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

Table of Contents 2
List of Tables 5
List of Figures 6
Abbreviations 7
Acknowledgements 10
Declaration/ Unique Contribution 11
Funding 14
Abstract 15
Structure of Thesis 17

Chapter 1  Introduction to the importance of early foetal factors ..........18

1.1 Preterm Birth.................................................................18
1.1.1 Causes and risk factors for preterm birth 21
1.1.2 Prevalence and survival rates for preterm birth globally 22
1.1.3 Perinatal morbidities as a result of preterm birth 23
1.2 Small for Gestational Age.............................................24
1.2.1 Differentiation of prematurity, low birthweight, and small for gestational age 24
1.2.2 Overlap between preterm birth and SGA 25
1.2.3 Causes and risk factors for SGA 26
1.2.4 Differences in SGA classification and references 26
1.2.5 Survival rates and perinatal morbidities as a result of SGA birth 29
1.3 Major disability as a result of preterm birth or SGA .................................................................29
1.4 Why does it matter? .....................................................30

Chapter 2  Theoretical models of the relationship of general cognitive, specific cognitive and behavioural outcomes..........................................................32

2.1 Rationale and History of IQ testing .......................32
2.2 Overview of Developmental Quotient and early IQ assessments and their associations with adult IQ .34
2.3 Executive Functioning models and its links to IQ ..36
2.4 Relationship between specific cognitive skills and attention............................................................40
2.5 Chapter Summary............................................................41

Chapter 3  The association of VP/VLBW and SGA with later cognitive and behavioural outcomes..........................................................42
3.1 General cognitive outcomes of VP/VLBW in infancy, childhood, and adulthood ........................................ 42
3.2 Specific neurocognitive and behavioural outcomes of preterms in both childhood and adulthood ............. 47
3.2.1 Executive Functions ........................................... 47
3.2.2 Attention ......................................................... 49
3.2.3 The Preterm Phenotype ........................................ 50
3.2.4 Summary of the links between preterm birth with cognitive and behavioural outcomes .................. 51
3.3 General cognitive outcomes of SGA individuals in infancy, childhood, and adulthood ......................... 52
3.4 Factors that must be accounted for in cognitive outcomes of VP/VLBW and SGA individuals ........... 55

Chapter 4 Outstanding issues and research questions ........................................ 57
4.1 Study 1 ........................................................................... 61
4.2 Study 2 ........................................................................... 61
4.3 Study 3 ........................................................................... 62

Chapter 5 Methodology .................................................................................. 63
5.1 Overview of the Bavarian Longitudinal Study .... 63
5.2 Overview of the EPICure study .............................. 65
5.3 Overview of RECAP-preterm and APIC cohorts 66
5.4 Measures ................................................................. 71
5.5 Statistical Analyses ....................................................... 75

Chapter 6 Intelligence of VP/VLBW individuals in adulthood and its antecedents: an IPD meta-analysis .............................................................. 76

Chapter 7 The Role of Executive and General Cognitive Functioning in the Attention Problems of Very and Extremely Preterm Adults .... 101

Chapter 8 Small for Gestational Age-Cognitive Performance from Infancy to Adulthood: An Observational Study ............................................. 123

Chapter 9 Overall Discussion ....................................................................... 142
9.1 Summary of Results: ................................................... 142
9.2 Integrated Discussion and Theoretical Implications .......................................................................... 143
9.3 Strengths and Limitations of the Research ................. 152
9.4 Future Directions and Practical Implications .......... 157
9.5 Concluding Remarks .............................................................................. 160
References 162

Appendices 219

Appendix 1: Childhood Neurosensory Impairment in VP/VLBW participants from each IPD cohort ..........220
Appendix 2: Newcastle Ottawa criteria and ratings for each IPD Cohort criteria : .......................................................223
Appendix 3: Linear mixed model demonstrating reducing Gestational Age by Birth Year among VP/VLBW participants .................................................................228
Appendix 4: Linear mixed model demonstrating reducing Birthweight by Birth Year among VP/VLBW participants .....229
Appendix 5: IQ and Demographic information of all participants from each IPD cohort.................................230
Appendix 6: Neonatal and demographic data for VP/VLBW participants from each IPD cohort ..................234
Appendix 7: Study Characteristics of VP/VLBW cohorts not included in the IPD meta-analysis .........................238
Appendix 8: Study 2 Cohort flow charts for the BLS and EPICure ..............................................................................239
Appendix 9: Diagram of the attention network task used in study 2 ........................................................................241
Appendix 10: Correlation Matrices for the BLS and EPICure ......243
Appendix 11. Flow chart of the Bavarian Longitudinal Study from birth to 26 years for study 3 .........................246
Appendix 12: Cognitive performance at 4 years composite-confirmatory factor analysis..........................247
Appendix 13: Description of the PIRI .................................................248
Appendix 14: Perinatal characteristic of the BLS participants according to SGA status and VP/VLBW (left SGA_N, right SGA_F) .................................................................250
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). ..................................................................................256
Appendix 16: GRIPP2 short form.................................................259
List of Tables

Table 1: Definitions for commonly used categories for foetal groups 19
Table 2: Summary of key research areas for this thesis 58
Table 3: Summary of the VP/VLBW cohorts included in Study 1 69
Table 4: Key measures of interest for this thesis 71
Table 5: Summary of cohorts included in the IPD analysis 84
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 91
Table 7: Univariate differences between VP/VLBW or EP participants and
controls 111
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 117
Table 9: Baseline characteristics of adults born SGA and AGA according to the
Neonatal based reference (SGAN) or Fetal Reference (SGAF) 130
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 136
List of Figures

Figure 1: SGA by gestational age, adapted from Pritchard et al. (2019) 28
Figure 2: Examples of the block design subtest, adapted from Lind et al. (2014) 34
Figure 3: Diamond’s model of executive functioning, adapted from Diamond (2013) 38
Figure 4: Anderson’s model of executive functioning, adapted from Anderson (2002) 39
Figure 5: Flow chart of studies included in the IPD and aggregate meta-analyses 95
Figure 6: Aggregate meta-analysis comparing IQ performance in IPD and Non-IPD VP/VLBW adult cohorts 96
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. 114
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 116
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. 133
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 134
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:

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

Contributions:
- Marina Mendonça
- Peter Bartmann
- Dieter Wolke

Reference:
One ancillary paper has also been published:


Funding

This PhD, as part of the RECAP preterm project, was funded by an EU Horizon 2020 grant, agreement number 733280.
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 alternative 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

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

---

**Figure 1: SGA by gestational age, adapted from Pritchard et al. (2019)**
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
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 (Guelles 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.

![Segmented and Unsegmented](image)

**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)
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 findings 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

<table>
<thead>
<tr>
<th>Area of Research</th>
<th>Very Preterm/Very low Birthweight Research</th>
<th>Small for Gestational Age Research</th>
</tr>
</thead>
<tbody>
<tr>
<td>IQ</td>
<td>Executive Functioning</td>
<td>Attention Problems</td>
</tr>
<tr>
<td>What is currently well known</td>
<td>VP/VLBW perform poorer than controls on early measures of developmental quotient and child IQ (Twilhaar et al., 2018)</td>
<td>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)</td>
</tr>
<tr>
<td>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)</td>
<td>SGA children perform poorer than controls on early measures of developmental quotient and child IQ (Bie et al., 2010)</td>
<td></td>
</tr>
<tr>
<td>IQ</td>
<td></td>
<td>IQ</td>
</tr>
</tbody>
</table>
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)?
<table>
<thead>
<tr>
<th>Where there is sparse or no evidence</th>
<th>What are the key neonatal or social variables that are universal risk and resiliency factors for adult IQ among VP/VLBW adults?</th>
<th>In adulthood, does executive functioning performance explain the difference between VP/VLBW adults and controls on measures of attention?</th>
<th>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?</th>
<th>Do SGA participants demonstrate “cognitive catch up” with smaller cognitive differences to controls in adulthood than in childhood?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Important for which research study in this thesis</td>
<td>Study 1</td>
<td>Study 2</td>
<td>Study 2</td>
<td>Study 3</td>
</tr>
</tbody>
</table>
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, Finland, 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

