, 95% CI -1.10, -0.76 vs -0.67 SD, 95% CI -0.80, -0.55)

IPD analysis determining antecedent risk factors among VP/VLBW participants

Table 6 shows the results of the IPD meta-analysis determining the association of neonatal factors, maternal education, and cohort factors with IQ among VP/VLBW participants. Associations that were significant in the multivariable analysis were gestational age, birthweight Z score, BPD, any IVH, and maternal education (Table 6). For example, among VP/VLBW participants, each extra week of gestation was associated with IQ scores increasing by 0.11 SD (1.7 IQ points). In contrast, neither sex nor single or multiple birth significantly altered adult IQ for VP/VLBW participants. Additionally, none of the cohort level factors were found to be significant in the multivariable analysis. Birth year was significant in the univariable analysis, which suggested cohorts born more recently had lower IQ scores in adulthood.

Table 6: Very Preterm/Very Low Birth Weight (VP/VLBW) analysis: univariable and multivariable associations of individual and cohort level factors with IQ Z scores

VP/VLBW only analysis, N=1068

			IQ Z Scores			
	1 Stage Univariable Estimate			*1 Stage Multivariable Estimate*		
Factors	*Estimates*	*CI*	*p*	*Estimate*	*CI*	*p*
Individual Level Factors						
Gestational age (weeks)	0.04	0.02, 0.06	**<0.001**	0.11	0.07, 0.14	**<0.001**
Male	0.07	-0.05, 0.20	0.26	0.09	-0.03, 0.20	0.16
Birthweight_Z score (per 1 SD)	0.05	-0.01, 0.11	0.08	0.21	0.14, 0.28	**<0.001**
Maternal education (1=Low, 2=Medium, 3= High)	0.25	0.17, 0.34	**<0.001**	0.26	0.17, 0.35	**<0.001**
Bronchopulmonary dysplasia (Reference: No BPD)[A]	-0.37	-0.51, -0.23	**<0.001**	-0.16	-0.30, -0.02	**0.02**
-Defined as oxygen after 28 days post birth[B]	-0.34	-0.56, -0.12	**0.003**	-	-	-

-Defined as oxygen after 36 weeks' postmenstrual age[B]	-0.40	-0.56, -0.23	**<0.001**	-	-	-
Any grade of intraventricular hemorrhage (Reference: No IVH)	-0.27	-0.40, -0.13	**<0.001**	-0.19	-0.33, -0.05	**0.007**
- IVH Grade 3 or 4 (Reference: All other grades)[C]	-0.66	-0.92, -0.41	**<0.001**	**-**	**-**	**-**
Multiple Birth (Reference: Singleton)	0.01	-0.13, 0.15	0.86	0.00	-0.13, 0.14	0.95
Cohort Level Factors						
VP/VLBW attrition %.	-0.00	-0.01, 0.00	0.08	-0.02	-0.06, 0.02	0.32
VP/VLBW with NSI attrition %	-0.00	-0.01, 0.00	0.25	0.02	-0.03, 0.08	0.39
Cohort age at IQ assessment	0.00	-0.01, 0.02	0.60	-0.10	-0.26, 0.06	0.28
Year of birth	-0.02	-0.03, -0.00	**0.02**	-0.03	-0.11, 0.05	0.48

Footnotes: Very preterm/Very Low Birthweight (VP/VLBW), Intelligence Quotient (IQ), Neurosensory Impairment (NSI), bronchopulmonary dysplasia (BPD), intraventricular hemorrhage (IVH).

^A Participants from the UCLH cohort were not included in the univariable estimate but had their BPD values imputed for the multivariable estimate

^B The AYLS, BLS, HESVA and NTNU used the criteria of 28 days post birth while EPICure, NZ VLBW and VICS used the criteria of 36 weeks' postmenstrual age.

^C NZ VLBW participants could not have IVH harmonized into Grade 3 or 4 and thus were not included for the sub-analysis

[A] Participants from the UCLH cohort were not included in the univariable estimate but had their BPD values imputed for the multivariable estimate

[B] The AYLS, BLS, HESVA and NTNU used the criteria of 28 days post birth while EPICure, NZ VLBW and VICS used the criteria of 36 weeks' postmenstrual age.

[C] NZ VLBW participants could not have IVH harmonized into Grade 3 or 4 and thus were not included for the sub-analysis

Sensitivity meta-analysis comparing IPD and Non-IPD VP/VLBW adult cohorts

The PubMed search of 413 records identified an additional four (Constable et al., 2013; Hallin et al., 2010; Lefebvre et al., 2005; Stålnacke et al., 2015) non RECAP/APIC cohorts with extractable IQ data. Adding the summary data from the Cleveland study (Hack et al., 2002) meant there were five non-IPD cohorts and eight IPD cohorts (Figure 5). Characteristics of the non-IPD cohorts are shown in Appendix 7.

Using aggregate data, the SMD between VP/VLBW and controls for non-IPD cohorts was -0.61 (95% CI -0.93, -0.29), and for IPD cohorts it was -0.84 (95% CI -0.97, -0.71) (Figure 6). According to Cochran's Q test, this suggested no significant differences between IPD and non-IPD cohorts (Q = 1.80, P = 0.18). However, the heterogeneity was larger among non-IPD cohorts ($I^2 = 75\%$) than within IPD cohorts ($I^2 = 41\%$)
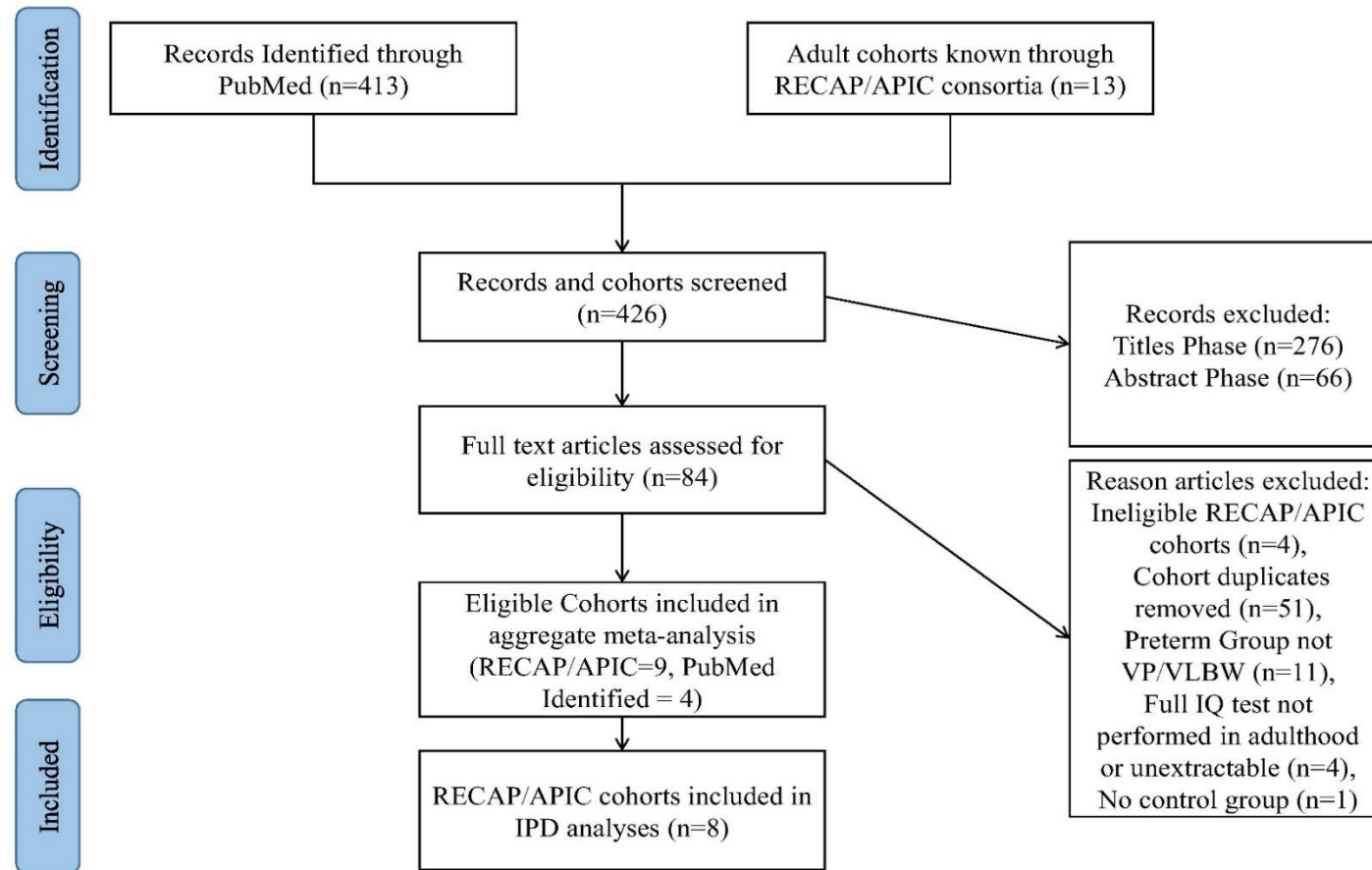
Figure 5: Flow chart of studies included in the IPD and aggregate meta-analyses
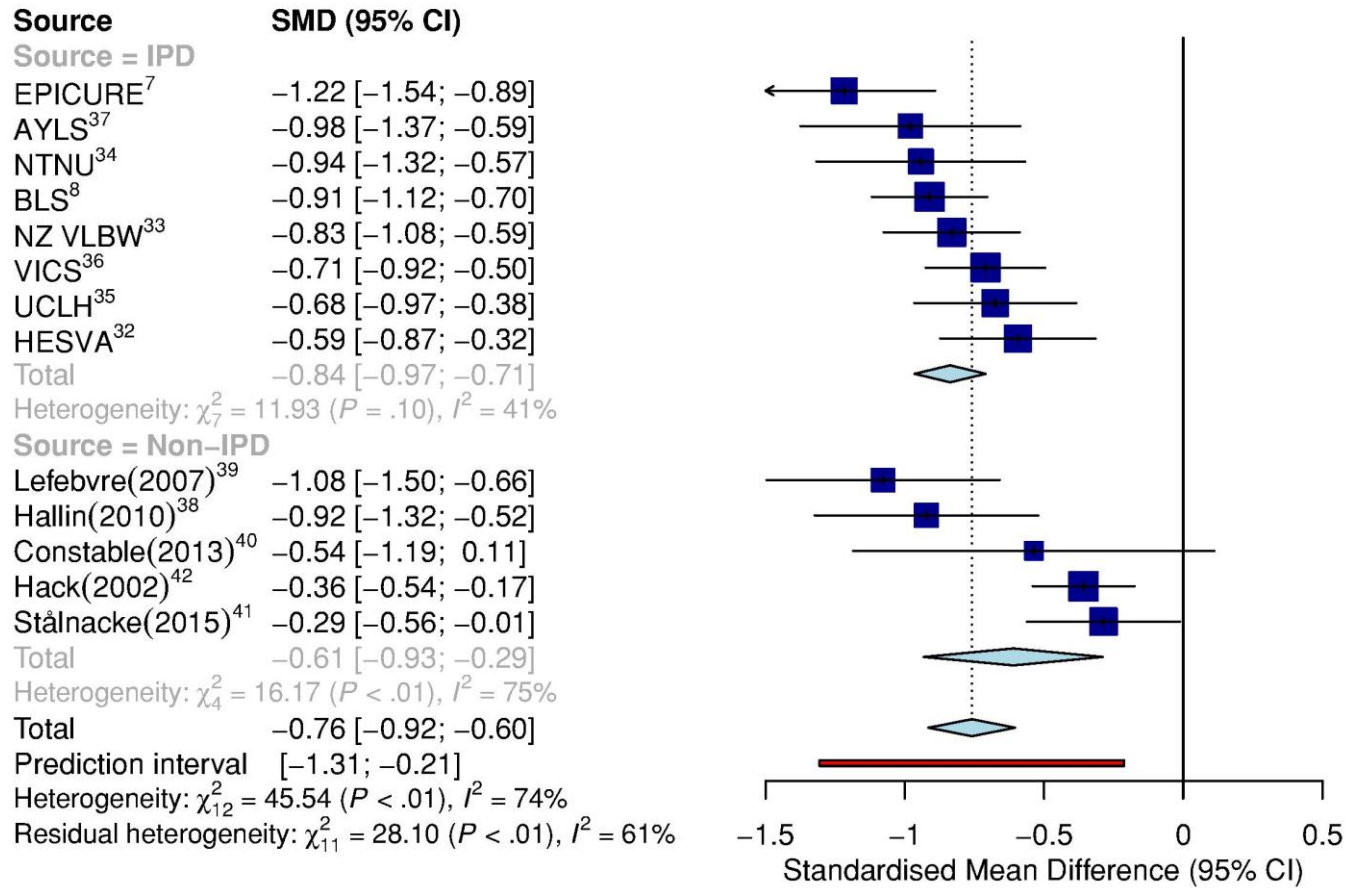
Figure 6: Aggregate meta-analysis comparing IQ performance in IPD and Non-IPD VP/VLBW adult cohorts

Discussion

Summary of evidence

This study examined the relationship between VP/VLBW birth and adult IQ. Among eight cohorts contributing IPD, VP/VLBW participants scored 0.78 SD lower than controls, equivalent to a between group difference of approximately 12 IQ points. This robust difference was marginally reduced after adjustment for sex and maternal education. Additionally, even when participants with prior classification of NSI in childhood were excluded, which removed those with low childhood IQ, the difference between VP/VLBW participants and controls was still 0.65 SD (9.8 IQ points). Adding further cohorts for whom IPD was not available did not alter findings significantly. Among the VP/VLBW participants, individual level factors associated with lower IQ were earlier gestational age, lower birthweight Z score, BPD, IVH, and being born to mothers with lower education.

The IQ difference of -0.78 SD between VP/VLBW and term-born controls is a larger standardized difference than reported for other functional outcomes ranging from mental to physical health and social functioning (Wolke et al., 2019). The IQ differences between VP/VLBW and term-born adults are also similar to those previously reported in childhood (Allotey et al., 2018; Brydges et al., 2018; Sentenac et al., 2020; Twilhaar et al., 2018). Three prospective studies previously reported that there is moderate to high stability of IQ-scores from childhood to adulthood for VP/VLBW individuals (Breeman et al., 2015; Darlow et al., 2020; Linsell et al., 2018). The adult cohorts in this IPD were followed for decades, resulting in a higher risk of selective attrition over such a long period. Individuals who are more socially disadvantaged or with NSI have been found to be lost to follow-up more often (Wolke et al., 1995) which may lead to a smaller difference in IQ between VP/VLBW participants and controls (Howe et al., 2013). However, the cohort differences in rates of attrition among all VP/VLBW adults or those specifically with childhood NSI were not associated with IQ, nor was the age at IQ assessment. Furthermore, the association between IQ and birth year was not significant after including individual level factors in the multivariable analysis. This suggests little

change in mean adult IQ scores for VP/VLBW individuals born between 1978 and 1995, after adjusting for the fact that individuals born more recently had on average lower birthweights and gestational ages. In more recent VP/VLBW cohorts, no improvements in childhood IQ by birth year have been reported (Twilhaar et al., 2018). Nor has IQ improvement been seen in successive extremely preterm child cohorts from the same regions (Cheong et al., 2017; Marlow et al., 2021). Thus, given the stability of IQ over time among these cohorts; changes in neonatal care and reduced mortality (Glass et al., 2015) do not appear to have translated into long term improvement in IQ for VP/VLBW individuals over this time frame.

For individual level factors (Table 6), even when adjusting for other neonatal factors and maternal education, there was a dose-response between gestational age and IQ, replicating individual study findings (Kroll et al., 2019; Wolke, Strauss, et al., 2015). That IVH and BPD are associated with lower IQ is also in concordance with prior meta-analyses in childhood and individual studies in adulthood (Breeman et al., 2017; Twilhaar et al., 2018). After controlling for other factors, BPD was associated with an IQ reduction of 2.4 IQ points (-0.16 SD) and presence of any IVH by 2.9 IQ points (-0.19 SD) (Table 6). Severe IVH (grade 3-4) was only investigated in a univariable analysis, not including all cohorts, but was associated with nearly a 10 point IQ deficit (-0.66 SD). However, this needs to be interpreted cautiously as this association is uncontrolled for other neonatal factors and maternal education. Birthweight Z scores showed a strong association with IQ after controlling for other factors. The multivariable model suggests that being born with a birth weight -2 SD for gestational age is associated with a 6 point IQ deficit (i.e. -2*0.21 SD= -0.42 SD) compared with being born at appropriate weight for gestation. This adds further evidence that being born small for gestational age is associated with lower IQ among VP/VLBW adults (Eves et al., 2020).

In contrast to neonatal factors, maternal education and similar factors have been largely overlooked in research on VP/VLBW and outcomes (Wolke, 2019). Compared with low maternal education, VP/VLBW adults with medium or highly educated mothers had IQ scores 0.26 SD and 0.52 SD (3.9 and 7.8 IQ points) higher on average, respectively. These associations are equivalent in magnitude to those

related to serious neonatal complications (e.g. BPD, IVH). The association of maternal education with adult IQ may reflect an amalgam of different factors. These may include genetic effects (Torres, 2013), maternal smoking (Rahu et al., 2010), breastfeeding rates (Brion et al., 2011), and parental behaviors (Breeman et al., 2017). Some could be modified postnatally and have been shown to have an impact on academic achievement and development in the general population and for VP/VLBW groups (R. Shah et al., 2016; Wolke et al., 2013).

Strengths and limitations

Strengths of the study are the harmonization and use of IPD regarding neonatal factors and maternal education, childhood NSI and IQ for eight VP/VLBW adult cohorts allowing for reliable comparison across cohorts. In addition, we tested a range of specific cohort factors which is challenging to do in an aggregate meta-analysis as these details are rarely available from published studies.

Limitations were cohort's differences regarding eligibility criteria, such as EPICure's stricter inclusion criterion of <26 weeks' gestational age, the use of maternal education rather than broader factors such as socioeconomic status or combined parental education and the different methods used for recruiting controls. Controls were typically recruited in infancy but in some cohorts this happened in childhood or adulthood where neonatal data were unavailable. Thus, we could not determine whether factors such as birthweight Z scores are similarly associated with IQ for term-born participants. Finally, as the mean age at assessment ranged from 18 to 30 years, these findings reflect IQ in young adulthood only.

Conclusions and Implications

VP/VLBW adults have IQ scores, on average, 12 points lower than term-born controls; similar to results reported in meta-analyses of IQ in childhood despite the greater risk of selective attrition due to a longer follow up period (Allotey et al., 2018; Breeman et al., 2015; Brydges et al., 2018; Linsell et al., 2018; Sentenac et al., 2020; Twilhaar et al., 2018). Antenatal and neonatal care that can reduce BPD and IVH (Gien & Kinsella, 2011; Taylor et al., 2013) and parenting or educational

interventions that can reduce the social disparities associated with maternal education (R. Shah et al., 2016) may improve intellectual outcomes in VP/VLBW adults.

Chapter 7   The Role of Executive and General Cognitive Functioning in the Attention Problems of Very and Extremely Preterm Adults

Abstract

Objective—To determine whether the attention problems in adults born very preterm/very low birthweight (VP/VLBW; <32 weeks' gestation/ <1500g) or extremely preterm (EP; <26 weeks' gestation) are associated with specific executive or general cognitive deficits.

Method— Cohorts of VP/VLBW (the Bavarian longitudinal study (BLS)) and EP (the EPICure Study) participants were followed from birth to early adulthood, each also following a respective control group. Adult ADHD symptoms were assessed via self-report in both cohorts and additionally by parent-report in the BLS. Participants in both cohorts also had their attention span rated by trained observers. Performed separately in each cohort, hierarchical regression analyses were used to assess whether the association between preterm birth status and attention problems remained after accounting for executive functioning (inhibitory control and working memory) in adulthood, childhood IQ or sex.

Results— In the discovery cohort of the BLS, significant differences were found between VP/VLBW adults and controls for parent-rated inattention ($p<0.001$). However, for self-reported measures of ADHD, no significant differences were found in the BLS or in the EPICure replication cohort. In both cohorts, observer-rated attention spans were lower for VP/VLBW and EP participants in comparison to their respective control groups ($p <0.001$). In final models for the BLS, inhibitory control and childhood IQ were significantly associated with parent-rated inattention symptoms (p<0.006). Whereas working memory and childhood IQ were significantly associated with observer-rated attention span ($p<0.001$). The effect of childhood IQ on observer-rated attention span was replicated in EPICure.

Conclusions—VP/VLBW and EP adults are at increased risk of observer-rated attention problems. These problems were predominantly associated with poorer

general cognitive ability in early childhood and somewhat with adult executive functioning.

Introduction

In comparison to term born controls, those born very preterm or at very low birthweight (<32 weeks' gestation or <1500g, VP/VLBW) have been found to have greater attention problems (Johnson & Wolke, 2017). In childhood, this has been found when assessed via parent report (Johnson et al., 2016), teacher rating (Nadeau et al., 2001) and observer rating of attention span (Breeman et al., 2016). VP/VLBW individuals are also at increased risk of Attention Deficit Hyperactivity Disorder (ADHD) diagnosis in childhood (Johnson & Wolke, 2017) and adulthood (Breeman et al., 2016). In particular, a preterm specific phenotype of ADHD, consisting of increased number of inattention symptoms (ADHD-I) with relatively few problems of hyperactivity/impulsivity (ADHD-H) (Johnson et al., 2016) has been proposed. While males are more likely to have ADHD symptoms or diagnosis in the general population, this sex difference has not been consistently found within VP/VLBW groups (Johnson & Wolke, 2017).

Attention problems have been primarily associated with deficits in executive functioning, a set of higher-order neurocognitive processes required for decision making and goal orienting (Willcutt et al., 2005). While there is discussion over which behaviours and tasks best measure executive functioning, Diamond's (2013) framework states that two main components are the ability to hold and manipulate information in mind - working memory - and the ability to selectively attend and suppress attention to stimuli - inhibitory control (Diamond, 2013). In comparison to controls, VP/VLBW children and adolescents show deficits on a range of executive functioning tasks (Burnett et al., 2013), which may explain the attention problems seen in VP/VLBW children. For example, working memory has been found to mediate the relationship between VP/VLBW birth and teacher-rated inattention (Nadeau et al., 2001). Similarly, impulse control, a component of inhibitory control, has been associated with attention scores in VP/VLBW children and controls (Aarnoudse-Moens et al., 2013). Thus, the greater childhood attention problems seen in VP/VLBW when compared to term born may be partly explained by executive functioning.

However, whether these specific executive functions explain differences in adulthood has not yet been explored.

Alternatively, it has been suggested that the differences in attention between VP/VLBW individuals and term born controls may be explained by VP/VLBW individuals having, on average, lower intelligence scores (IQ) (Johnson et al., 2016). However, scores on tests of IQ and executive function are correlated with poor executive functioning being partially responsible for poor IQ scores (Engelhardt et al., 2016). This is especially true for adult IQ tests that have working memory as a subtest for the calculation of full-scale IQ, meaning the two constructs are not independent. To reduce this issue, childhood IQ can be used to control for general cognitive ability while being less correlated with current abilities in executive function. Overall, if adult inattention is primarily a result of specifically poor executive function, then concurrent measures of executive function should provide the best ability to explain differences in attention between groups, over and above the effect of childhood IQ scores.

The aim of this study was to investigate whether the greater attention problems seen in VP/VLBW as compared to term born adults are best explained by specific executive functioning deficits, general cognitive abilities or sex. The discovery sample is the Bavarian Longitudinal Study (BLS) and replication was conducted in the EPICure study of extremely preterm participants (EP, <26 weeks' gestation). It was hypothesised that the poorer attention seen in VP/VLBW and EP adults would be significantly associated with poor executive functioning, as measured by inhibitory control and working memory, and that these effects would remain after controlling for other potential risk factors of low childhood IQ and male sex.

Method

Participants Bavarian Longitudinal Study (BLS)

Details of the design of the BLS have been previously reported (Wolke, Schmid, et al., 2009), as have the details of the assessments at 26 years of age (Eryigit Madzwamuse et al., 2015). Briefly, of 682 VP/VLBW infants born alive between January 1985 and March 1986 in Southern Bavaria, Germany, and who required

admission to a children's hospital within the first 10 days after birth, 411 were alive and eligible for the 26-year follow-up assessment. 260 participated (63%) with 194 (47%) completing measures of self-reported ADHD and experimental measures of executive functioning. Three hundred and fifty eligible healthy term-born controls born in the same hospitals, matched for sex and socioeconomic status, served as controls and were also followed from birth. In adulthood, 308 controls were eligible for inclusion, 229 (74%) participated with 197 (64%) completing self-reported ADHD and executive functioning measures at 26 years and are thus included in this study. Of the 194 VP/VLBW participants and 197 controls, 172 (89%) and 181 (93%) also had data available for parent-reported ADHD symptoms at 26 years of age. The participant flow chart for the BLS is presented in Appendix 8. Informed consent was obtained from parents and participants, ethical approval was obtained from University Hospital Bonn Ethical Committee.

EPICure

Details of the design of EPICure have been previously reported (Costeloe et al., 2000) as have the details of the assessments at 19 years of age (Linsell et al., 2018). Briefly, EPICure included EP infants who were born in the United Kingdom and Ireland from March through to December 1995. Of the 315 alive at hospital discharge, 306 EP participants were eligible for the 19-year follow-up assessment of which 129 (42%) participated. Of these, 107 (35%) completed measures of self-reported ADHD symptoms and tests of executive functioning. A stratified comparison group of 160 children were initially recruited at age 6 with 43 further recruited at 11 years. Of the full-term control group at 11 years (N: 153), 65 (42%) took part at 19 years of age, with 60 (39%) completing measures of self-reported ADHD symptoms and tests of executive functioning. The participant flow chart for EPICure is presented in Appendix 8. Informed consent was obtained from participants, ethical approval was obtained from the South Central – Hampshire A Research Ethics Committee.

Measures

Adult ADHD Symptoms

Both EPICure and BLS participants completed Kooij's DSM-IV based ADHD adult rating scale (Kooij et al., 2005). This 23 item scale is considered a valid and reliable measure of ADHD in adulthood (Kooij et al., 2005). The scale determines a participant as having a symptom if the participant responds 'often' or 'very often' to items such as 'I fail to give close attention to details in work'. Two subscores assessing 9 ADHD-I symptoms and 9 ADHD-H symptoms, ranging from 0 (no ADHD sub score symptoms present) to 9 (maximum number of ADHD sub score symptoms present) are calculated with the combined ADHD symptoms (ADHD-C) calculated by totalling the two sub scores. In both cohorts, the self-reported ADHD scales had good internal reliability (BLS α= 0.75, EPICure α= 0.85). In the BLS cohort only, parents also assessed their child's ADHD symptoms using the same questionnaire, with a similarly good internal reliability (α= 0.88). All ADHD-I, ADHD-H and ADHD-C symptom scores were then converted into Z scores based upon the mean and standard deviation of each cohort's respective control group.

Tester Rating of Adult Behaviour - Attention Span (TRAB-AS)

In both cohorts, psychologists rated the individual's attention on a scale from 1 (very short attention span) to 9 (very long attention span) (Wolke, 2012). Assessments were made three times across the assessment day: (1) during the cognitive assessment, (2) during the afternoon session, and (3) at the end of the assessment day. The means of these three time points were then combined to produce an overall assessment of attention span which were then converted into Z scores based upon the mean and standard deviation of each cohort's respective control group. Within the BLS, Tester Rating of Adult Behaviour - Attention Span (TRAB-AS) showed moderate inter-rater reliability (Kappa=0.67). For EPICure, all assessments were made by a single psychologist.

Adult Executive Functioning: Inhibitory control

Inhibitory control was measured using the Attention Network Task (ANT) (Fan et al., 2002). The ANT measures alerting, orienting and executive control. For this study, executive control was of interest as a measure of inhibitory control. Consisting of 128 trials, the ANT requires participants to determine the direction of a central target arrow as accurately and as quickly as possible while ignoring flanker arrows. Inhibitory control was calculated by taking the mean reaction time on trials when the flanker arrows were incongruent and subtracting the mean reaction time when the flanker arrows were congruent. Scores were measured in milliseconds with a larger inhibitory control score indicating greater difficulty with inhibiting extraneous stimuli. See Appendix 9 for a diagram demonstrating the sequence of events in an ANT trial and a detailed description of how the ANT was performed in both cohorts using identical procedure.

Adult Executive Functioning: Working Memory

For BLS participants, the working memory assessment comprised a Letter-Number Sequencing task, a subtest of Wechsler Adult Intelligence Scale III (von Aster et al., 2006). Participants heard sequences of numbers and letters and then repeated back the numbers in ascending order and the letters in alphabetical order. EPICure participants partook in a different verbal working memory assessment, the backwards digit recall task a subtest of Wechsler Adult Intelligence Scale IV (Wechsler, 2008). Participants listened to sequences of numbers and then repeated them back in reverse order, a working memory assessment found to be closely related to the Letter-Number Sequencing task (Crowe, 2000). Scores in both cohorts were standardised based upon each cohort's respective control group with a mean of 100 and a standard deviation of 15.

Childhood IQ

At 6 years of age, the IQ of participants was assessed with the Kaufman Assessment Battery for Children Mental Processing Component, comprising of 8 subtests, 5 subtests to measure simultaneous processing and 3 subtests to sequential processing (Kaufman & Kaufman, 1983; Marlow et al., 2007; Wolke & Meyer, 1999). Scores in

both cohorts were standardised based upon each cohort's respective control group with a mean of 100 and a standard deviation of 15. If IQ data were missing at 6 years, IQ scores from the next available cognitive assessment at either 8 years (BLS) or 11 years (EPICure) were used (N:41, 7% of all participants).

Statistical Analysis

SPSS version 24 (IBM Corp., Armonk, NY) and R version 3.4.2 were used to analyse the data. The comparison of demographic data in VP/VLBW or EP and control samples were assessed using chi-squared tests in both cohorts. Participants with complete data for measures of executive functioning, self-reported ADHD symptoms and TRAB-AS were included for analysis. All analyses were performed separately for each cohort; first in the BLS and then subsequently replicated in EPICure, allowing for the robustness of findings to be explored.

To test for differences between VP/VLBW participants or EP participants and controls, independent samples t-tests were first used to compare self-reported ADHD symptoms, parent-reported ADHD symptoms (BLS only), TRAB-AS, inhibitory control, working memory and IQ at 6 years for each cohort. Adjustment for multiple comparisons were made using Hochberg's procedure (Hochberg, 1988). Effect sizes are reported as Cohen's d: 0.20 = small, 0.50 = medium, 0.80 = large (Cohen, 1992).

When significant differences in attention problems were found between VP/VLBW or EP participants and controls, hierarchical regressions were performed to identify which factors reduced and explained these differences. This was performed first in the discovery sample of the BLS and replicated when possible in EPICure. Hierarchical regressions were used to determine whether deficits in executive function explained the greater attention problems in VP/VLBW and EP individuals, above and beyond the effect of IQ or sex. Each hierarchical regression added at step 1 the binary variable of birth group (VP/VLBW or control for BLS, EP or control for EPICure). At step 2, measures of executive function were added. IQ at 6 years was added at step 3 while male sex, a common risk factor for attention problems, was added at step 4. At each

step in the hierarchical regression, the importance of each variable was assessed in two ways. Firstly, by the R-square change of the overall model fit for the ADHD-I symptoms or TRAB-AS outcome, determining how each step improves the prediction of attention problems in adulthood. At step 4, the final model was assessed to determine the predictive ability of each variable upon consideration of all other variables in the model and the total variance explained. Additionally, the estimated adjusted means for VP/VLBW(or EP) and controls were calculated at each step in the hierarchical regression. This assessed the importance of inhibitory control, working memory, IQ at 6 years and sex by their effect on the differences in means between the VP/VLBW(or EP) groups and their respective controls. If for example, the reason for poor attention in VP/VLBW and EP adults was a result of poor executive functioning, then the adding of executive functioning measures at step 2 should cause the difference in estimated adjusted means between VP/VLBW and controls to diminish, becoming no longer statistically significant.

Results

Demographic Data and Drop-out Analysis

Information regarding demographic data and loss to follow-up into adulthood have been reported previously for the BLS (Eryigit Madzwamuse et al., 2015) and in EPICure (Linsell et al., 2018). VP/VLBW and EP participants in both cohorts were more likely to be of higher socioeconomic status than dropouts from their respective cohorts ($p = 0.003$ in BLS, $p = 0.004$ in EPICure). Participating EPICure EP individuals were also more likely to be female than EP participants lost to follow up ($p = 0.039$). The only significant difference within both cohorts comparing demographic data of VP/VLBW and EP to controls was that BLS controls were more likely to have higher socioeconomic status than BLS VP/VLBW individuals ($p = 0.030$).

Differences between EP/VP/VLBW adults and controls in ADHD symptoms, executive function and IQ

Between group differences in ADHD symptoms, attention span, executive function and IQ are shown in Table 7. In the discovery sample, the BLS, VP/VLBW participants did not self-report significantly higher ADHD-I, ADHD-H or ADHD-C symptoms than controls. Similarly, after adjustments for multiple comparisons were made,(Hochberg, 1988) there were no significant differences in self-reported ADHD between EP and controls in the replication sample of EPICure. Parents of the BLS VP/VLBW participants reported their adult children as having significantly higher ADHD-C symptoms than controls, which was primarily due to differences in ADHD-I symptoms rather than ADHD-H symptoms. Finally, in the BLS VP/VLBW participants were found to have considerably shorter attention spans than controls when rated by observers using the TRAB-AS, which was replicated in EPICure (Table 7).

For executive function, BLS's VP/VLBW participants demonstrated poorer performance in both domains, with larger response times for inhibitory control and lower working memory scores in comparison to controls. On the measure of IQ at 6 years of age, VP/VLBW participants scored considerably lower than their respective control group. In the replication sample of EPICure, a robustly similar set of findings regarding executive and general cognitive functions were found. However, the magnitude of difference between the EP participants and controls was slightly larger than the difference found between the VP/VLBW and controls in the BLS (Table 7). A correlation matrix for attention measures, executive functioning and general cognitive functioning is also provided in Appendix 10.