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

<table>
<thead>
<tr>
<th>Measure</th>
<th>Age the measure was collected</th>
<th>Description of Measure</th>
<th>Used as a Predictor or as an Outcome</th>
<th>Used in study 1, 2 or 3 and in which cohorts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gestational Age</td>
<td>Birth</td>
<td>Weeks the foetus has been carried for as estimated by ultrasound estimate, last menstrual period or neonatal examination:</td>
<td>Predictor</td>
<td>1,2,3 - All cohorts</td>
</tr>
<tr>
<td>Birth Weight</td>
<td>Birth</td>
<td>Measured in grams as taken from the neonatal records</td>
<td>Predictor</td>
<td>1,2,3 - All cohorts</td>
</tr>
<tr>
<td>SGA- Foetal</td>
<td>Birth</td>
<td>Birthweight percentile below the 10th centile in comparison to the</td>
<td>Predictor</td>
<td>3 - BLS</td>
</tr>
<tr>
<td>Predictor</td>
<td>Birth</td>
<td>estimated weight of foetuses at equivalent gestation</td>
<td>Predictor</td>
<td>3 - BLS</td>
</tr>
<tr>
<td>-----------</td>
<td>-------</td>
<td>------------------------------------------------------</td>
<td>-----------</td>
<td>---------</td>
</tr>
<tr>
<td>SGA- Neonatal</td>
<td>Birth</td>
<td>Birthweight percentile below the 10th centile in comparison to other neonates of equivalent gestation</td>
<td>Predictor</td>
<td>1 - All cohorts</td>
</tr>
<tr>
<td>Birthweight Z score</td>
<td>Birth</td>
<td>A continuous, relative measure of the weight of infant for their gestation</td>
<td>Predictor</td>
<td>1 - All cohorts</td>
</tr>
<tr>
<td>Intraventricular Haemorrhage (IVH)</td>
<td>Birth</td>
<td>Categorical measure of whether bleeding occurred inside or around the ventricles</td>
<td>Predictor</td>
<td>1 - All cohorts</td>
</tr>
<tr>
<td>Bronchopulmonary Dysplasia (BPD)</td>
<td>Birth</td>
<td>Categorical measure of long-term breathing problems. Defined by the use of oxygen after a certain time period (post menstrual age or following birth)</td>
<td>Predictor</td>
<td>1 - All cohorts</td>
</tr>
<tr>
<td>Parental Socio-economic Status/ Maternal Education</td>
<td>Birth</td>
<td>a weighted composite score of maternal/paternal</td>
<td>Predictor</td>
<td>1,2,3 - All cohorts</td>
</tr>
<tr>
<td>Predictor</td>
<td>Outcome</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------</td>
<td>---------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parent-Infant Relationship</td>
<td>Scale derived from concerns regarding the mother and infant or father and infant relationship</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Developmental Quotient</td>
<td>Griffiths ordinal measure of early developmental abilities</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Childhood IQ</td>
<td>IQ tests used to measure full scale child intellectual abilities (test used depends on age of assessment)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adult IQ</td>
<td>WAIS/WASI used to measure full scale adult intellectual abilities (test used depends on cohort)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adult ADHD Symptoms-</td>
<td>Self</td>
<td>19/26 years (EPICure and BLS)</td>
<td>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.</td>
<td>Outcome</td>
</tr>
<tr>
<td>---------------------</td>
<td>------</td>
<td>-----------------------------</td>
<td>-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>---------</td>
</tr>
<tr>
<td>Parent Reported</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adult ADHD Symptoms-</td>
<td>Parent</td>
<td>26 years (BLS)</td>
<td>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.</td>
<td>Outcome</td>
</tr>
<tr>
<td>Tester Rating of Adult Behaviour – Attention span (TRAB -AS)</td>
<td>19/26 years (EPICure and BLS)</td>
<td>Psychologists rated the participant’s attention on a scale from very short span to very long</td>
<td>Outcomes</td>
<td>2 - BLS and EPICure</td>
</tr>
<tr>
<td>Working Memory</td>
<td>19/26 years</td>
<td>Letter Number Sequencing task (BLS) or backwards digit recall task (EPICure)</td>
<td>Predictor</td>
<td>2 - BLS and EPICure</td>
</tr>
<tr>
<td>----------------</td>
<td>-------------</td>
<td>---------------------------------------------------------------------------------</td>
<td>-----------</td>
<td>---------------------</td>
</tr>
<tr>
<td>Inhibitory Control</td>
<td>19/26 years</td>
<td>Attention Network Task – Inhibitory Control</td>
<td>Predictor</td>
<td>2 - BLS and EPICure</td>
</tr>
</tbody>
</table>

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 I² while a subgroup analysis differentiated between IPD and non-IPD cohorts as to test for selection bias.

**Results**

**Study Selection and IPD Obtained**

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>
<thead>
<tr>
<th>Cohort</th>
<th>Birth Year</th>
<th>Country</th>
<th>Mean Age</th>
<th>IQ test</th>
<th>Initial Eligibility</th>
<th>Initial VP/ VLBW Surviving to Discharge</th>
<th>Eligible Adult VP/ VLBW (N)</th>
<th>VP/ VLBW Attrition %</th>
<th>VP/ VLBW+ NSI Attrition %</th>
<th>VP/ VLBW with IQ scores (N)</th>
<th>Control Group (N)</th>
<th>Harmonization Issues</th>
<th>VP/ VLBW Neonatal / Maternal Education Data</th>
<th>Imputed %</th>
</tr>
</thead>
<tbody>
<tr>
<td>AYLS</td>
<td>1985-1986</td>
<td>Finland</td>
<td>26</td>
<td>WAIS</td>
<td>Preterm</td>
<td>108</td>
<td>68</td>
<td>59</td>
<td>80</td>
<td>28</td>
<td>303</td>
<td>None</td>
<td>Recruited Infant</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(Heinonen et al., 2018)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BLS</td>
<td>1985-1986</td>
<td>Germany</td>
<td>26</td>
<td>WAIS</td>
<td>VP/ VLBW</td>
<td>510</td>
<td>411</td>
<td>51</td>
<td>76</td>
<td>203</td>
<td>192</td>
<td>None</td>
<td>Recruited Infant</td>
<td>&lt;1%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(Eryigit Madzwa)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study</td>
<td>Year</td>
<td>Country</td>
<td>Age</td>
<td>Test</td>
<td>Score at Birth</td>
<td>Score at Follow-Up</td>
<td>IQ</td>
<td>Recruitment Age</td>
<td>Deficit</td>
<td>Notes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------</td>
<td>---------</td>
<td>---------</td>
<td>------</td>
<td>----------</td>
<td>----------------</td>
<td>-------------------</td>
<td>----</td>
<td>----------------</td>
<td>---------</td>
<td>-------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EPICure</td>
<td>1995</td>
<td>UK &amp; Ireland</td>
<td>19</td>
<td>WAIS EP (&lt;26 weeks)</td>
<td>315</td>
<td>306</td>
<td>59</td>
<td>69</td>
<td>124</td>
<td>64- Recruited at ages 6 or 11</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>None</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HESVA</td>
<td>1978-1985</td>
<td>Finland</td>
<td>25</td>
<td>WAIS VLBW (&lt;1500 g)</td>
<td>334</td>
<td>254</td>
<td>57</td>
<td>69</td>
<td>109</td>
<td>98-Recruited in Adulthood</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1985</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Maternal education measured in adulthood. NSI did not include IQ &lt;70 and could not differentiate ambulatory/non-</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
ambulatory cerebral palsy.

<table>
<thead>
<tr>
<th>Country</th>
<th>Year</th>
<th>Gender</th>
<th>Sample Size</th>
<th>WASI</th>
<th>VLBW (&lt;1500 g)</th>
<th>WASI-II</th>
<th>VLBW (&lt;1500 g)</th>
<th>Maternal Education</th>
<th>Recruitment Stage</th>
</tr>
</thead>
<tbody>
<tr>
<td>NTNU</td>
<td>1986-1988</td>
<td>Norway</td>
<td>26</td>
<td>WASI</td>
<td>VLBW (1999)</td>
<td>86</td>
<td>82</td>
<td>37</td>
<td>60</td>
</tr>
<tr>
<td>NZ VLBW</td>
<td>1986</td>
<td>New Zealand</td>
<td>28</td>
<td>WASI-II</td>
<td>VLBW (2011)</td>
<td>338</td>
<td>323</td>
<td>30</td>
<td>64</td>
</tr>
<tr>
<td>Study</td>
<td>Country</td>
<td>Date</td>
<td>Sample Size</td>
<td>Intelligence Scale</td>
<td>Classification</td>
<td>Age at Recruitment</td>
<td>IQ Cut-off</td>
<td>Maternal Education</td>
<td>Follow-up</td>
</tr>
<tr>
<td>-------</td>
<td>---------</td>
<td>------</td>
<td>-------------</td>
<td>-------------------</td>
<td>----------------</td>
<td>-------------------</td>
<td>------------</td>
<td>-------------------</td>
<td>-----------</td>
</tr>
<tr>
<td>UCLH</td>
<td>UK</td>
<td>1979-1984</td>
<td>30</td>
<td>WAIS II 302A</td>
<td>VP(&lt;33 weeks, reduced to VP/VLBW)</td>
<td>220A</td>
<td>46</td>
<td>85</td>
<td>104</td>
</tr>
<tr>
<td>VICS</td>
<td>Australia</td>
<td>1991-1992</td>
<td>18</td>
<td>WAIS II 299</td>
<td>EP/ELBW(&lt;28 weeks/ &lt;1000 g)</td>
<td>277</td>
<td>19</td>
<td>37</td>
<td>224</td>
</tr>
</tbody>
</table>

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 than 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

<table>
<thead>
<tr>
<th>Factors</th>
<th>1 Stage Univariable Estimate</th>
<th>CI</th>
<th>p</th>
<th>1 Stage Multivariable Estimate</th>
<th>CI</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individual Level Factors</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gestational age (weeks)</td>
<td>0.04</td>
<td>0.02, 0.06</td>
<td>&lt;0.001</td>
<td>0.11</td>
<td>0.07, 0.14</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Male</td>
<td>0.07</td>
<td>-0.05, 0.20</td>
<td>0.26</td>
<td>0.09</td>
<td>-0.03, 0.20</td>
<td>0.16</td>
</tr>
<tr>
<td>Birthweight_Z score (per 1 SD)</td>
<td>0.05</td>
<td>-0.01, 0.11</td>
<td>0.08</td>
<td>0.21</td>
<td>0.14, 0.28</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Maternal education (1=Low, 2=Medium, 3= High)</td>
<td>0.25</td>
<td>0.17, 0.34</td>
<td>&lt;0.001</td>
<td>0.26</td>
<td>0.17, 0.35</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Bronchopulmonary dysplasia (Reference: No BPD)&lt;sup&gt;A&lt;/sup&gt;</td>
<td>-0.37</td>
<td>-0.51, -0.23</td>
<td>&lt;0.001</td>
<td>-0.16</td>
<td>-0.30, -0.02</td>
<td>0.02</td>
</tr>
<tr>
<td>-Defined as oxygen after 28 days post birth&lt;sup&gt;B&lt;/sup&gt;</td>
<td>-0.34</td>
<td>-0.56, -0.12</td>
<td>0.003</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>
- Defined as oxygen after 36 weeks’ postmenstrual age\textsuperscript{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, & - 0.007 \\

- IVH Grade 3 or 4 (Reference: All other grades)\textsuperscript{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.
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

<table>
<thead>
<tr>
<th>Source</th>
<th>SMD (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source = IPD</td>
<td></td>
</tr>
<tr>
<td>EPICURE</td>
<td>-1.22 [-1.54; -0.89]</td>
</tr>
<tr>
<td>AYLS</td>
<td>-0.98 [-1.37; -0.59]</td>
</tr>
<tr>
<td>NTNU</td>
<td>-0.94 [-1.32; -0.57]</td>
</tr>
<tr>
<td>BLS</td>
<td>-0.91 [-1.12; -0.70]</td>
</tr>
<tr>
<td>NZ VLBW</td>
<td>-0.83 [-1.08; -0.59]</td>
</tr>
<tr>
<td>VICS</td>
<td>-0.71 [-0.92; -0.50]</td>
</tr>
<tr>
<td>UCLH</td>
<td>-0.68 [-0.97; -0.38]</td>
</tr>
<tr>
<td>HESVA</td>
<td>-0.59 [-0.87; -0.32]</td>
</tr>
<tr>
<td>Total</td>
<td>-0.84 [-0.97; -0.71]</td>
</tr>
<tr>
<td>Heterogeneity: $\chi^2 = 11.93 \ (P = .10), \ I^2 = 41%$</td>
<td></td>
</tr>
<tr>
<td>Source = Non-IPD</td>
<td></td>
</tr>
<tr>
<td>Lefebvre(2007)</td>
<td>-1.08 [-1.50; -0.66]</td>
</tr>
<tr>
<td>Hallin(2010)</td>
<td>-0.92 [-1.32; -0.52]</td>
</tr>
<tr>
<td>Constable(2013)</td>
<td>-0.54 [-1.19; 0.11]</td>
</tr>
<tr>
<td>Hack(2002)</td>
<td>-0.36 [-0.54; -0.17]</td>
</tr>
<tr>
<td>Stålnacke(2015)</td>
<td>-0.29 [-0.56; -0.01]</td>
</tr>
<tr>
<td>Total</td>
<td>-0.61 [-0.93; -0.29]</td>
</tr>
<tr>
<td>Heterogeneity: $\chi^2 = 16.17 \ (P &lt; .01), \ I^2 = 75%$</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>-0.76 [-0.92; -0.60]</td>
</tr>
<tr>
<td>Prediction interval</td>
<td>[-1.31; -0.21]</td>
</tr>
<tr>
<td>Heterogeneity: $\chi^2 = 45.54 \ (P &lt; .01), \ I^2 = 74%$</td>
<td></td>
</tr>
<tr>
<td>Residual heterogeneity: $\chi^2 = 28.10 \ (P &lt; .01), \ I^2 = 61%$</td>
<td></td>
</tr>
</tbody>
</table>
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 postnataally 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 Madzwanuse 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 $\alpha = 0.75$, EPICure $\alpha = 0.85$). In the BLS cohort only, parents also assessed their child’s ADHD symptoms using the same questionnaire, with a similarly good internal reliability ($\alpha = 0.88$). 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