Table 7: Univariate differences between VP/VLBW or EP participants and controls

	Bavarian Longitudinal Study				EPICure			
	Mean difference (VP/VLBW-Control)	Mean difference 95% CI	Adjusted P-Value	Cohen's D	Mean difference (EP- Control)	Mean difference 95% CI	Adjusted P-Value	Cohen's D
ADHD- Inattention Self-Reported symptoms – Z scored	0.12	[-0.09, 0.34]	0.522	0.11	0.39	[0.03, 0.75]	0.084	0.34
ADHD- Hyperactivity/impulsivity Self-Reported– Z scored	-0.16	[-0.36, 0.03]	0.340	-0.17	-0.06	[-0.40, 0.29]	0.739	-0.05
ADHD- Combined Self-Reported –Z scored	-0.05	[-0.26, 0.15]	0.597	-0.05	0.19	[-0.16, 0.54]	0.543	0.17
ADHD- Inattention Parent Reported – Z scored	0.95	[0.49, 1.41]	<0.001	0.44	-	-	-	-
ADHD-Hyperactivity/impulsivity Parent Reported – Z scored	0.20	[-0.05, 0.44]	0.34	0.17	-	-	-	-
ADHD- Combined Parent Reported – Z scored	0.51	[0.19, 0.84]	0.01	0.33	-	-	-	-

Observer rating of attention span(TRAB-AS) – Z scored	-0.48	[-0.70, -0.25]	<0.001	-0.42	-1.14	[-1.73,-0.55]	0.001	-0.62
Inhibitory Control (ms)	27.53	[17.04, 38.01]	<0.001	0.52	41.86	[22.4, 61.33]	<0.001	0.69
Working Memory	-8.98	[-12.72, -5.24]	<0.001	-0.48	-10.37	[-14.77,-5.96]	<0.001	-0.75
IQ at 6 years	-16.49	[-19.81, -13.17]	<0.001	-0.99	-26.24	[-31.69, -20.79]	<0.001	-1.54

Note: ADHD(attention deficit hyperactivity disorder). Inhibitory Control as measured by the Attention Network Task. Working memory as measured by the letter number sequencing task in the BLS and backwards digit recall task in EPICure. IQ at 6 years as measured by the K-ABC task. P values are Adjusted using Hochberg's correction. Z- scored indicates that raw scores are standardised based upon the mean and standard deviation of the respective control group.

Hierarchical regressions explaining TRAB-AS and ADHD-I symptoms differences in VP/VLBW or EP adults and controls

For TRAB-AS in the BLS, the estimated adjusted means between groups at each hierarchical step are shown in figure 7. Initially at step 1, the VP/VLBW groups' attention span ratings were z= -0.48 (-0.70, -0.25) lower than controls. At step 2, both inhibitory control and working memory were found to be significantly associated with TRAB-AS rating, with the difference in adjusted means between groups reducing to z=-0.21 (-0.43, 0.01) and no longer statistically significant. At step 3, IQ at 6 years old was also found to be significantly associated with TRAB-AS rating, further reducing the estimated adjusted means to a difference of z=-0.04(-0.26, 0.19). While at step 1, the difference in estimated adjusted means between VP/VLBW and controls was found to be 0.48, this reduced to 0.04 at step 4 (see figure 7). The final model for predicting TRAB-AS in the BLS explained 23% of the variance with working memory and IQ at 6 years old the only factors remaining significantly associated with attention span rating (Table 8).

For TRAB-AS in EPICure, the estimated adjusted means between groups at each hierarchical step are shown in figure 7. Initially at step 1, the EP groups' attention span ratings were z= -1.14 (-1.73, -0.55) lower than controls. At step 2, working memory and inhibitory controls significantly diminished the effect of birth group on attention span rating to z= -0.58(-1.21,0.06). At step 3, adding the measure of IQ at 6 years old, both executive functioning variables were no longer statistically significant and resulted in controls having an adjusted attention span of z=0.14 (-0.55, 0.83) lower than EP participants. While at step 1, the estimated difference in adjusted means found the EP group to have a deficit of z= -1.14, at step 4 with sex also introduced the difference had switched to controls having a deficit of z= 0.11 (see figure 7). The final model for TRAB-AS in EPICure explained 26% of the variance, with IQ at 6 years of age being the only remaining significant predictor (Table 8)
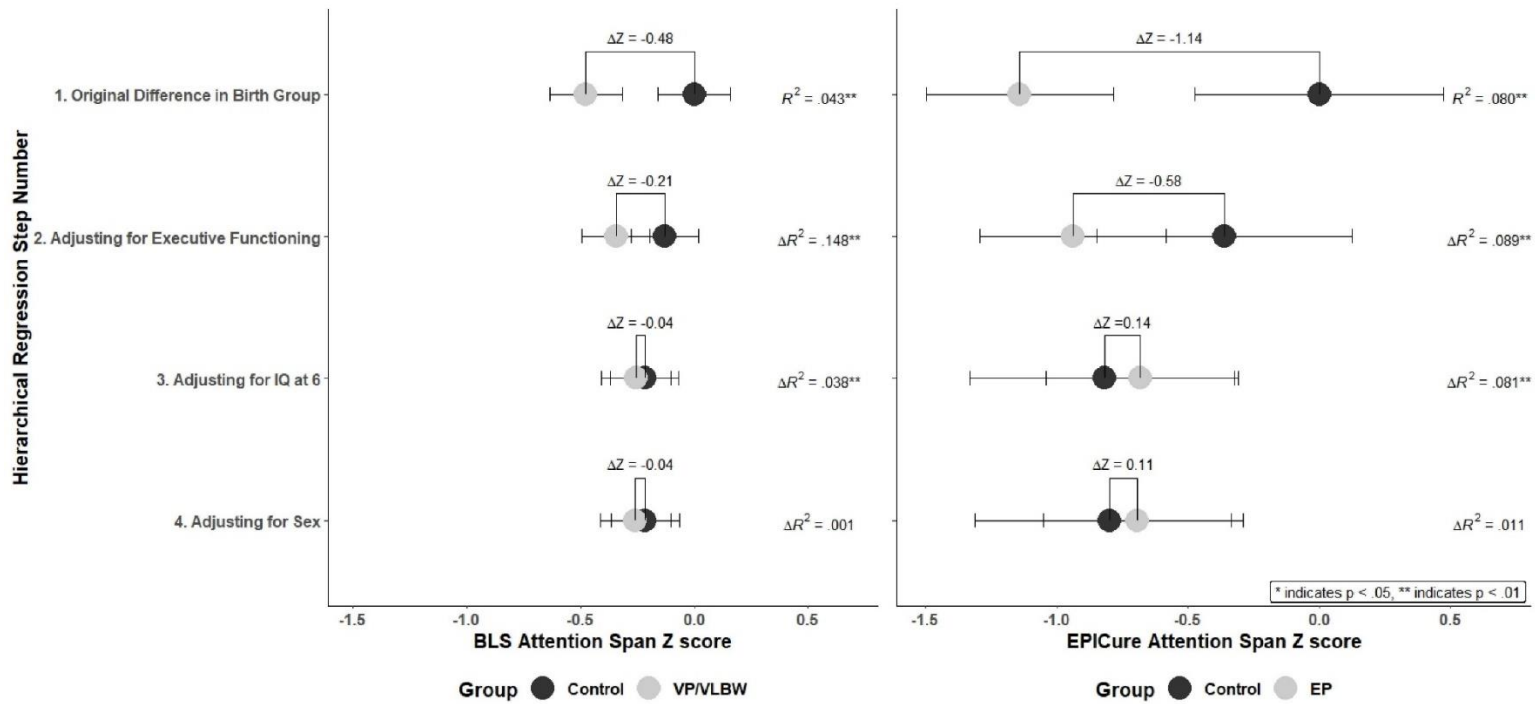
Figure 7: Differences in Tester Rating of Adult Behaviour-Attention span (TRAB-AS) between VP/VLBW and EP with their respective control group at each step of the hierarchical regression for the Bavarian Longitudinal Study and EPICure.

For BLS parent-reported ADHD-I symptoms, the estimated adjusted means for VP/VLBW and controls at each hierarchical step are shown in figure 8. Initially at step 1, the VP/VLBW group had an ADHD-I symptom z score 0.95 greater than the controls, 95% confidence interval 0.49 to 1.41. When inhibitory control and working memory were entered at step 2, both executive functioning measures were significantly associated with ADHD-I symptoms, with the difference in estimated adjusted means between VP/VLBW and controls reducing to z=0.50 (0.04, 0.95). It was not until step 3, when IQ at age 6 years was added, that the estimated mean differences between groups became statistically insignificant, reducing to a difference of z=0.03 (-0.43, 0.50). At step 4, the variable of sex did not significantly increase R^2 and only minimally influenced the estimated adjusted means 0.01(-0.46, 0.48). From the initial differences between VP/VLBW and controls at step 1 being z=0.95, the difference in estimated adjusted means between VP/VLBW and controls in the final model was reduced to a difference of z=0.01. The final model for BLS parent-reported ADHD-I symptoms explained 22% of the variance and was predominantly explained by IQ at 6 years of age and inhibitory control in adulthood (Table 8).
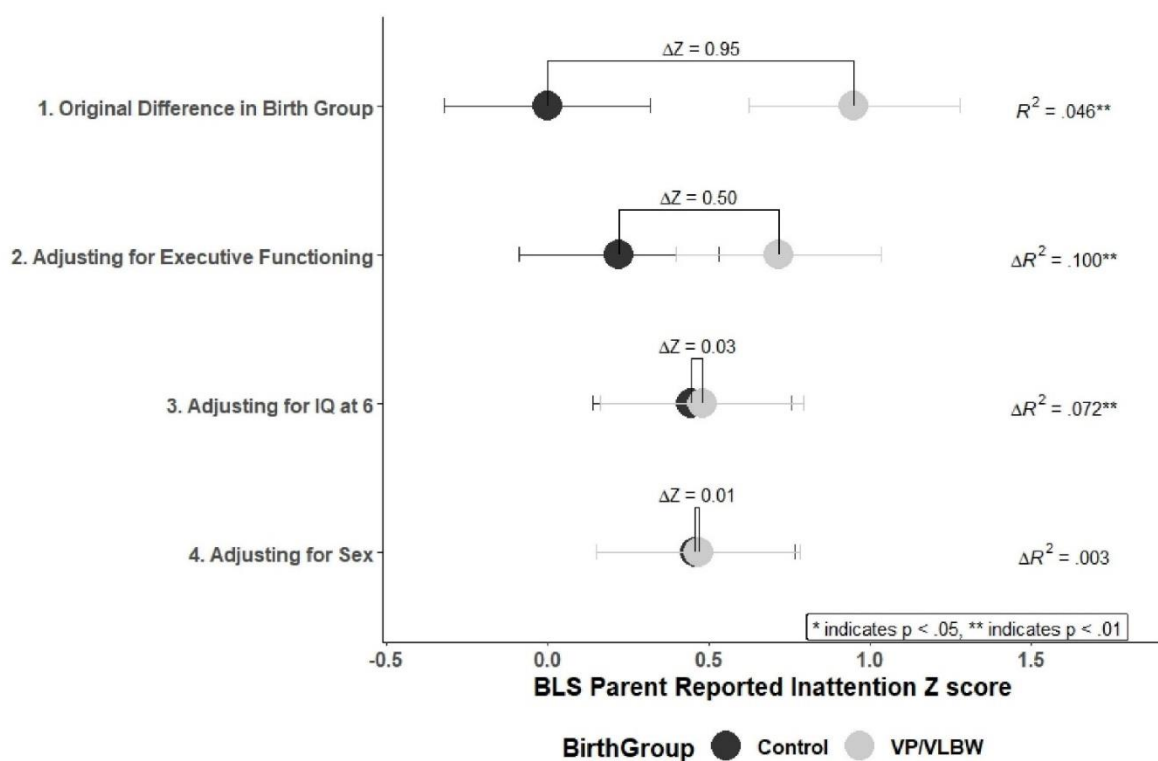
Figure 8: Differences in parent reported ADHD-Inattention symptomology between VP/VLBW and controls at each step of the hierarchical regression for the Bavarian Longitudinal Study

Table 8: Final multiple regression models (step 4) predicting standardised parent reported ADHD-I symptoms and TRAB-AS ratings in the Bavarian Longitudinal Study (BLS) and EPICure

Predictor	BLS ADHD-I PR		BLS TRAB-AS		EPICure TRAB-AS	
	Beta	P-Value	Beta	P-Value	Beta	P-Value
Birth Group(0 = Control, 1 =EP/VP/VLBW)	0.00	0.971	-0.02	0.712	0.03	0.759
Inhibitory Control	0.14	0.006	-0.07	0.149	-0.11	0.114
Working Memory	-0.07	0.213	0.24	<0.001	0.12	0.165
IQ at 6 years	-0.35	<0.001	0.26	<0.001	0.39	<0.001
Sex (0 = Female, 1=Male)	0.06	0.218	0.03	0.566	-0.11	0.119
Total R^2	0.22		0.23		0.26	

Note: ADHD-I PR: Parent reported ADHD-inattention symptoms , TRAB-AS: observer rating of attention span. Inhibitory Control as measured by the Attention Network Task, working memory as measured by the letter number sequencing task in the BLS and backwards digit memory task in EPICure. IQ at 6 years as measured by the K-ABC task.

Discussion

In the discovery sample of the BLS, we observed evidence of greater attention problems for VP/VLBW adults, as demonstrated by poorer observed attention span in comparison to controls, further validated by greater parent-reported ADHD-I symptoms. In contrast, we found no self-reported difference in ADHD between VP/VLBW and controls. These results were found to be robust, being replicated in the EPICure sample in which EP adults had shorter observer rated attention span but no self-reported differences in ADHD either. Our hypothesis, that differences in attention would be explained by executive functioning was only partially supported. In the BLS, measures of inhibitory control and working memory in adulthood partially explained the effect of VP/VLBW birth. However, after childhood IQ was accounted for, inhibitory control only remained significantly associated with parent-reported ADHD-I symptoms, while working memory only remained significantly associated with TRAB-AS ratings. For EPICure, while the effect of EP birth on TRAB-AS rating was explained by inhibitory control and working memory, neither factor remained significant after accounting for childhood IQ. The results from both cohorts indicate that while specific executive functioning measures can aid in explaining why VP/VLBW or EP adults show more attention problems than controls, childhood IQ explains a larger amount of the difference between groups.

The pattern of results from adulthood is largely in concordance with past research looking at attention problems in preterm children, suggesting specific problems of inattention rather than hyperactivity/impulsivity. Additionally, the greater relative differences found between EP and controls in EPICure than between the VP/VLBW and controls in the BLS may result from a "gestational gradient", whereby the risk of attention problems increases as gestational age at birth decreases (Johnson & Wolke, 2017). The EPICure EP group were born on average 6 weeks more preterm than the BLS VP/VLBW group. Also consistent with this interpretation is the relatively poorer performances on measures of executive functioning and the larger deficits in general cognitive ability between EPICure's EP adults and controls than between BLS's VP/VLBW adults and controls. Alternatively, or additionally, year of birth (1985 vs

1995) and age of assessment (26 vs 19 years old) differed between the discovery sample (BLS) and the replication sample (EPICure). Regarding era of birth, previous studies (Cheong et al., 2017; Twilhaar et al., 2018) found that while survival of very preterm born babies has increased, there is little evidence of improved cognitive outcome across eras. Age of assessment may also be important if deficit in executive function and attention is due to developmental delay that may narrow with age. As the BLS' VP/VLBW participants were older than EPICure's EP participants, they may have had more time to 'catch up' in comparison to their respective control group. Nevertheless, our results were remarkably similar across cohorts despite differences in degree of prematurity and age of assessment, indicating generalisability of findings.

Within the general population and in VP/VLBW children, attention problems have been primarily associated with deficits in executive functioning (Aarnoudse-Moens et al., 2013; Retzler et al., 2018; Willcutt et al., 2005), however, we found inconsistent evidence for this after we controlled for childhood IQ. Our results are in line with Willcutt, Doyle and Nigg et al's (2005) postulation that deficits in executive function are important but are not the sole factor causing ADHD symptoms (Willcutt et al., 2005). Alternatively, as our VP/VLBW and EP participants demonstrated a behaviourally distinct phenotype, composed primarily of inattention rather than hyperactivity/impulsivity, it may be that this phenotype has a different primary factor. The attention problems of VP/VLBW and EP adults, as shown here, would appear to be due to a general cognitive deficit rather than the specific executive functioning deficit seen in the general population. However, if inattention is a result of a specific executive functioning deficit it is also possible that our measures were not sensitive to those specific deficits. In childhood, inattention within the general population but also in VP/VLBW and EP participants has been found to be more closely related to visuo-spatial working memory rather than verbal working memory (Martinussen et al., 2005; Mulder et al., 2011; Retzler et al., 2018). As our measures of working memory were verbal, it may be that we failed to assess the correct specific measures of executive functioning. While future studies should look to address this, the current results are in line with recent research suggesting the limited efficacy of working memory interventions on attention and working memory performance itself for VP/VLBW

children (Anderson et al., 2018). If verbal working memory is both impervious to intervention and only partially related to inattention in VP/VLBW and EP adults, it suggests that interventions for VP/VLBW and EP children may be focused elsewhere.

The fact that childhood IQ was significantly related to attention problems in adulthood in both cohorts, regardless of how attention was assessed, and partially explained the effect of being born VP/VLBW or EP is pertinent. Intelligence is unlikely to be assessed independent of executive function in childhood. For example, the IQ test used (the K-ABC) , has some tasks that are related to executive functioning. However, the K-ABC is strongly correlated with the widely used Wechsler Intelligence Scale for Children, at r=.79 and .70 throughout childhood (Kaufman & Kaufman, 1983; Zins & Barnett, 1984). Thus, our results are unlikely to differ depending on the child IQ test used. Regardless, failing to control for general cognitive ability might lead to the potentially erroneous conclusion that a specific executive functioning is responsible for attention problems when it is instead part of a more general cognitive deficit. If early identification of VP/VLBW or EP children at risk of long-term attention problems is of primary importance, then IQ testing appears a relatively straightforward approach to do so. VP/VLBW and EP individuals have been found to be at increased risk of brain injury, such as reduced cholinergic basal forebrain integrity and decreased white and grey matter, which has been found to mediate the relationship between preterm birth and poorer IQ (Grothe et al., 2017; Nosarti et al., 2008). It may be that IQ scores in childhood act as an indicator of overall poor brain growth. This poor brain growth may result in long term behavioural deficits in domains such as inattention, but less so for behaviours regarding hyperactivity and impulsivity. The finding of a strong association between general cognitive ability and inattention are consistent with evidence from EPICure in childhood (Johnson et al., 2016), as well as other research finding strong links between general cognitive performance and behavioural difficulties for VP/VLBW children (Aarnoudse-Moens et al., 2013; Burnett et al., 2019).

Another important finding is that the method for assessing attention problems is key, with non-significant differences by self-report but larger differences when assessed

through parent report or observer rating. When BLS VP/VLBW behaviour was rated by their parents or observer, more attention problems were found but this was not found for self-report. In EPICure parent report was unavailable but the results found a similar disparity between self-report and observer ratings. Overall, our results support other research into attention in extremely low birthweight adults and controls, finding no significant difference for self-reported ADHD of any subtype (M. H. Boyle et al., 2011; Johnson et al., 2019). We can speculate that the VP/VLBW group's reporting of fewer symptoms as compared to parents is compatible with Festinger's theory of social comparison (Festinger, 1954). VP/VLBW and EP adults have been found to have a lower educational level and are more likely to be in manual employment (Mathiasen et al., 2009). An individual's primary comparison is with those they socialise with mostly, i.e. peers. Compared to peers in their social circle, VP/VLBW and EP adults may not consider themselves to have attention problems. In contrast, parents are more likely to compare their offspring to their birth cohort (i.e. all adults) and thus use a different comparison level and report more attention problems, similar to observation measures of attention. Regardless of why EP and VP/VLBW adults under report their own symptoms, these results are in concordance with studies in the general population. In both childhood and into adulthood, there is substantive evidence that individuals with attention problems report less symptoms than their parents or independent observers do (Knouse et al., 2005; Owens & Hoza, 2003). Overall, self-report measures of ADHD may underestimate symptoms in VP/VLBW and EP adults and as such multi informants should be assessed.

There are clear strengths to this study. These include the use of two prospectively studied cohorts allowing for replication of findings. The use of identical measures for ADHD symptoms, observer rating of attention span, inhibitory control and child IQ in both cohorts reduces the influence of methodological issues in interpreting results. However, there are also limitations. Firstly, the rate of attrition was moderate to high, with remaining participants found to be of higher socioeconomic status in both cohorts. This potential bias is unlikely to have had an impact on our results, as regressions models may be only marginally affected by selective dropout (Wolke, Waylen, et al., 2009); nevertheless, bias cannot be excluded. The lack of parent report

in EPICure and the difference in working memory assessments limited direct replication of some of the findings from the BLS. Though the two measures of verbal working memory have been found to be closely related (Crowe, 2000), the letter number sequencing task may be more associated with attention ratings due to its greater complexity (Engle, 2010). Future research should look to address the importance of task complexity as well as assessing visuo-spatial working memory, which as previously noted may be more linked to attention deficits. Finally, while our study was able to assess multiple possible predictors of inattention, it had the limitation that we were unable to directly assess other important cognitive factors such as processing speed equivalently for both cohorts, as it has been noted as a core deficit for inattention in the general population and VP/VLBW children (Diamond, 2005; Mulder et al., 2011). While working memory performance is thought to be at least partially reliant on processing speed (Fry & Hale, 2000), directly testing whether this lower level ability is key to adult inattention could be pivotal for future interventions.

To conclude, this study provides further evidence for specific attention problems in early adulthood for VP/VLBW and EP in comparison to controls, replicating findings from childhood. While we found that adult executive functioning measures were associated with attention problems in adulthood, childhood IQ was a stronger and more consistent predictor in both the discovery and replication sample. Early assessment of cognitive ability would allow for early identification of VP/VLBW and EP children at risk for long term attention problems.

Chapter 8   **Small for Gestational Age-Cognitive Performance from Infancy to Adulthood: An Observational Study**

Abstract

Objective: To determine whether cognitive performance from infancy to adulthood is affected by being born SGA, and if this depends on the SGA reference used. Furthermore, to determine SGA's effect while considering the effects of very preterm/very low birth weight (VP/VLBW), socioeconomic status (SES), and parent-infant relationship.

Design, setting and population: 414 participants (197 Term-Born, 217 VP/VLBW) of the Bavarian Longitudinal Study

Methods: SGA was classified using neonatal or foetal growth references. SES and the parent-infant relationship were assessed before 5 months old.

Main outcome measures: Developmental (DQ) and IQ tests assessed cognitive performance on 6 occasions, from 5-months to 26-years old.

Results: The foetal reference classified more infants as SGA (<10th percentile) than the neonatal reference (N=138, 33% Vs N=75,18%). Using linear mixed models, SGA was associated with IQ -8 points lower than AGA, regardless of reference used (CI [-13.66, -0.64] and [-13.75,-1.98]). This difference narrowed minimally into adulthood. Being VP/VLBW was associated with IQ -16 [CI -21.01,-10.04] points lower than term-born participants. Low SES was associated with IQ -14 [CI -18.55, -9.06] points lower than high SES. A poor parent-infant relationship was associated with IQ -10 points lower than those with a good relationship [CI -13.91,-6.47]

Conclusions: SGA is associated with lower IQ throughout development, independent of VP/VLBW birth, low SES or poor parent-child relationship. Social factors have comparable effects on IQ than SGA and should be considered for interventions.

Introduction

Small for Gestational Age (SGA, birthweight <10% for gestation) has been consistently associated with lower cognitive performance in childhood, as demonstrated by lower scores on developmental and intelligence tests (DQ and IQ) (Bie et al., 2010). However, there are contradictory findings of long-term effects of SGA on IQ (Eryigit Madzwamuse et al., 2015; Østgård et al., 2014; Pyhala et al., 2011), suggesting diminishing cognitive differences between SGA and Appropriate for Gestational Age (AGA) adults. To investigate this, studies assessing cognitive performance throughout development are needed, from infancy to adulthood. This contrasts to previous studies that have reported on different individuals and their DQ/IQ scores cross-sectionally.

Additionally, it has been hypothesised that SGA's effects on IQ differ depending on whether one is born at term or preterm, with SGA's effects being disproportionately larger for those born very preterm (P. Shah & Kingdom, 2011). However, the few very preterm/very low birthweight (VP/VLBW; <32 weeks gestation and/or <1500g) cohorts that have tested whether the factors interact have not found significant interactions (Eryigit Madzwamuse et al., 2015; McCarton et al., 1996). Rather, results suggest that SGA has an independent adverse effect (Bie et al., 2010; McCarton et al., 1996) or no effect on IQ for VP/VLBW groups (Eryigit Madzwamuse et al., 2015; Pyhala et al., 2011).

Determining SGA's effect on IQ is complicated due to no universally accepted reference for classifying SGA. Longitudinal cohorts have traditionally used SGA references based on large datasets of neonates (Mikolajczyk et al., 2011). However, as preterm infants are more likely to be growth restricted (Mikolajczyk et al., 2011), neonatal references classifying only the lowest 10% as SGA may underdiagnose the true rates of growth restriction in the VP/VLBW population (Mikolajczyk et al., 2011). Alternatively, references comparing birthweight to estimated foetal weight can be used. Foetal references are prone to measurement error (Mayer & Joseph, 2013) but have been found to be superior in predicting infant mortality, especially

for VP/VLBW infants (Ding et al., 2013). For predicting cognitive outcomes, foetal references may have superior sensitivity than neonatal references (Neta et al., 2011), however this has not been further supported (Charkaluk et al., 2012).

Finally, socio-environmental factors that influence cognitive performance must also be considered (Wolke, 2019). Low familial socioeconomic status (SES) has been consistently associated with lower IQ (Linsell et al., 2015). However, the mechanisms of how SES affects IQ scores are not well understood and are likely multifactorial. One pivotal factor may be the parent-infant relationship (Wolke, 2019), as it has been found to predict long-term cognitive outcomes (Breeman et al., 2017).

The primary objective of this study was to determine whether the IQ of SGA and AGA participants differed over time, classified using either foetal or neonatal references. The second objective was to determine whether SGA's effects were disproportionately larger for participants born VP/VLBW than at term. Finally, to consider whether SGA's effects persisted once socio-environmental risk factors were controlled for.

Methods

Participants

The BLS is a geographically defined prospective whole population sample of VP/VLBW(<32 weeks gestation and/or <1500g) and term-born children born in Southern Bavaria (Germany) between January 1985 and March 1986. The VP/VLBW group were admitted to one of 16 children's hospitals within the first 10 days after birth (Wolke, Schmid, et al., 2009). Of the initial 682 VP/VLBW, 411 were alive and eligible for the 26-year follow-up assessment with 260 (63%) VP/VLBW participants participating. 203 undertook IQ testing in adulthood while a further 14 participants had severe impairments and were unable to undertake adult assessment and were given a proxy IQ score at 26 years of age (Eryigit Madzwamuse et al., 2015). This resulted in 217 (53%) VP/VLBW participants included. Infants who were born at term in the same obstetric hospitals were

recruited as controls. Of the initial 916 term-born controls alive at 6 years, 350 were randomly selected within the stratification variables of sex and family SES as to be comparable to the VP/VLBW sample. Of these, 308 were eligible for the 26-year follow-up assessment, 229 participated (74%) with 197 (64%) completing cognitive assessments in adulthood. A full flow chart can be found in Appendix 11.

Ethical approval for this study was granted by the Ethical Board of the University Hospital Bonn (date of approval: 19th august 2009, reference 159/09). Informed consent was provided by parents and adult participants. The BLS was supported by grants PKE24, JUG14, 01EP9504 and 01ER0801 from the German Federal Ministry of Education and Science, which underwent peer review but did not involve public involvement, see Appendix 16 for the completed GRIPP2 form. The funders had no role in the study design, data collection, data analysis, data interpretation, or writing of the report. While the BLS has measured other outcomes of interest (behavioural outcomes, mental health etc.), cognitive outcomes were the sole focus of the current study.

<u>Gestational age and birth weight</u>

Gestational age was determined from maternal dates of the last menstrual period and serial ultrasounds during pregnancy. Where gestational age estimates from these methods differed by less than 2 weeks, maternal dates were used (Riegel et al., 1995). Birth weight was documented from birth records.

<u>Small for Gestational Age classification</u>

Two references were used to classify SGA, a neonatal reference (SGA_N) and a foetal reference (SGA_F). SGA_N used Voigt's data from 2.3 million live and still singleton births in Germany from 1995 – 2000 with a gestational age from 20 – 43 weeks (Voigt et al., 2006), allowing for sex-specific weight percentiles to be calculated. SGA_F instead compared the birthweight of our participants to the expected foetal weight of healthy developing foetuses using Mikolajczyk et al's model (Mikolajczyk et al., 2011). This method is based on Hadlock's growth equation which used

ultrasound measurements to calculate weight percentiles from 10- 41 weeks gestation (Hadlock et al., 1991). A key aspect of Hadlock's equation is that a healthy developing foetus should reach a final weight of 3705g at 40.5 weeks gestation, however this uses USA data and is non-sex specific (Hadlock et al., 1991). Mikolajczyk et al. adjusts Hadlock's equation from 3705g to a country specific average birthweight at 40 weeks. We therefore took the 50th percentile for a German infant at 40 weeks from Voigt's data as reference, while also doing this separately for males (3624g) and females (3473g) so that both SGA references were country and sex specific. For both references, SGA/AGA status was determined if the respective weight percentile was below or above the 10th percentile (SD = -1.282).

Cognitive assessments

The cognitive assessments used in the BLS have been previously reported.(Breeman et al., 2015)

The Griffiths Mental Development Scale measures DQ at 5 and 20 months old, age corrected for prematurity. It assesses 5 dimensions of development: locomotor, personal-social development, hearing and speech, hand and eye coordination, and performance (Breeman et al., 2015).

IQ at 4 years was assessed through a composite of cognitive tasks: the Columbia Mental Maturity Scale, the Active Vocabulary Test, and the Beery-Buktenica Developmental Test of Visual-Motor Integration (Breeman et al., 2015). Using confirmatory factor analysis a composite score was constructed using the standardized scores from all participants who undertook cognitive testing at 4 years of age, for more information see Appendix 12.

IQ at 6 and 8 years was assessed with the German version of the Kaufmann Assessment Battery for Children. A total IQ score was calculated from the sequential (3 subtests) and simultaneous (5 subtests) processing scales (Breeman et al., 2015).

IQ at 26 years was assessed with a German version of the Wechsler Adult Intelligence Scale. The 6 subtests were vocabulary, similarities, letter number-

sequence, block design, matrix reasoning, and digit symbol coding. Cognitive functioning scores of the subtests were converted into a Full-Scale IQ score (Breeman et al., 2015).

For brevity, all differences in DQ and IQ scores will be simply referred to as differences in IQ scores. IQ scores were standardized at each time point based upon the mean and standard deviation of the most optimal birth group, participants who were born both at term and AGA. Therefore, while the term and AGA group would demonstrate a flat cognitive group score (i.e. mean IQ scores equal to 100 and a standard deviation of 15 at each time point), scores for all other participants reflect catch up or deterioration in relation to the most optimal birth group over time.

Socio-environmental Factors

Family SES data was obtained by standard interviews with the infants' parents in the first 10 days of life. SES was computed as a weighted composite score of maternal highest educational qualification, paternal highest educational qualification, and occupation of the head of family and grouped as low, middle or high (Bauer, 1988).

The Parent-Infants Relationship Index (PIRI) is an 8-item based scale derived from concerns regarding the mother and infant or father and infant relationship (Breeman et al., 2017). 5 items were derived from questions during an interview with the parents in the neonatal ward or at 5 months of age. The final 3 items were assessed by the study nurse in the neonatal ward. Items were coded as 0=No concern or 1= Concern. Final scores on the PIRI were dichotomized into good parent-infant relationship (0, all items= 0) or poor parent-infant relationship (1, at least one item=1), see Appendix 13 for more detail on the PIRI.

Data Analyses

Participants with IQ scores at 26 years of age were included for analysis. The missing datapoints from IQ tests (4.8%) in childhood were imputed using the R package mice (Buuren & Groothuis-Oudshoorn, 2011). Mixed modelling was used to investigate IQ trajectories from infancy to adulthood, using maximum likelihood for parameter estimation. Advantages of the mixed modelling method is the use of a

structure that allows multiple assessments across time to be nested within a single individual, with variation between individuals treated as a random effect. This method then allows for trajectories of IQ scores to be predicted as a linear function of fixed effects, such as SGA or VP/VLBW status.

Analyses using either SGA_N or SGA_F references were performed separately. Initial unadjusted mixed models (Model 1- SGA_N and Model 1 - SGA_F) added two fixed effects based on birth group: an SGA or AGA group variable and a VP/VLBW or term-born group variable. Additionally, interactions between these variables and age (measured in years) were considered as to determine if being VP/VLBW interacted with the effect of SGA or if their effects changed over time. Secondly, models were adjusted for sex and SES, both as fixed effects (Model 2- SGA_N and Model 2 - SGA_F). Thirdly, the effect of the parent-infant relationship measured with the PIRI was considered as another fixed effect (Model 3- SGA_N and Model 3 - SGA_F). In regard to random effects, all models allowed for the intercept and slope of IQ trajectories to vary by individual. In total, this resulted in 3 models for each SGA classification reference and therefore 6 models altogether.

Results

Participants

Baseline characteristics of SGA and AGA participants according to SGA_N and SGA_F are shown in in Table 9. The SGA_N reference classified 61 (28%) VP/VLBW and 14 (7%) term-born participants as SGA. In contrast, the SGA_F reference classified 116 (53%) VP/VLBW and 22 (11%) term-born participants as SGA. Further perinatal information can be found in Appendix 14.

Table 9: Baseline characteristics of adults born SGA and AGA according to the Neonatal based reference (SGA$_N$) or Fetal Reference (SGA$_F$)

	SGA using Neonatal Reference				SGA using Fetal Reference			
	AGA$_N$ + VP/VLBW (n=156)	SGA$_N$ + VP/VLBW (n=61)	AGA$_N$ + Term-Born (n=183)	SGA$_N$ + Term-Born (n=14)	AGA$_F$ + VP/VLBW (n=101)	SGA$_F$ + VP/VLBW (n=116)	AGA$_F$ + Term-Born (n=175)	SGA$_F$ + Term-Born (n=22)
Birthweight (g)								
Mean (SD)	1390 (313)	1110 (239)	3430 (398)	2540 (235)	1490 (325)	1150 (219)	3460 (395)	2700 (292)
Gestational Age (weeks)								
Mean (SD)	29.7 (1.49)	32.0 (2.40)	39.7 (1.13)	38.8 (1.25)	29.6 (1.56)	31.0 (2.17)	39.7 (1.14)	39.6 (1.40)
Sex								
Male	86 (55.1%)	28 (45.9%)	86 (47.0%)	8 (57.1%)	60 (59.4%)	54 (46.6%)	83 (47.4%)	11 (50.0%)
Female	70 (44.9%)	33 (54.1%)	97 (53.0%)	6 (42.9%)	41 (40.6%)	62 (53.4%)	92 (52.6%)	11 (50.0%)

Socioeconomic Status								
High	34 (21.8%)	14 (23.0%)	64 (35.0%)	5 (35.7%)	23 (22.8%)	25 (21.6%)	61 (34.9%)	8 (36.4%)
Middle	75 (48.1%)	27 (44.3%)	77 (42.1%)	6 (42.9%)	47 (46.5%)	55 (47.4%)	74 (42.3%)	9 (40.9%)
Low	47 (30.1%)	20 (32.8%)	42 (23.0%)	3 (21.4%)	31 (30.7%)	36 (31.0%)	40 (22.9%)	5 (22.7%)
Parent Infant Relationship								
Poor	73 (46.8%)	29 (47.5%)	50 (27.3%)	3 (21.4%)	48 (47.5%)	54 (46.6%)	47 (26.9%)	6 (27.3%)
Good	78 (50.0%)	28 (45.9%)	133 (72.7%)	11 (78.6%)	49 (48.5%)	57 (49.1%)	128 (73.1%)	16 (72.7%)
Missing	5 (3.2%)	4 (6.6%)	0 (0%)	0 (0%)	4 (4.0%)	5 (4.3%)	0 (0%)	0 (0%)

IQ trajectories by SGA$_N$ status and VP/VLBW status

Observed and predicted group trajectories are shown in Figure 9 while the results of the mixed models are displayed in Table 10. In the initial unadjusted model utilizing the SGA$_N$ it was found that in comparison to AGA participants, being born SGA resulted in an initial -7.89 IQ points decrease (95% CI [-14.62, -1.17]), (see Model 1-SGA$_N$). The effect of being VP/VLBW resulted in scores -21 IQ points [CI -25.98, -16.28] lower than being term-born. The interaction between being term-born and SGA was not significant, indicating that the effect of being born SGA was similarly detrimental on IQ for both term and VP/VLBW participants[CI -12.14,15.98]. Trajectories were found to be relatively stable with age and similar across the birth groups. However, there was a trend indicating SGA participants caught up in comparison to AGA participants, at approximately 0.23 IQ points [CI -0.02,0.48] per year, meaning Model 1- SGA$_N$ estimated the difference between SGA and AGA participants to reduce from 7.89 to just 2.01 IQ points by 26 years old.

Figure 9: Observed and predicted (Model 3- SGA$_N$) cognitive trajectories stratified by small for gestational age (SGA) status and very preterm/very low birthweight (VP/VLBW) status.

IQ trajectories by SGA$_F$ status and VP/VLBW status

Observed and predicted group trajectories are shown in Figure 10 while the results of the mixed models are displayed in Table 10. In the unadjusted model utilizing SGA$_F$, it was found that the effect of being born SGA resulted in an initial -8.38 IQ points [CI -14.46,-2.29] decrease in comparison to being born AGA, (see Model 1- SGA$_F$). The effect of being VP/VLBW resulted in a decrease in -19 IQ points [CI -13.46, -24.63] in comparison to being born at term. The interaction between being term-born and SGA was not significant, indicating that the effect of being born SGA was similar regardless of VP/VLBW status [CI -7.98,15.62]. Trajectories were found to be stable with age and similar across the birth groups. A minimal trend indicated SGA participants demonstrated an approximate 0.18 IQ points [CI -0.04,0.41] per year catch-up in comparison to AGA participants. Thus, Model 1- SGA$_F$ estimated

the difference between SGA and AGA participants to reduce from 8.38 to 3.78 IQ points by 26 years old. In regard to rates of minor (IQ<85) and major (IQ<70) impairment, the percentage at each time point for both SGA references are provided in Appendix 15.
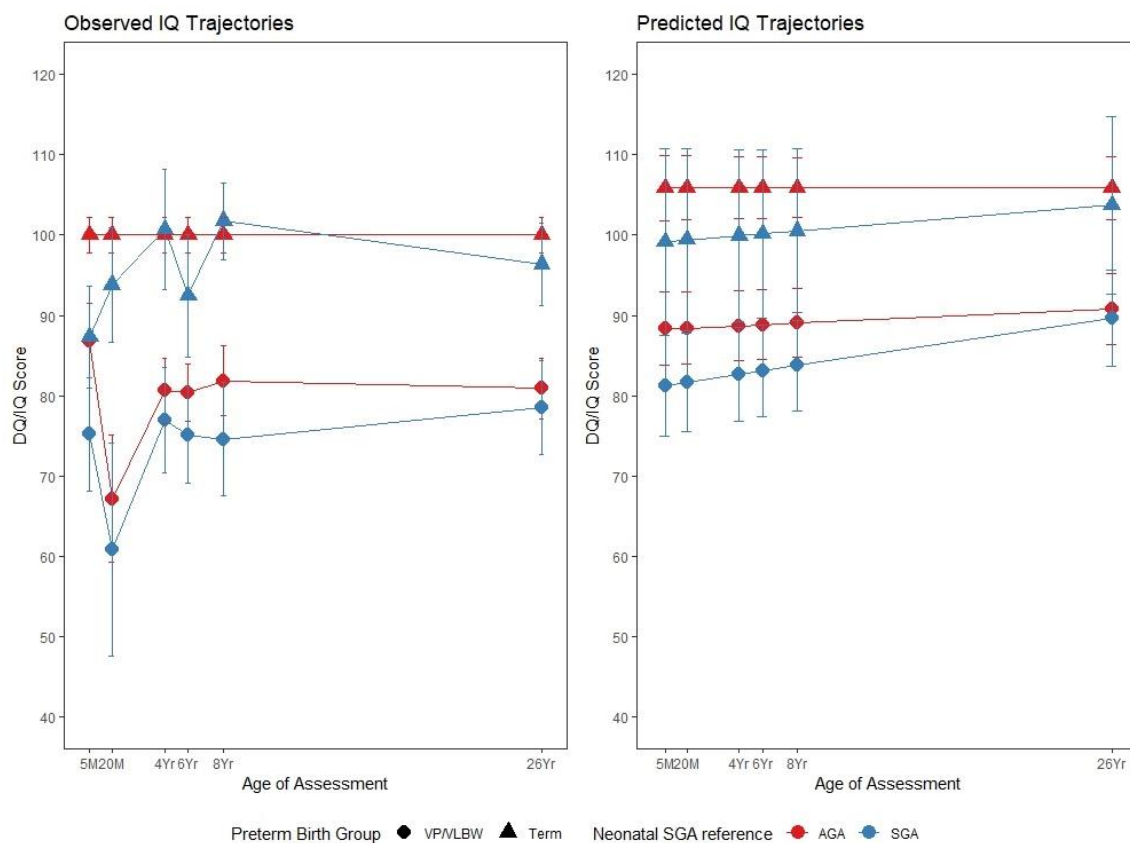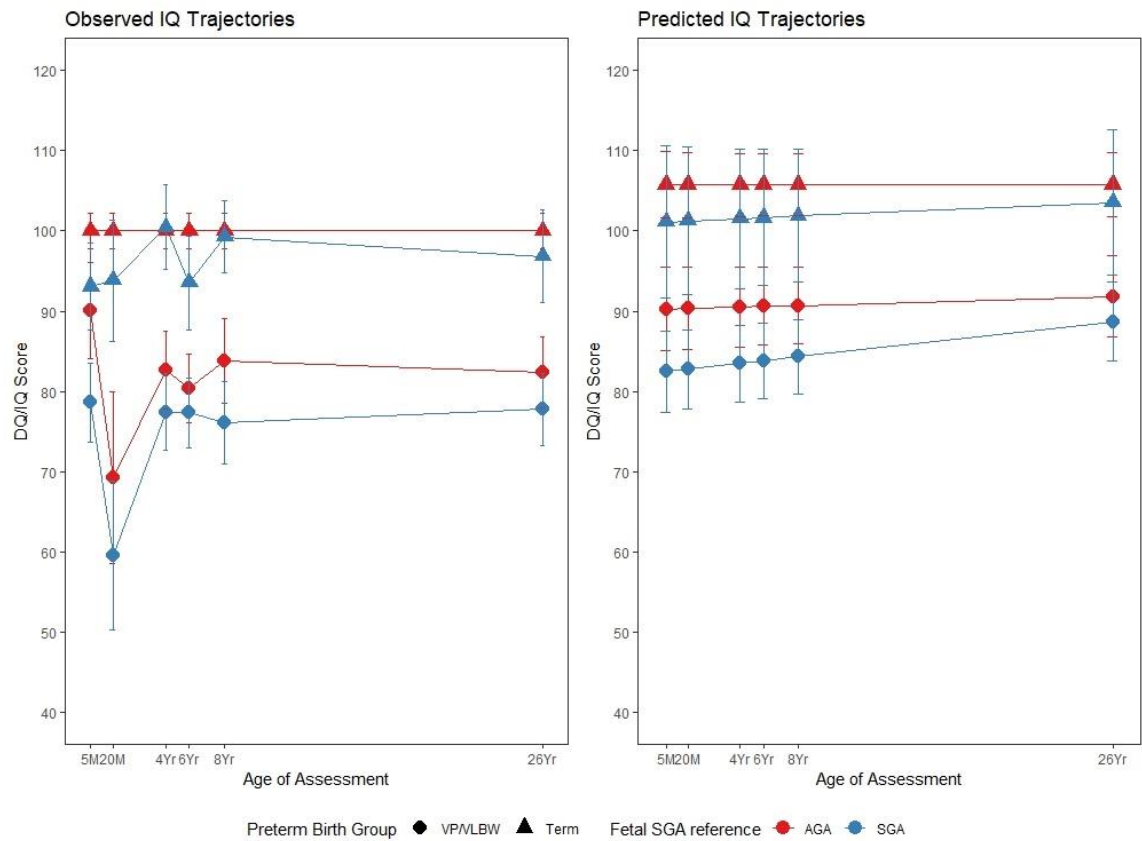

Figure 10: Observed and predicted cognitive trajectories (Model 3- SGA_F) stratified by small for gestational age (SGA) status and very preterm/very low birthweight(VP/VLBW) status

Effects of sex, socioeconomic status and the parent-infant relationship on IQ
trajectories

The addition of sex did not have a significant effect in either of the adjusted models
(Model 2- SGA$_N$, CI [-5.60,1.63]; Model 2 – SGA$_F$, CI [-5.37,1.92]). SES was found
to be significantly associated with IQ scores in both models, with those born into a
low SES family having on average -14 IQ points less than those from a high SES
family (Model 2- SGA$_N$ CI [-19.17,-9.41], Model 2- SGA$_F$ CI [-19.16, -9.36]).
Finally, the effect of having a poor parent-infant relationship (PIRI) was found to
have a large effect of approximately -10 IQ points in both final models (Model 3-
SGA$_N$ CI [-13.80, -6.39]; Model 3- SGA$_F$ CI [-13.91,-6.47]). The inclusion of sex,
SES and PIRI did not significantly change the effects of VP/VLBW status or SGA
status on IQ scores in either final model. Additionally all models demonstrated
evidence of individual differences in intercept and slope, as demonstrated by the
random effects. On average, intercepts had a standard deviation of approximately 20
IQ points and slopes varied with a standard deviation of 0.31 IQ points per years.
See Table 10 for all models.

Table 10: Estimated mean differences in cognitive test scores from linear mixed model analyses using either Neonatal (SGAN) or Fetal (SGAF) references for SGA classification

Parameters	SGA_N_Model 1		SGA_N_Model 2		SGA_N_Model 3		SGA_F Model 1		SGA_F Model 2		SGA_F Model 3	
Fixed effects	*Est*	*CI*	*Est*	*CI*	*Est*	*CI*	*Est*	*CI*	*Est*	*CI*	*Est*	*CI*
(Intercept)	78.87	**[75.32,82.43]**	90.53	**[83.40,97.65]**	93.15	**[86.18,100.13]**	80.95	**[76.51,85.39]**	92.03	**[84.52,99.53]**	94.70	**[87.36,102.04]**
SGA(AGA=0, SGA=1)	-7.89	**[-14.62,-1.17]**	-7.68	**[-14.38,-0.99]**	-7.14	**[-13.64,-0.64]**	-8.38	**[-14.46,-2.29]**	-8.02	**[-14.09,-1.96]**	-7.86	**[-13.75,-1.98]**
Control(VP/VLBW=0, Term = 1)	21.13	**[16.28,25.98]**	19.71	**[14.84,24.57]**	17.61	**[12.83,22.39]**	19.05	**[13.46,24.64]**	17.75	**[12.15,23.34]**	15.52	**[10.04,21.01]**
Age	0.10	[-0.04,0.23]	0.10	[-0.04,0.23]	0.10	[-0.04,0.23]	0.06	[-0.11,0.23]	0.06	[-0.11,0.23]	0.06	[-0.11,0.23]
SGA*Age	0.23	[-0.02,0.48]	0.23	[-0.02,0.49]	0.23	[-0.02,0.49]	0.18	[-0.05,0.42]	0.18	[-0.05,0.42]	0.18	[-0.05,0.42]
SGA*Control	1.92	[-12.14,15.98]	1.36	[-12.63,15.36]	0.36	[-13.23,13.95]	3.82	[-7.98,15.62]	3.28	[-8.47,15.03]	3.20	[-8.19,14.59]
Control*Age	-0.10	[-0.28,0.09]	-0.10	[-0.28,0.09]	-0.10	[-0.28,0.09]	-0.06	[-0.27,0.15]	-0.06	[-0.28,0.15]	-0.06	[-0.28,0.15]
Control*SGA*Age	-0.05	[-0.58,0.48]	-0.05	[-0.59,0.48]	-0.05	[-0.59,0.48]	-0.09	[-0.54,0.36]	-0.09	[-0.54,0.37]	-0.09	[-0.54,0.37]

	Est	CI	Est	CI	Est	CI	Est	CI	Est	CI	Est	CI
Sex(Male=1,Female = 2)	-	-	-1.98	[-5.60,1.63]	-0.65	[-4.18,2.88]	-	-	-1.72	[-5.37,1.92]	-0.35	[-3.90,3.21]
Middle SES at birth (High- as reference)	-	-	-9.31	**[-13.66,-4.95]**	-9.10	**[-13.32,-4.89]**	-	-	-9.19	**[-13.56,-4.81]**	-8.97	**[-13.21,-4.74]**
Low SES at birth (High as reference)	-	-	-14.29	**[-19.17,-9.41]**	-13.86	**[-18.58,-9.14]**	**-**	-	-14.26	**[-19.16,-9.36]**	-13.81	**[-18.55,-9.06]**
PIRI (0= No Concern, 1 = Concern)	-	-	-	-	-10.10	**[-13.80,-6.39]**	-	-	-	-	-10.19	**[-13.91,-6.47]**
Random Effects												
sd(Intercept)	20.65	[19.03,22.42]	20.52	[18.90,22.28]	19.79	[18.20,21.53]	20.66	[19.01,22.45]	20.53	[18.90,22.30]	19.78	[18.17,21.53]
sd(Age)	0.29	[0.21,0.40]	0.31	[0.24,0.42]	0.32	[0.24,0.42]	0.29	[0.21,0.40]	0.33	[0.25,0.43]	0.32	[0.24,0.43]
cor(Intercept, Age)	-0.83	[-0.94,-0.55]	-0.90	[-0.97,-0.67]	-0.90	[-0.97,-0.67]	-0.85	[-0.95,-0.55]	-0.90	[-0.97,-0.68]	-0.89	[-0.97,-0.66]
BIC	22071		22059		22039		22160		22150		22130	

Note P <0.05 signified in bold, BIC (Bayesian Information Criterion), PIRI (Parent-Infant Relationship Index), Est (Estimate from linear mixed model), CI (95% confidence interval).

Discussion

<u>Main Findings</u>

In this longitudinal study, regardless of the SGA reference used, SGA was associated with lower IQ scores that had not fully diminished into adulthood. This was found despite the fact that the foetal SGA reference classified more participants as SGA than the neonatal reference, with almost double the number of VP/VLBW participants classified as SGA. We found the effect of SGA was additive but not interactive with VP/VLBW, indicating that both factors are important for cognitive development. We also found large effects of early socio-environmental factors. Being born into a low SES family and having a poor parent-infant relationship were both strong and independent risk factors, associated with lower IQ scores throughout development.

<u>Strengths and Limitations</u>

Our study has several strengths. It assessed cognitive performance 6 times, from infancy to adulthood. We controlled for important covariates such as SES and the parent-infant relationship and assessed their independent effects on IQ. Limitations are that we were unable to differentiate between SGA participants with intrauterine growth restriction and those who are constitutionally small; future research should address this potential measurement error. Finally, there was a lack of moderately/late preterm participants while the number of participants born both SGA and at term was small. This limited the ability to determine whether SGA has a disproportionately larger effect on the VP/VLBW or preterms generally.

<u>Interpretation</u>

Past cross-sectional research has found larger effects of SGA on IQ in childhood than in adulthood (Bie et al., 2010; Eryigit Madzwamuse et al., 2015; Østgård et al., 2014; Pyhala et al., 2011). This resulted in our hypothesis that SGA participants may demonstrate cognitive catch up in comparison to AGA participants. Our longitudinal study found SGA participants continued to have lower IQ scores in adulthood,

however, the difference with AGA participants had diminished from approximately 8 IQ points to 2.01 or 3.78 IQ points depending on the SGA reference used. Rather than between childhood and adulthood, it was between infancy and childhood that the largest temporal change in cognitive performance was seen, between the DQ assessment at 20 months and IQ at 4 years. This may in part be due to the difference in tasks, such as the greater reliance on sensorimotor skill for DQ assessments (Aylward, 2009). The neurological mechanism that underpins the relationship between SGA and continued lower IQ is unclear. As IQ is a general measure of intelligence, it may not be linked to any specific cortical or subcortical deficit (Østgård et al., 2014) but instead global differences between the brains of SGA and AGA participants, such as reduced cortical thickness or lower brain volume (De Bie et al., 2011; Østgård et al., 2014).

We also found that the effect of VP/VLBW on IQ was stable across lifespan, in concordance with past research of cognitive trajectories of preterm individuals (Breeman et al., 2015; Linsell et al., 2018). Our hypothesis that the effect of SGA on IQ would be disproportionately larger for the VP/VLBW participants was not supported, with the interaction term failing to reach significance. Despite the fact that VP/VLBW infants are also more likely to be growth restricted (Ding et al., 2013), the current research suggests that both SGA and VP/VLBW have significant, independent effects on IQ.

The use of neonatal references to classify SGA in longitudinal cohorts has been debated, especially for VP/VLBW participants (Neta et al., 2011). However, we did not find large differences between references for their subsequent effects on IQ. This may be surprising considering the foetal reference almost doubled the number of participants classified as SGA. Charkaluk et al. (2012) suggested that if foetal SGA references simply lower the threshold to be classified as SGA then the inclusion of less at-risk individuals should reduce the effect of SGA on IQ (Charkaluk et al., 2012). Instead, in both the current and Charkaluk's study, those classified as SGA using the foetal SGA reference have a similar risk as those classified using the neonatal SGA reference (Charkaluk et al., 2012). Future research should investigate how references differ on birthweight percentiles as well as the binary SGA vs AGA

cut-offs used here, as the relationship between birthweight percentile and IQ is likely continuous (McEwen et al., 2018). Additionally, other anthropological measurements, such as head circumference, may provide better utility in predicting long term cognitive outcomes (Jaekel et al., 2019).

Our study also considered the effect of socio-environmental measures on IQ, finding both to be significant factors on IQ. Low familial SES was associated with IQ scores 14 points lower than those born into high SES, a considerably larger effect than SGA. However, SES is a multifactorial construct with the specific mechanisms that influence IQ needing to be elucidated (Wolke, 2019). Parenting behaviour has been previously linked to SES (Wolke, 2019). However, the parent-infant relationship measure also had a significant independent effect on IQ, with a slightly larger effect than SGA. As the effect of socio-environmental factors on IQ appear to be of similar magnitude but have less measurement error than the effect of SGA, more focus should be spent on optimizing these potentially modifiable factors through intervention.

Conclusion

To conclude, SGA is associated with lower IQ into adulthood, regardless of how SGA is determined. The effect of SGA on IQ is similar regardless of whether the infant is born at term or very preterm, or after controlling for socio-environmental factors. Familial SES and the parent-infant relationship have similar but potentially more modifiable effects on IQ than SGA, suggesting interventions in parent-infant relationship in the perinatal period may be beneficial.

Chapter 9   **Overall Discussion**

The overall aim of this thesis was to determine the foetal origins of adult neurocognitive performance. It looked to expand on current knowledge by investigating the universality of outcomes across different VP/VLBW cohorts, by investigating how general and specific cognitive abilities affect attention problems and how cognitive performance changes throughout development in relation to being born VP/VLBW or SGA. In this final chapter, a general summary of the findings will be provided first, followed by an integrated discussion. The future research directions and the consequences for identifying individuals at risk and identifying interventions will be subsequently proposed.

9.1   Summary of Results:

Study 1 had three aims. First, what are the differences in IQ scores between VP/VLBW adults and term born controls in eight international cohorts providing IPD. Second, what are the individual and cohort level factors that influence the IQ scores of VP/VLBW adults. Finally, are scores from cohorts that provide IPD representative of all published literature on IQ performance for VP/VLBW adults.

It was found that VP/VLBW's adult IQ scores are significantly lower than term born controls, at 0.78 SD (equivalent to 12 IQ points). This was found consistently across cohort, indicating a universal finding. It was additionally found that a number of individual level risk factors were associated with lower IQ scores for VP/VLBW adults. These include IVH, BPD, lower gestational age, lower birthweight Z score and lower maternal education.

Study 2 investigated whether the greater attention problems seen in VP/VLBW and EP individuals as compared to term-born adults are best explained by specific executive functioning deficits, general cognitive abilities, or sex. Additionally, it was investigated whether attention problems and the factors associated with them are

replicated across cohorts or are dependent on the informant used (parent, self or observer rating). It was found that in both the BLS and EPICure cohorts, VP/VLBW or EP adults had lower observer rated attention span than controls. Similarly, parents rated VP/VLBW individuals as having more ADHD-I symptoms than controls as assessed in the BLS. In contrast, no differences in VP/VLBW's or EP's self-report of their ADHD-I symptoms compared to controls were found. The differences in attention problems reported by observers or parents were only inconsistently associated with executive function. Instead, IQ measured in childhood was strongly correlated with inattention in adulthood across both cohorts.

Study 3 investigated whether the IQ of SGA and AGA participants differed over time, classified using either foetal or neonatal references. It also considered whether the effects of SGA were disproportionately larger for participants born VP/VLBW than at term. Finally, it determined if the effects of SGA persisted once socio-environmental risk factors were controlled for. It was found that individuals classified as SGA had IQ scores 8 points lower than those classified AGA and the gap narrowed slightly from the earliest DQ measurement to the last IQ measurement in adulthood. This was found regardless of whether a foetal or neonatal reference was used, despite the large differences in the number of individuals classified using either reference. In addition, SGA's effect was similar regardless of whether a participants was born at term or VP/VLBW, refuting the hypothesis that the two factors should significantly interact to cause disproportionately lower IQ. Finally, as well as the strong main effects of VP/VLBW and SGA, there were additional strong independent main effects for socio-environmental factors of parental socioeconomic status and the parent infant relationship.

9.2 Integrated Discussion and Theoretical Implications

The findings highlight the consistency of findings across VP/VLBW cohorts and the stability of lowered cognitive performance from infancy to adulthood. In spite of the fact that VP/VLBW cohorts differed in birth year, country, and inclusion criteria;

there was consistently lower cognitive performance and similar attention problems across cohorts. Furthermore, in both studies 1 and 3, relatively low birthweight for gestational age, measured either as a continuous Z score or as a categorical variable of SGA/AGA, was associated with lower IQ. Finally, while the results from all three studies indicate the importance of early neonatal factors on cognitive and behavioural outcomes, the findings also point to the importance of the environment and in particular, parental factors. These findings naturally build on and extend what is currently known regarding the long-term outcomes of being born VP/VLBW and SGA. Despite differences across cohorts regarding initial inclusion criteria, rates of participants lost to follow up, nationality, birth year and age at adult assessment; this did not appear to result in drastically different cognitive and behavioural outcomes for those born VP/VLBW.

In study 1, the findings highlight the universality of cognitive outcomes and their links with neonatal factors for VP/VLBW adults among countries where data is currently available. However, there is a large period of time that passed in between the birth of the VP/VLBW individuals to when they were assessed in adulthood. Thus, it is overly simplistic and deterministic to assume that no other life factors across childhood may have additionally contributed to cognitive development (Wolke, 2019). In fact, in both studies 1 and 3, it was found that societal or parental factors were also significantly related to cognitive performance throughout development, in concordance with prior research in VP/VLBW children and adults (Breeman et al., 2017; Linsell et al., 2015). However, parental SES or maternal education are relatively blunt or simplistic measures when there are many other potential environmental and genetic factors that likely contribute to cognitive development of the child (Duncan & Magnuson, 2012). These may include a diversity of correlated factors such as breastfeeding (Brion et al., 2011), air pollution (Loftus et al., 2019), parental behaviours (Raviv et al., 2004), or genetic effects (Torres, 2013).

It has been proposed that maternal education or SES may simply be a proxy measure for maternal IQ and as such reflect a genetic effect due to the high correlation between

the constructs (Deary et al., 2007). However, in analyses looking at the moderating or mediating effects of both socio-economic status and maternal IQ on child IQ, both were found to have overlapping and independent main effects (Torres, 2013). While this thesis looked at how maternal education or socio-economic status was linked to their child's IQ, research has also found that maternal IQ predicts the IQ of VP/VLBW children, although the heritability effect may be weaker than in maternal-term born child dyads (Lean et al., 2018). As well as understanding which specifics aspects of maternal education, such as maternal IQ, influence the cognitive development of VP/VLBW individuals (Wolke, 2019), it is additionally important to consider how these environmental or genetic influences may change over time.

It was found that IQ scores were relatively stable with age, regardless of whether the individual was born SGA or VP/VLBW. This is generally in line with research from the general population and other VP/VLBW cohorts finding that IQ is a largely stable trait across the lifespan (Breeman et al., 2015; Darlow et al., 2020; Deary et al., 2004; Linsell et al., 2018). It has already been alluded to that maternal education effects may be at least partially a genetic, heritable effect via maternal IQ. This however raises an interesting avenue of research when in combination with the "Wilson effect" (Bouchard, 2013). The Wilson effect posits that the heritability of IQ surprisingly increases, not decreases, with age. This effect may be somewhat reduced for VP/VLBW individuals, where admittedly heritability of IQ appears lower with shared environment instead explaining much more variance in IQ performance (Koeppen-Schomerus et al., 2000; Lean et al., 2018). However, if VP/VLBW individuals also demonstrate the Wilson effect then it would be expected that VP/VLBW individuals with higher IQ parents would show some degree of IQ catch up while those with lower IQ parents may show continued IQ deterioration into adulthood. Further research is therefore needed looking at how maternal and paternal IQ influences the cognitive trajectories of VP/VLBW or SGA groups into early adulthood and beyond. By being able to control for potential heritability effects, the relative importance of VP/VLBW, SGA or environmental factors could also be more accurately considered.

In study 2 the importance of specific and general cognitive performance on attention problems was considered. This is important to investigate as multiple studies have now reported on increased attention problems in VP/VLBW individuals but much less research has considered the reasons behind why they come about. Both theory and research from the general population has linked specific executive functioning deficits to attention problems (Pennington & Ozonoff, 1996; Willcutt et al., 2005). In study 2, there was only weak evidence that specific executive function deficits were more important to attention problems than the effects of general IQ for VP/VLBW adults.

These findings from study 2 have clear implications for VP/VLBW research groups which utilise interventions developed specifically for ADHD individuals from the general population. If VP/VLBW individuals do not show either the same symptomology or the same underlying deficits, it appears unlikely they will benefit from interventions developed for a different population. This is supported by recent research that has demonstrated the limited efficacy of executive functioning-based interventions for VP/VLBW children with attention problems (Anderson et al., 2018; Jaekel et al., 2020; van Houdt et al., 2020). Interventions aimed to improve executive functioning in VP/VLBW children would ideally bring a range of benefits including reduced attention problems. This was thought to be an area with much promise, due to research suggesting executive functioning could be improved in children, especially for those with particularly poor executive functioning (Diamond & Lee, 2011). In three randomised control trials, computerized training was given to children born very and extremely preterm in order to improve executive functioning, attention, and a number of other linked constructs. Disappointingly, none of the studies found long term improvements in executive functioning, parent and teacher rated attention, selective and sustained attention, IQ, behaviour, or school performance problems (Anderson et al., 2018; Jaekel et al., 2020; van Houdt et al., 2020). Therefore, computerised trials do not appear to be a successful way of reducing the attention problems of VP/VLBW children. However, these results may need caveating, as the sole use of computerised tasks has been noted as a limited way to improve EF, with broader interventions needed (Diamond & Lee, 2011). If instead the general cognitive functioning of VP/VLBW is the most pivotal factors for later attention problems, then

broad interventions based around parental behaviours and education may be more fruitful, as evidenced by the results of both study 1 and study 3.

While the findings from study 2 did not provide strong corroborating evidence for EF as the key factor for VP/VLBW attention problems, the findings did support other aspects of past research of VP/VLBW and attention outcomes. The finding that inattention symptoms are far more common than hyperactivity/impulsivity for VP/VLBW individuals is in concordance with research from childhood (Franz et al., 2018; Indredavik et al., 2004; Johnson et al., 2016). This also supports the "preterm phenotype" hypothesis, indicating that the comorbidities for those born VP/VLBW are somewhat unique and differ from those with ADHD in the general population (Burnett et al., 2019; Johnson & Marlow, 2017; Lahat et al., 2014). While VP/VLBW adults may show a somewhat distinct phenotype, the results from study 2 were in concordance with some other aspects of ADHD adults from the general population. In particular the agreement, or lack of, between self-reported ADHD symptoms and parent reported ADHD (Du Rietz et al., 2016). Study 2 found VP/VLBW adults were less likely to self-report their own inattention problems, in stark contrast to either experimenter rating or parent rating. This lack of agreement between self and parent reports for attention problems has previously been found in the general population (Du Rietz et al., 2016). Although self-report is often used in adult assessments due to its ease and greater practicality, it appears to be subject to bias (Manor et al., 2012). The results and research therefore from other VP/VLBW adult cohorts appear to indicate that solely using self-reported measures may lead to the erroneous conclusion that attention problems have subsided by adulthood (Lahat et al., 2014). It will be important when investigating ADHD symptoms to have data from different data sources and consider the different perspectives reported by different data sources when comparing findings across studies. This needs to be considered for future studies considering only using self-reported measures of attention in VP/VLBW adult populations.

As well as considering how attention should be assessed, this thesis also investigated other methodological questions, such as how certain neonatal risk factors should be

operationalised. In studies 1 and 3, a number of different SGA references or measures of relative birthweight were used. This included foetal and neonatal references, made into categorical variables above or below the 10th percentile or used as a continuous Z score. While it was consistently found that being born with relatively low birth weight for gestation was significantly associated with lower cognitive performance, there were not strong differences between foetal and neonatal references in relation to cognitive outcome. This is especially surprising for VP/VLBW individuals, considering that foetal references can classify almost double the number of VP/VLBW individuals as SGA, as seen in study 3 and elsewhere (Charkaluk et al., 2012; Pritchard et al., 2019). The findings therefore show that while SGA is associated with lower cognitive performance, there is a certain degree of measurement error. SGA is thought to be a proxy indicator of IUGR, defined as the disruption of growth in the womb as determined by repeated ultrasounds (Schlaudecker et al., 2017). However, it may be that no SGA references, that are based on measurements solely made at birth, can reliably identify those with IUGR (Hutcheon et al., 2008).

Further research is needed to investigate whether certain SGA references are more closely associated with IUGR than others, as it is likely to offer superior specificity when looking at later cognitive performance. While in higher income countries, the more frequent use and collection of ultrasound data may mean IUGR measures can be used more often, this is unlikely to be practical in low-income countries. Relatedly, understanding the universal cognitive outcomes following SGA birth, and how important the reference used is to this, requires investigation. Study 3 found no significant differences between foetal or neonatal SGA references for predicting cognitive development. However, this was limited to just one relatively small cohort from one high income country. In addition, this thesis did not directly compare differences between a global or local reference or whether it is customised based on maternal factors such as ethnicity or height. Therefore, determining the optimal SGA reference for predicting cognitive performance, either foetal or neonatal, either global or local, customised on maternal factors or not, remains a pressing issue that requires further investigation.

Another important factor to consider regarding the SGA research performed in this thesis is the fact that only linear effects for birthweight Z scores on IQ were tested. It is possible that a difference of a birthweight Z score from -1.3 to -0.3 (SGA to AGA) may be more beneficial to IQ than going from 0.3 to 1.3 (AGA to LGA). In the general population and according to standardised growth charts, there is generally an inherent assumption that a birthweight Z score of 0, meaning the birthweight percentile is exactly at the 50th percentile, is best (Vasak et al., 2015). This is likely somewhat driven by the fact that as an infant becomes larger and larger it imposes greater risk to the mother's life (Vasak et al., 2015) and to the health of the infant (Norris et al., 2015). However, it may be that having a somewhat higher relative birthweight is beneficial for the infant's survival and later cognition, especially a VP infant.

Within term born groups, research has found birthweight Z scores above the 97th percentile has been linked to a higher initial risk of neonatal mortality in comparison to average birthweight Z scores (25th percentile to 75th percentile) (Altman et al., 2012). Thus, if a pattern of increased mortality risk is in concordance with increased risk of cognitive impairment, then high birthweight Z scores should also be linked to cognitive impairment. Surprisingly, despite the increased mortality risk, term born children with birthweight Z scores above the 90th percentile were found to outperform those born with birthweight Z scores either between the 50th and 90th percentile, the 10th and 50th percentile or below the 10th percentile on IQ at 6 years of age (Yang et al., 2010). Whether these findings are also found within the preterm population requires investigation.