<table>
<thead>
<tr>
<th></th>
<th>Bavarian Longitudinal Study</th>
<th></th>
<th></th>
<th>EPI Cure</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean difference</td>
<td>Adjusted P-Value</td>
<td>Cohen’s D</td>
<td>Mean difference</td>
<td>Adjusted P-Value</td>
<td>Cohen’s D</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(VP/VLBW-Control)</td>
<td>(EP-Control)</td>
<td></td>
<td>95% CI</td>
<td></td>
<td></td>
<td>95% CI</td>
</tr>
<tr>
<td>ADHD- Inattention Self-Reported symptoms – Z scored</td>
<td>0.12</td>
<td>[-0.09, 0.34]</td>
<td>0.522</td>
<td>0.11</td>
<td></td>
<td></td>
<td>0.39</td>
</tr>
<tr>
<td>ADHD- Hyperactivity/impulsivity Self-Reported – Z scored</td>
<td>-0.16</td>
<td>[-0.36, 0.03]</td>
<td>0.340</td>
<td>-0.17</td>
<td>-0.06</td>
<td>[-0.40, 0.29]</td>
<td>0.739</td>
</tr>
<tr>
<td>ADHD- Combined Self-Reported – Z scored</td>
<td>-0.05</td>
<td>[-0.26, 0.15]</td>
<td>0.597</td>
<td>-0.05</td>
<td>0.19</td>
<td>[-0.16, 0.54]</td>
<td>0.543</td>
</tr>
<tr>
<td>ADHD- Inattention Parent Reported – Z scored</td>
<td>0.95</td>
<td>[0.49, 1.41]</td>
<td>&lt;0.001</td>
<td>0.44</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>ADHD-Hyperactivity/impulsivity Parent Reported – Z scored</td>
<td>0.20</td>
<td>[-0.05, 0.44]</td>
<td>0.34</td>
<td>0.17</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>ADHD- Combined Parent Reported – Z scored</td>
<td>0.51</td>
<td>[0.19, 0.84]</td>
<td>0.01</td>
<td>0.33</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------------------------</td>
<td>--------</td>
<td>----------</td>
<td>--------</td>
<td>-------</td>
<td>----------</td>
<td>-------</td>
<td></td>
</tr>
<tr>
<td>Observer rating of attention span (TRAB-AS) – Z scored</td>
<td>-0.48</td>
<td>[-0.70, -0.25]</td>
<td>&lt;0.001</td>
<td>-0.42</td>
<td>-1.14</td>
<td>[-1.73,-0.55]</td>
<td>0.001</td>
</tr>
<tr>
<td>Inhibitory Control (ms)</td>
<td>27.53</td>
<td>[17.04, 38.01]</td>
<td>&lt;0.001</td>
<td>0.52</td>
<td>41.86</td>
<td>[22.4, 61.33]</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Working Memory</td>
<td>-8.98</td>
<td>[-12.72, -5.24]</td>
<td>&lt;0.001</td>
<td>-0.48</td>
<td>-10.37</td>
<td>[-14.77,-5.96]</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>IQ at 6 years</td>
<td>-16.49</td>
<td>[-19.81, -13.17]</td>
<td>&lt;0.001</td>
<td>-0.99</td>
<td>-26.24</td>
<td>[-31.69,-20.79]</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

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 R² 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

<table>
<thead>
<tr>
<th>Predictor</th>
<th>BLS ADHD-I PR</th>
<th>BLS TRAB-AS</th>
<th>EPICure TRAB-AS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Beta</td>
<td>P-Value</td>
<td>Beta</td>
</tr>
<tr>
<td>Birth Group (0 = Control, 1 = EP/VP/VLBW)</td>
<td>0.00</td>
<td>0.971</td>
<td>-0.02</td>
</tr>
<tr>
<td>Inhibitory Control</td>
<td>0.14</td>
<td>0.006</td>
<td>-0.07</td>
</tr>
<tr>
<td>Working Memory</td>
<td>-0.07</td>
<td>0.213</td>
<td>0.24</td>
</tr>
<tr>
<td>IQ at 6 years</td>
<td>-0.35</td>
<td>&lt;0.001</td>
<td>0.26</td>
</tr>
<tr>
<td>Sex (0 = Female, 1 = Male)</td>
<td>0.06</td>
<td>0.218</td>
<td>0.03</td>
</tr>
<tr>
<td>Total R²</td>
<td>0.22</td>
<td>0.23</td>
<td>0.26</td>
</tr>
</tbody>
</table>

*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
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 Madzwanuse 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.

Gestational age and birth weight

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.

Small for Gestational Age classification

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. Finally 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 SGAN or SGAF references were performed separately. Initial unadjusted mixed models (Model 1- SGAN and Model 1 - SGAF) 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- SGAN and Model 2 - SGAF). Thirdly, the effect of the parent-infant relationship measured with the PIRI was considered as another fixed effect (Model 3- SGAN and Model 3 - SGAF). 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 SGAN and SGAF are shown in Table 9. The SGAN reference classified 61 (28%) VP/VLBW and 14 (7%) term-born participants as SGA. In contrast, the SGAF 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 (SGAN) or Fetal Reference (SGAF)

<table>
<thead>
<tr>
<th></th>
<th>SGA using Neonatal Reference</th>
<th>SGA using Fetal Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AGAN + VP/VLBW (n=156)</td>
<td>AGAN + Term-Born (n=183)</td>
</tr>
<tr>
<td>Birthweight (g)</td>
<td>1390 (313)</td>
<td>3430 (398)</td>
</tr>
<tr>
<td></td>
<td>1110 (239)</td>
<td>2540 (235)</td>
</tr>
<tr>
<td>Gestational Age (weeks)</td>
<td>Mean (SD)</td>
<td>29.7 (1.49)</td>
</tr>
<tr>
<td></td>
<td>30 (2.40)</td>
<td>39.7 (1.13)</td>
</tr>
<tr>
<td></td>
<td>28 (47.0%)</td>
<td>38.8 (1.25)</td>
</tr>
<tr>
<td>Sex</td>
<td>Male</td>
<td>Female</td>
</tr>
<tr>
<td></td>
<td>86 (55.1%)</td>
<td>70 (44.9%)</td>
</tr>
<tr>
<td></td>
<td>28 (45.9%)</td>
<td>33 (54.1%)</td>
</tr>
<tr>
<td></td>
<td>86 (47.0%)</td>
<td>97 (53.0%)</td>
</tr>
<tr>
<td></td>
<td>8 (57.1%)</td>
<td>6 (42.9%)</td>
</tr>
<tr>
<td>Socioeconomic Status</td>
<td>34 (21.8%)</td>
<td>14 (23.0%)</td>
</tr>
<tr>
<td>----------------------</td>
<td>------------</td>
<td>------------</td>
</tr>
<tr>
<td>High</td>
<td>75 (48.1%)</td>
<td>27 (44.3%)</td>
</tr>
<tr>
<td>Middle</td>
<td>47 (30.1%)</td>
<td>20 (32.8%)</td>
</tr>
<tr>
<td>Low</td>
<td>73 (46.8%)</td>
<td>29 (47.5%)</td>
</tr>
<tr>
<td>Parent Infant Relationship</td>
<td>78 (50.0%)</td>
<td>28 (45.9%)</td>
</tr>
<tr>
<td>Poor</td>
<td>5 (3.2%)</td>
<td>4 (6.6%)</td>
</tr>
<tr>
<td>Good</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Missing</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
IQ trajectories by SGA 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 status 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). 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 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- SGAş) cognitive trajectories stratified by small for gestational age (SGA) status and very preterm/very low birthweight (VP/VLBW) status.

IQ trajectories by SGAş 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 SGAş, 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- SGAş). 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- SGAş 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_ref) 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