VP infants do appear to show a somewhat similar pattern regarding higher birthweight Z scores and lower mortality. As the average birthweight of a VP infant is far from the average for a term born infant, it is thought that having a relatively high birthweight for gestation is beneficial for a VP infant's survival (Draper et al., 1999). This is further supported by research showing an increased survival for VP infants with a birthweight percentile between 75% and 97% relative to those with a birthweight percentile between 25% and 75% (Evans et al., 2007) However, for VP infants above the 97th percentile, there was a lowered chance of survival, again indicating a nonlinear

effect of birthweight percentile on survival, similar to results seen in term born infants (Altman et al., 2012; Evans et al., 2007). Whether similar effects are found on IQ for VP requires investigation. As studies in this thesis did not test nonlinear effects of birthweight Z scores on IQ, it was not possible to determine this potential resiliency factor or determine particularly sensitive areas in the distribution where birthweight Z score are particularly important for later IQ scores.

Furthermore, the results of both studies 1 and 3 demonstrate that while lower gestational age clearly was an important factor for cognitive performance, a number of other factors were also found to be independently associated with cognitive development. These included being born with relative lower birthweight/SGA, the parent – infant relationship and SES/maternal education. When considered in multivariate models, these factors were still independently associated with cognitive development. Therefore, when a VP/VLBW individual is born SGA, comes from a low SES family or has a poor parent-infant relationship then greater support and care is likely needed. As these factors were not found to be strongly overlapping but instead significant main effects, it suggests that an accumulation of these factors is likely to result in particularly low cognitive performance.

The research studies also demonstrate the importance of child IQ as an early marker for adult outcomes. Study 3 showed the stability of IQ from infancy to adulthood, with differences maintaining particularly stable from 4 years of age to 26 years old. Additionally, as seen in study 2, child IQ was consistently a predictor of adult inattention. Thus, child IQ appears to be a key factor that can predict not just future cognitive performance but behavioural problems too. As preterm children are thought to suffer from long term brain differences, identifying these children is of critical importance (Ment et al., 2009). While brain function could be more directly measured via magnetic resonance imaging or other neuroimaging techniques, this is expensive and time consuming (Crosson et al., 2010). Instead, it is possible that child IQ may act as a reliable marker of brain growth and explains a large amount of variance in other behavioural or cognitive domains. Indeed, as well as attention problems, child IQ has been found to largely explain language differences between EP and term born children,

suggesting no specific cognitive or brain deficit (Wolke et al., 2008). The strong association between lower IQ and global brain differences is supported by the finding that 70% of the variance in IQ data can be explained by total white matter volume and corpus callosum area for preterm adolescents (Northam et al., 2011). Overall, this would therefore suggest that psychologists interested in reliably identifying preterms at risk for long term cognitive and behavioural problems should take confidence in the use of a child IQ test as a quick and relatively low-cost screening tool.

The findings regarding SGA, VP/VLBW and IQ also have potential theoretical implications, especially for models such as allostatic load model or developmental origins of health and disease hypothesis (DOHaD). The allostatic load model would suggest that prematurity or foetal growth restriction stresses the infant and increases the "wear and tear" on the body for future outcomes (Juster et al., 2010). If there is an accumulation of risk factors this can cause allostatic overload or breakdown, resulting in a nonlinear decline in cognitive performance, greater than the individual sum of each risk factor (Guidi et al., 2020). Pertinent to these models are therefore whether risk factors for IQ act independently (statistical main effects), moderate one another (statistical interaction), or demonstrate high degrees of overlap/collinearity. The strongest interpretation of the allostatic load model would hypothesise that being born VP/VLBW and SGA should interact to cause disproportionately lower cognitive performance due to the increased stress imposed by each factor causing allostatic overload (Olson et al., 2015). When adding in further stressors, such as low parental SES or a poor parent-infant relationship, the allostatic load model would predict a further increase in the risk for poor cognitive outcomes due to the increasing accumulation of stressors. Thus, the results from study 3 partially go against a "strong" interpretation of the allostatic load model as the interactions between SGA and VP/VLBW were not significant, suggesting complete allostatic overload/breakdown did not occur. However, the results do somewhat support the allostatic load model, due to the fact that accumulating risk factors did all have independent association with IQ, indicating each factor increased the "wear and tear".

In contrast, the DOHaD hypothesis suggests that factors such as foetal growth restriction lead to the development of a thrifty phenotype, adapting the foetus or making the foetus more sensitive to a specific external environment post birth (Wadhwa et al., 2009). Often, the supposed external environment adapted for does not match the actual external environment, resulting in a mismatch and thus pathology (Pluess & Belsky, 2011). The DOHaD hypothesis is therefore not a traditional model of diathesis stress, where the presence of both a vulnerability (e.g. SGA) and a life stress (e.g. low SES) interact to result in a more negative outcome (e.g. low IQ) (Monk et al., 2019). Instead, the DOHaD hypothesis is more closely related to a "for better or for worse" differential susceptibility model, where those born SGA may be more sensitive to a certain type of external environment (e.g. SES level) than those born AGA (Pluess & Belsky, 2011). Thus, if SES is low, SGA children may perform poorer than AGA children who are also from low SES families. However, if SES is high, then SGA children may outperform AGA children as they are more sensitive and so can benefit more from the beneficial external environment. Differential susceptibility effects regarding SGA has been previously found, with SGA individuals outperforming AGA individuals on wealth outcomes when they both had mothers with high maternal sensitivity (Nichols et al., 2020) or on reading performance when both given specific interventions (van der Kooy-Hofland et al., 2012). This thesis did not explicitly test interactions between SGA and measures of environmental optimality, meaning models such as DOHaD or allostatic load were not comparatively tested. However, the results do generally support the DOHaD and allostatic load models as both models would predict that foetal factors are strongly associated with long term development.

9.3   Strengths and Limitations of the Research

There are a range of strengths to the current research. Firstly, the combination of multiple international cohorts in both studies 1 and 2 gives greater confidence in the

findings due to the replication of findings and the greater statistical power. This is superior to the majority of past research which has reported results from a single cohort. When past research has looked to combine data from multiple VP/VLBW cohorts, it has largely either been in childhood or has been on an aggregate level, i.e. non IPD meta-analyses (Mendonça et al., 2019; Wolke, Baumann, et al., 2015). The ability to combine adult data at the individual level allows to investigate individual level risk factors and determine their long-term consequences. Importantly, the ability to harmonise risk factors and outcomes across cohorts also resulted in greater validity of cohort comparison (Tierney et al., 2015). In particular, aggregate meta-analyses are susceptible to aggregation bias when looking at risk factors. For example, the mean level of maternal education at a cohort level, does not reflect the importance of maternal education at an individual level, within cohorts (Lambert et al., 2002). By looking at risk factors at an individual level, their importance can be more reliably determined. This is potentially best demonstrated by the fact that a meta-regression looking at the child IQ of VP/VLBW individuals did not find a cohort's mean maternal education level to be a significant factor (Twilhaar et al., 2018), despite it being one of the most important predictors in the IPD meta-analysis performed in this thesis. By using IPD, as the "gold standard" for systematic review (Tierney et al., 2015), this thesis expanded on the current knowledge base for the cognitive performance of VP/VLBW adults and which specific risk factors are universally important. In addition, the IPD meta-analysis investigated each predictor's relative importance as part of a multivariable analysis, so that potentially important confounding could be considered.

This thesis also had the advantage of looking at cognitive development from infancy to adulthood for both SGA and AGA adults. While the vast majority of research has assessed outcomes at one timepoint, very few have looked at how cognitive performance changes over time and into adulthood (Bie et al., 2010). This is important as when assessing at only one point in time, especially in adulthood, there is the possibility of systematic loss to follow up biasing the findings. This is potentially why past research appears to have found greater IQ differences in childhood than in adulthood (Eryigit Madzwamuse et al., 2015; Gutbrod et al., 2000; Strauss, 2000). In

study 3, by only including participants with IQ scores in adulthood, any smaller cognitive differences at later time points cannot be due to selective attrition but instead due to later cognitive catch up. While the remaining SGA adult participants may have been higher functioning than the full starting SGA group in infancy, it was still found SGA was a long-term risk factor for IQ, providing a better understanding of the long-term cognitive outcomes for SGA individuals.

There are also limitations, firstly, all data came from adult studies. VP/VLBW infants born today receive different care to what the participants in these cohorts received 25 or more years ago. This may mean the findings are no longer, or at least less, applicable for VP/VLBW populations nowadays. However, research into VP/VLBW child populations born in the 2000s show that there has been no improvement in cognitive outcome compared to the differences to term-born controls in more recently born cohorts (Cheong et al., 2017; Marlow et al., 2021; Twilhaar et al., 2018). Thus, despite improvements in survival there is no evidence in improvement in quality of survival in relation to IQ for cohorts born more recently, suggesting the adult findings reported here are likely highly applicable to VP/VLBW infants born more recently.

The unavailability of certain perinatal indicators or cohort information is another limitation of using historical data, especially when important confounding variables may be missing. For example, infants with chronic lung disease (BPD) are often given postnatal corticosteroids as a form of treatment (DeMauro et al., 2014; Doyle et al., 2014). However, the use of postnatal corticosteroids has also been linked to elements of neurosensory impairment such as cerebral palsy (Barrington, 2001; DeMauro et al., 2014). Due to a lack of harmonizable data from each cohort on steroid use, the estimate for BPD being associated with lower IQ may therefore be confounded or influenced by whether postnatal corticosteroids were administered. Overall, the thesis was able to determine the universal association of some but not all key neonatal factors for long term VP/VLBW cognitive performance. The degree to which these neonatal factors should be thought of as "causal" should naturally be limited, due to the use of observational data rather than a randomised control trials where factors such as steroid use can be systematically tested (Doyle et al., 2014).

Potentially important missing variables were also seen in study 3. As previously discussed, the ability to determine IUGR would have potentially increased specificity for linking growth restriction and long-term cognitive outcomes. In study 2, rather than missing perinatal information, it would have been beneficial to have had a measure of processing speed for both cohorts, as it has been indicated as a potential key risk factor for attention problems of VP/VLBW individuals (Mulder et al., 2011). While executive functioning has been suggested as a building block for IQ, some have postulated that processing speed is a key lower cognitive process for both executive functions and IQ (Salthouse, 1996; Sheppard & Vernon, 2008). Processing speed has been found to be slower in VP/VLBW children and has been suggested to be an important explanatory factor for VP/VLBW's lower executive functioning performance and academic functioning (Aarnoudse-Moens et al., 2012; Mulder et al., 2010). While in the BLS, adult measures of processing speed had smaller effect sizes than IQ (Eryigit Madzwamuse et al., 2015), there was not an equivalent measure in EPICure. Therefore, processing speed could not be included as a predictor in study 2. Overall, it may be that some specific cognitive factors are key to adult VP/VLBW attention problems, but they were unfortunately not able to be tested in the study. Overall, these are limitations of secondary data analysis and as such cannot be easily avoided.

It must also be considered that the cohorts used do not represent all VP/VLBW adults. Firstly, the high loss to follow up in almost all of the cohort studies is potentially a limitation. As attrition is often systematic, this may mean that the reported differences between VP/VLBW adults and controls is an underestimation of the true differences between groups. However, the effect this would have on the regression models is likely minimal (Wolke, Waylen, et al., 2009) and there was no effect after controlling for rates of VP/VLBW attrition in study 1. Secondly, all cohorts were from high income countries, and the majority from western and northern Europe. Whether the findings are applicable to VP/VLBW individuals born in countries with less advanced healthcare systems is debatable (Chawanpaiboon et al., 2019). To ensure findings are as representative as possible, work must continue to aim for the highest possible retention rates, potentially improved through new techniques such as contacting

participants via social media (Haikerwal et al., 2020), and look to investigate VP/VLBW outcomes in lower and middle income countries. Statistically, some degree of missing data can also be dealt with by using imputation. This was performed in all three research chapters using multiple imputation by chained equations (MICE) (Buuren & Groothuis-Oudshoorn, 2011). Whether this is the optimal statistical approach for dealing with missing data, especially longitudinal data, is an area of continued debate (Genolini et al., 2013; Zhang, 2016).

It could also be argued that as the research analysed data from VP/VLBW and term born controls, the impact of being born moderately/mildly preterm on cognitive and behavioural outcomes was not considered. As arguments regarding the dose-response relationship between gestational age and IQ were made, it would have been advantageous to have the full gestational age distribution when testing this. While studies have found throughout the entire gestational age distribution, each week gained is beneficial for factors such as mortality and education (Deb-Rinker et al., 2015; D'Onofrio et al., 2013; Quigley et al., 2012), the research in this thesis specifically focused on those born VP/VLBW and thus were not able to ascertain this regarding cognition. In addition, only linear effects for gestational age on IQ were tested. However, this may be an oversimplification with periods where a week gained in gestational age may be particularly beneficial. For example, nearer the threshold of viability between 24 and 25 weeks, there may be greater benefit to IQ per week gained than between 32 and 33 weeks or 39 and 40 weeks. In one study from childhood, it was found that using piecewise regressions, there was a dose response between gestational age and IQ but this was only for gestations from 23 to 33 weeks, with no dose response for gestations greater than 33 weeks (Wolke, Strauss, et al., 2015). However, there is substantial evidence that those born moderately preterm have lower cognitive functioning than term born controls in childhood but that evidence within adulthood is sparser and inconsistent (de Jong et al., 2012). In addition, a linear effect is also likely an oversimplification as it would suggest that those born post term should outperform those born at term. However, research has found that post term birth is associated with both a higher risk of infant mortality and lower child IQ relative to

those born at 40 weeks, indicating a non-linear effect (Altman et al., 2012; Yang et al., 2010).

9.4   Future Directions and Practical Implications

This thesis contributes and expands the knowledge of the long term outcomes for those born VP/VLBW and SGA. The research highlights the universality of cognitive performance, the stability of cognitive performance throughout development, and the subsequent implications of lower cognitive performance for adult inattention. In light of the current findings from this thesis, there are a number of likely fruitful avenues of future research to explore.

Firstly, the findings of differences in parent, self and experimenter reported attention problems from study 2 need further validation for VP/VLBW adults, ideally through IPD meta-analyses of multiple cohorts. Until further validation is provided, it appears that self-report should not be the only method used to assess the attention problems of VP/VLBW adults as it may lack construct validity by not accurately identifying those with attention problems (Cronbach & Meehl, 1955). In addition, investigating a wider battery of executive functioning measures, such as visuo-spatial working memory rather than verbal working memory, or processing speed would allow for greater confidence in the argument that lower general cognitive performance is a greater predictor of attention problems than any specific executive functioning measure.

While SGA was found to be a significant factor associated with IQ, it may be that other early anthropometric measurements offer superior ability to predict long term cognitive outcomes. As previously noted, low birthweight has been historically used as an indication of a foetus being at risk (Camerota & Bollen, 2016), which has been further differentiated into preterm birth or SGA (Hughes et al., 2017). It may be that other anthropometric factors have greater importance and need to be integrated, especially for cognition. For example, catch up growth in childhood has suggested to be an important factor for the long term cognitive performance of SGA individuals

(Brandt et al., 2003), however this has not been further validated (Beyerlein et al., 2010). Alternatively, head circumference within infancy and childhood has been associated with long term cognitive development for VP/VLBW individuals (Jaekel et al., 2019; Sammallahti et al., 2014). This is potentially more theoretically valid than using birth weight. As cognitive performance is naturally a result of brain functioning, it is logical that head circumference may act as an accurate proxy indicator for early brain growth (Gale et al., 2006). Whether this is solely an indicator for later IQ or if it is also predictive of other cognitive, behavioural, or social outcomes should be further considered. In order to perform this research, future VP/VLBW cohorts will need to collect measurements of head circumference not just at birth but regularly at follow up, as this data was not available in all cohorts included in the IPD analysis.

Further research into the long term cognitive and attention outcomes of being born VP/VLBW or SGA would also benefit from the greater use of sibling, twin and genetic studies (Lindström et al., 2011), potentially under a differential susceptibility framework. Twins are naturally matched on gestational age but can differ on birthweight (Torche & Echevarría, 2011). This means the relative importance of SGA status and birthweight Z scores can be more accurately determined by comparing twin differences on birthweight and a long term outcomes such as IQ or attention. This has the benefit of controlling for a number of socioenvironmental factors, superior to comparing participants in different families, raised in potentially vastly different environments. For example, by controlling for environmental factors in a more systematic way, twin studies have found a 1SD increase in birthweight Z scores is associated with a 15% increase in maths performance (Torche & Echevarría, 2011) or having a birthweight 20% higher is associated with 0.10 to 0.24 SD reduction in attention problems (Groen-Blokhuis et al., 2011). While both of these studies have found that lower birthweight is associated with poorer outcomes, it may be that low birthweight should instead be thought of as a marker for being more sensitive to their environment, under a "for better or for worse" differential susceptibility model (Pluess & Belsky, 2011). Thus, it should be considered that when the environment is more optimal, does the lower birthweight twin outperform the higher birthweight twin on measures such as IQ or attention.

While the research in this thesis has included data from infancy, childhood and early adulthood, there is the possibility that this is still not the complete picture for long term cognitive performance of VP/VLBW individuals. In the general field of ageing, there is evidence that those with perinatal complications, including being born preterm or SGA, were found to subjectively look older and objectively have shorter telomeres - the "biological clock of cellular aging" (Shalev et al., 2014). Specifically on brain health, initial evidence suggests that the "brain age" of VP/VLBW adults is older than term born controls, indicating they may show earlier or more sustained cognitive decline into later adulthood (Hedderich et al., 2021; Karolis et al., 2017). This is further supported by the findings that lower ponderal index at birth, a similar measure to BMI, is associated with smaller brain volume and less white matter at 75 years of age (Muller et al., 2014). Identifying how VP/VLBW adults perform cognitively relative to the general population into later adulthood is imperative. While the economic costs associated with VP/VLBW largely focus on neonatal and early educational costs, this may be an underestimation of the cost across the lifespan if the cognitive decline is either more severe or earlier (Mangham et al., 2009).

Regarding practical implications from this thesis, some have already been mentioned in the overall discussion. These include the avoidance of solely EF based interventions for VP/VLBW individuals with attention problems, the use of child IQ as a simple screening tool for adult cognitive and behavioural problems, and the pitfalls of solely using self-report for assessing adult attention problems. Another important implication from this research is how it is decided which preterm children specifically should be the focus for early neurodevelopmental screening. In the UK, the national institute for health and care excellence (NICE) currently provides guidelines and criteria regarding this (National Institute for Health and Care Excellence, 2017). The NICE guidelines currently suggest that those born <30 weeks gestation or those <36 weeks gestation with a neonatal risk factor (e.g., IVH grade 3 or 4) should be screened at 2 years of age. While the NICE guidelines acknowledge that SGA and maternal SES are both independent risk factors for intellectual disability, these factors are not then used in the criteria for whether a screening at 2 years should take place. In light of the findings

from this thesis, especially the relatively large effect sizes of maternal education/SES for cognitive development, this should potentially be reconsidered.

9.5   Concluding Remarks

This thesis aimed to accurately identify the foetal origins of adult neurocognitive performance and their relative importance. In particular, it looked to identify the universality of findings, how cognition changes over time and what consequences cognitive performance has for other behavioural outcomes. The work in this thesis demonstrated the pervasive nature that early neonatal factors have on long term cognitive and behavioural outcomes. The effects of VP/VLBW or SGA on cognitive performance are large, universal, and has consequences for a number of life course outcomes. In addition, while these two factors do not appear to moderate one another, they do both have independent effects that in combination result in significantly lower cognitive performance. When in combination with socio-environmental risk factors, the accumulation of all these risk factors on cognitive development is substantial. The stability and reliability of these findings indicate the need to accurately identify infants at most risk and to optimise their external environment. This may include parenting or school interventions.

Low birthweight was historically used as the key marker to determine an infant at neonatal risk. In this thesis it has been shown that two subgroups of low birthweight infants, those born VP/VLBW or SGA, are at continued risk into early adulthood for lower cognitive and behavioural outcomes. Crucially, there is variability both between and within these two subgroups. The variability between these groups indicates the continued need to include both factors in analyses as they have been found to both have independent effects on IQ. The variability within these subgroups may be partially explained by parental and environmental factors which may improve cognitive outcomes. Further unpicking the synergies between VP/VLBW, SGA, and the environment for long term cognitive performance is most certainly warranted.

References

Aanes, S., Bjuland, K. J., Skranes, J., & Løhaugen, G. C. C. (2015). Memory function and hippocampal volumes in preterm born very-low-birth-weight (VLBW) young adults. *NeuroImage*, *105*, 76–83. https://doi.org/10.1016/j.neuroimage.2014.10.023

Aarnoudse-Moens, C. S. H., Duivenvoorden, H. J., Weisglas-Kuperus, N., Goudoever, J. B. V., & Oosterlaan, J. (2012). The profile of executive function in very preterm children at 4 to 12 years. *Developmental Medicine & Child Neurology*, *54*(3), 247–253. https://doi.org/10.1111/j.1469-8749.2011.04150.x

Aarnoudse-Moens, C. S. H., Weisglas-Kuperus, N., Duivenvoorden, H. J., Goudoever, J. B. van, & Oosterlaan, J. (2013). Executive Function and IQ Predict Mathematical and Attention Problems in Very Preterm Children. *PLOS ONE*, *8*(2), e55994. https://doi.org/10.1371/journal.pone.0055994

Aarnoudse-Moens, C. S. H., Weisglas-Kuperus, N., Goudoever, J. B. van, & Oosterlaan, J. (2009). Meta-Analysis of Neurobehavioral Outcomes in Very Preterm and/or Very Low Birth Weight Children. *Pediatrics*, *124*(2), 717–728. https://doi.org/10.1542/peds.2008-2816

Alberry, M., & Soothill, P. (2007). Management of fetal growth restriction. *Archives of Disease in Childhood. Fetal and Neonatal Edition*, *92*(1), F62–F67. https://doi.org/10.1136/adc.2005.082297

Albu, A., Horhoianu, I., Dumitrascu, M., & Horhoianu, V. (2014). Growth assessment in diagnosis of Fetal Growth Restriction. Review. *Journal of Medicine and Life*, *7*(2), 150–154.

Allotey, J., Zamora, J., Cheong-See, F., Kalidindi, M., Arroyo-Manzano, D., Asztalos, E., Van Der Post, J., Mol, B. W., Moore, D., & Birtles, D. (2018). Cognitive, motor, behavioural and academic performances of children born preterm: A meta-analysis and systematic review involving 64 061 children. *BJOG: An International Journal of Obstetrics & Gynaecology*, *125*(1), 16–25.

Alloway, T. P. (2010). Working memory and executive function profiles of individuals with borderline intellectual functioning. *Journal of Intellectual Disability Research*, *54*(5), 448–456. https://doi.org/10.1111/j.1365-2788.2010.01281.x

Alloway, T. P., & Alloway, R. G. (2010). Investigating the predictive roles of working memory and IQ in academic attainment. *Journal of Experimental Child Psychology*, *106*(1), 20–29. https://doi.org/10.1016/j.jecp.2009.11.003

Altman, M., Edstedt Bonamy, A.-K., Wikström, A.-K., & Cnattingius, S. (2012). Cause-specific infant mortality in a population-based Swedish study of term and post-term births: The contribution of gestational age and birth weight. *BMJ Open*, *2*(4), e001152. https://doi.org/10.1136/bmjopen-2012-001152

American College of Obstetricians and Gynecologists. (2017). Methods for estimating the due date. Committee Opinion No. 700. *Obstet Gynecol*, *129*(5), e150–e154.

Ananth, C. V., & Vintzileos, A. M. (2009). Distinguishing pathological from constitutional small for gestational age births in population-based studies. *Early Human Development*, *85*(10), 653–658. https://doi.org/10.1016/j.earlhumdev.2009.09.004

Anderson, P. J. (2002). Assessment and Development of Executive Function (EF) During Childhood. *Child Neuropsychology*, *8*(2), 71–82. https://doi.org/10.1076/chin.8.2.71.8724

Anderson, P. J. (2003). Neurobehavioral Outcomes of School-age Children Born Extremely Low Birth Weight or Very Preterm in the 1990s. *JAMA*, *289*(24), 3264. https://doi.org/10.1001/jama.289.24.3264

Anderson, P. J., & Doyle, L. W. (2008). Cognitive and Educational Deficits in Children Born Extremely Preterm. *Seminars in Perinatology*, *32*(1), 51–58. https://doi.org/10.1053/j.semperi.2007.12.009

Anderson, P. J., Lee, K. J., Roberts, G., Spencer-Smith, M. M., Thompson, D. K., Seal, M. L., Nosarti, C., Grehan, A., Josev, E. K., Gathercole, S., Doyle, L. W., & Pascoe, L. (2018). Long-Term Academic Functioning Following Cogmed Working Memory Training for Children Born Extremely Preterm: A Randomized Controlled Trial. *The Journal of Pediatrics*, *202*, 92-97.e4. https://doi.org/10.1016/j.jpeds.2018.07.003

Archambeau, K., & Gevers, W. (2018). Chapter 16—(How) Are Executive Functions Actually Related to Arithmetic Abilities? In A. Henik & W. Fias (Eds.), *Heterogeneity of Function in Numerical Cognition* (pp. 337–357). Academic Press. https://doi.org/10.1016/B978-0-12-811529-9.00016-9

Ardila, A., Pineda, D., & Rosselli, M. (2000). Correlation between intelligence test scores and executive function measures. *Archives of Clinical Neuropsychology*, *15*(1), 31–36.

Arpi, E., & Ferrari, F. (2013). Preterm birth and behaviour problems in infants and preschool-age children: A review of the recent literature. *Developmental Medicine & Child Neurology*, *55*(9), 788–796.

Aylward, G. P. (2009). Developmental Screening and Assessment: What Are We Thinking? *Journal of Developmental*, *30*(2), 169–173. https://doi.org/10.1097/DBP.0b013e31819f1c3e

Back, S. A., & Miller, S. P. (2014). Brain Injury in Premature Neonates: A Primary Cerebral Dysmaturation Disorder? *Annals of Neurology*, *75*(4), 469–486. https://doi.org/10.1002/ana.24132

Balduzzi, S., Rücker, G., & Schwarzer, G. (2019). How to perform a meta-analysis with R: A practical tutorial. *Evidence Based Mental Health*, *22*(4), 153–160. https://doi.org/10.1136/ebmental-2019-300117

Barker, D. J., Osmond, C., & Law, C. M. (1989). The intrauterine and early postnatal origins of cardiovascular disease and chronic bronchitis. *Journal of Epidemiology and Community Health*, *43*(3), 237–240.

Barrera, M. E., Rosenbaum, P. L., & Cunningham, C. E. (1987). Corrected and uncorrected Bayley scores: Longitudinal developmental patterns in low and high birth weight preterm infants. *Infant Behavior and Development*, *10*(3), 337–346. https://doi.org/10.1016/0163-6383(87)90021-X

Barrington, K. J. (2001). The adverse neuro-developmental effects of postnatal steroids in the preterm infant: A systematic review of RCTs. *BMC Pediatrics*, *1*(1), 1. https://doi.org/10.1186/1471-2431-1-1

Basten, M., Jaekel, J., Johnson, S., Gilmore, C., & Wolke, D. (2015). Preterm Birth and Adult Wealth: Mathematics Skills Count. *Psychological Science*, *26*(10), 1608–1619. https://doi.org/10.1177/0956797615596230

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bauer, A. (1988). Ein Verfahren zur Messung des für das Bildungsverhalten relevanten Status (BRSS) [A procedure for the measurement of social status related to educational behaviour (BRSS)]. *Frankfurt: Deutsches Institut Für Internationale Pädagogische Forschung*.

Beeckman, K., Louckx, F., Downe, S., & Putman, K. (2013). The relationship between antenatal care and preterm birth: The importance of content of care. *European Journal of Public Health*, *23*(3), 366–371. https://doi.org/10.1093/eurpub/cks123

Beery, K. E., & Buktenica, N. A. (1982). *Administration, scoring, and teaching manual for the Developmental Test of Visual-Motor Integration*. Modern Curriculum Press Cleveland, OH.

Benavente-Fernández, I., Synnes, A., Grunau, R. E., Chau, V., Ramraj, C., Glass, T., Cayam-Rand, D., Siddiqi, A., & Miller, S. P. (2019). Association of Socioeconomic Status and Brain Injury With Neurodevelopmental Outcomes of Very Preterm Children. *JAMA Network Open*, *2*(5), e192914. https://doi.org/10.1001/jamanetworkopen.2019.2914

Benavides, A., Metzger, A., Tereschenko, S., Conrad, A., Bell, E. F., Spencer, J., Ross-Sheehy, S., Georgieff, M., Magnotta, V., & Nopoulos, P. (2019). Sex

Specific Alterations in Preterm Brain. *Pediatric Research*, *85*(1), 55–62. https://doi.org/10.1038/s41390-018-0187-5

Best, J. R., & Miller, P. H. (2010). A Developmental Perspective on Executive Function. *Child Development*, *81*(6), 1641–1660. https://doi.org/10.1111/j.1467-8624.2010.01499.x

Beyerlein, A., Ness, A. R., Streuling, I., Hadders-Algra, M., & von Kries, R. (2010). Early rapid growth: No association with later cognitive functions in children born not small for gestational age. *The American Journal of Clinical Nutrition*, *92*(3), 585–593. https://doi.org/10.3945/ajcn.2009.29116

Bhutta, A. T., Cleves, M. A., Casey, P. H., Cradock, M. M., & Anand, K. J. S. (2002). Cognitive and Behavioral Outcomes of School-Aged Children Who Were Born Preterm: A Meta-analysis. *JAMA*, *288*(6), 728–737. https://doi.org/10.1001/jama.288.6.728

Bickle Graz, M., Tolsa, J.-F., & Fischer Fumeaux, C. J. (2015). Being Small for Gestational Age: Does it Matter for the Neurodevelopment of Premature Infants? A Cohort Study. *PLoS ONE*, *10*(5). https://doi.org/10.1371/journal.pone.0125769

Bie, H. M. A. de, Oostrom, K. J., & Waal, H. A. D. de. (2010). Brain Development, Intelligence and Cognitive Outcome in Children Born Small for Gestational Age. *Hormone Research in Paediatrics*, *73*(1), 6–14. https://doi.org/10.1159/000271911

Bilgin, A., Mendonca, M., & Wolke, D. (2018). Preterm Birth/Low Birth Weight and Markers Reflective of Wealth in Adulthood: A Meta-analysis. *Pediatrics*, *142*(1). https://doi.org/10.1542/peds.2017-3625

Binet, A., & Simon, T. (1916). *New methods for the diagnosis of the intellectual level of subnormals.(L'Année Psych., 1905, pp. 191-244).*

Blencowe, H., Cousens, S., Chou, D., Oestergaard, M., Say, L., Moller, A.-B., Kinney, M., & Lawn, J. (2013). Born Too Soon: The global epidemiology of 15 million preterm births. *Reproductive Health*, *10*(Suppl 1), S2. https://doi.org/10.1186/1742-4755-10-S1-S2

Blencowe, H., Lee, A. C., Cousens, S., Bahalim, A., Narwal, R., Zhong, N., Chou, D., Say, L., Modi, N., Katz, J., Vos, T., Marlow, N., & Lawn, J. E. (2013). Preterm birth–associated neurodevelopmental impairment estimates at regional and global levels for 2010. *Pediatric Research*, *74*(Suppl 1), 17–34. https://doi.org/10.1038/pr.2013.204

Blondel, B., Morin, I., Platt, R. W., Kramer, M. S., Usher, R., & Bréart, G. (2002). Algorithms for combining menstrual and ultrasound estimates of gestational age: Consequences for rates of preterm and postterm birth. *BJOG: An International Journal of Obstetrics and Gynaecology*, *109*(6), 718–720. https://doi.org/10.1016/S1470-0328(02)01068-6

Boake, C. (2002). From the Binet–Simon to the Wechsler–Bellevue: Tracing the history of intelligence testing. *Journal of Clinical and Experimental Neuropsychology*, *24*(3), 383–405.

Bolisetty, S., Dhawan, A., Abdel-Latif, M., Bajuk, B., Stack, J., Oei, J.-L., & Lui, K. (2014). *Intraventricular Hemorrhage and Neurodevelopmental Outcomes in Extreme Preterm Infants*. *133*(1), 10.

Bonamy, A. K. E., Zeitlin, J., Piedvache, A., Maier, R. F., Heijst, A. van, Varendi, H., Manktelow, B. N., Fenton, A., Mazela, J., Cuttini, M., Norman, M., Petrou, S.,

Reempts, P. V., Barros, H., & Draper, E. S. (2019). Wide variation in severe neonatal morbidity among very preterm infants in European regions. *Archives of Disease in Childhood - Fetal and Neonatal Edition*, *104*(1), F36–F45. https://doi.org/10.1136/archdischild-2017-313697

Bouchard, T. J. (2013). The Wilson Effect: The Increase in Heritability of IQ With Age. *Twin Research and Human Genetics*, *16*(5), 923–930. https://doi.org/10.1017/thg.2013.54

Boyle, A. K., Rinaldi, S. F., Norman, J. E., & Stock, S. J. (2017). Preterm birth: Inflammation, fetal injury and treatment strategies. *Journal of Reproductive Immunology*, *119*, 62–66. https://doi.org/10.1016/j.jri.2016.11.008

Boyle, M. H., Miskovic, V., Lieshout, R. V., Duncan, L., Schmidt, L. A., Hoult, L., Paneth, N., & Saigal, S. (2011). Psychopathology in young adults born at extremely low birth weight. *Psychological Medicine*, *41*(8), 1763–1774. https://doi.org/10.1017/S0033291710002357

Brandt, I., Sticker, E. J., & Lentze, M. J. (2003). Catch-up growth of head circumference of very low birth weight, small for gestational age preterm infants and mental development to adulthood. *The Journal of Pediatrics*, *142*(5), 463–470. https://doi.org/10.1067/mpd.2003.149

Breeman, L. D., Jaekel, J., Baumann, N., Bartmann, P., & Wolke, D. (2015). Preterm Cognitive Function Into Adulthood. *Pediatrics*, *136*(3), 415–423. https://doi.org/10.1542/peds.2015-0608

Breeman, L. D., Jaekel, J., Baumann, N., Bartmann, P., & Wolke, D. (2016). Attention problems in very preterm children from childhood to adulthood: The Bavarian

Longitudinal Study. *Journal of Child Psychology and Psychiatry*, *57*(2), 132–140.