<table>
<thead>
<tr>
<th>Parameters</th>
<th>SGA&lt;sub&gt;N&lt;/sub&gt; Model 1</th>
<th>SGA&lt;sub&gt;N&lt;/sub&gt; Model 2</th>
<th>SGA&lt;sub&gt;N&lt;/sub&gt; Model 3</th>
<th>SGA&lt;sub&gt;F&lt;/sub&gt; Model 1</th>
<th>SGA&lt;sub&gt;F&lt;/sub&gt; Model 2</th>
<th>SGA&lt;sub&gt;F&lt;/sub&gt; Model 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixed effects</td>
<td>Est</td>
<td>CI</td>
<td>Est</td>
<td>CI</td>
<td>Est</td>
<td>CI</td>
</tr>
<tr>
<td>(Intercept)</td>
<td>78.87</td>
<td>[75.32, 82.43]</td>
<td>90.53</td>
<td>[83.40, 97.65]</td>
<td>93.15</td>
<td>[86.18, 100.13]</td>
</tr>
<tr>
<td>Age</td>
<td>0.10</td>
<td>[-0.04, 0.23]</td>
<td>0.10</td>
<td>[-0.04, 0.23]</td>
<td>0.10</td>
<td>[-0.04, 0.23]</td>
</tr>
<tr>
<td>SGA*Age</td>
<td>0.23</td>
<td>[-0.02, 0.48]</td>
<td>0.23</td>
<td>[-0.02, 0.49]</td>
<td>0.23</td>
<td>[-0.02, 0.49]</td>
</tr>
<tr>
<td>Control*Age</td>
<td>-0.10</td>
<td>[-0.28, 0.09]</td>
<td>-0.10</td>
<td>[-0.28, 0.09]</td>
<td>-0.10</td>
<td>[-0.28, 0.09]</td>
</tr>
<tr>
<td>Control<em>SGA</em>Age</td>
<td>-0.05</td>
<td>[-0.58, 0.48]</td>
<td>-0.05</td>
<td>[-0.59, 0.48]</td>
<td>-0.05</td>
<td>[-0.59, 0.48]</td>
</tr>
<tr>
<td>Sex(Male=1,Female=2)</td>
<td>-1.98 [-5.60,1.63]</td>
<td>-0.65 [-4.18,2.88]</td>
<td>-1.72 [-5.37,1.92]</td>
<td>-0.35 [-3.90,3.21]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------------------</td>
<td>---------------------</td>
<td>---------------------</td>
<td>---------------------</td>
<td>---------------------</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Random Effects**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>sd(Age)</td>
<td>0.29 [0.21,0.40]</td>
<td>0.31 [0.24,0.42]</td>
<td>0.32 [0.24,0.42]</td>
<td>0.29 [0.21,0.40]</td>
</tr>
<tr>
<td>cor(Intercept, Age)</td>
<td>-0.83 [-0.94,-0.55]</td>
<td>-0.90 [-0.97,-0.67]</td>
<td>-0.85 [-0.95,-0.67]</td>
<td>-0.90 [-0.97,-0.68]</td>
</tr>
<tr>
<td>cor(Age, Age)</td>
<td>0.67</td>
<td>0.67</td>
<td>0.67</td>
<td>0.67</td>
</tr>
<tr>
<td>BIC</td>
<td>22071</td>
<td>22059</td>
<td>22039</td>
<td>2216</td>
</tr>
</tbody>
</table>

**Notes:**
- The values in square brackets represent the 95% confidence intervals.
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

Main Findings

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.

Strengths and Limitations

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.

Interpretation

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 participant 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”.

151
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 Madzwmuse 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


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


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


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


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


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


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

<table>
<thead>
<tr>
<th></th>
<th>AYLS</th>
<th>BLS</th>
<th>EPICURE</th>
<th>HESVA</th>
<th>NTNU</th>
<th>NZVLBW</th>
<th>UCLH</th>
<th>VICS</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Evidence of Severe NSI</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>3 (10.7%)</td>
<td>22 (10.8%)</td>
<td>14 (11.3%)</td>
<td>5 (4.6%)</td>
<td>4 (7.8%)</td>
<td>9 (4.0%)</td>
<td>3 (2.9%)</td>
<td>27 (12.1%)</td>
<td>87 (8.1%)</td>
</tr>
<tr>
<td>No</td>
<td>25 (89.3%)</td>
<td>181 (89.2%)</td>
<td>110 (88.7%)</td>
<td>104 (95.4%)</td>
<td>47 (92.2%)</td>
<td>216 (96.0%)</td>
<td>101 (97.1%)</td>
<td>197 (87.9%)</td>
<td>981 (91.9%)</td>
</tr>
</tbody>
</table>

| **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%) | 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%) | 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%) |

<p>| <strong>Non-Ambulatory Cerebral Palsy</strong> |       |       |         |       |       |         |       |       |         |
| 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%) |</p>
<table>
<thead>
<tr>
<th></th>
<th>Yes</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0 (0%)</td>
<td>7 (3.4%)</td>
<td>2 (1.6%)</td>
<td>5 (4.6%)</td>
<td>2 (3.9%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>2 (3.9%)</td>
<td>18 (1.7%)</td>
</tr>
<tr>
<td>Missing</td>
<td>0 (0%)</td>
<td>1 (0.5%)</td>
<td>7 (5.6%)</td>
<td>3 (2.8%)</td>
<td>0 (0%)</td>
<td>6 (2.7%)</td>
<td>104 (100%)</td>
<td>0 (0%)</td>
<td>121 (11.3%)</td>
</tr>
<tr>
<td>Child IQ &lt;70</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>22 (78.6%)</td>
<td>168 (82.8%)</td>
<td>108 (87.1%)</td>
<td>0 (0%)</td>
<td>39 (76.5%)</td>
<td>212 (94.2%)</td>
<td>99 (95.2%)</td>
<td>194 (86.6%)</td>
<td>842 (78.8%)</td>
</tr>
<tr>
<td>Yes</td>
<td>3 (10.7%)</td>
<td>18 (8.9%)</td>
<td>13 (10.5%)</td>
<td>0 (0%)</td>
<td>3 (5.9%)</td>
<td>7 (3.1%)</td>
<td>3 (2.9%)</td>
<td>25 (11.2%)</td>
<td>72 (6.7%)</td>
</tr>
<tr>
<td>Missing</td>
<td>3 (10.7%)</td>
<td>17 (8.4%)</td>
<td>3 (2.4%)</td>
<td>109 (100%)</td>
<td>9 (17.6%)</td>
<td>6 (2.7%)</td>
<td>2 (1.9%)</td>
<td>5 (2.2%)</td>
<td>154 (14.4%)</td>
</tr>
</tbody>
</table>
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) Demonstration that outcome of interest was not present at start of study
(Was adult cognitive performance known when the participants were recruited?)
A) yes ~
B) no

Comparability
1) Comparability of cohorts on the basis of the design or analysis
A) study controls for maternal education (select the most important factor) ~
B) study controls for any additional factor (sex)

Outcome
1) Assessment of outcome (Did the study use a standardised full-scale IQ assessment?)
A) independent blind assessment
B) record linkage
C) self report
D) no description
2) Was follow-up long enough for outcomes to occur (Did the cohort assess adult IQ outcomes?)

A) yes (17 years or greater)
B) no

3) Adequacy of follow up of cohorts. (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

<table>
<thead>
<tr>
<th>Cohort</th>
<th>Representativeness of the exposed cohort</th>
<th>Selectivity of the non exposed cohort</th>
<th>Ascertainment of exposure</th>
<th>Demonstration that outcome of interest was not present at start of study</th>
<th>Comparability of cohorts on the basis of the design or analysis</th>
<th>Assessment of outcome</th>
<th>Was follow-up long enough for outcomes to occur</th>
<th>Adequacy of follow up of cohorts (above or below 50%)</th>
<th>Overall Cohort Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>AYLS (Matinolli et al., 2017)</td>
<td>A (regional)</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>C</td>
<td>8</td>
</tr>
<tr>
<td>BLS (Eryigit Madzwamuse et al., 2015; Wolke &amp; Meyer, 1999)</td>
<td>A (regional)</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>C</td>
<td>8</td>
</tr>
<tr>
<td>EPICure (Costello et al., 2000; Linsell et al., 2018)</td>
<td>A (national)</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>C</td>
<td>7</td>
</tr>
<tr>
<td>Study</td>
<td>Region</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>C</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>------------------------------</td>
<td>------------</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>HESVA (Riikka Pyhälä, 2012)</td>
<td>A (regional)</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>C</td>
<td>8</td>
</tr>
<tr>
<td>NTNU (Lærum et al., 2019)</td>
<td>A (regional)</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>7</td>
</tr>
<tr>
<td>NZ_VLBW darlow</td>
<td>A (national)</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>8</td>
</tr>
<tr>
<td>UCLH (Kroll et al., 2017; Stewart et al., 1987)</td>
<td>A (regional)</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>8</td>
</tr>
<tr>
<td>VICS (Victorian Infant Collaborative Study Group, 1997)</td>
<td>A (regional)</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>9</td>
</tr>
</tbody>
</table>
Appendix 3: Linear mixed model demonstrating reducing Gestational Age by Birth Year among VP/VLBW participants

**VP/VLBW only analysis**

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Estimates</th>
<th>CI</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept – Estimate for 1978)</td>
<td>32.07</td>
<td>29.66 – 34.49</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Birth year – per year post 1978</td>
<td>-0.32</td>
<td>-0.60 – -0.04</td>
<td>0.025</td>
</tr>
</tbody>
</table>