Breeman, L. D., Jaekel, J., Baumann, N., Bartmann, P., & Wolke, D. (2017). Neonatal predictors of cognitive ability in adults born very preterm: A prospective cohort study. *Developmental Medicine & Child Neurology*, *59*(5), 477–483.

Bridgett, D. J., & Walker, M. E. (2006). Intellectual functioning in adults with ADHD: A meta-analytic examination of full scale IQ differences between adults with and without ADHD. *Psychological Assessment*, *18*(1), 1–14. https://doi.org/10.1037/1040-3590.18.1.1

Brion, M.-J. A., Lawlor, D. A., Matijasevich, A., Horta, B., Anselmi, L., Araújo, C. L., Menezes, A. M. B., Victora, C. G., & Smith, G. D. (2011). What are the causal effects of breastfeeding on IQ, obesity and blood pressure? Evidence from comparing high-income with middle-income cohorts. *International Journal of Epidemiology*, *40*(3), 670–680. https://doi.org/10.1093/ije/dyr020

Brydges, C. R., Landes, J. K., Reid, C. L., Campbell, C., French, N., & Anderson, M. (2018). Cognitive outcomes in children and adolescents born very preterm: A meta-analysis. *Developmental Medicine & Child Neurology*, *60*(5), 452–468. https://doi.org/10.1111/dmcn.13685

Burnett, A. C., Scratch, S. E., & Anderson, P. J. (2013). Executive function outcome in preterm adolescents. *Early Human Development*, *89*(4), 215–220.

Burnett, A. C., Youssef, G., Anderson, P. J., Duff, J., Doyle, L. W., & Cheong, J. L. Y. (2019). Exploring the "Preterm Behavioral Phenotype" in Children Born Extremely Preterm: *Journal of Developmental & Behavioral Pediatrics*, *40*(3), 200–207. https://doi.org/10.1097/DBP.0000000000000646

Buuren, S. van, & Groothuis-Oudshoorn, K. (2011). mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software*, *45*(1), 1–67. https://doi.org/10.18637/jss.v045.i03

Camerota, M., & Bollen, K. A. (2016). Birth Weight, Birth Length, and Gestational Age as Indicators of Favorable Fetal Growth Conditions in a US Sample. *PloS One*, *11*(4), e0153800. https://doi.org/10.1371/journal.pone.0153800

Capper, A. (1928). THE FATE AND DEVELOPMENT OF THE IMMATURE AND OF THE PREMATURE CHILD: A CLINICAL STUDY Review of the Literature and Study of Cerebral Hemorrhage in the New-Born Infant. *American Journal of Diseases of Children*, *35*(3), 443–491. https://doi.org/10.1001/archpedi.1928.01920210088010

Cattell, R. B. (1971). *Abilities: Their structure, growth, and action*. Houghton Mifflin.

Charkaluk, M.-L., Marchand-Martin, L., Ego, A., Zeitlin, J., Arnaud, C., Burguet, A., Marret, S., Rozé, J.-C., Vieux, R., Kaminski, M., Ancel, P.-Y., & Pierrat, V. (2012). The Influence of Fetal Growth Reference Standards on Assessment of Cognitive and Academic Outcomes of Very Preterm Children. *The Journal of Pediatrics*, *161*(6), 1053-1058.e1. https://doi.org/10.1016/j.jpeds.2012.05.037

Chauhan, S. P., Rice, M. M., Grobman, W. A., Bailit, J., Reddy, U. M., Wapner, R. J., Varner, M. W., Thorp, J. M., Leveno, K. J., Caritis, S. N., Prasad, M., Tita, A. T. N., Saade, G., Sorokin, Y., Rouse, D. J., & Tolosa, J. E. (2017). Neonatal Morbidity of Small- and Large-for-Gestational-Age Neonates Born at Term in Uncomplicated Pregnancies. *Obstetrics and Gynecology*, *130*(3), 511–519. https://doi.org/10.1097/AOG.0000000000002199

Chawanpaiboon, S., Vogel, J. P., Moller, A.-B., Lumbiganon, P., Petzold, M., Hogan, D., Landoulsi, S., Jampathong, N., Kongwattanakul, K., & Laopaiboon, M. (2019). Global, regional, and national estimates of levels of preterm birth in 2014: A systematic review and modelling analysis. *The Lancet Global Health*, *7*(1), e37–e46.

Cheong, J. L. Y., Anderson, P. J., Burnett, A. C., Roberts, G., Davis, N., Hickey, L., Carse, E., Doyle, L. W., & Group, for the V. I. C. S. (2017). Changing Neurodevelopment at 8 Years in Children Born Extremely Preterm Since the 1990s. *Pediatrics*. https://doi.org/10.1542/peds.2016-4086

Claas, M. J., Bruinse, H. W., Heide-Jalving, M. van der, Termote, J. U. M., & Vries, L. S. de. (2010). Changes in Survival and Neonatal Morbidity in Infants with a Birth Weight of 750 g or Less. *Neonatology*, *98*(3), 278–288. https://doi.org/10.1159/000285715

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155.

Constable, R. T., Vohr, B. R., Scheinost, D., Benjamin, J. R., Fulbright, R. K., Lacadie, C., Schneider, K. C., Katz, K. H., Zhang, H., & Papademetris, X. (2013). A left cerebellar pathway mediates language in prematurely-born young adults. *Neuroimage*, *64*, 371–378.

Costeloe, K. L., Hennessy, E., Gibson, A. T., Marlow, N., & Wilkinson, A. R. (2000). The EPICure study: Outcomes to discharge from hospital for infants born at the threshold of viability. *Pediatrics*, *106*(4), 659–671.

Costeloe, K. L., Hennessy, E. M., Haider, S., Stacey, F., Marlow, N., & Draper, E. S. (2012). Short term outcomes after extreme preterm birth in England:

Comparison of two birth cohorts in 1995 and 2006 (the EPICure studies). *BMJ*, *345*, e7976. https://doi.org/10.1136/bmj.e7976

Coutinho, T. V., Reis, S. P. S., da Silva, A. G., Miranda, D. M., & Malloy-Diniz, L. F. (2018). Deficits in Response Inhibition in Patients with Attention-Deficit/Hyperactivity Disorder: The Impaired Self-Protection System Hypothesis. *Frontiers in Psychiatry*, *8*. https://doi.org/10.3389/fpsyt.2017.00299

Cowan, N. (2008). What are the differences between long-term, short-term, and working memory? *Progress in Brain Research*, *169*, 323–338.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*(4), 281–302. https://doi.org/10.1037/h0040957

Crosson, B., Ford, A., McGregor, K. M., Meinzer, M., Cheshkov, S., Li, X., Walker-Batson, D., & Briggs, R. W. (2010). Functional Imaging and Related Techniques: An Introduction for Rehabilitation Researchers. *Journal of Rehabilitation Research and Development*, *47*(2), vii–xxxiv.

Crowe, S. F. (2000). Does the Letter Number Sequencing Task Measure Anything More Than Digit Span? *Assessment*, *7*(2), 113–117. https://doi.org/10.1177/107319110000700202

Dammann, O., Leviton, A., Gappa, M., & Dammann, C. E. L. (2005). Lung and brain damage in preterm newborns, and their association with gestational age, prematurity subgroup, infection/inflammation and long term outcome. *BJOG: An International Journal of Obstetrics & Gynaecology*, *112*(s1), 4–9. https://doi.org/10.1111/j.1471-0528.2005.00576.x

Darlow, B. A., Horwood, L. J., Woodward, L. J., Elliott, J. M., Troughton, R. W., Elder, M. J., Epton, M. J., Stanton, J. D., Swanney, M. P., Keenan, R., Melzer, T. R., McKelvey, V. A., Levin, K., Meeks, M. G., Espiner, E. A., Cameron, V. A., & Martin, J. (2015). The New Zealand 1986 very low birth weight cohort as young adults: Mapping the road ahead. *BMC Pediatrics*, *15*, 90. https://doi.org/10.1186/s12887-015-0413-9

Darlow, B. A., Woodward, L. J., Levin, K. J., Melzer, T., & Horwood, L. J. (2020). Perinatal and childhood predictors of general cognitive outcome at 28 years in a very-low-birthweight national cohort. *Dev Med Child Neurol*, *n/a*(n/a). https://doi.org/10.1111/dmcn.14649

De Bie, H. M. A., Oostrom, K. J., Boersma, M., Veltman, D. J., Barkhof, F., Delemarre-van de Waal, H. A., & van den Heuvel, M. P. (2011). Global and regional differences in brain anatomy of young children born small for gestational age. *PloS One*, *6*(9), e24116. https://doi.org/10.1371/journal.pone.0024116

De Jesus, L. C., Pappas, A., Shankaran, S., Li, L., Das, A., Bell, E. F., Stoll, B. J., Laptook, A. R., Walsh, M. C., Hale, E. C., Newman, N. S., Bara, R., & Higgins, R. D. (2013). Outcomes of Small for Gestational Age Infants Born at <27 Weeks' Gestation. *The Journal of Pediatrics*, *163*(1), 55-60.e3. https://doi.org/10.1016/j.jpeds.2012.12.097

de Jong, M., Verhoeven, M., & van Baar, A. L. (2012). School outcome, cognitive functioning, and behaviour problems in moderate and late preterm children and adults: A review. *Seminars in Fetal and Neonatal Medicine*, *17*(3), 163–169. https://doi.org/10.1016/j.siny.2012.02.003

de Kieviet, J. F., van Elburg, R. M., Lafeber, H. N., & Oosterlaan, J. (2012). Attention problems of very preterm children compared with age-matched term controls at school-age. *The Journal of Pediatrics*, *161*(5), 824–829. https://doi.org/10.1016/j.jpeds.2012.05.010

de Kleine, M. J. K., den Ouden, A. L., Kollée, L. A. A., Ilsen, A., van Wassenaer, A. G., Brand, R., & Verloove-Vanhorick, S. P. (2007). Lower mortality but higher neonatal morbidity over a decade in very preterm infants. *Paediatric and Perinatal Epidemiology*, *21*(1), 15–25. https://doi.org/10.1111/j.1365-3016.2007.00780.x

de Onis, M., & Habicht, J. P. (1996). Anthropometric reference data for international use: Recommendations from a World Health Organization Expert Committee. *The American Journal of Clinical Nutrition*, *64*(4), 650–658. https://doi.org/10.1093/ajcn/64.4.650

Deary, I. J., Strand, S., Smith, P., & Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence*, *35*(1), 13–21. https://doi.org/10.1016/j.intell.2006.02.001

Deary, I. J., Whiteman, M. C., Starr, J. M., Whalley, L. J., & Fox, H. C. (2004). The Impact of Childhood Intelligence on Later Life: Following Up the Scottish Mental Surveys of 1932 and 1947. *Journal of Personality and Social Psychology*, *86*(1), 130–147. https://doi.org/10.1037/0022-3514.86.1.130

Deb-Rinker, P., León, J. A., Gilbert, N. L., Rouleau, J., Andersen, A.-M. N., Bjarnadóttir, R. I., Gissler, M., Mortensen, L. H., Skjærven, R., Vollset, S. E., Zhang, X., Shah, P., Sauve, R. S., Kramer, M. S., Joseph, K. S., & for the Canadian Perinatal Surveillance System (Public Health Agency of Canada).

(2015). Differences in perinatal and infant mortality in high-income countries: Artifacts of birth registration or evidence of true differences? *BMC Pediatrics*, *15*(1), 112. https://doi.org/10.1186/s12887-015-0430-8

DeMauro, S. B., Dysart, K., & Kirpalani, H. (2014). Stopping the Swinging Pendulum of Postnatal Corticosteroid Use. *The Journal of Pediatrics*, *164*(1), 9–11. https://doi.org/10.1016/j.jpeds.2013.10.020

Dennis, M., Francis, D. J., Cirino, P. T., Schachar, R., Barnes, M. A., & Fletcher, J. M. (2009). Why IQ is not a covariate in cognitive studies of neurodevelopmental disorders. *Journal of the International Neuropsychological Society: JINS*, *15*(3), 331–343. https://doi.org/10.1017/S1355617709090481

Deputy, N. P., Nguyen, P. H., Pham, H., Nguyen, S., Neufeld, L., Martorell, R., & Ramakrishnan, U. (2017). Validity of gestational age estimates by last menstrual period and neonatal examination compared to ultrasound in Vietnam. *BMC Pregnancy and Childbirth*, *17*(1), 25. https://doi.org/10.1186/s12884-016-1192-5

Deter, R. L., Lee, W., Yeo, L., Erez, O., Ramamurthy, U., Naik, M., & Romero, R. (2018). Individualized Growth Assessment: Conceptual Framework and Practical Implementation for the Evaluation of Fetal and Neonatal Growth. *American Journal of Obstetrics and Gynecology*, *218*(2 Suppl), S656–S678. https://doi.org/10.1016/j.ajog.2017.12.210

Diamond, A. (2005). Attention-deficit disorder (attention-deficit/ hyperactivity disorder without hyperactivity): A neurobiologically and behaviorally distinct disorder from attention-deficit/hyperactivity disorder (with hyperactivity).

Development and Psychopathology, *17*(3), 807–825. https://doi.org/10.1017/S0954579405050388

Diamond, A. (2013). Executive Functions. *Annual Review of Psychology*, *64*(1), 135–168. https://doi.org/10.1146/annurev-psych-113011-143750

Diamond, A., & Lee, K. (2011). Interventions shown to Aid Executive Function Development in Children 4–12 Years Old. *Science (New York, N.Y.)*, *333*(6045), 959–964. https://doi.org/10.1126/science.1204529

Ding, G., Tian, Y., Zhang, Y., Pang, Y., Zhang, J., & Zhang, J. (2013). Application of *A global reference for fetal-weight and birthweight percentiles* in predicting infant mortality. *BJOG: An International Journal of Obstetrics & Gynaecology*, *120*(13), 1613–1621. https://doi.org/10.1111/1471-0528.12381

D'Onofrio, B. M., Class, Q. A., Rickert, M. E., Larsson, H., Långström, N., & Lichtenstein, P. (2013). Preterm birth and mortality and morbidity: A population-based quasi-experimental study. *JAMA Psychiatry*, *70*(11), 1231–1240.

Doyle, L. W., Cheong, J. L. Y., Burnett, A. C., Roberts, G., Lee, K. J., Anderson, P. J., & for the Victorian Infant Collaborative Study Group. (2015). Biological and Social Influences on Outcomes of Extreme-Preterm/Low-Birth Weight Adolescents. *PEDIATRICS*, *136*(6), e1513–e1520. https://doi.org/10.1542/peds.2015-2006

Doyle, L. W., Faber, B., Callanan, C., & Morley, R. (2003). Blood Pressure in Late Adolescence and Very Low Birth Weight. *Pediatrics*, *111*(2), 252–257. https://doi.org/10.1542/peds.111.2.252

Doyle, L. W., Halliday, H. L., Ehrenkranz, R. A., Davis, P. G., & Sinclair, J. C. (2014). An update on the impact of postnatal systemic corticosteroids on mortality and cerebral palsy in preterm infants: Effect modification by risk of bronchopulmonary dysplasia. *The Journal of Pediatrics*, *165*(6), 1258–1260. https://doi.org/10.1016/j.jpeds.2014.07.049

Draper, E. S., Manktelow, B., Field, D. J., & James, D. (1999). Prediction of survival for preterm births by weight and gestational age: Retrospective population based study. *BMJ*, *319*(7217), 1093–1097. https://doi.org/10.1136/bmj.319.7217.1093

Du Rietz, E., Cheung, C. H. M., McLoughlin, G., Brandeis, D., Banaschewski, T., Asherson, P., & Kuntsi, J. (2016). Self-report of ADHD shows limited agreement with objective markers of persistence and remittance. *Journal of Psychiatric Research*, *82*, 91–99. https://doi.org/10.1016/j.jpsychires.2016.07.020

Dubowitz, L. M. S., Dubowitz, V., & Goldberg, C. (1970). Clinical assessment of gestational age in the newborn infant. *The Journal of Pediatrics*, *77*(1), 1–10. https://doi.org/10.1016/S0022-3476(70)80038-5

Duncan, G. J., & Magnuson, K. (2012). Socioeconomic status and cognitive functioning: Moving from correlation to causation. *Wiley Interdisciplinary Reviews: Cognitive Science*, *3*(3), 377–386.

Eggert, D. (1972). Die Columbia Mental Maturity Scale als Individualtest für normalentwickelte Kinder im Alter von 3–10 Jahren. *Diagnose Der Minderbegabung. Weinheim: Beltz Verlag*.

Engelhardt, L. E., Mann, F. D., Briley, D. A., Church, J. A., Harden, K. P., & Tucker-Drob, E. M. (2016). Strong genetic overlap between executive functions and intelligence. *Journal of Experimental Psychology. General*, *145*(9), 1141–1159. https://doi.org/10.1037/xge0000195

Engle, R. W. (2010). Role of Working-Memory Capacity in Cognitive Control. *Current Anthropology*, *51*(S1), S17–S26. https://doi.org/10.1086/650572

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics*, *16*(1), 143–149. https://doi.org/10.3758/BF03203267

Eryigit Madzwamuse, S., Baumann, N., Jaekel, J., Bartmann, P., & Wolke, D. (2015). Neuro-cognitive performance of very preterm or very low birth weight adults at 26 years. *Journal of Child Psychology and Psychiatry*, *56*(8), 857–864. https://doi.org/10.1111/jcpp.12358

Evans, N., Hutchinson, J., Simpson, J. M., Donoghue, D., Darlow, B. A., & Henderson-Smart, D. (2007). Prenatal predictors of mortality in very preterm infants cared for in the Australian and New Zealand Neonatal Network. *Archives of Disease in Childhood - Fetal and Neonatal Edition*, *92*(1), F34–F40. https://doi.org/10.1136/adc.2006.094169

Eves, R., Mendonça, M., Bartmann, P., & Wolke, D. (2020). Small for gestational age—cognitive performance from infancy to adulthood: An observational study. *BJOG: An International Journal of Obstetrics & Gynaecology*, *127*(13), 1598–1606. https://doi.org/10.1111/1471-0528.16341

Fan, J., McCandliss, B. D., Sommer, T., Raz, A., & Posner, M. I. (2002). Testing the Efficiency and Independence of Attentional Networks. *Journal of Cognitive Neuroscience*, *14*(3), 340–347. https://doi.org/10.1162/089892902317361886

Farrant, B. M., Fletcher, J., & Maybery, M. T. (2014, January 12). *Cognitive Flexibility, Theory of Mind, and Hyperactivity/Inattention* [Research Article]. Child Development Research; Hindawi. https://doi.org/10.1155/2014/741543

Fenton, T. R., & Kim, J. H. (2013). A systematic review and meta-analysis to revise the Fenton growth chart for preterm infants. *BMC Pediatrics*, *13*(1), 59. https://doi.org/10.1186/1471-2431-13-59

Festinger, L. (1954). A theory of social comparison processes. *Human Relations*, *7*(2), 117–140.

Fitzpatrick, A., Carter, J., & Quigley, M. A. (2016). Association of Gestational Age With Verbal Ability and Spatial Working Memory at Age 11. *Pediatrics*, *138*(6). https://doi.org/10.1542/peds.2016-0578

Flanagan, D. P., & Dixon, S. G. (2014). The Cattell-Horn-Carroll Theory of Cognitive Abilities. In *Encyclopedia of Special Education*. American Cancer Society. https://doi.org/10.1002/9781118660584.ese0431

Flensborg-Madsen, T., & Mortensen, E. L. (2015). Infant developmental milestones and adult intelligence: A 34-year follow-up. *Early Human Development*, *91*(7), 393–400. https://doi.org/10.1016/j.earlhumdev.2015.04.006

Flensborg-Madsen, T., & Mortensen, E. L. (2018). Associations of Early Developmental Milestones With Adult Intelligence. *Child Development*, *89*(2), 638–648. https://doi.org/10.1111/cdev.12760

Flynn, J. R. (2007). *What is intelligence? Beyond the Flynn effect*. Cambridge University Press. https://doi.org/10.1017/CBO9780511605253

Franz, A. P., Bolat, G. U., Bolat, H., Matijasevich, A., Santos, I. S., Silveira, R. C., Procianoy, R. S., Rohde, L. A., & Moreira-Maia, C. R. (2018). Attention-Deficit/Hyperactivity Disorder and Very Preterm/Very Low Birth Weight: A Meta-analysis. *Pediatrics*, *141*(1). https://doi.org/10.1542/peds.2017-1645

Friedman, N. P., Miyake, A., Corley, R. P., Young, S. E., DeFries, J. C., & Hewitt, J. K. (2006). Not All Executive Functions Are Related to Intelligence. *Psychological Science*, *17*(2), 172–179. https://doi.org/10.1111/j.1467-9280.2006.01681.x

Fry, A. F., & Hale, S. (2000). Relationships among processing speed, working memory, and fluid intelligence in children. *Biological Psychology*, *54*(1), 1–34. https://doi.org/10.1016/S0301-0511(00)00051-X

Gale, C. R., O'Callaghan, F. J., Bredow, M., Martyn, C. N., & Team, the A. L. S. of P. and C. S. (2006). The Influence of Head Growth in Fetal Life, Infancy, and Childhood on Intelligence at the Ages of 4 and 8 Years. *Pediatrics*, *118*(4), 1486–1492. https://doi.org/10.1542/peds.2005-2629

Ganzevoort, W., Thornton, J. G., Marlow, N., Thilaganathan, B., Arabin, B., Prefumo, F., Lees, C., & Wolf, H. (2020). Comparative analysis of 2-year outcomes in GRIT and TRUFFLE trials. *Ultrasound in Obstetrics & Gynecology*, *55*(1), 68–74. https://doi.org/10.1002/uog.20354

Gardosi, J., Francis, A., Turner, S., & Williams, M. (2018). Customized growth charts: Rationale, validation and clinical benefits. *American Journal of Obstetrics & Gynecology*, *218*(2), S609–S618. https://doi.org/10.1016/j.ajog.2017.12.011

Genolini, C., Écochard, R., & Jacqmin-Gadda, H. (2013). Copy Mean: A New Method to Impute Intermittent Missing Values in Longitudinal Studies. *Open Journal of Statistics*, *3*(4), 26–40. https://doi.org/10.4236/ojs.2013.34A004

Gien, J., & Kinsella, J. P. (2011). Pathogenesis and Treatment of Bronchopulmonary Dysplasia. *Current Opinion in Pediatrics*, *23*(3), 305. https://doi.org/10.1097/MOP.0b013e328346577f

Girault, J. B., Langworthy, B. W., Goldman, B. D., Stephens, R. L., Cornea, E., Steven Reznick, J., Fine, J., & Gilmore, J. H. (2018). The predictive value of developmental assessments at 1 and 2 for intelligence quotients at 6. *Intelligence*, *68*, 58–65. https://doi.org/10.1016/j.intell.2018.03.003

Glass, H. C., Costarino, A. T., Stayer, S. A., Brett, C., Cladis, F., & Davis, P. J. (2015). Outcomes for Extremely Premature Infants. *Anesthesia and Analgesia*, *120*(6), 1337–1351. https://doi.org/10.1213/ANE.0000000000000705

Goldenberg, R., Culhane, J. F., Iams, J. D., & Romero, R. (2008). Epidemiology and causes of preterm birth. *The Lancet*, *371*(9606), 75–84. https://doi.org/10.1016/S0140-6736(08)60074-4

Goldenberg, R., Cutter, G. R., Hoffman, H. J., Foster, J. M., Nelson, K. G., & Hauth, J. C. (1989). Intrauterine growth retardation: Standards for diagnosis. *American Journal of Obstetrics and Gynecology*, *161*(2), 271–277. https://doi.org/10.1016/0002-9378(89)90497-3

Gottfried, A. W., Gottfried, A. E., & Guerin, D. W. (2006). The Fullerton Longitudinal Study: A Long-Term Investigation of Intellectual and Motivational Giftedness. *Journal for the Education of the Gifted*, *29*(4), 430–450. https://doi.org/10.4219/jeg-2006-244

Grant, D. A., & Berg, E. (1948). A behavioral analysis of degree of reinforcement and ease of shifting to new responses in a Weigl-type card-sorting problem. *Journal of Experimental Psychology*, *38*(4), 404–411. https://doi.org/10.1037/h0059831

Griffiths, R. (1970). *The abilities of young children. A study in mental measurement*. London: University of London Press.

Griffiths, R., & Brandt, I. (1983). *[Entwicklungsskalen (GES) zur Beurteilung der Entwicklung in den ersten beiden Lebensjahren]; Griffiths Entwicklungsskalen (GES) zur Beurteilung der Entwicklung in den ersten beiden Lebensjahren.* Beltz.

Groen-Blokhuis, M. M., Middeldorp, C. M., Beijsterveldt, C. E. M. van, & Boomsma, D. I. (2011). Evidence for a Causal Association of Low Birth Weight and Attention Problems. *Journal of the American Academy of Child & Adolescent Psychiatry*, *50*(12), 1247-1254.e2. https://doi.org/10.1016/j.jaac.2011.09.007

Grothe, M. J., Scheef, L., Bäuml, J., Meng, C., Daamen, M., Baumann, N., Zimmer, C., Teipel, S., Bartmann, P., Boecker, H., Wolke, D., Wohlschläger, A., & Sorg, C. (2017). Reduced Cholinergic Basal Forebrain Integrity Links Neonatal Complications and Adult Cognitive Deficits After Premature Birth. *Biological Psychiatry*, *82*(2), 119–126. https://doi.org/10.1016/j.biopsych.2016.12.008

Guellec, I., Lapillonne, A., Renolleau, S., Charlaluk, M.-L., Roze, J.-C., Marret, S., Vieux, R., Monique, K., Ancel, P.-Y., & Group, the E. S. (2011). Neurologic Outcomes at School Age in Very Preterm Infants Born With Severe or Mild

Growth Restriction. *Pediatrics*, *127*(4), e883–e891. https://doi.org/10.1542/peds.2010-2442

Guidi, J., Lucente, M., Sonino, N., & Fava, G. A. (2020). Allostatic Load and Its Impact on Health: A Systematic Review. *Psychother Psychosom*, 17.

Gutbrod, T., Wolke, D., Soehne, B., Ohrt, B., & Riegel, K. (2000). Effects of gestation and birth weight on the growth and development of very low birthweight small for gestational age infants: A matched group comparison. *Archives of Disease in Childhood - Fetal and Neonatal Edition*, *82*(3), F208–F214. https://doi.org/10.1136/fn.82.3.F208

Hack, M. (2006). Young adult outcomes of very-low-birth-weight children. *Seminars in Fetal and Neonatal Medicine*, *11*(2), 127–137. https://doi.org/10.1016/j.siny.2005.11.007

Hack, M., Flannery, D. J., Schluchter, M., Cartar, L., Borawski, E., & Klein, N. (2002). Outcomes in Young Adulthood for Very-Low-Birth-Weight Infants. *New England Journal of Medicine*, *346*(3), 149–157. https://doi.org/10.1056/NEJMoa010856

Hadlock, F. P., Harrist, R. B., & Martinez-Poyer, J. (1991). In utero analysis of fetal growth: A sonographic weight standard. *Radiology*, *181*(1), 129–133. https://doi.org/10.1148/radiology.181.1.1887021

Haikerwal, A., Doyle, L. W., Patton, G., Wark, J. D., Garland, S. M., Burnett, A. C., & Cheong, J. L. Y. (2020). Using Facebook to Improve Participation Among 25-Year-Olds Enrolled in a Longitudinal Preterm Birth Cohort Study. *Academic Pediatrics*, *20*(7), 1029–1036.

Hallin, A.-L., Hellström-Westas, L., & Stjernqvist, K. (2010). Follow-up of adolescents born extremely preterm: Cognitive function and health at 18 years of age. *Acta Paediatrica (Oslo, Norway: 1992)*, *99*(9), 1401–1406. https://doi.org/10.1111/j.1651-2227.2010.01850.x

Harms, M. B., Zayas, V., Meltzoff, A. N., & Carlson, S. M. (2014). Stability of executive function and predictions to adaptive behavior from middle childhood to pre-adolescence. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.00331

Harrison, M. S., & Goldenberg, R. (2016). Global burden of prematurity. *Seminars in Fetal and Neonatal Medicine*, *21*(2), 74–79.

Hedderich, D. M., Menegaux, A., Schmitz-Koep, B., Nuttall, R., Zimmermann, J., Schneider, S. C., Bäuml, J. G., Daamen, M., Boecker, H., Wilke, M., Zimmer, C., Wolke, D., Bartmann, P., Sorg, C., & Gaser, C. (2021). Increased brain age gap estimate (BrainAGE) in young adults after premature birth. *Frontiers in Aging Neuroscience*, *13*. https://doi.org/10.3389/fnagi.2021.653365

Hegelund, E. R., Flensborg-Madsen, T., Dammeyer, J., & Mortensen, E. L. (2018). Low IQ as a predictor of unsuccessful educational and occupational achievement: A register-based study of 1,098,742 men in Denmark 1968–2016. *Intelligence*, *71*, 46–53. https://doi.org/10.1016/j.intell.2018.10.002

Heinonen, K., Lahti, J., Sammallahti, S., Wolke, D., Lano, A., Andersson, S., Pesonen, A.-K., Eriksson, J. G., Kajantie, E., & Raikkonen, K. (2018). Neurocognitive outcome in young adults born late-preterm. *Developmental Medicine and Child Neurology*, *60*(3), 267–274. https://doi.org/10.1111/dmcn.13616

Hirvonen, M., Ojala, R., Korhonen, P., Haataja, P., Eriksson, K., Gissler, M., Luukkaala, T., & Tammela, O. (2018). Visual and Hearing Impairments After Preterm Birth. *Pediatrics*, *142*(2). https://doi.org/10.1542/peds.2017-3888

Hochberg, Y. (1988). A sharper Bonferroni procedure for multiple tests of significance. *Biometrika*, *75*(4), 800–802. https://doi.org/10.1093/biomet/75.4.800

Houdt, C. A. van, Oosterlaan, J., Wassenaer-Leemhuis, A. G. van, Kaam, A. H. van, & Aarnoudse-Moens, C. S. H. (2019). Executive function deficits in children born preterm or at low birthweight: A meta-analysis. *Developmental Medicine & Child Neurology*, *61*(9), 1015–1024. https://doi.org/10.1111/dmcn.14213

Hovi, P., Andersson, S., Eriksson, J. G., Järvenpää, A.-L., Strang-Karlsson, S., Mäkitie, O., & Kajantie, E. (2007). Glucose Regulation in Young Adults with Very Low Birth Weight. *New England Journal of Medicine*, *356*(20), 2053–2063. https://doi.org/10.1056/NEJMoa067187

Howe, L. D., Tilling, K., Galobardes, B., & Lawlor, D. A. (2013). Loss to Follow-up in Cohort Studies: Bias in Estimates of Socioeconomic Inequalities. *Epidemiology*, *24*(1), 1–9. https://doi.org/10.1097/EDE.0b013e31827623b1

Hughes, M. M., Black, R. E., & Katz, J. (2017). 2500-g Low Birth Weight Cutoff: History and Implications for Future Research and Policy. *Maternal and Child Health Journal*, *21*(2), 283–289. https://doi.org/10.1007/s10995-016-2131-9

Hutcheon, J. A., Zhang, X., Cnattingius, S., Kramer, M. S., & Platt, R. W. (2008). Customised birthweight percentiles: Does adjusting for maternal characteristics matter? *BJOG: An International Journal of Obstetrics &*

Gynaecology, *115*(11), 1397–1404. https://doi.org/10.1111/j.1471-0528.2008.01870.x

Hutton, J. L., Pharoah, P. O. D., Cooke, R. W. I., & Stevenson, R. C. (1997). Differential effects of preterm birth and small gestational age on cognitive and motor development. *Archives of Disease in Childhood - Fetal and Neonatal Edition*, *76*(2), F75–F81. https://doi.org/10.1136/fn.76.2.F75

Iliodromiti, S., Mackay, D. F., Smith, G. C., Pell, J. P., Sattar, N., Lawlor, D. A., & Nelson, S. M. (2017). Customised and noncustomised birth weight centiles and prediction of stillbirth and infant mortality and morbidity: A cohort study of 979,912 term singleton pregnancies in Scotland. *PLoS Medicine*, *14*(1).

Indredavik, M. S., Vik, T., Heyerdahl, S., Kulseng, S., Fayers, P., & Brubakk, A.-M. (2004). Psychiatric symptoms and disorders in adolescents with low birth weight. *Archives of Disease in Childhood - Fetal and Neonatal Edition*, *89*(5), F445–F450. https://doi.org/10.1136/adc.2003.038943

Jacobsson, B., Ahlin, K., Francis, A., Hagberg, G., Hagberg, H., & Gardosi, J. (2008). Cerebral palsy and restricted growth status at birth: Population-based case–control study. *BJOG: An International Journal of Obstetrics & Gynaecology*, *115*(10), 1250–1255. https://doi.org/10.1111/j.1471-0528.2008.01827.x

Jaekel, J., Heuser, K. M., Zapf, A., Roll, C., Nuñez, F. B., Bartmann, P., Wolke, D., Felderhoff-Mueser, U., & Huening, B. (2020). Preterm children's long-term academic performance after adaptive computerized training: An efficacy and process analysis of a randomized controlled trial. *Pediatric Research*, 1–8. https://doi.org/10.1038/s41390-020-01114-w

Jaekel, J., Pluess, M., Belsky, J., & Wolke, D. (2015). Effects of maternal sensitivity on low birth weight children's academic achievement: A test of differential susceptibility versus diathesis stress. *Journal of Child Psychology and Psychiatry*, *56*(6), 693–701. https://doi.org/10.1111/jcpp.12331

Jaekel, J., Sorg, C., Baeuml, J., Bartmann, P., & Wolke, D. (2019). Head Growth and Intelligence from Birth to Adulthood in Very Preterm and Term Born Individuals. *Journal of the International Neuropsychological Society*, *25*(1), 48–56. https://doi.org/10.1017/S135561771800084X

Jensen, E. A., Dysart, K., Gantz, M. G., McDonald, S., Bamat, N. A., Keszler, M., Kirpalani, H., Laughon, M. M., Poindexter, B. B., Duncan, A. F., Yoder, B. A., & Eichenwald, E. C. (2019). The Diagnosis of Bronchopulmonary Dysplasia in Very Preterm Infants. An Evidence-based Approach. *American Journal of Respiratory and Critical Care Medicine*, *200*(6), 751–759. https://doi.org/10.1164/rccm.201812-2348OC

Johnson, S., Kochhar, P., Hennessy, E., Marlow, N., Wolke, D., & Hollis, C. (2016). Antecedents of Attention-Deficit/Hyperactivity Disorder symptoms in children born extremely preterm. *Journal of Developmental and Behavioral Pediatrics : JDBP*, *37*(4), 285–297. https://doi.org/10.1097/DBP.0000000000000298

Johnson, S., & Marlow, N. (2011, May 1). *Preterm Birth and Childhood Psychiatric Disorders* [Special Features]. Pediatric Research. https://doi.org/10.1203/PDR.0b013e318212faa0

Johnson, S., & Marlow, N. (2017). Early and long-term outcome of infants born extremely preterm. *Archives of Disease in Childhood*, *102*(1), 97–102. https://doi.org/10.1136/archdischild-2015-309581

Johnson, S., O'Reilly, H., Ni, Y., Wolke, D., & Marlow, N. (2019). Psychiatric symptoms and disorders in extremely preterm young adults at 19 years of age and longitudinal findings from middle childhood. *Journal of the American Academy of Child & Adolescent Psychiatry*, *58*(8), 820–826. https://doi.org/10.1016/j.jaac.2019.02.020

Johnson, S., & Wolke, D. (2017). Prematurity and low birthweight. In B. Hopkins, E. Geangu, & S. Linkenauger (Eds.), *The Cambridge Encyclopedia of Child Development* (2nd ed., pp. 705–716). Cambridge University Press. https://doi.org/10.1017/9781316216491.112

Juster, R.-P., McEwen, B. S., & Lupien, S. J. (2010). Allostatic load biomarkers of chronic stress and impact on health and cognition. *Neuroscience & Biobehavioral Reviews*, *35*(1), 2–16.

Kajantie, E., Hovi, P., Räikkönen, K., Pesonen, A.-K., Heinonen, K., Järvenpää, A.-L., Eriksson, J. G., Strang-Karlsson, S., & Andersson, S. (2008). Young adults with very low birth weight: Leaving the parental home and sexual relationships--Helsinki Study of Very Low Birth Weight Adults. *Pediatrics*, *122*(1), e62-72. https://doi.org/10.1542/peds.2007-3858

Karolis, V. R., Froudist-Walsh, S., Kroll, J., Brittain, P. J., Tseng, C.-E. J., Nam, K.-W., Reinders, A. A. T. S., Murray, R. M., Williams, S. C. R., Thompson, P. M., & Nosarti, C. (2017). Volumetric grey matter alterations in adolescents and adults born very preterm suggest accelerated brain maturation.

NeuroImage, *163*, 379–389. https://doi.org/10.1016/j.neuroimage.2017.09.039

Katz, J., Lee, A. C., Kozuki, N., Lawn, J. E., Cousens, S., Blencowe, H., Ezzati, M., Bhutta, Z. A., Marchant, T., Willey, B. A., Adair, L., Barros, F., Baqui, A. H., Christian, P., Fawzi, W., Gonzalez, R., Humphrey, J., Huybregts, L., Kolsteren, P., … Black, R. E. (2013). Mortality risk in preterm and small-for-gestational-age infants in low-income and middle-income countries: A pooled country analysis. *The Lancet*, *382*(9890), 417–425. https://doi.org/10.1016/S0140-6736(13)60993-9

Kaufman, A. S., & Kaufman, N. L. (1983). *K-ABC: Kaufman assessment battery for children: Interpretive manual*. American Guidance Service.

Kerr-Wilson, C. O., Mackay, D. F., Smith, G. C. S., & Pell, J. P. (2012). Meta-analysis of the association between preterm delivery and intelligence. *Journal of Public Health*, *34*(2), 209–216. https://doi.org/10.1093/pubmed/fdr024

Kiese, C., & Kozielski, P. M. (1979). *Aktiver Wortschatztest für drei-bis sechsjährige Kinder. AWST 3–6. Beltz*. Weinheim.

Kim, E. T., Singh, K., Moran, A., Armbruster, D., & Kozuki, N. (2018). Obstetric ultrasound use in low and middle income countries: A narrative review. *Reproductive Health*, *15*. https://doi.org/10.1186/s12978-018-0571-y

Kiserud, T., Piaggio, G., Carroli, G., Widmer, M., Carvalho, J., Jensen, L. N., Giordano, D., Cecatti, J. G., Aleem, H. A., Talegawkar, S. A., Benachi, A., Diemert, A., Kitoto, A. T., Thinkhamrop, J., Lumbiganon, P., Tabor, A., Kriplani, A., Perez, R. G., Hecher, K., … Platt, L. D. (2017). The World Health Organization Fetal Growth Charts: A Multinational Longitudinal Study of

Ultrasound Biometric Measurements and Estimated Fetal Weight. *PLOS Medicine*, *14*(1), e1002220. https://doi.org/10.1371/journal.pmed.1002220

Knouse, L. E., Bagwell, C. L., Barkley, R. A., & Murphy, K. R. (2005). Accuracy of self-evaluation in adults with ADHD: Evidence from a driving study. *Journal of Attention Disorders*, *8*(4), 221–234.

Koeppen-Schomerus, G., Eley, T. C., Wolke, D., Gringras, P., & Plomin, R. (2000). The interaction of prematurity with genetic and environmental influences on cognitive development in twins. *The Journal of Pediatrics*, *137*(4), 527–533. https://doi.org/10.1067/mpd.2000.108445

Kooij, J. J. S., Buitelaar, J. K., van den Oord, E. J., Furer, J. W., Rijnders, C. A. T., & Hodiamont, P. P. G. (2005). Internal and external validity of attention-deficit hyperactivity disorder in a population-based sample of adults. *Psychological Medicine*, *35*(6), 817–827.

Kormos, C. E., Wilkinson, A. J., Davey, C. J., & Cunningham, A. J. (2014). Low birth weight and intelligence in adolescence and early adulthood: A meta-analysis. *Journal of Public Health*, *36*(2), 213–224. https://doi.org/10.1093/pubmed/fdt071

Kozuki, N., Katz, J., Lee, A. C., Vogel, J. P., Silveira, M. F., Sania, A., Stevens, G. A., Cousens, S., Caulfield, L. E., Christian, P., Huybregts, L., Roberfroid, D., Schmiegelow, C., Adair, L. S., Barros, F. C., Cowan, M., Fawzi, W., Kolsteren, P., Merialdi, M., … Black, R. E. (2015). Short Maternal Stature Increases Risk of Small-for-Gestational-Age and Preterm Births in Low- and Middle-Income Countries: Individual Participant Data Meta-Analysis and

Population Attributable Fraction. *The Journal of Nutrition*, *145*(11), 2542–2550. https://doi.org/10.3945/jn.115.216374

Kramer, M. S. (1987). Determinants of low birth weight: Methodological assessment and meta-analysis. *Bulletin of the World Health Organization*, *65*(5), 663–737.

Kroll, J., Karolis, V., Brittain, P. J., Tseng, C.-E. J., Froudist-Walsh, S., Murray, R. M., & Nosarti, C. (2017). Real-Life Impact of Executive Function Impairments in Adults Who Were Born Very Preterm. *Journal of the International Neuropsychological Society*, *23*(5), 381–389. https://doi.org/10.1017/S1355617717000169

Kroll, J., Karolis, V., Brittain, P. J., Tseng, C.-E. J., Froudist-Walsh, S., Murray, R. M., & Nosarti, C. (2019). Systematic assessment of perinatal and socio-demographic factors associated with IQ from childhood to adult life following very preterm birth. *Intelligence*, *77*, 101401. https://doi.org/10.1016/j.intell.2019.101401

Laerum, A. M. W., Reitan, S. K., Evensen, K. A. I., Lydersen, S., Brubakk, A. M., Skranes, J., & Indredavik, M. S. (2019). Psychiatric symptoms and risk factors in adults born preterm with very low birthweight or born small for gestational age at term. *BMC Psychiatry*, *19*(1), 223. https://doi.org/10.1186/s12888-019-2202-8

Lærum, A. M. W., Reitan, S. K., Evensen, K. A. I., Lydersen, S., Brubakk, A.-M., Skranes, J., & Indredavik, M. S. (2019). Psychiatric symptoms and risk factors in adults born preterm with very low birthweight or born small for gestational age at term. *BMC Psychiatry*, *19*(1), 223. https://doi.org/10.1186/s12888-019-2202-8

Lahat, A., Van Lieshout, R. J., Saigal, S., Boyle, M. H., & Schmidt, L. A. (2014). ADHD among young adults born at extremely low birth weight: The role of fluid intelligence in childhood. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.00446

Lambert, P. C., Sutton, A. J., Abrams, K. R., & Jones, D. R. (2002). A comparison of summary patient-level covariates in meta-regression with individual patient data meta-analysis. *Journal of Clinical Epidemiology*, *55*(1), 86–94. https://doi.org/10.1016/S0895-4356(01)00414-0

Lang, J. M., Lieberman, E., & Cohen, A. (1996). A Comparison of Risk Factors for Preterm Labor and Term Small-for-Gestational-Age Birth. *Epidemiology*, *7*(4), 369–376.

Lean, R. E., Paul, R. A., Smyser, C. D., & Rogers, C. E. (2018). Maternal intelligence quotient (IQ) predicts IQ and language in very preterm children at age 5 years. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *59*(2), 150–159. https://doi.org/10.1111/jcpp.12810

Lee, A. C., Panchal, P., Folger, L., Whelan, H., Whelan, R., Rosner, B., Blencowe, H., & Lawn, J. E. (2017). Diagnostic Accuracy of Neonatal Assessment for Gestational Age Determination: A Systematic Review. *Pediatrics*, *140*(6). https://doi.org/10.1542/peds.2017-1423

Lefebvre, F., Mazurier, E., & Tessier, R. (2005). Cognitive and educational outcomes in early adulthood for infants weighing 1000 grams or less at birth. *Acta Paediatrica (Oslo, Norway: 1992)*, *94*(6), 733–740. https://doi.org/10.1111/j.1651-2227.2005.tb01973.x

Lind, S. E., Bowler, D. M., & Raber, J. (2014). Spatial navigation, episodic memory, episodic future thinking, and theory of mind in children with autism spectrum disorder: Evidence for impairments in mental simulation? *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.01411

Lindström, K., Lindblad, F., & Hjern, A. (2011). Preterm birth and attention-deficit/hyperactivity disorder in schoolchildren. *Pediatrics*, *127*(5), 858–865. https://doi.org/10.1542/peds.2010-1279

Linsell, L., Johnson, S., Wolke, D., O'Reilly, H., Morris, J. K., Kurinczuk, J. J., & Marlow, N. (2018). Cognitive trajectories from infancy to early adulthood following birth before 26 weeks of gestation: A prospective, population-based cohort study. *Archives of Disease in Childhood*, *103*(4), 363–370. https://doi.org/10.1136/archdischild-2017-313414

Linsell, L., Malouf, R., Morris, J., Kurinczuk, J. J., & Marlow, N. (2015). Prognostic Factors for Poor Cognitive Development in Children Born Very Preterm or With Very Low Birth Weight: A Systematic Review. *JAMA Pediatrics*, *169*(12), 1162–1172. https://doi.org/10.1001/jamapediatrics.2015.2175

Loftus, C. T., Hazlehurst, M. F., Szpiro, A. A., Ni, Y., Tylavsky, F. A., Bush, N. R., Sathyanarayana, S., Carroll, K. N., Karr, C. J., & LeWinn, K. Z. (2019). Prenatal air pollution and childhood IQ: Preliminary evidence of effect modification by folate. *Environmental Research*, *176*, 108505.

Løhaugen, G. C. C., Gramstad, A., Evensen, K. A. I., Martinussen, M., Lindqvist, S., Indredavik, M., Vik, T., Brubakk, A.-M., & Skranes, J. (2010). Cognitive profile in young adults born preterm at very low birthweight. *Developmental*

Medicine and Child Neurology, *52*(12), 1133–1138. https://doi.org/10.1111/j.1469-8749.2010.03743.x

Løhaugen, G. C. C., Østgård, H. F., Andreassen, S., Jacobsen, G. W., Vik, T., Brubakk, A.-M., Skranes, J., & Martinussen, M. (2013). Small for gestational age and intrauterine growth restriction decreases cognitive function in young adults. *The Journal of Pediatrics*, *163*(2), 447-453. e1.

Lundgren, E. M., Cnattingius, S., Jonsson, B., & Tuvemo, T. (2001). Intellectual and Psychological Performance in Males Born Small for Gestational Age With and Without Catch-Up Growth. *Pediatric Research*, *50*(1), 91. https://doi.org/10.1203/00006450-200107000-00017

Luttikhuizen dos Santos, E. S., de Kieviet, J. F., Königs, M., van Elburg, R. M., & Oosterlaan, J. (2013). Predictive value of the Bayley Scales of Infant Development on development of very preterm/very low birth weight children: A meta-analysis. *Early Human Development*, *89*(7), 487–496. https://doi.org/10.1016/j.earlhumdev.2013.03.008

Macaulay, S., Buchmann, E. J., Dunger, D. B., & Norris, S. A. (2019). Reliability and validity of last menstrual period for gestational age estimation in a low-to-middle-income setting. *Journal of Obstetrics and Gynaecology Research*, *45*(1), 217–225. https://doi.org/10.1111/jog.13801

Mangham, L. J., Petrou, S., Doyle, L. W., Draper, E. S., & Marlow, N. (2009). The Cost of Preterm Birth Throughout Childhood in England and Wales. *Pediatrics*, *123*(2), e312–e327. https://doi.org/10.1542/peds.2008-1827

Manor, I., Vurembrandt, N., Rozen, S., Gevah, D., Weizman, A., & Zalsman, G. (2012). Low self-awareness of ADHD in adults using a self-report screening

questionnaire. *European Psychiatry*, *27*(5), 314–320. https://doi.org/10.1016/j.eurpsy.2010.08.013

Marlow, N., Hennessy, E. M., Bracewell, M. A., Wolke, D., & for the EPICure Study Group. (2007). Motor and Executive Function at 6 Years of Age After Extremely Preterm Birth. *PEDIATRICS*, *120*(4), 793–804. https://doi.org/10.1542/peds.2007-0440

Marlow, N., Ni, Y., Lancaster, R., Suonpera, E., Bernardi, M., Fahy, A., Larsen, J., Trickett, J., Hurst, J. R., Morris, J., Wolke, D., & Johnson, S. (2021). No change in neurodevelopment at 11 years after extremely preterm birth. *Archives of Disease in Childhood - Fetal and Neonatal Edition*. https://doi.org/10.1136/archdischild-2020-320650

Marlow, N., Wolke, D., Bracewell, M. A., & Samara, M. (2005). Neurologic and Developmental Disability at Six Years of Age after Extremely Preterm Birth. *New England Journal of Medicine*, *352*(1), 9–19. https://doi.org/10.1056/NEJMoa041367

Martel, M., Nikolas, M., & Nigg, J. T. (2007). Executive Function in Adolescents With ADHD. *Journal of the American Academy of Child & Adolescent Psychiatry*, *46*(11), 1437–1444. https://doi.org/10.1097/chi.0b013e31814cf953

Martinussen, R., Hayden, J., Hogg-johnson, S., & Tannock, R. (2005). A Meta-Analysis of Working Memory Impairments in Children With Attention-Deficit/Hyperactivity Disorder. *Journal of the American Academy of Child & Adolescent Psychiatry*, *44*(4), 377–384. https://doi.org/10.1097/01.chi.0000153228.72591.73

Mathiasen, R., Hansen, B. M., Anderson, A.-M. N., & Greisen, G. (2009). Socio-economic achievements of individuals born very preterm at the age of 27 to 29 years: A nationwide cohort study. *Developmental Medicine & Child Neurology*, *51*(11), 901–908. https://doi.org/10.1111/j.1469-8749.2009.03331.x

Matinolli, H. M., Hovi, P., Levalahti, E., Kaseva, N., Silveira, P. P., Hemio, K., Jarvenpaa, A. L., Eriksson, J. G., Andersson, S., Lindstrom, J., Mannisto, S., & Kajantie, E. (2017). Neonatal nutrition predicts energy balance in young adults born preterm at very low birth weight. *Nutrients*, *9 (12) (no pagination)*(1282). Embase. https://doi.org/10.3390/nu9121282

Mayer, C., & Joseph, K. S. (2013). Fetal growth: A review of terms, concepts and issues relevant to obstetrics. *Ultrasound in Obstetrics & Gynecology*, *41*(2), 136–145. https://doi.org/10.1002/uog.11204

McCarton, C. M., Wallace, I. F., Divon, M., & Vaughan, H. G. (1996). Cognitive and Neurologic Development of the Premature, Small for Gestational Age Infant Through Age 6: Comparison by Birth Weight and Gestational Age. *Pediatrics*, *98*(6), 1167–1178.

McCowan, L., & Horgan, R. P. (2009). Risk factors for small for gestational age infants. *Best Practice & Research Clinical Obstetrics & Gynaecology*, *23*(6), 779–793. https://doi.org/10.1016/j.bpobgyn.2009.06.003

McEwen, E. C., Guthridge, S. L., He, V. YF., McKenzie, J. W., Boulton, T. J., & Smith, R. (2018). What birthweight percentile is associated with optimal perinatal mortality and childhood education outcomes? *American Journal of*

Obstetrics and Gynecology, *218*(2, Supplement), S712–S724. https://doi.org/10.1016/j.ajog.2017.11.574

McKinnon, B., Yang, S., Kramer, M. S., Bushnik, T., Sheppard, A. J., & Kaufman, J. S. (2016). Comparison of black–white disparities in preterm birth between Canada and the United States. *CMAJ : Canadian Medical Association Journal*, *188*(1), E19–E26. https://doi.org/10.1503/cmaj.150464

Melchers, P., & Preuß, U. (1991). K-ABC: Kaufman Battery for Children: Deutschsprachige Fassung. *Frankfurt, AM: Swets & Zeitlinger*.

Mendez-Figueroa, H., Truong, V. T. T., Pedroza, C., & Chauhan, S. P. (2017). Morbidity and Mortality in Small-for-Gestational-Age Infants: A Secondary Analysis of Nine MFMU Network Studies. *American Journal of Perinatology*, *34*(4), 323–332. https://doi.org/10.1055/s-0036-1586502

Mendonça, M., Bilgin, A., & Wolke, D. (2019). Association of Preterm Birth and Low Birth Weight With Romantic Partnership, Sexual Intercourse, and Parenthood in Adulthood: A Systematic Review and Meta-analysis. *JAMA Network Open*, *2*(7), e196961–e196961. https://doi.org/10.1001/jamanetworkopen.2019.6961

Ment, L. R., Hirtz, D., & Hüppi, P. S. (2009). Imaging biomarkers of outcome in the developing preterm brain. *The Lancet Neurology*, *8*(11), 1042–1055. https://doi.org/10.1016/S1474-4422(09)70257-1

Mikolajczyk, R. T., Zhang, J., Betran, A. P., Souza, J. P., Mori, R., Gülmezoglu, A. M., & Merialdi, M. (2011). A global reference for fetal-weight and birthweight percentiles. *The Lancet*, *377*(9780), 1855–1861. https://doi.org/10.1016/S0140-6736(11)60364-4

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology*, *41*(1), 49–100.

Monden, C. W. S., & Smits, J. (2009). Maternal height and child mortality in 42 developing countries. *American Journal of Human Biology*, *21*(3), 305–311. https://doi.org/10.1002/ajhb.20860

Monk, C., Lugo-Candelas, C., & Trumpff, C. (2019). Prenatal Developmental Origins of Future Psychopathology: Mechanisms and Pathways. *Annual Review of Clinical Psychology*, *15*, 317–344. https://doi.org/10.1146/annurev-clinpsy-050718-095539

Muhihi, A., Sudfeld, C. R., Smith, E. R., Noor, R. A., Mshamu, S., Briegleb, C., Bakari, M., Masanja, H., Fawzi, W., & Chan, G. J.-Y. (2016). Risk factors for small-for-gestational-age and preterm births among 19,269 Tanzanian newborns. *BMC Pregnancy and Childbirth*, *16*. https://doi.org/10.1186/s12884-016-0900-5

Mulder, H., Pitchford, N. J., & Marlow, N. (2010). Processing speed and working memory underlie academic attainment in very preterm children. *Archives of Disease in Childhood - Fetal and Neonatal Edition*, *95*(4), F267–F272. https://doi.org/10.1136/adc.2009.167965

Mulder, H., Pitchford, N. J., & Marlow, N. (2011). Inattentive behaviour is associated with poor working memory and slow processing speed in very pre-term children in middle childhood. *British Journal of Educational Psychology*, *81*(1), 147–160. https://doi.org/10.1348/000709910X505527

Muller, M., Sigurdsson, S., Kjartansson, O., Jonsson, P. V., Garcia, M., Bonsdorff, M. B. von, Gunnarsdottir, I., Thorsdottir, I., Harris, T. B., Buchem, M. van, Gudnason, V., & Launer, L. J. (2014). Birth Size and Brain Function 75 Years Later. *Pediatrics*, *134*(4), 761–770. https://doi.org/10.1542/peds.2014-1108

Murray, E., Fernandes, M., Fazel, M., Kennedy, S. H., Villar, J., & Stein, A. (2015). Differential effect of intrauterine growth restriction on childhood neurodevelopment: A systematic review. *BJOG: An International Journal of Obstetrics & Gynaecology*, *122*(8), 1062–1072. https://doi.org/10.1111/1471-0528.13435

Nadeau, L., Boivin, M., Tessier, R., Lefebvre, F., & Robaey, P. (2001). Mediators of behavioral problems in 7-year-old children born after 24 to 28 weeks of gestation. *Journal of Developmental and Behavioral Pediatrics: JDBP*, *22*(1), 1–10.

National Institute for Health and Care Excellence. (2017). *Recommendations | Developmental follow-up of children and young people born preterm | Guidance | NICE.* NICE. https://www.nice.org.uk/guidance/ng72/chapter/Recommendations

Neta, G., Grewal, J., Mikolajczyk, R., Klebanoff, M., & Zhang, J. (2011). Does the individualized reference outperform a simple ultrasound-based reference applied to birth weight in predicting child neurodevelopment? *Ultrasound in Obstetrics & Gynecology*, *38*(1), 62–66. https://doi.org/10.1002/uog.8902

Nichols, T., Jaekel, J., Bartmann, P., & Wolke, D. (2020). Differential susceptibility effects of maternal sensitivity in childhood on small for gestational age adults' wealth. *Development and Psychopathology*, *32*(1), 197–203.

Nohr, E. A., & Liew, Z. (2018). How to investigate and adjust for selection bias in cohort studies. *Acta Obstetricia et Gynecologica Scandinavica*, *97*(4), 407–416. https://doi.org/10.1111/aogs.13319

Norman, M., Hallberg, B., Abrahamsson, T., Björklund, L. J., Domellöf, M., Farooqi, A., Bruun, C. F., Gadsbøll, C., Hellström-Westas, L., Ingemansson, F., Källén, K., Ley, D., Maršál, K., Normann, E., Serenius, F., Stephansson, O., Stigson, L., Um-Bergström, P., & Håkansson, S. (2019). Association Between Year of Birth and 1-Year Survival Among Extremely Preterm Infants in Sweden During 2004-2007 and 2014-2016. *JAMA*, *321*(12), 1188–1199. https://doi.org/10.1001/jama.2019.2021

Norris, T., Johnson, W., Farrar, D., Tuffnell, D., Wright, J., & Cameron, N. (2015). Small-for-gestational age and large-for-gestational age thresholds to predict infants at risk of adverse delivery and neonatal outcomes: Are current charts adequate? An observational study from the Born in Bradford cohort. *BMJ Open*, *5*(3), e006743. https://doi.org/10.1136/bmjopen-2014-006743

Northam, G. B., Liégeois, F., Chong, W. K., Wyatt, J. S., & Baldeweg, T. (2011). Total brain white matter is a major determinant of IQ in adolescents born preterm. *Annals of Neurology*, *69*(4), 702–711. https://doi.org/10.1002/ana.22263

Nosarti, C., Giouroukou, E., Healy, E., Rifkin, L., Walshe, M., Reichenberg, A., Chitnis, X., Williams, S. C. R., & Murray, R. M. (2008). Grey and white matter distribution in very preterm adolescents mediates neurodevelopmental outcome. *Brain*, *131*(1), 205–217. https://doi.org/10.1093/brain/awm282

Nosarti, C., Giouroukou, E., Micali, N., Rifkin, L., Morris, R. G., & Murray, R. M. (2007). Impaired executive functioning in young adults born very preterm. *Journal of the International Neuropsychological Society*, *13*(4), 571–581. https://doi.org/10.1017/S1355617707070725

Nyberg, D. A., Abuhamad, A., & Ville, Y. (2004). Ultrasound assessment of abnormal fetal growth. *Seminars in Perinatology*, *28*(1), 3–22. https://doi.org/10.1053/j.semperi.2003.10.010

Olson, D. M., Severson, E. M., Verstraeten, B. S. E., Ng, J. W. Y., McCreary, J. K., & Metz, G. A. S. (2015). Allostatic Load and Preterm Birth. *International Journal of Molecular Sciences*, *16*(12), 29856–29874. https://doi.org/10.3390/ijms161226209

Oskoui, M., Coutinho, F., Dykeman, J., Jetté, N., & Pringsheim, T. (2013). An update on the prevalence of cerebral palsy: A systematic review and meta-analysis. *Developmental Medicine & Child Neurology*, *55*(6), 509–519. https://doi.org/10.1111/dmcn.12080

Østgård, H. F., Løhaugen, G. C. C., Bjuland, K. J., Rimol, L. M., Brubakk, A.-M., Martinussen, M., Vik, T., Håberg, A. K., & Skranes, J. (2014). Brain Morphometry and Cognition in Young Adults Born Small for Gestational Age at Term. *The Journal of Pediatrics*, *165*(5), 921-927.e1. https://doi.org/10.1016/j.jpeds.2014.07.045

Owens, J. S., & Hoza, B. (2003). The role of inattention and hyperactivity/impulsivity in the positive illusory bias. *Journal of Consulting and Clinical Psychology*, *71*(4), 680.

Papile, L.-A., Burstein, J., Burstein, R., & Koffler, H. (1978). Incidence and evolution of subependymal and intraventricular hemorrhage: A study of infants with birth weights less than 1,500 gm. *The Journal of Pediatrics*, *92*(4), 529–534. https://doi.org/10.1016/S0022-3476(78)80282-0

Pennington, B. F., & Ozonoff, S. (1996). Executive functions and developmental psychopathology. *Journal of Child Psychology and Psychiatry*, *37*(1), 51–87.

Peterson, B. S. (2003). Brain Imaging Studies of the Anatomical and Functional Consequences of Preterm Birth for Human Brain Development. *Annals of the New York Academy of Sciences*, *1008*(1), 219–237. https://doi.org/10.1196/annals.1301.023

Pluess, M., & Belsky, J. (2011). Prenatal programming of postnatal plasticity? *Development and Psychopathology*, *23*(1), 29–38. https://doi.org/10.1017/S0954579410000623

Pritchard, N. L., Hiscock, R. J., Lockie, E., Permezel, M., McGauren, M. F. G., Kennedy, A. L., Green, B., Walker, S. P., & Lindquist, A. C. (2019). Identification of the optimal growth charts for use in a preterm population: An Australian state-wide retrospective cohort study. *PLOS Medicine*, *16*(10), e1002923. https://doi.org/10.1371/journal.pmed.1002923

Prunet, C., Delnord, M., Saurel-Cubizolles, M.-J., Goffinet, F., & Blondel, B. (2017). Risk factors of preterm birth in France in 2010 and changes since 1995: Results from the French National Perinatal Surveys. *Journal of Gynecology Obstetrics and Human Reproduction*, *46*(1), 19–28. https://doi.org/10.1016/j.jgyn.2016.02.010

Pyhala, R., Lahti, J., Heinonen, K., Pesonen, A.-K., Strang-Karlsson, S., Hovi, P., Jarvenpaa, A.-L., Eriksson, J. G., Andersson, S., Kajantie, E., & Raikkonen, K. (2011). Neurocognitive abilities in young adults with very low birth weight. *Neurology*, *77*(23), 2052–2060. https://doi.org/10.1212/WNL.0b013e31823b473e

Pyhälä, R., Lahti, J., Heinonen, K., Pesonen, A.-K., Strang-Karlsson, S., Hovi, P., Järvenpää, A.-L., Eriksson, J. G., Andersson, S., Kajantie, E., & Räikkönen, K. (2011). Neurocognitive abilities in young adults with very low birth weight. *Neurology*, *77*(23), 2052–2060. https://doi.org/10.1212/WNL.0b013e31823b473e

Pyhälä, Riikka. (2012). *Psychological and psychophysiological functioning of young adults born preterm: The Helsinki Study of Very Low Birth Weight Adults*.

Quigley, M. A., Poulsen, G., Boyle, E., Wolke, D., Field, D., Alfirevic, Z., & Kurinczuk, J. J. (2012). Early term and late preterm birth are associated with poorer school performance at age 5 years: A cohort study. *Archives of Disease in Childhood. Fetal and Neonatal Edition*, *97*(3), F167-173. https://doi.org/10.1136/archdischild-2011-300888

R Core Team. (2013). *R: A language and environment for statistical computing*. Vienna, Austria.

Rahu, K., Rahu, M., Pullmann, H., & Allik, J. (2010). Effect of birth weight, maternal education and prenatal smoking on offspring intelligence at school age. *Early Human Development*, *86*(8), 493–497. https://doi.org/10.1016/j.earlhumdev.2010.06.010

Räikkönen, K., & Pesonen, A. (2009). Early life origins of psychological development and mental health. *Scandinavian Journal of Psychology*, *50*(6), 583–591.

Raviv, T., Kessenich, M., & Morrison, F. J. (2004). A mediational model of the association between socioeconomic status and three-year-old language abilities: The role of parenting factors. *Early Childhood Research Quarterly*, *19*(4), 528–547. https://doi.org/10.1016/j.ecresq.2004.10.007

Retzler, J., Johnson, S., Groom, M., Hollis, C., Budge, H., & Cragg, L. (2018). Cognitive predictors of parent-rated inattention in very preterm children: The role of working memory and processing speed. *Child Neuropsychology*, *0*(0), 1–19. https://doi.org/10.1080/09297049.2018.1510908

Riegel, K., Ohrt, B., Wolke, D., & Österlund, K. (1995). *Die Entwicklung gefährdet geborener Kinder bis zum fünften Lebensjahr: Die Arvo-Ylppö-Neugeborenen-Nachfolgestudie in Südbayern und Südfinnland;[Prof. Dr. Dr. Hc Klaus Betke zum 80. Geburtstag]*. Enke.

Ritter, B. C., Nelle, M., Perrig, W., Steinlin, M., & Everts, R. (2013). Executive functions of children born very preterm—Deficit or delay? *European Journal of Pediatrics*, *172*(4), 473–483.

Rommelse, N. N. J., Altink, M. E., de Sonneville, L. M. J., Buschgens, C. J. M., Buitelaar, J., Oosterlaan, J., & Sergeant, J. A. (2007). Are Motor Inhibition and Cognitive Flexibility Dead Ends in ADHD? *Journal of Abnormal Child Psychology*, *35*(6), 957–967. https://doi.org/10.1007/s10802-007-9146-z

Sacchi, C., Marino, C., Nosarti, C., Vieno, A., Visentin, S., & Simonelli, A. (2020). Association of Intrauterine Growth Restriction and Small for Gestational Age

Status With Childhood Cognitive Outcomes: A Systematic Review and Meta-analysis. *JAMA Pediatrics*. https://doi.org/10.1001/jamapediatrics.2020.1097

Saigal, S., & Doyle, L. W. (2008). An overview of mortality and sequelae of preterm birth from infancy to adulthood. *The Lancet*, *371*(9608), 261–269. https://doi.org/10.1016/S0140-6736(08)60136-1

Saigal, S., Ferro, M. A., Van Lieshout, R. J., Schmidt, L. A., Morrison, K. M., & Boyle, M. H. (2016). Health-Related Quality of Life Trajectories of Extremely Low Birth Weight Survivors into Adulthood. *The Journal of Pediatrics*, *179*, 68-73.e1. https://doi.org/10.1016/j.jpeds.2016.08.018

Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychological Review*, *103*(3), 403.