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**

<table>
<thead>
<tr>
<th>Birthweight (g)</th>
<th>Estimates</th>
<th>CI</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predictors</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Intercept – Estimate for 1978)</td>
<td>1464.87</td>
<td>1211.59 – 1718.14</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Birth year – per year post 1978</td>
<td>-29.85</td>
<td>-58.78 – -0.91</td>
<td>0.043</td>
</tr>
<tr>
<td>Observations</td>
<td>1068</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Marginal R² / Conditional R²</td>
<td>0.164 / 0.411</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
# Appendix 5: IQ and Demographic Information of all Participants from each IPD Cohort

<table>
<thead>
<tr>
<th>Group</th>
<th>AYLS</th>
<th>BLS</th>
<th>EPICURE</th>
<th>HESVA</th>
<th>NTNU</th>
<th>NZVLBW</th>
<th>UCLH</th>
<th>VICS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Con</td>
<td>VP/</td>
<td>Con</td>
<td>VP/</td>
<td>Con</td>
<td>VP/</td>
<td>VP/</td>
<td>Con</td>
<td>VP/</td>
</tr>
<tr>
<td>s</td>
<td>VLB</td>
<td>s</td>
<td>VLB</td>
<td>s</td>
<td>VLB</td>
<td>VLB</td>
<td>s</td>
<td>VLB</td>
</tr>
<tr>
<td>n=28</td>
<td>W</td>
<td>n=</td>
<td>n=124</td>
<td>n=9</td>
<td>n=7</td>
<td>n=5</td>
<td>n=5</td>
<td>n=146</td>
</tr>
<tr>
<td>303</td>
<td></td>
<td>192</td>
<td></td>
<td>8</td>
<td>5</td>
<td>100</td>
<td>89</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

## IQ Z Score

<table>
<thead>
<tr>
<th>Mean (SD)</th>
<th>Mean (SD)</th>
<th>Mean (SD)</th>
<th>Mean (SD)</th>
<th>Mean (SD)</th>
<th>Mean (SD)</th>
<th>Mean (SD)</th>
<th>Mean (SD)</th>
<th>Mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>(-0.95)</td>
<td>(-0.83)</td>
<td>(-1.06)</td>
<td>(-0.57)</td>
<td>(-0.86)</td>
<td>(-0.78)</td>
<td>(-0.64)</td>
<td>(-0.67)</td>
<td></td>
</tr>
<tr>
<td>(1.21)</td>
<td>(1.04)</td>
<td>(0.96)</td>
<td>(1.06)</td>
<td>(1.10)</td>
<td>(1.00)</td>
<td>(1.00)</td>
<td>(1.01)</td>
<td></td>
</tr>
<tr>
<td>(5)</td>
<td>(64)</td>
<td>(96)</td>
<td>(106)</td>
<td>(110)</td>
<td>(100)</td>
<td>(100)</td>
<td>(101)</td>
<td></td>
</tr>
<tr>
<td>Cohort</td>
<td>AYLS</td>
<td>BLS</td>
<td>EPICURE</td>
<td>HESVA</td>
<td>NTNU</td>
<td>NZVLBW</td>
<td>UCLH</td>
<td>VICS</td>
</tr>
<tr>
<td>--------</td>
<td>------</td>
<td>-----</td>
<td>---------</td>
<td>-------</td>
<td>------</td>
<td>--------</td>
<td>------</td>
<td>-----</td>
</tr>
<tr>
<td>Group</td>
<td>Con</td>
<td>VP/</td>
<td>Con</td>
<td>VP/</td>
<td>Con</td>
<td>VP/</td>
<td>Con</td>
<td>VP/</td>
</tr>
<tr>
<td></td>
<td>s</td>
<td>VLB</td>
<td>s</td>
<td>VLB</td>
<td>s</td>
<td>VLB</td>
<td>s</td>
<td>VLB</td>
</tr>
<tr>
<td>n=</td>
<td>n=28</td>
<td>n=</td>
<td>n=124</td>
<td>n=5</td>
<td>n=7</td>
<td>n=5</td>
<td>n=225</td>
<td>n=</td>
</tr>
<tr>
<td>303</td>
<td>192</td>
<td>203</td>
<td>8</td>
<td>5</td>
<td>1</td>
<td>100</td>
<td>89</td>
<td>146</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>134</td>
<td>16</td>
<td>92</td>
<td>108</td>
<td>25</td>
<td>56</td>
<td>42</td>
<td>47</td>
</tr>
<tr>
<td></td>
<td>(44.</td>
<td>(57.1</td>
<td>(47.</td>
<td>(53.2</td>
<td>(39.1</td>
<td>(45.2%</td>
<td>(42.</td>
<td>(43.1</td>
</tr>
<tr>
<td>2%)</td>
<td>(%)</td>
<td>(%)</td>
<td>(%)</td>
<td>(%)</td>
<td>(%)</td>
<td>(%)</td>
<td>(%)</td>
<td>(%)</td>
</tr>
<tr>
<td>Female</td>
<td>169</td>
<td>12</td>
<td>100</td>
<td>95</td>
<td>39</td>
<td>68</td>
<td>56</td>
<td>62</td>
</tr>
<tr>
<td></td>
<td>(55.</td>
<td>(42.9</td>
<td>(52.</td>
<td>(46.8</td>
<td>(60.9</td>
<td>(54.8%</td>
<td>(57.</td>
<td>(56.9</td>
</tr>
<tr>
<td>8%)</td>
<td>(%)</td>
<td>1%</td>
<td>(%)</td>
<td>(%)</td>
<td>(%)</td>
<td>(%)</td>
<td>(%)</td>
<td>(%)</td>
</tr>
</tbody>
</table>

**Maternal Education Level**
<table>
<thead>
<tr>
<th>Group</th>
<th>Cohort</th>
<th>AYLS</th>
<th>BLS</th>
<th>EPICURE</th>
<th>HESVA</th>
<th>NTNU</th>
<th>NZVLBW</th>
<th>UCLH</th>
<th>VICS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>s</td>
<td>VP/</td>
<td>s</td>
<td>Con</td>
<td>VP/</td>
<td>Con</td>
<td>VP/</td>
<td>Con</td>
<td>VP/</td>
</tr>
<tr>
<td>n=303</td>
<td>VLB</td>
<td>W</td>
<td>n=28</td>
<td>n=124</td>
<td>VLB</td>
<td>W</td>
<td>n=64</td>
<td>W</td>
<td>VLB</td>
</tr>
<tr>
<td></td>
<td>n=203</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mediu n</td>
<td>101</td>
<td>9</td>
<td>72</td>
<td>112</td>
<td>48</td>
<td>90</td>
<td>56</td>
<td>64</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td>(33.3)</td>
<td>(32.1)</td>
<td>(37.0)</td>
<td>(55.2)</td>
<td>(75.0)</td>
<td>(72.6%)</td>
<td>(57.3)</td>
<td>(58.7)</td>
<td>(41.4)</td>
</tr>
<tr>
<td></td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>148</td>
<td>12</td>
<td>32</td>
<td>27</td>
<td>12</td>
<td>4</td>
<td>29</td>
<td>26</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td>(48.8)</td>
<td>(42.9)</td>
<td>(16.4)</td>
<td>(13.3)</td>
<td>(18.8)</td>
<td>(3.2%)</td>
<td>(29.0)</td>
<td>(23.9)</td>
<td>(37.1)</td>
</tr>
<tr>
<td></td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
</tr>
</tbody>
</table>