Sammallahti, S., Pyhälä, R., Lahti, M., Lahti, J., Pesonen, A.-K., Heinonen, K., Hovi, P., Eriksson, J. G., Strang-Karlsson, S., Andersson, S., Järvenpää, A.-L., Kajantie, E., & Räikkönen, K. (2014). Infant Growth after Preterm Birth and Neurocognitive Abilities in Young Adulthood. *The Journal of Pediatrics*, *165*(6), 1109-1115.e3. https://doi.org/10.1016/j.jpeds.2014.08.028

Savitz, D. A., Terry, J. W., Dole, N., Thorp, J. M., Siega-Riz, A. M., & Herring, A. H. (2002). Comparison of pregnancy dating by last menstrual period, ultrasound scanning, and their combination. *American Journal of Obstetrics and Gynecology*, *187*(6), 1660–1666. https://doi.org/10.1067/mob.2002.127601

Schlaudecker, E. P., Munoz, F. M., Bardají, A., Boghossian, N. S., Khalil, A., Mousa, H., Nesin, M., Nisar, M. I., Pool, V., Spiegel, H. M. L., Tapia, M. D., Kochhar, S., & Black, S. (2017). Small for gestational age: Case definition & guidelines for data collection, analysis, and presentation of maternal immunisation safety

data. *Vaccine*, *35*(48Part A), 6518–6528. https://doi.org/10.1016/j.vaccine.2017.01.040

Schneider, W., Niklas, F., & Schmiedeler, S. (2014). Intellectual development from early childhood to early adulthood: The impact of early IQ differences on stability and change over time. *Learning and Individual Differences*, *32*, 156–162. https://doi.org/10.1016/j.lindif.2014.02.001

Sentenac, M., Boutron, I., Draper, E. S., Kajantie, E., Maier, R. F., Wolke, D., & Zeitlin, J. (2020). Defining Very Preterm Populations for Systematic Reviews With Meta-analyses. *JAMA Pediatrics*. https://doi.org/10.1001/jamapediatrics.2020.0956

Shah, P., & Kingdom, J. (2011). Long-term neurocognitive outcomes of SGA/IUGR infants. *Obstetrics, Gynaecology & Reproductive Medicine*, *21*(5), 142–146. https://doi.org/10.1016/j.ogrm.2011.02.004

Shah, R., Kennedy, S., Clark, M. D., Bauer, S. C., & Schwartz, A. (2016). Primary Care–Based Interventions to Promote Positive Parenting Behaviors: A Meta-analysis. *Pediatrics*, *137*(5). https://doi.org/10.1542/peds.2015-3393

Shalev, I., Caspi, A., Ambler, A., Belsky, D. W., Chapple, S., Cohen, H. J., Israel, S., Poulton, R., Ramrakha, S., Rivera, C. D., Sugden, K., Williams, B., Wolke, D., & Moffitt, T. E. (2014). Perinatal Complications and Aging Indicators by Midlife. *Pediatrics*, *134*(5), e1315–e1323. https://doi.org/10.1542/peds.2014-1669

Sharma, D., Shastri, S., & Sharma, P. (2016). Intrauterine Growth Restriction: Antenatal and Postnatal Aspects. *Clinical Medicine Insights. Pediatrics*, *10*, 67–83. https://doi.org/10.4137/CMPed.S40070

Sheppard, L. D., & Vernon, P. A. (2008). Intelligence and speed of information-processing: A review of 50 years of research. *Personality and Individual Differences*, *44*(3), 535–551. https://doi.org/10.1016/j.paid.2007.09.015

Silva, P. A., McGee, R., & Williams, S. (1984). A longitudinal study of the intelligence and behavior of preterm and small for gestational age children. *Journal of Developmental and Behavioral Pediatrics*, *5*(1), 1–5. https://doi.org/10.1097/00004703-198402000-00001

Sommerfelt, K. (2000). Cognitive development of term small for gestational age children at five years of age. *Archives of Disease in Childhood*, *83*(1), 25–30. https://doi.org/10.1136/adc.83.1.25

Spittle, A. J., Cameron, K., Doyle, L. W., Cheong, J. L., & Group, for the V. I. C. S. (2018). Motor Impairment Trends in Extremely Preterm Children: 1991–2005. *Pediatrics*, *141*(4). https://doi.org/10.1542/peds.2017-3410

Stålnacke, J., Lundequist, A., Böhm, B., Forssberg, H., & Smedler, A.-C. (2015). Individual cognitive patterns and developmental trajectories after preterm birth. *Child Neuropsychology: A Journal on Normal and Abnormal Development in Childhood and Adolescence*, *21*(5), 648–667. https://doi.org/10.1080/09297049.2014.958071

Sternberg, R. J. (2005). Intelligence. In *The Cambridge handbook of thinking and reasoning* (pp. 751–773). Cambridge University Press.

Stewart, A. L., Reynolds, E. O. R., Hope, P. L., Hamilton, P. A., Baudin, J., Costello, A. M. de L., Bradford, B. C., & Wyatt, J. S. (1987). Probabilité of Neurodevelopmental Disorders Estimated from Ultrasound Appearance of

Brains of Very Preterm Infants. *Developmental Medicine & Child Neurology*, *29*(1), 3–11. https://doi.org/10.1111/j.1469-8749.1987.tb02101.x

Stoelhorst, G. M., Rijken, M., Martens, S. E., van Zwieten, P. H., Feenstra, J., Zwinderman, A. H., Wit, J.-M., & Veen, S. (2003). Developmental outcome at 18 and 24 months of age in very preterm children: A cohort study from 1996 to 1997. *Early Human Development*, *72*(2), 83–95.

Stoll, B. J., Hansen, N. I., Bell, E. F., Walsh, M. C., Carlo, W. A., Shankaran, S., Laptook, A. R., Sánchez, P. J., Van Meurs, K. P., Wyckoff, M., Das, A., Hale, E. C., Ball, M. B., Newman, N. S., Schibler, K., Poindexter, B. B., Kennedy, K. A., Cotten, C. M., Watterberg, K. L., … Higgins, R. D. (2015). Trends in Care Practices, Morbidity, and Mortality of Extremely Preterm Neonates, 1993-2012. *JAMA*, *314*(10), 1039. https://doi.org/10.1001/jama.2015.10244

Strang-Karlsson, S., Räikkönen, K., Pesonen, A.-K., Kajantie, E., Paavonen, E. J., Lahti, J., Hovi, P., Heinonen, K., Järvenpää, A.-L., Eriksson, J. G., & Andersson, S. (2008). Very Low Birth Weight and Behavioral Symptoms of Attention Deficit Hyperactivity Disorder in Young Adulthood: The Helsinki Study of Very-Low-Birth-Weight Adults. *American Journal of Psychiatry*, *165*(10), 1345–1353. https://doi.org/10.1176/appi.ajp.2008.08010085

Strauss, R. S. (2000). Adult Functional Outcome of Those Born Small for Gestational Age: Twenty-six–Year Follow-up of the 1970 British Birth Cohort. *JAMA*, *283*(5), 625–632. https://doi.org/10.1001/jama.283.5.625

Strenze, T. (2007). Intelligence and socioeconomic success: A meta-analytic review of longitudinal research. *Intelligence*, *35*(5), 401–426. https://doi.org/10.1016/j.intell.2006.09.004

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*(6), 643–662. https://doi.org/10.1037/h0054651

Suikkanen, J., Miettola, S., Heinonen, K., Vääräsmäki, M., Tikanmäki, M., Sipola, M., Matinolli, H.-M., Järvelin, M.-R., Räikkönen, K., Hovi, P., & Kajantie, E. (2020). Reaction times, learning, and executive functioning in adults born preterm. *Pediatric Research*, 1–7. https://doi.org/10.1038/s41390-020-0851-4

Taylor, L. K., Lee, Y. Y. C., Lim, K., Simpson, J. M., Roberts, C. L., & Morris, J. (2013). Potential prevention of small for gestational age in Australia: A population-based linkage study. *BMC Pregnancy and Childbirth*, *13*, 210. https://doi.org/10.1186/1471-2393-13-210

Tierney, J. F., Vale, C., Riley, R., Smith, C. T., Stewart, L., Clarke, M., & Rovers, M. (2015). Individual Participant Data (IPD) Meta-analyses of Randomised Controlled Trials: Guidance on Their Use. *PLOS Medicine*, *12*(7), e1001855. https://doi.org/10.1371/journal.pmed.1001855

Torche, F., & Echevarría, G. (2011). The effect of birthweight on childhood cognitive development in a middle-income country. *International Journal of Epidemiology*, *40*(4), 1008–1018. https://doi.org/10.1093/ije/dyr030

Torres, D. D. (2013). Understanding How Family Socioeconomic Status Mediates the Maternal Intelligence–Child Cognitive Outcomes Relationship: A Moderated Mediation Analysis. *Biodemography and Social Biology*, *59*(2), 157–177. https://doi.org/10.1080/19485565.2013.833804

Trahan, L., Stuebing, K. K., Hiscock, M. K., & Fletcher, J. M. (2014). The Flynn Effect: A Meta-analysis. *Psychological Bulletin*, *140*(5), 1332–1360. https://doi.org/10.1037/a0037173

Twilhaar, E. S., Kieviet, J. F. D., Elburg, R. M. V., & Oosterlaan, J. (2020). Neurocognitive processes underlying academic difficulties in very preterm born adolescents. *Child Neuropsychology*, *26*(2), 274–287. https://doi.org/10.1080/09297049.2019.1639652

Twilhaar, E. S., Wade, R. M., Kieviet, J. F. de, Goudoever, J. B. van, Elburg, R. M. van, & Oosterlaan, J. (2018). Cognitive Outcomes of Children Born Extremely or Very Preterm Since the 1990s and Associated Risk Factors: A Meta-analysis and Meta-regression. *JAMA Pediatrics*, *172*(4), 361–367. https://doi.org/10.1001/jamapediatrics.2017.5323

UNESCO Institute for Statistics. (2012). *International standard classification of education: ISCED 2011*. UNESCO Institute for Statistics Montreal.

van Baar, A. L., van Wassenaer, A. G., Briët, J. M., Dekker, F. W., & Kok, J. H. (2005). Very Preterm Birth is Associated with Disabilities in Multiple Developmental Domains. *Journal of Pediatric Psychology*, *30*(3), 247–255. https://doi.org/10.1093/jpepsy/jsi035

van der Kooy-Hofland, V. A. C., van der Kooy, J., Bus, A. G., van IJzendoorn, M. H., & Bonsel, G. J. (2012). Differential susceptibility to early literacy intervention in children with mild perinatal adversities: Short- and long-term effects of a randomized control trial. *Journal of Educational Psychology*, *104*(2), 337–349. https://doi.org/10.1037/a0026984

van Houdt, C. A., van Wassenaer-Leemhuis, A. G., Oosterlaan, J., Königs, M., Koopman-Esseboom, C., Laarman, A. R. C., van Kaam, A. H., & Aarnoudse-Moens, C. S. H. (2020). Executive function training in very preterm children:

A randomized controlled trial. *European Child & Adolescent Psychiatry*. https://doi.org/10.1007/s00787-020-01561-0

Vandenbroucke, L., Verschueren, K., & Baeyens, D. (2017). The development of executive functioning across the transition to first grade and its predictive value for academic achievement. *Learning and Instruction*, *49*, 103–112. https://doi.org/10.1016/j.learninstruc.2016.12.008

Vasak, B., Koenen, S. V., Koster, M. P. H., Hukkelhoven, C. W. P. M., Franx, A., Hanson, M. A., & Visser, G. H. A. (2015). Human fetal growth is constrained below optimal for perinatal survival. *Ultrasound in Obstetrics & Gynecology*, *45*(2), 162–167. https://doi.org/10.1002/uog.14644

Verkauskiene, R., Figueras, F., Deghmoun, S., Chevenne, D., Gardosi, J., & Levy-Marchal, M. (2008). Birth Weight and Long-Term Metabolic Outcomes: Does the Definition of Smallness Matter? *Hormone Research in Paediatrics*, *70*(5), 309–315. https://doi.org/10.1159/000157878

Victorian Infant Collaborative Study Group. (1997). Outcome at 2 years of children 23–27 weeks' gestation born in Victoria in 1991–92. *Journal of Paediatrics and Child Health*, *33*(2), 161–165.

Villar, J., Ismail, L. C., Victora, C. G., Ohuma, E. O., Bertino, E., Altman, D. G., Lambert, A., Papageorghiou, A. T., Carvalho, M., Jaffer, Y. A., Gravett, M. G., Purwar, M., Frederick, I. O., Noble, A. J., Pang, R., Barros, F. C., Chumlea, C., Bhutta, Z. A., & Kennedy, S. H. (2014). International standards for newborn weight, length, and head circumference by gestational age and sex: The Newborn Cross-Sectional Study of the INTERGROWTH-21st Project.

The Lancet, *384*(9946), 857–868. https://doi.org/10.1016/S0140-6736(14)60932-6

Vincer, M. J., Allen, A. C., Allen, V. M., Baskett, T. F., & O'Connell, C. M. (2014). Trends in the prevalence of cerebral palsy among very preterm infants (<31 weeks' gestational age). *Paediatrics & Child Health*, *19*(4), 185–189.

Voigt, M., Fusch, C., Olbertz, D., Hartmann, K., Rochow, N., Renken, C., & Schneider, K. (2006). Analyse des Neugeborenenkollektivs der Bundesrepublik Deutschland. *Geburtshilfe und Frauenheilkunde*, *66*(10), 956–970. https://doi.org/10.1055/s-2006-924458

Vollmer, B., & Edmonds, C. J. (2019). School Age Neurological and Cognitive Outcomes of Fetal Growth Retardation or Small for Gestational Age Birth Weight. *Frontiers in Endocrinology*, *10*. https://doi.org/10.3389/fendo.2019.00186

von Aster, M., Neubauer, A., & Horn, R. von. (2006). *Wechsler Intelligenztest für Erwachsene: WIE; Übersetzung und Adaption der WAIS-III*. Harcourt Test Services.

von Beckerath, A.-K., Kollmann, M., Rotky-Fast, C., Karpf, E., Lang, U., & Klaritsch, P. (2013). Perinatal complications and long-term neurodevelopmental outcome of infants with intrauterine growth restriction. *American Journal of Obstetrics and Gynecology*, *208*(2), 130.e1-130.e6. https://doi.org/10.1016/j.ajog.2012.11.014

Wadhwa, P. D., Buss, C., Entringer, S., & Swanson, J. M. (2009). Developmental Origins of Health and Disease: Brief History of the Approach and Current

Focus on Epigenetic Mechanisms. *Seminars in Reproductive Medicine*, *27*(5), 358–368. https://doi.org/10.1055/s-0029-1237424

Walker, D.-M., & Marlow, N. (2008). Neurocognitive outcome following fetal growth restriction. *Archives of Disease in Childhood - Fetal and Neonatal Edition*, *93*(4), F322–F325. https://doi.org/10.1136/adc.2007.120485

Wechsler, D. (2008). *Wechsler Adult Intelligence Scale–Fourth Edition (WAIS–IV)*. San Antonio, TX: The Psychological Corporation.

Weisglas-Kuperus, N., Hille, E. T. M., Duivenvoorden, H. J., Finken, M. J. J., Wit, J. M., van Buuren, S., van Goudoever, J. B., Verloove-Vanhorick, S. P., & for the Dutch POPS-19 Collaborative Study Group. (2008). Intelligence of very preterm or very low birthweight infants in young adulthood. *Archives of Disease in Childhood - Fetal and Neonatal Edition*, *94*(3), F196–F200. https://doi.org/10.1136/adc.2007.135095

Weiss, L. G., Saklofske, D. H., Holdnack, J. A., & Prifitera, A. (2015). *WISC-V assessment and interpretation: Scientist-practitioner perspectives*. Academic Press.

Wells, G., Shea, B., O'Connell, D., Peterson, J., Welch, V., Losos, M., & Tugwell, P. (2012). The Newcastle-Ottawa Scale (NOS) for assessing the quality if nonrandomized studies in meta-analyses. Http. *Www. Ohric a/Progr Ams/Clini Cal_epide Miolo Gy/Oxfor Dasp*.

Werner, E. E., & Bayley, N. (1966). The Reliability of Bayley's Revised Scale of Mental and Motor Development during the First Year of Life. *Child Development*, *37*(1), 39–50. JSTOR. https://doi.org/10.2307/1126427

White, T., Su, S., Schmidt, M., Kao, C.-Y., & Sapiro, G. (2010). The Development of Gyrification in Childhood and Adolescence. *Brain and Cognition*, *72*(1), 36. https://doi.org/10.1016/j.bandc.2009.10.009

Willcutt, E. G., Doyle, A. E., Nigg, J. T., Faraone, S. V., & Pennington, B. F. (2005). Validity of the Executive Function Theory of Attention-Deficit/Hyperactivity Disorder: A Meta-Analytic Review. *Biological Psychiatry*, *57*(11), 1336–1346. https://doi.org/10.1016/j.biopsych.2005.02.006

Wolke, D. (2012). *TRAB tester's rating of adult behavior (in German)*. Available from author upon request.

Wolke, D. (2019). Is Social Inequality in Cognitive Outcomes Increased by Preterm Birth–Related Complications? *JAMA Network Open*, *2*(5), e192902–e192902. https://doi.org/10.1001/jamanetworkopen.2019.2902

Wolke, D., Baumann, N., Strauss, V., Johnson, S., & Marlow, N. (2015). Bullying of Preterm Children and Emotional Problems at School Age: Cross-Culturally Invariant Effects. *The Journal of Pediatrics*, *166*(6), 1417–1422. https://doi.org/10.1016/j.jpeds.2015.02.055

Wolke, D., Jaekel, J., Hall, J., & Baumann, N. (2013). Effects of Sensitive Parenting on the Academic Resilience of Very Preterm and Very Low Birth Weight Adolescents. *Journal of Adolescent Health*, *53*(5), 642–647. https://doi.org/10.1016/j.jadohealth.2013.06.014

Wolke, D., Johnson, S., & Mendonça, M. (2019). The Life Course Consequences of Very Preterm Birth. *Annual Review of Developmental Psychology*, *1*(1), 69–92. https://doi.org/10.1146/annurev-devpsych-121318-084804

Wolke, D., & Meyer, R. (1999). Cognitive status, language attainment, and prereading skills of 6-year-old very preterm children and their peers: The Bavarian Longitudinal Study. *Developmental Medicine and Child Neurology*, *41*(2), 94–109.

Wolke, D., Ratschinski, G., Ohrt, B., & Riegel, K. (1994). The cognitive 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*, *153*(12), 906–915. https://doi.org/10.1007/BF01954744

Wolke, D., Samara, M., Bracewell, M., & Marlow, N. (2008). Specific Language Difficulties and School Achievement in Children Born at 25 Weeks of Gestation or Less. *The Journal of Pediatrics*, *152*(2), 256-262.e1. https://doi.org/10.1016/j.jpeds.2007.06.043

Wolke, D., Schmid, G., Schreier, A., & Meyer, R. (2009). Crying and Feeding Problems in Infancy and Cognitive Outcome in Preschool Children Born at Risk: A Prospective Population Study. *Journal of Developmental & Behavioral Pediatrics*, *30*(3), 226. https://doi.org/10.1097/DBP.0b013e3181a85973

Wolke, D., Söhne, B., Ohrt, B., & Riegel, K. (1995). Follow-up of preterm children: Important to document dropouts. *Lancet (London, England)*, *345*(8947), 447. https://doi.org/10.1016/s0140-6736(95)90425-5

Wolke, D., Strauss, V. Y.-C., Johnson, S., Gilmore, C., Marlow, N., & Jaekel, J. (2015). Universal Gestational Age Effects on Cognitive and Basic Mathematic Processing: 2 Cohorts in 2 Countries. *The Journal of Pediatrics*, *166*(6), 1410-1416.e2. https://doi.org/10.1016/j.jpeds.2015.02.065

Wolke, D., Waylen, A., Samara, M., Steer, C., Goodman, R., Ford, T., & Lamberts, K. (2009). Selective drop-out in longitudinal studies and non-biased prediction of behaviour disorders. *The British Journal of Psychiatry*, *195*(3), 249–256. https://doi.org/10.1192/bjp.bp.108.053751

Woltering, S., Liu, Z., Rokeach, A., & Tannock, R. (2013). Neurophysiological differences in inhibitory control between adults with ADHD and their peers. *Neuropsychologia*, *51*(10), 1888–1895. https://doi.org/10.1016/j.neuropsychologia.2013.06.023

Wood, N. S., Marlow, N., Costeloe, K. L., Gibson, A. T., & Wilkinson, A. R. (2000). Neurologic and Developmental Disability after Extremely Preterm Birth. *New England Journal of Medicine*, *343*(6), 378–384. https://doi.org/10.1056/NEJM200008103430601

Yang, S., Platt, R. W., & Kramer, M. S. (2010). Variation in Child Cognitive Ability by Week of Gestation Among Healthy Term Births. *American Journal of Epidemiology*, *171*(4), 399–406. https://doi.org/10.1093/aje/kwp413

Yanney, M., & Marlow, N. (2004). Paediatric consequences of fetal growth restriction. *Seminars in Fetal and Neonatal Medicine*, *9*(5), 411–418. https://doi.org/10.1016/j.siny.2004.03.005

Zeitlin, J., Szamotulska, K., Drewniak, N., Mohangoo, A. D., Chalmers, J., Sakkeus, L., Irgens, L., Gatt, M., Gissler, M., & Blondel, B. (2013). Preterm birth time trends in Europe: A study of 19 countries. *BJOG: An International Journal of Obstetrics & Gynaecology*, *120*(11), 1356–1365. https://doi.org/10.1111/1471-0528.12281

Zhang, Z. (2016). Missing data imputation: Focusing on single imputation. *Annals of Translational Medicine*, *4*(1). https://doi.org/10.3978/j.issn.2305-5839.2015.12.38

Zins, J. E., & Barnett, D. W. (1984). A validity study of the K-ABC, the WISC-R, and the Stanford-Binet with nonreferred children. *Journal of School Psychology*, *22*(4), 369–371. https://doi.org/10.1016/0022-4405(84)90024-4

Appendices

Appendix 1: Childhood Neurosensory Impairment in VP/VLBW participants from each IPD cohort

	AYLS	BLS	EPICURE	HESVA	NTNU	NZVLBW	UCLH	VICS	Overall
	VP/VLBW (n=28)	VP/VLBW (n=203)	VP/VLBW (n=124)	VP/VLBW (n=109)	VP/VLBW (n=51)	VP/VLBW (n=225)	VP/VLBW (n=104)	VP/VLBW (n=224)	VP/VLBW (n=1068)
Evidence of Severe NSI									
Yes	3 (10.7%)	22 (10.8%)	14 (11.3%)	5 (4.6%)	4 (7.8%)	9 (4.0%)	3 (2.9%)	27 (12.1%)	87 (8.1%)
No	25 (89.3%)	181 (89.2%)	110 (88.7%)	104 (95.4%)	47 (92.2%)	216 (96.0%)	101 (97.1%)	197 (87.9%)	981 (91.9%)
Visual Impairment									
No	26 (92.9%)	200 (98.5%)	117 (94.4%)	107 (98.2%)	37 (72.5%)	218 (96.9%)	0 (0%)	224 (100%)	929 (87.0%)
Yes	0 (0%)	2 (1.0%)	0 (0%)	0 (0%)	0 (0%)	1 (0.4%)	0 (0%)	0 (0%)	3 (0.3%)
Missing	2 (7.1%)	1 (0.5%)	7 (5.6%)	2 (1.8%)	14 (27.5%)	6 (2.7%)	104 (100%)	0 (0%)	136 (12.7%)
Hearing Impairment									
No	26 (92.9%)	201 (99.0%)	116 (93.5%)	107 (98.2%)	37 (72.5%)	217 (96.4%)	0 (0%)	223 (99.6%)	927 (86.8%)
Yes	0 (0%)	1 (0.5%)	1 (0.8%)	0 (0%)	0 (0%)	2 (0.9%)	0 (0%)	1 (0.4%)	5 (0.5%)
Missing	2 (7.1%)	1 (0.5%)	7 (5.6%)	2 (1.8%)	14 (27.5%)	6 (2.7%)	104 (100%)	0 (0%)	136 (12.7%)
Non-Ambulatory Cerebral Palsy									
No	28 (100%)	195 (96.1%)	115 (92.7%)	101 (92.7%)	49 (96.1%)	219 (97.3%)	0 (0%)	222 (99.1%)	929 (87.0%)

Appendix 1: Childhood Neurosensory Impairment in VP/VLBW participants from each IPD cohort

Yes	0 (0%)	7 (3.4%)	2 (1.6%)	5 (4.6%)	2 (3.9%)	0 (0%)	0 (0%)	2 (0.9%)	18 (1.7%)
Missing	0 (0%)	1 (0.5%)	7 (5.6%)	3 (2.8%)	0 (0%)	6 (2.7%)	104 (100%)	0 (0%)	121 (11.3%)
Child IQ <70								
No	22 (78.6%)	168 (82.8%)	108 (87.1%)	0 (0%)	39 (76.5%)	212 (94.2%)	99 (95.2%)	194 (86.6%)	842 (78.8%)
Yes	3 (10.7%)	18 (8.9%)	13 (10.5%)	0 (0%)	3 (5.9%)	7 (3.1%)	3 (2.9%)	25 (11.2%)	72 (6.7%)
Missing	3 (10.7%)	17 (8.4%)	3 (2.4%)	109 (100%)	9 (17.6%)	6 (2.7%)	2 (1.9%)	5 (2.2%)	154 (14.4%)

Appendix 2: Newcastle Ottawa criteria and ratings for each IPD Cohort criteria :

Newcastle Ottawa

Rating Scale   http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp

Selection

1) Representativeness of the exposed

cohort

A) truly representative of the average ____**VP/VLBW (not a sub-selection such as just those**

with BPD or only males)_____ in the community ‾

B) somewhat representative of the average _____ in the community ‾

C) selected group of users eg nurses, volunteers

D) no description of the derivation of the cohort

2) Selection of the non exposed cohort

A) drawn from the same community as the exposed cohort ‾

B) drawn from a different

source

C) no description of the derivation of the non exposed cohort

3) Ascertainment of exposure

A) secure record (eg surgical records) ‾

B) structured interview ‾

C) written self

report

D) no description

4) <u>Demonstration that outcome of interest was not present at start of</u>

<u>study</u> **(Was adult cognitive performance known when the**

participants were recruited?)

A) yes ⁻

B) no

Comparability

1) <u>Comparability of cohorts on the basis of the design or</u>

<u>analysis</u>

A) study controls for ____**maternal education**_____ (select the most important factor) ⁻

B) study controls for any additional factor (**sex**)

Outcome

1) <u>Assessment of outcome</u> (**Did the study use a standardised**

full-scale IQ assessment?)

A) independent blind assessment

B) record linkage

C) self report

D) no description

2) <u>Was follow-up long enough for outcomes to occur</u> **(Did the cohort assess adult**

IQ outcomes?)

A) yes (17 years or greater)

B) no

3) <u>Adequacy of follow up of cohorts</u>**. (Of the potential VP/VLBW participants eligible in**

adulthood, were over 50% of them assessed?)

A) complete follow up - all subjects accounted for ⁻

B) subjects lost to follow up unlikely to introduce bias - small number lost - > **__50__** %

follow up, or description provided of those lost) ⁻

C) follow up rate < **__50__**% and no description of those lost

D) no statement

Appendix 2(continued) : Newcastle Ottawa criteria and ratings for each IPD Cohort

Criteria:

Cohort	Representativeness of the exposed cohort	Selection of the non exposed cohort	Ascertainment of exposure	Demonstration that outcome of interest was not present at start of study	Comparability of cohorts on the basis of the design or analysis	Assessment of outcome	Was follow-up long enough for outcomes to occur	Adequacy of follow up of cohorts (above or below 50%)	Overall Cohort Score
AYLS(Matinolli et al., 2017)	A (regional)	A	A	A	A	A	A	C	8
BLS(Eryigit Madzwamuse et al., 2015; Wolke & Meyer, 1999)	A (regional)	A	A	A	A	A	A	C	8
EPICure(Costeloe et al., 2000; Linsell et al., 2018)	A (national)	B	A	A	A	A	A	C	7

HESVA(Riikka Pyhälä, 2012)	A(regional)	A	A	A	A	A	A	C	8
NTNU(Lærum et al., 2019)	A (regional)	B	A	A	A	A	A	B	7
NZ_VLBW darlow	A (national)	B	A	A	A	A	A	B	8
UCLH(Kroll et al., 2017; Stewart et al., 1987)	A (regional)	B	A	A	A	A	A	B	8
VICS(Victorian Infant Collaborative Study Group, 1997)	A (regional)	A	A	A	A	A	A	B	9

Appendix 3: Linear mixed model demonstrating reducing Gestational Age by Birth Year among VP/VLBW participants

VP/VLBW only analysis

Predictors	Gestational Age (weeks)		
	Estimates	*CI*	*p*
(Intercept – Estimate for 1978)	32.07	29.66 – 34.49	**<0.001**
Birth year – per year post 1978	-0.32	-0.60 – -0.04	**0.025**
Observations	1068		
Marginal R^2 / Conditional R^2	0.222 / 0.488		

Appendix 4: Linear mixed model demonstrating reducing Birthweight by Birth Year among VP/VLBW participants

VP/VLBW only analysis

Predictors	Birthweight (g)		
	Estimates	*CI*	*p*
(Intercept – Estimate for 1978)	1464.87	1211.59 – 1718.14	**<0.001**
Birth year – per year post 1978	-29.85	-58.78 – -0.91	**0.043**
Observations	1068		
Marginal R^2 / Conditional R^2	0.164 / 0.411		

Appendix 5: IQ and Demographic information of all participants from each IPD cohort

Cohort	AYLS		BLS		EPICURE		HESVA		NTNU		NZVLBW		UCLH		VICS	
Group	Cons n=303	VP/VLBW n=28	Cons n=192	VP/VLBW n=203	Cons n=64	VP/VLBW n=124	Cons n=98	VP/VLBW n=109	Cons n=75	VP/VLBW n=51	Cons n=100	VP/VLBW n=225	Cons n=89	VP/VLBW n=104	Cons n=146	VP/VLBW n=224
IQ Z Score																
Mean (SD)	0.00 (0.94)	-0.95 (1.21)	0.00 (0.75)	-0.83 (1.04)	0.00 (0.64)	-1.06 (0.96)	0.00 (0.84)	-0.57 (1.06)	0.00 (0.70)	-0.86 (1.15)	0.00 (0.78)	-0.78 (1.00)	0.00 (0.89)	-0.64 (1.00)	0.00 (0.84)	-0.67 (1.01)

Cohort	AYLS		BLS		EPICURE		HESVA		NTNU		NZVLBW		UCLH		VICS	
Group	Cons n=303	VP/VLBW n=28	Cons n=192	VP/VLBW n=203	Cons n=64	VP/VLBW n=124	Cons n=98	VP/VLBW n=109	Cons n=75	VP/VLBW n=51	Cons n=100	VP/VLBW n=225	Cons n=89	VP/VLBW n=104	Cons n=146	VP/VLBW n=224
Sex																
Male	134 (44.2%)	16 (57.1%)	92 (47.9%)	108 (53.2%)	25 (39.1%)	56 (45.2%)	42 (42.9%)	47 (43.1%)	33 (44.0%)	25 (49.0%)	37 (37.0%)	100 (44.4%)	42 (47.2%)	63 (60.6%)	56 (38.4%)	96 (42.9%)
Female	169 (55.8%)	12 (42.9%)	100 (52.1%)	95 (46.8%)	39 (60.9%)	68 (54.8%)	56 (57.1%)	62 (56.9%)	42 (56.0%)	26 (51.0%)	63 (63.0%)	125 (55.6%)	47 (52.8%)	41 (39.4%)	90 (61.6%)	128 (57.1%)

Maternal

Education Level

Cohort	AYLS		BLS		EPICURE		HESVA		NTNU		NZVLBW		UCLH		VICS	
Group	Cons n=303	VP/VLBW n=28	Cons n=192	VP/VLBW n=203	Cons n=64	VP/VLBW n=124	Cons n=98	VP/VLBW n=109	Cons n=75	VP/VLBW n=51	Cons n=100	VP/VLBW n=225	Cons n=89	VP/VLBW n=104	Cons n=146	VP/VLBW n=224
Low	52 (17.2%)	7 (25.0%)	87 (45.3%)	61 (30.0%)	4 (6.2%)	23 (18.5%)	13 (13.3%)	17 (15.6%)	2 (2.7%)	2 (3.9%)	4 (4.0%)	85 (37.8%)	3 (3.4%)	3 (2.9%)	9 (6.2%)	47 (21.0%)
Medium	101 (33.3%)	9 (32.1%)	72 (37.5%)	112 (55.2%)	48 (75.0%)	90 (72.6%)	56 (57.1%)	64 (58.7%)	31 (41.3%)	22 (43.1%)	33 (33.0%)	68 (30.2%)	23 (25.8%)	40 (38.5%)	34 (23.3%)	57 (25.4%)
High	148 (48.8%)	12 (42.9%)	32 (16.7%)	27 (13.3%)	12 (18.8%)	4 (3.2%)	29 (29.6%)	26 (23.9%)	28 (37.3%)	16 (31.4%)	63 (63.0%)	64 (28.4%)	39 (43.8%)	20 (19.2%)	26 (17.8%)	24 (10.7%)

Cohort	AYLS		BLS		EPICURE		HESVA		NTNU		NZVLBW		UCLH		VICS	
Group	Cons n=303	VP/VLBW n=28	Cons n=192	VP/VLBW n=203	Cons n=64	VP/VLBW n=124	Cons n=98	VP/VLBW n=109	Cons n=75	VP/VLBW n=51	Cons n=100	VP/VLBW n=225	Cons n=89	VP/VLBW n=104	Cons n=146	VP/VLBW n=224
Missing	2 (0.7%)	0 (0%)	1 (0.5%)	3 (1.5%)	0 (0%)	7 (5.6%)	0 (0%)	2 (1.8%)	14 (18.7%)	11 (21.6%)	0 (0%)	8 (3.6%)	24 (27.0%)	41 (39.4%)	77 (52.7%)	96 (42.9%)

Appendix 6: Neonatal and demographic data for VP/VLBW participants from each IPD cohort

	AYLS VP/ VLBW (n=28)	BLS VP/VLBW (n=203)	EPICURE VP/VLBW (n=124)	HESVA VP/VLBW (n=109)	NTNU VP/VLBW (n=51)	NZVLBW VP/VLBW (n=225)	UCLH VP/ VLBW (n=104)	VICS VP/VLBW (n=224)	Overall VP/ VLBW (n=1068)
Gestational Age (weeks)									
Mean (SD)	29.6 (2.09)	30.4 (2.05)	24.5 (0.748)	29.3 (2.33)	29.0 (2.49)	29.3 (2.50)	28.8 (2.00)	26.6 (1.99)	28.3 (2.81)
Birthweight Z Score									
Mean (SD)	-0.00903 (1.08)	-0.603 (1.20)	0.230 (0.822)	-0.421 (1.00)	-0.182 (1.08)	-0.607 (1.07)	-0.0826 (0.930)	-0.167 (1.07)	-0.311 (1.09)
Multiple Birth									

	AYLS	BLS	EPICURE	HESVA	NTNU	NZVLBW	UCLH	VICS	Overall
	VP/ VLBW (n=28)	VP/VLBW (n=203)	VP/VLBW (n=124)	VP/VLBW (n=109)	VP/VLBW (n=51)	VP/VLBW (n=225)	VP/ VLBW (n=104)	VP/VLBW (n=224)	VP/ VLBW (n=1068)
Singleton	25 (89.3%)	149 (73.4%)	83 (66.9%)	92 (84.4%)	41 (80.4%)	169 (75.1%)	81 (77.9%)	150 (67.0%)	790 (74.0%)
Multiple	3 (10.7%)	54 (26.6%)	40 (32.3%)	17 (15.6%)	10 (19.6%)	56 (24.9%)	19 (18.3%)	74 (33.0%)	273 (25.6%)
Missing	0 (0%)	0 (0%)	1 (0.8%)	0 (0%)	0 (0%)	0 (0%)	4 (3.8%)	0 (0%)	5 (0.5%)
Intraventricular Haemorrhage									
No Grade	23 (82.1%)	161 (79.3%)	44 (35.5%)	69 (63.3%)	38 (74.5%)	158 (70.2%)	48 (46.2%)	152 (67.9%)	693 (64.9%)
Any Grade	5 (17.9%)	41 (20.2%)	79 (63.7%)	14 (12.8%)	5 (9.8%)	52 (23.1%)	55 (52.9%)	72 (32.1%)	323 (30.2%)
Missing	0 (0%)	1 (0.5%)	1 (0.8%)	26 (23.9%)	8 (15.7%)	15 (6.7%)	1 (1.0%)	0 (0%)	52 (4.9%)

	AYLS	BLS	EPICURE	HESVA	NTNU	NZVLBW	UCLH	VICS	Overall
	VP/ VLBW (n=28)	VP/VLBW (n=203)	VP/VLBW (n=124)	VP/VLBW (n=109)	VP/VLBW (n=51)	VP/VLBW (n=225)	VP/ VLBW (n=104)	VP/VLBW (n=224)	VP/ VLBW (n=1068)
Bronchopulmonary Dysplasia Diagnosed									
No	27 (96.4%)	101 (49.8%)	35 (28.2%)	80 (73.4%)	39 (76.5%)	181 (80.4%)	0 (0%)	138 (61.6%)	601 (56.3%)
Yes	1 (3.6%)	102 (50.2%)	89 (71.8%)	25 (22.9%)	10 (19.6%)	44 (19.6%)	0 (0%)	86 (38.4%)	357 (33.4%)
Missing	0 (0%)	0 (0%)	0 (0%)	4 (3.7%)	2 (3.9%)	0 (0%)	104 (100%)	0 (0%)	110 (10.3%)
ISCED Maternal Education									
Low	7 (25.0%)	61 (30.0%)	23 (18.5%)	17 (15.6%)	2 (3.9%)	85 (37.8%)	3 (2.9%)	47 (21.0%)	245 (22.9%)

	AYLS	BLS	EPICURE	HESVA	NTNU	NZVLBW	UCLH	VICS	Overall
	VP/ VLBW (n=28)	VP/VLBW (n=203)	VP/VLBW (n=124)	VP/VLBW (n=109)	VP/VLBW (n=51)	VP/VLBW (n=225)	VP/ VLBW (n=104)	VP/VLBW (n=224)	VP/ VLBW (n=1068)
Medium	9 (32.1%)	112 (55.2%)	90 (72.6%)	64 (58.7%)	22 (43.1%)	68 (30.2%)	40 (38.5%)	57 (25.4%)	462 (43.3%)
High	12 (42.9%)	27 (13.3%)	4 (3.2%)	26 (23.9%)	16 (31.4%)	64 (28.4%)	20 (19.2%)	24 (10.7%)	193 (18.1%)
Missing	0 (0%)	3 (1.5%)	7 (5.6%)	2 (1.8%)	11 (21.6%)	8 (3.6%)	41 (39.4%)	96 (42.9%)	168 (15.7%)
Cohort Mean Birth Year									
Mean	1985.3	1985.2	1995.0	1982.4	1987.2	1986.0	1982.1	1991.6	1987.4
Cohort Mean Age Assessed									
Mean	25.8	26.2	19.3	24.5	26.3	28.4	30.5	17.9	24.4

Appendix 7: Study Characteristics of VP/VLBW cohorts not included in the IPD meta-

    analysis

			VP/VLBW		Controls		
Cohort	Birth year	IQ Test	IQ, *M (SD)*	*n*	IQ, *M (SD)*	*n*	Age at assessment, *M (SD)*
Constable (2013)(Constable et al., 2013)	1990	WISC TIQ	91.7 (12.4)	19	100.4 (18.7)	19	20.1 (0.9)
Hack (2002)(Hack et al., 2002)	1977	WAIS-R	86.87(14.23)[A]	236	92(14.4)	231	20
Hallin 2010)(Hallin et al., 2010)	1985	WAIS-III	93 (15.4)	52	106 (12.5)	54	18.3
Lefebvre (2005)(Lefebvre et al., 2005)	1976	WAIS-R	94(12)	59	108(14)	44	18.4
Stålnacke (2015)(Stålnacke et al., 2015)	1988	WISC-III	-0.315(1.165)[B]	118	0(1)	91	18

A = Derived from weighted average of the male and female reported scores.