<p>| Low       | 52     | 7    | 87  | 61     | 4     | 23   | 13     | 17   | 2    |
|           | (17.2)| (25.0)| (45.0)| (30.0)| (6.2%)| (18.5%)| (13.1)| (15.6)| (2.7) |
|           | %      | %    | %   | %       | %     | %    | %      | %    | %    |
| Mediu n   | 101    | 9    | 72  | 112     | 48    | 90   | 56     | 64   | 31   |
|           | (33.3)| (32.1)| (37.0)| (55.2)| (75.0)| (72.6%)| (57.3)| (58.7)| (41.4)|
|           | %      | %    | %   | %       | %     | %    | %      | %    | %    |
| High      | 148    | 12   | 32  | 27      | 12    | 4    | 29     | 26   | 28   |
|           | (48.8)| (42.9)| (16.4)| (13.3)| (18.8)| (3.2%)| (29.0)| (23.9)| (37.1)|
|           | %      | %    | %   | %       | %     | %    | %      | %    | %    |</p>
<table>
<thead>
<tr>
<th>Cohort</th>
<th>Group</th>
<th>Con</th>
<th>VP/ VL</th>
<th>Con</th>
<th>VP/ VLB</th>
<th>Con</th>
<th>VP/ VLB</th>
<th>Con</th>
<th>VP/ VLB</th>
<th>Con</th>
<th>VP/ VLB</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AYLS</td>
<td>s</td>
<td>n=303</td>
<td>28</td>
<td>n=192</td>
<td>64</td>
<td>124</td>
<td>8</td>
<td>5</td>
<td>100</td>
<td>89</td>
<td>104</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BLS</td>
<td>s</td>
<td>n=303</td>
<td>28</td>
<td>n=192</td>
<td>64</td>
<td>124</td>
<td>8</td>
<td>5</td>
<td>100</td>
<td>89</td>
<td>104</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EPICURE</td>
<td>Con</td>
<td>n=64</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HESVA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NTNU</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NZVLBW</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UCLH</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VICS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Missing</td>
<td>2</td>
<td>(0.7)</td>
<td>0</td>
<td>(0.5)</td>
<td>1</td>
<td>(0.5)</td>
<td>3</td>
<td>(1.5)</td>
<td>0</td>
<td>(0.5)</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(%)</td>
<td></td>
<td>(%)</td>
<td>(%)</td>
<td></td>
<td>(%)</td>
<td>(%)</td>
<td>(%)</td>
<td>(%)</td>
<td>(%)</td>
</tr>
</tbody>
</table>
Appendix 6: Neonatal and demographic data for VP/VLBW participants from each IPD cohort

<table>
<thead>
<tr>
<th></th>
<th>AYLS</th>
<th>BLS</th>
<th>EPICURE</th>
<th>HESVA</th>
<th>NTNU</th>
<th>NZVLBW</th>
<th>UCLH</th>
<th>VICS</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>VP/VLBW (n=28)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VP/VLBW (n=203)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VP/VLBW (n=124)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VP/VLBW (n=109)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VP/VLBW (n=51)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VP/VLBW (n=225)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VP/VLBW (n=104)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VP/VLBW (n=224)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VP/VLBW (n=1068)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Gestational Age (weeks)**
- Mean (SD)
  - AYLS: 29.6 (2.09)
  - BLS: 30.4 (2.05)
  - EPICURE: 24.5 (0.748)
  - HESVA: 29.3 (2.33)
  - NTNU: 29.0 (2.49)
  - NZVLBW: 29.3 (2.50)
  - UCLH: 28.8 (2.00)
  - VICS: 26.6 (1.99)
  - Overall: 28.3 (2.81)

**Birthweight Z Score**
- Mean (SD)
  - AYLS: -0.00903 (1.08)
  - BLS: -0.603 (1.20)
  - EPICURE: 0.230 (0.822)
  - HESVA: -0.421 (1.00)
  - NTNU: -0.182 (1.08)
  - NZVLBW: -0.607 (1.07)
  - UCLH: -0.0826 (0.930)
  - VICS: -0.167 (1.07)
  - Overall: -0.311 (1.09)

**Multiple Birth**
<table>
<thead>
<tr>
<th></th>
<th>AYLS VP/ VLBW (n=28)</th>
<th>BLS VP/ VLBW (n=203)</th>
<th>EPICURE VP/ VLBW (n=124)</th>
<th>HESVA VP/ VLBW (n=109)</th>
<th>NTNU VP/ VLBW (n=51)</th>
<th>NZVLBW VP/ VLBW (n=225)</th>
<th>UCLH VP/ VLBW (n=104)</th>
<th>VICS VP/ VLBW (n=224)</th>
<th>Overall VP/ VLBW (n=1068)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Singleton</td>
<td>25 (89.3%)</td>
<td>149 (73.4%)</td>
<td>83 (66.9%)</td>
<td>92 (84.4%)</td>
<td>41 (80.4%)</td>
<td>169 (75.1%)</td>
<td>81 (77.9%)</td>
<td>150 (67.0%)</td>
<td>790 (74.0%)</td>
</tr>
<tr>
<td>Multiple</td>
<td>3 (10.7%)</td>
<td>54 (26.6%)</td>
<td>40 (32.3%)</td>
<td>17 (15.6%)</td>
<td>10 (19.6%)</td>
<td>56 (24.9%)</td>
<td>19 (18.3%)</td>
<td>74 (33.0%)</td>
<td>273 (25.6%)</td>
</tr>
<tr>
<td>Missing</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>1 (0.8%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>4 (3.8%)</td>
<td>0 (0%)</td>
<td>5 (0.5%)</td>
<td></td>
</tr>
</tbody>
</table>

**Intraventricular Haemorrhage**

<table>
<thead>
<tr>
<th></th>
<th>AYLS No Grade (n=28)</th>
<th>BLS No Grade (n=203)</th>
<th>EPICURE No Grade (n=124)</th>
<th>HESVA No Grade (n=109)</th>
<th>NTNU No Grade (n=51)</th>
<th>NZVLBW No Grade (n=225)</th>
<th>UCLH No Grade (n=104)</th>
<th>VICS No Grade (n=224)</th>
<th>Overall No Grade (n=1068)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Grade</td>
<td>23 (82.1%)</td>
<td>161 (79.3%)</td>
<td>44 (35.5%)</td>
<td>69 (63.3%)</td>
<td>38 (74.5%)</td>
<td>158 (70.2%)</td>
<td>48 (46.2%)</td>
<td>152 (67.9%)</td>
<td>693 (64.9%)</td>
</tr>
<tr>
<td>Any Grade</td>
<td>5 (17.9%)</td>
<td>41 (20.2%)</td>
<td>79 (63.7%)</td>
<td>14 (12.8%)</td>
<td>5 (9.8%)</td>
<td>52 (23.1%)</td>
<td>55 (52.9%)</td>
<td>72 (32.1%)</td>
<td>323 (30.2%)</td>
</tr>
<tr>
<td>Missing</td>
<td>0 (0%)</td>
<td>1 (0.5%)</td>
<td>1 (0.8%)</td>
<td>26 (23.9%)</td>
<td>8 (15.7%)</td>
<td>15 (6.7%)</td>
<td>1 (1.0%)</td>
<td>0 (0%)</td>
<td>52 (4.9%)</td>
</tr>
<tr>
<td></td>
<td>AYLS (n=28)</td>
<td>BLS (n=203)</td>
<td>EPICURE (n=124)</td>
<td>HESVA (n=109)</td>
<td>NTNU (n=51)</td>
<td>NZVLBW (n=225)</td>
<td>UCLH (n=104)</td>
<td>VICS (n=224)</td>
<td>Overall (n=1068)</td>
</tr>
<tr>
<td>----------------</td>
<td>-------------</td>
<td>-------------</td>
<td>-----------------</td>
<td>--------------</td>
<td>-------------</td>
<td>----------------</td>
<td>--------------</td>
<td>-------------</td>
<td>-----------------</td>
</tr>
<tr>
<td>VP/ VLBW</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bronchopulmonary Dysplasia Diagnosed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>27 (96.4%)</td>
<td>101 (49.8%)</td>
<td>35 (28.2%)</td>
<td>80 (73.4%)</td>
<td>39 (76.5%)</td>
<td>181 (80.4%)</td>
<td>0 (0%)</td>
<td>138 (61.6%)</td>
<td>601 (56.3%)</td>
</tr>
<tr>
<td>Yes</td>
<td>1 (3.6%)</td>
<td>102 (50.2%)</td>
<td>89 (71.8%)</td>
<td>25 (22.9%)</td>
<td>10 (19.6%)</td>
<td>44 (19.6%)</td>
<td>0 (0%)</td>
<td>86 (38.4%)</td>
<td>357 (33.4%)</td>
</tr>
<tr>
<td>Missing</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>4 (3.7%)</td>
<td>2 (3.9%)</td>
<td>0 (0%)</td>
<td>104 (100%)</td>
<td>0 (0%)</td>
<td>110 (10.3%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ISCED Maternal Education</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>7 (25.0%)</td>
<td>61 (30.0%)</td>
<td>23 (18.5%)</td>
<td>17 (15.6%)</td>
<td>2 (3.9%)</td>
<td>85 (37.8%)</td>
<td>3 (2.9%)</td>
<td>47 (21.0%)</td>
<td>245 (22.9%)</td>
</tr>
<tr>
<td>Cohort</td>
<td>AYLS VP/VLBW (n=28)</td>
<td>BLS VP/VLBW (n=203)</td>
<td>EPICURE VP/VLBW (n=124)</td>
<td>HESVA VP/VLBW (n=109)</td>
<td>NTNU VP/VLBW (n=51)</td>
<td>NZVLBW VP/VLBW (n=225)</td>
<td>UCLH VP/VLBW (n=104)</td>
<td>VICS VP/VLBW (n=224)</td>
<td>Overall VP/VLBW (n=1068)</td>
</tr>
<tr>
<td>--------</td>
<td>---------------------</td>
<td>----------------------</td>
<td>-------------------------</td>
<td>------------------------</td>
<td>---------------------</td>
<td>------------------------</td>
<td>-----------------------</td>
<td>----------------------</td>
<td>------------------------</td>
</tr>
<tr>
<td>Medium</td>
<td>9 (32.1%)</td>
<td>112 (55.2%)</td>
<td>90 (72.6%)</td>
<td>64 (58.7%)</td>
<td>22 (43.1%)</td>
<td>68 (30.2%)</td>
<td>40 (38.5%)</td>
<td>57 (25.4%)</td>
<td>462 (43.3%)</td>
</tr>
<tr>
<td>High</td>
<td>12 (42.9%)</td>
<td>27 (13.3%)</td>
<td>4 (3.2%)</td>
<td>26 (23.9%)</td>
<td>16 (31.4%)</td>
<td>64 (28.4%)</td>
<td>20 (19.2%)</td>
<td>24 (10.7%)</td>
<td>193 (18.1%)</td>
</tr>
<tr>
<td>Missing</td>
<td>0 (0%)</td>
<td>3 (1.5%)</td>
<td>7 (5.6%)</td>
<td>2 (1.8%)</td>
<td>11 (21.6%)</td>
<td>8 (3.6%)</td>
<td>41 (39.4%)</td>
<td>96 (42.9%)</td>
<td>168 (15.7%)</td>
</tr>
</tbody>
</table>