B = Derived from the combined Z score for verbal and non-verbal ability

Appendix 8: Study 2 Cohort flow charts for the BLS and EPICure

```
┌─────────────────────────────────────────────┐
│              The EPICure Study                │
└─────────────────────────────────────────────┘
```

Initial EP infant Sample, N = 315	EPICure Control Participants recruited at age 6, N= 160
Potential Adult Sample, N= 306	Potential Adult Sample, N= 153
Assessed at age 19, N = 129	Assessed at age 19, N = 65
Self-reported ADHD and Executive functioning data available, N=107	Self-reported ADHD and Executive functioning data available N=60
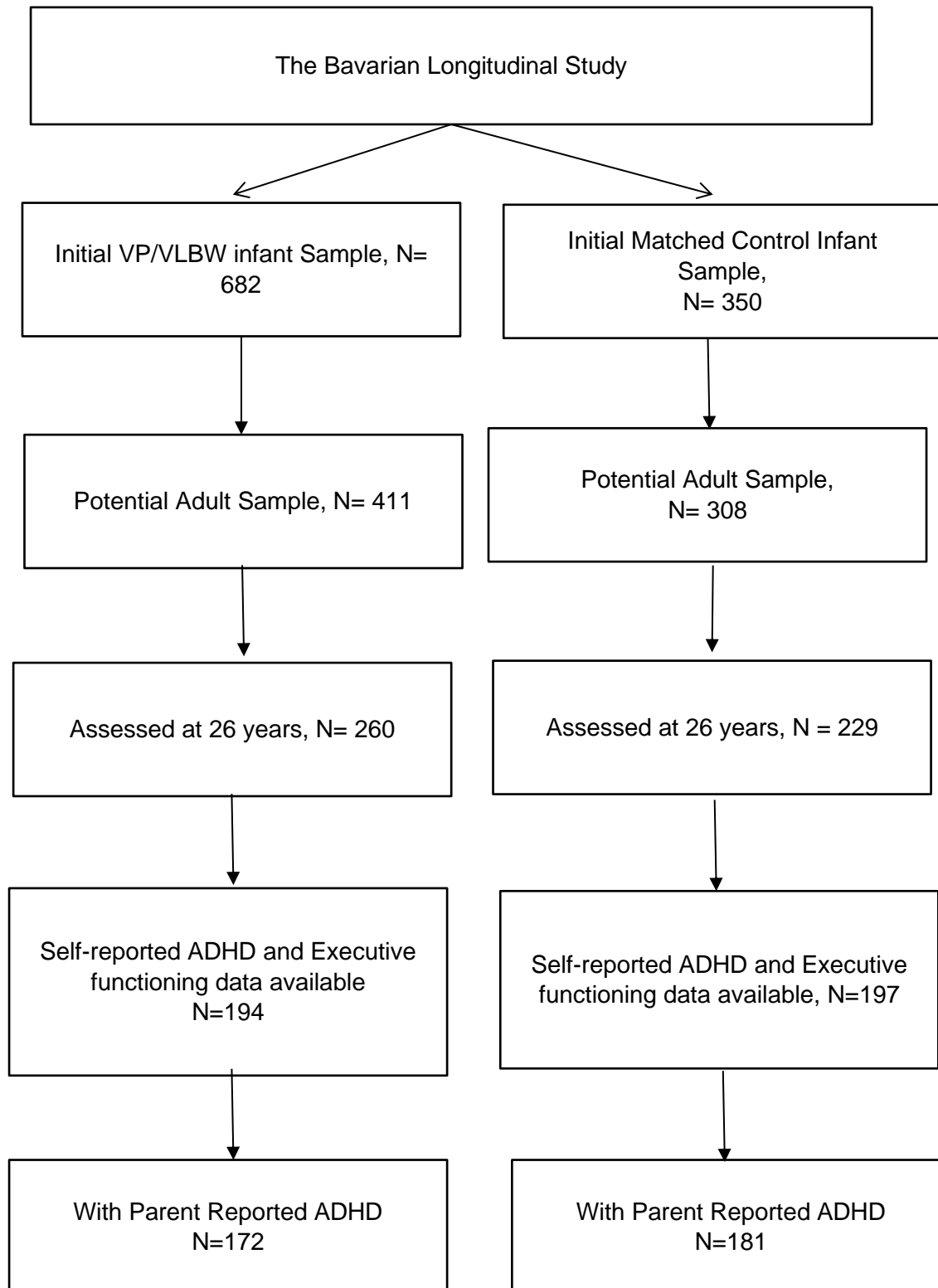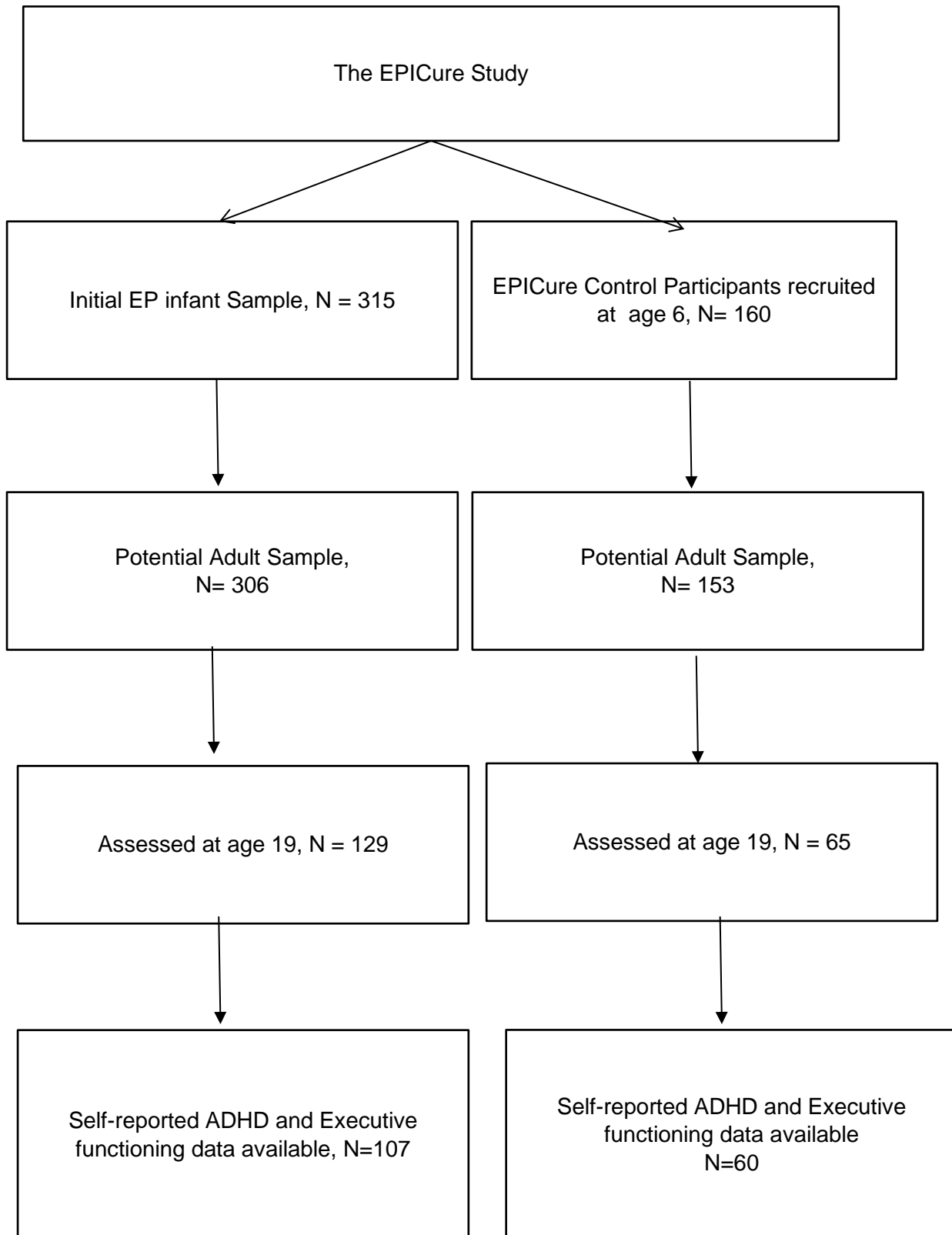
Appendix 9: Diagram of the attention network task used in study 2
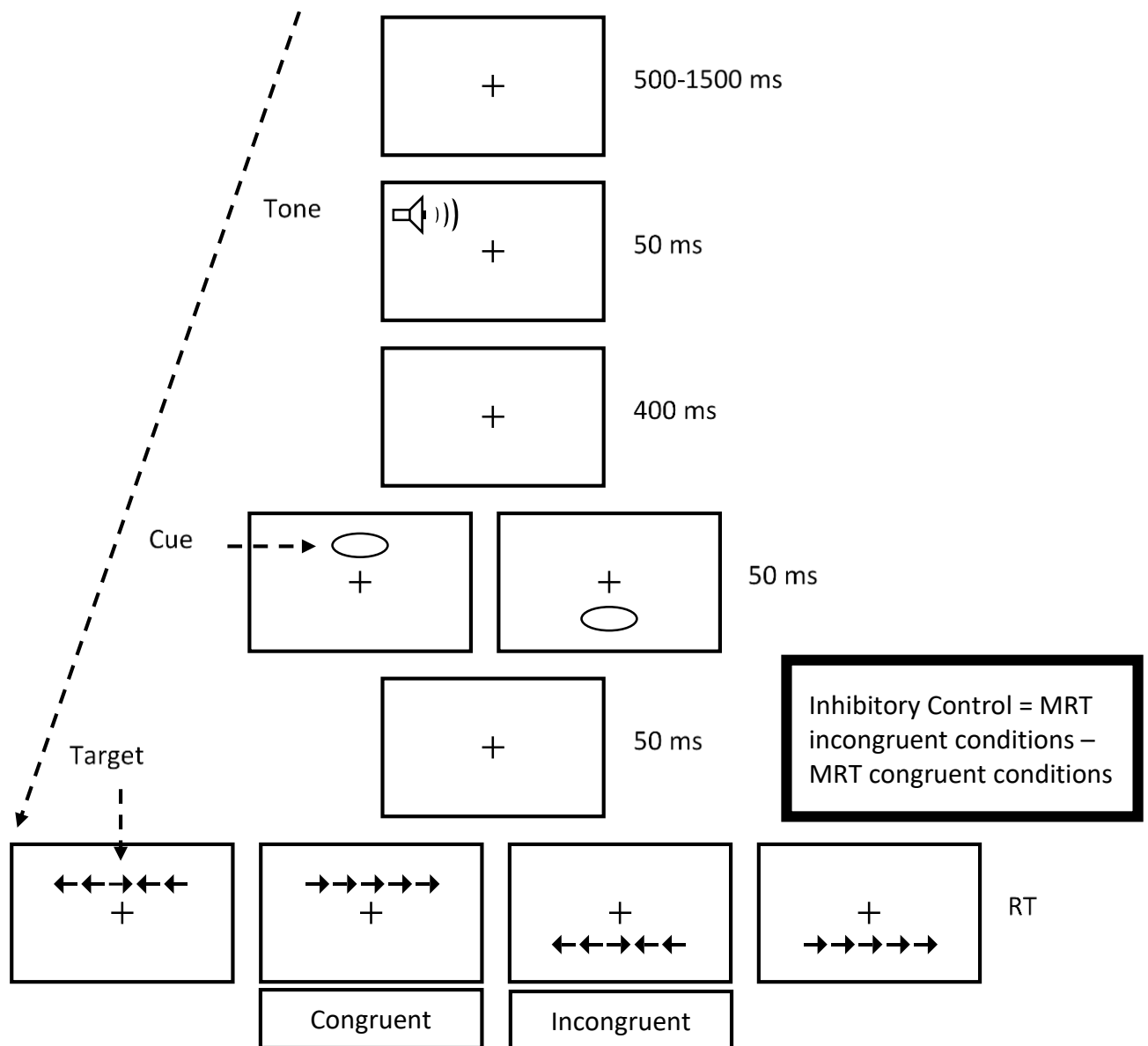
Figure 1. *Sequence of events in the ANT. A tone (present or absent) was followed by a spatial cue (top or bottom). The subsequent target arrow in the middle was either at the cued or uncued location and surrounded by congruent or incongruent flanker arrows. MRT = mean reaction time*

The ANT (Fan et al., 2002) was presented utilizing identical computers in both cohorts. Stimuli were presented on a 19" LCD monitor at approximately 57 cm and responses were recorded using the left and right arrow keys of a computer keyboard. Stimuli consisted of lines (thickness: 0.18° visual angle) and triangles drawn in grey (RGB values: 128, 128, 128)

on a black background. The sequence of events in each trial is depicted in the Figure above. Each trial started with the presentation of a fixation cross (1.5°) at the centre of the screen. After a random duration of 500 to 1500 ms, an auditory tone (~400Hz) was either played for 50 MS or not played. 400 MS later the spatial cue – a horizontal non-filled oval (1.5° x 0.75°) – was presented 5.4° above or below fixation for 50 Ms. After a short gap of 50 MS, five arrows (2.25° x 1.06°) were presented also 5.4° above or below fixation. The target arrow in the middle (i.e., aligned with fixation) was enclosed by flanker arrows 5.4° and 2.7° to the left and to the right of the target (see Figure above). The participant's task was to indicate the direction of the middle arrow by pressing the corresponding key. All stimuli were removed after the participant responded, and feedback was given after an erroneous response by presenting "error" for 1000 ms. Participants were instructed to respond as quickly and accurately as possible making less than 5% errors overall. The inter-trial interval was 1s.

Appendix 10: Correlation Matrices for the BLS and EPICure

BLS Correlation Matrix of measures

	Self-Reported Inattention	Self-Reported Hyperactivity	Parent-Reported Inattention	Parent-Reported Hyperactivity	Observer Rating of Attention	Inhibitory Control	Working Memory
Self-Reported Inattention							
Self-Reported Hyperactivity	0.43****						
Parent-Reported Inattention	0.26****	0.08					
Parent-Reported Hyperactivity	0.20***	0.19***	0.61****				
Observer Rating of Attention	-0.10	-0.03	-0.32****	-0.17**			

Inhibitory Control	0.07	-0.03	0.29****	0.11*	-0.23****		
Working Memory	-0.13*	0.05	-0.29****	-0.15**	0.40****	-0.24****	
IQ at 6 Years	-0.09	0.03	-0.44****	-0.22****	0.43****	-0.37****	0.54****

p < .0001****, p < .001*** , p < .01**, p < .05*

EPICure Correlation Matrix of measures

	Self-Reported Inattention	Self-Reported Hyperactivity	Observer Rating of Attention	Inhibitory Control	Working Memory
Self-Reported Inattention					
Self-Reported Hyperactivity	0.58****				
Observer Rating of Attention	-0.33****	-0.31***			
Inhibitory Control	0.12	0.13	-0.21**		
Working Memory	-0.11	-0.11	0.35****	-0.10	
IQ at 6 Years	-0.28***	-0.11	0.48****	-0.23**	0.57****

p < .0001****, p < .001*** , p < .01**, p < .05*

Appendix 11. Flow chart of the Bavarian Longitudinal Study from birth to 26 years for study 3.
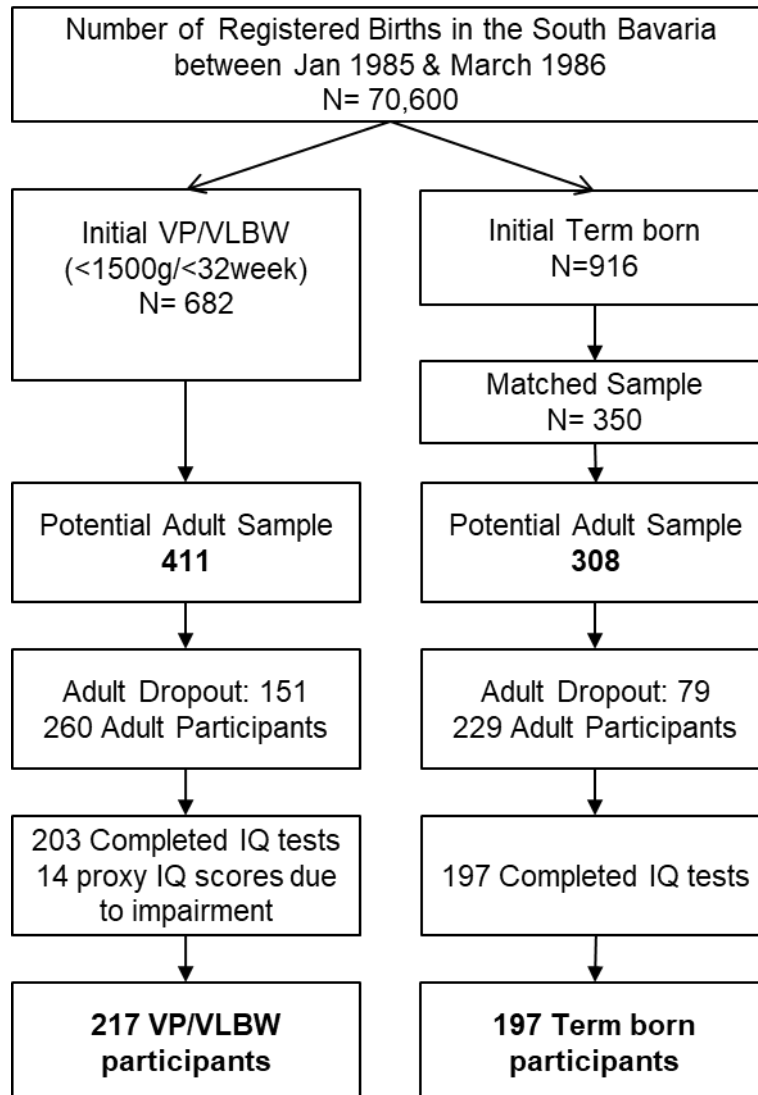
Appendix 12: Cognitive performance at 4 years composite- confirmatory factor analysis.

Scale Descriptions

Columbia Mental Maturity Scale: assesses the general reasoning ability of children between the ages of 3 and 10 years. The CMM scale consists of eight age-specific levels, each contains between 51 and 65 pictorial and figural classification items. The child has to select from a series of drawings the one drawing that is out of place.

Active Vocabulary Test: The AWST evaluates the expressive vocabulary of preschool children. It was developed for German-speaking countries and is similar to the widely used and valid Peabody Picture Vocabulary Test. The AWST consists of 82 drawings, and the child has to name the presented item.

Beery-Buktenica Developmental Test of Visual-Motor Integration: the BEERY measures the integration of visual and motor abilities. In the short version 15 drawings of geometric forms are arranged in order of increasing difficulty that the child is asked to copy. Each drawing is evaluated using predefined scoring criteria, i.e., task solved versus not solved, and a sum score is computed, ranging from 0 to 15. A higher score indicates better performance.

Factor Loadings onto the IQ at 4 years latent variable :

Factor	B	SE	Z	p-value	Beta
CMM	11.363	0.191	59.465	<0.001	0.828
AWST	11.137	0.212	52.481	<0.001	0.728
BEERY	1.707	0.036	47.266	<0.001	0.654
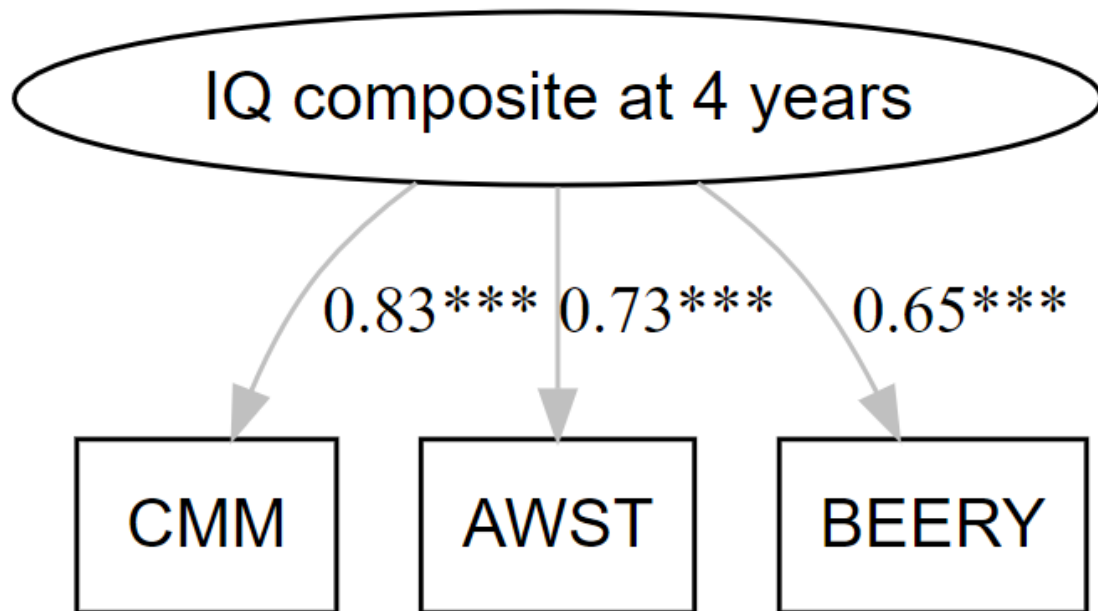
Appendix 13: Description of the PIRI.

Standardized interview with parents (SI) and research nurses' observations (NO) of attachment-related parental concerns, feelings, and behavior. All research nurses were trained in advance, but inter-rater agreement was not assessed. The scale comprised 8 items of yes (1) or no (0) ratings on the following items: (1) mother does not yet know the infant (SI), (2)

mother visits the infant once a week or less (SI), (3) father visits the infant less than once a week (SI), (4) mother is insecure when taking care of the child at home (SI), (5) mother shows little pleasure when interacting with the child (NO), (6) father shows little pleasure when interacting with the child (NO), (7) the probability that these parents develop problems in taking care of the infant is high (NO), (8) mother has trouble building a relationship with the child (SI). First, a sum score was calculated by adding one point for each 'yes' answer. As the resulting sum score did not show a normal distribution as most parents reported and demonstrated a good relationship with their infant, the sum score was recoded into no concerns for the parent–infant relationship and some degree of concern for the parent–infant relationship

Appendix 14: Perinatal characteristic of the BLS participants according to SGA status and VP/VLBW (left SGA$_N$, right SGA$_F$)

Appendix 14: Perinatal characteristic of the BLS participants according to SGA status and VP/VLBW (left SGAN, right SGAF)

	SGA using Neonatal Reference				SGA using Foetal Reference			
	AGA+ VP/VLBW (n=156)	SGA+ VP/VLBW (n=61)	AGA+ Term Born (n=183)	SGA+ Term Born (n=14)	AGA+ VP/VLBW (n=101)	SGA+ VP/VLBW (n=116)	AGA+ Term-Born (n=175)	SGA+ Term-Born (n=22)
Maternal Age								
Mean (SD)	29.1 (4.85)	28.5 (4.42)	29.3 (4.66)	28.8 (4.79)	29.3 (4.84)	28.6 (4.64)	29.5 (4.77)	27.5 (3.23)
Mother smoked during pregnancy								
Did not smoke	128 (82.1%)	52 (85.2%)	168 (91.8%)	14 (100%)	86 (85.1%)	94 (81.0%)	161 (92.0%)	21 (95.5%)
Did smoke	23 (14.7%)	7 (11.5%)	15 (8.2%)	0 (0%)	12 (11.9%)	18 (15.5%)	14 (8.0%)	1 (4.5%)
Missing	5 (3.2%)	2 (3.3%)	0 (0%)	0 (0%)	3 (3.0%)	4 (3.4%)	0 (0%)	0 (0%)
Parity								
Multiparous	70 (44.9%)	17 (27.9%)	78 (42.6%)	7 (50.0%)	50 (49.5%)	37 (31.9%)	79 (45.1%)	6 (27.3%)

Primiparous	86 (55.1%)	44 (72.1%)	105 (57.4%)	7 (50.0%)	51 (50.5%)	79 (68.1%)	96 (54.9%)	16 (72.7%)
Supplemental oxygen duration (days)								
Mean (SD)	6.98 (7.47)	7.41 (8.90)	NA (NA)	NA (NA)	6.34 (7.40)	7.77 (8.25)	NA (NA)	NA (NA)
Cerebral Palsy Grade 3 or 4								
No	133 (85.3%)	49 (80.3%)	183 (100%)	14 (100%)	87 (86.1%)	95 (81.9%)	175 (100%)	22 (100%)
Yes	10 (6.4%)	1 (1.6%)	0 (0%)	0 (0%)	5 (5.0%)	6 (5.2%)	0 (0%)	0 (0%)
Missing	13 (8.3%)	11 (18.0%)	0 (0%)	0 (0%)	9 (8.9%)	15 (12.9%)	0 (0%)	0 (0%)
Visual Impairment								
Not Blind	140 (89.7%)	50 (82.0%)	183 (100%)	14 (100%)	90 (89.1%)	100 (86.2%)	175 (100%)	22 (100%)

Blind	3 (1.9%)	0 (0%)	0 (0%)	0 (0%)	2 (2.0%)	1 (0.9%)	0 (0%)	0 (0%)
Missing	13 (8.3%)	11 (18.0%)	0 (0%)	0 (0%)	9 (8.9%)	15 (12.9%)	0 (0%)	0 (0%)

Hearing Impairment

Not deaf	143 (91.7%)	50 (82.0%)	182 (99.5%)	14 (100%)	92 (91.1%)	101 (87.1%)	174 (99.4%)	22 (100%)
Deaf	0 (0%)	0 (0%)	1 (0.5%)	0 (0%)	0 (0%)	0 (0%)	1 (0.6%)	0 (0%)
Missing	13 (8.3%)	11 (18.0%)	0 (0%)	0 (0%)	9 (8.9%)	15 (12.9%)	0 (0%)	0 (0%

IVH

none, stage 1 or stage 2	140 (89.7%)	59 (96.7%)	183 (100%)	14 (100%)	94 (93.1%)	105 (90.5%)	175 (100%)	22 (100%)
stage 3 or stage 4	16 (10.3%)	2 (3.3%)	0 (0%)	0 (0%)	7 (6.9%)	11 (9.5%)	0 (0%)	0 (0%)

Prenatal infectious disease								
No	135 (86.5%)	52 (85.2%)	151 (82.5%)	13 (92.9%)	86 (85.1%)	101 (87.1%)	145 (82.9%)	19 (86.4%)
Yes	20 (12.8%)	9 (14.8%)	32 (17.5%)	1 (7.1%)	14 (13.9%)	15 (12.9%)	30 (17.1%)	3 (13.6%)
Premature rupture of membrane								
No	78 (50.0%)	38 (62.3%)	149 (81.4%)	12 (85.7%)	47 (46.5%)	69 (59.5%)	143 (81.7%)	18 (81.8%)
Yes	64 (41.0%)	15 (24.6%)	30 (16.4%)	2 (14.3%)	43 (42.6%)	36 (31.0%)	28 (16.0%)	4 (18.2%)
Missing	14 (9.0%)	8 (13.1%)	4 (2.2%)	0 (0%)	11 (10.9%)	11 (9.5%)	4 (2.3%)	0 (0%)

Appendix 15: Percentage of mild (DQ/IQ <85) and major (DQ/IQ < 70) cognitive impairment of the BLS participants according to VP/VLBW and SGA status (Top SGAF-Fetal reference, Bottom SGAN-Neonatal Reference).

Appendix 15: Percentage of mild (DQ/IQ <85) and major (DQ/IQ < 70) cognitive impairment of the BLS participants according to VP/VLBW and SGA status (Top SGA$_F$-Fetal reference, Bottom SGA-Neonatal Reference).

SGA/ AGA	Birth Group	n	5M DQ <70 %	5M DQ <85 %	20M DQ <70 %	20M DQ <85 %	4Y IQ < 70 %	4Y IQ <85 %	6Y IQ <70 %	6Y IQ <85 %	8Y IQ <70 %	8Y IQ <85 %	26Y IQ <70 %	26Y IQ <85 %
AGA$_F$	VP/ VLBW	101	17	32	29	56	27	43	24	58	20	44	22	48
SGA$_F$	VP/ VLBW	116	36	57	53	75	34	54	33	55	32	60	32	51
AGA$_F$	Term	175	3	17	1	21	5	14	3	13	3	15	3	15
SGA$_F$	Term	22	0	32	9	23	0	9	9	23	0	14	5	23
AGA$_N$	VP/ VLBW	156	22	40	38	57	30	46	26	54	24	42	27	49

SGA$_N$	VP/ VLBW	61	41	59	53	66	36	56	34	64	33	59	28	51
AGA$_N$	Term	183	3	16	2	20	5	14	3	14	3	11	3	15
SGA$_N$	Term	14	0	43	7	21	0	14	7	21	0	7	0	21

Note: Due to the small number of term born SGA participants, the rates of impairment in this sample should be interpreted with caution.

Appendix 16: GRIPP2 short form

Section and topic	Item	Reported on page No
1: Aim	Report the aim of PPI in the study	The BLS underwent peer review but did not involve public involvement. Page 6.
2: Methods	Provide a clear description of the methods used for PPI in the study	None
3: Study results	Outcomes—Report the results of PPI in the study, including both positive and negative outcomes	None
4: Discussion and conclusions	Outcomes—Comment on the extent to which PPI influenced	As there was no PPI, there is the potential difference between the constructs of interests measured by the researchers and those that would be of interest to the public

	the study overall. Describe positive and negative effects	
5: Reflections/critical perspective	Comment critically on the study, reflecting on the things that went well and those that did not, so others can learn from this experience	As the BLS is a long term, older study it was established when PPI was not a requirement or crucial part of research. The funding that supported this analysis (RECAP-preterm) has Parent involvement via the European Foundation for the Care of Newborn Infants (EFCNI) – and cognitive outcome and associated life chances were considered as a major outcome by parents of preterm children

PPI=patient and public involvement

END