**Cohort Mean Birth Year**

**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

<table>
<thead>
<tr>
<th>Cohort</th>
<th>Birth year</th>
<th>IQ Test</th>
<th>VP/VLBW IQ, $M\ (SD)$</th>
<th>$n$</th>
<th>Controls IQ, $M\ (SD)$</th>
<th>$n$</th>
<th>Age at assessment, $M\ (SD)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constable (2013)(Constable et al., 2013)</td>
<td>1990</td>
<td>WISC-TIQ</td>
<td>91.7 (12.4)</td>
<td>19</td>
<td>100.4 (18.7)</td>
<td>19</td>
<td>20.1 (0.9)</td>
</tr>
<tr>
<td>Hallin 2010)(Hallin et al., 2010)</td>
<td>1985</td>
<td>WAIS-III</td>
<td>93 (15.4)</td>
<td>52</td>
<td>106 (12.5)</td>
<td>54</td>
<td>18.3</td>
</tr>
<tr>
<td>Lefebvre (2005)(Lefebvre et al., 2005)</td>
<td>1976</td>
<td>WAIS-R</td>
<td>94(12)</td>
<td>59</td>
<td>108(14)</td>
<td>44</td>
<td>18.4</td>
</tr>
<tr>
<td>Stålnacke (2015)(Stålnacke et al., 2015)</td>
<td>1988</td>
<td>WISC-III</td>
<td>0.315(1.165)$^B$</td>
<td>118</td>
<td>0(1)</td>
<td>91</td>
<td>18</td>
</tr>
</tbody>
</table>

$^A$ = 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 Bavarian Longitudinal Study

- Initial VP/VLBW infant Sample, N= 682
  - Potential Adult Sample, N= 411
    - Assessed at 26 years, N= 260
      - Self-reported ADHD and Executive functioning data available, N=194
        - With Parent Reported ADHD, N=172
  - Initial Matched Control Infant Sample, N= 350
    - Potential Adult Sample, N= 308
      - Assessed at 26 years, N = 229
        - Self-reported ADHD and Executive functioning data available, N=197
        - With Parent Reported ADHD, N=181
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

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).

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

Inhibitory Control = MRT incongruent conditions – MRT congruent conditions
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

<table>
<thead>
<tr>
<th></th>
<th>Self-Reported Inattention</th>
<th>Self-Reported Hyperactivity</th>
<th>Parent-Reported Inattention</th>
<th>Parent-Reported Hyperactivity</th>
<th>Observer Rating of Attention</th>
<th>Inhibitory Control</th>
<th>Working Memory</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self-Reported Inattention</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-Reported Hyperactivity</td>
<td>0.43****</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parent-Reported Inattention</td>
<td>0.26****</td>
<td>0.08</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parent-Reported Hyperactivity</td>
<td>0.20***</td>
<td>0.19***</td>
<td>0.61****</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observer Rating of Attention</td>
<td>-0.10</td>
<td>-0.03</td>
<td>-0.32****</td>
<td>-0.17**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.07</td>
<td>-0.03</td>
<td>0.29****</td>
<td>0.11*</td>
<td>-0.23***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------------------------</td>
<td>------</td>
<td>-------</td>
<td>----------</td>
<td>-------</td>
<td>----------</td>
<td>---------------------</td>
<td></td>
</tr>
<tr>
<td><strong>Inhibitory Control</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Working Memory</strong></td>
<td>-0.13*</td>
<td>0.05</td>
<td>-0.29****</td>
<td>-0.15**</td>
<td>0.40***</td>
<td>-0.24****</td>
<td></td>
</tr>
<tr>
<td><strong>IQ at 6 Years</strong></td>
<td>-0.09</td>
<td>0.03</td>
<td>-0.44****</td>
<td>-0.22****</td>
<td>0.43***</td>
<td>-0.37****</td>
<td>0.54***</td>
</tr>
</tbody>
</table>

p < .0001****, p < .001***, p < .01**, p < .05*
<table>
<thead>
<tr>
<th></th>
<th>Self-Reported Inattention</th>
<th>Self-Reported Hyperactivity</th>
<th>Observer Rating of Attention</th>
<th>Inhibitory Control</th>
<th>Working Memory</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self-Reported Inattention</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-Reported Hyperactivity</td>
<td>0.58****</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observer Rating of Attention</td>
<td>-0.33****</td>
<td>-0.31***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inhibitory Control</td>
<td>0.12</td>
<td>0.13</td>
<td>-0.21**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Working Memory</td>
<td>-0.11</td>
<td>-0.11</td>
<td>0.35****</td>
<td>-0.10</td>
<td></td>
</tr>
<tr>
<td>IQ at 6 Years</td>
<td>-0.28****</td>
<td>-0.11</td>
<td>0.48****</td>
<td>-0.23**</td>
<td>0.57****</td>
</tr>
</tbody>
</table>

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.

Number of Registered Births in the South Bavaria between Jan 1985 & March 1986
N= 70,600

Initial VP/VLBW (<1500g/32week)
N= 682

Potential Adult Sample
411

Adult Dropout: 151
260 Adult Participants

203 Completed IQ tests
14 proxy IQ scores due to impairment

217 VP/VLBW participants

Initial Term born
N= 916

Matched Sample
N= 350

Potential Adult Sample
308

Adult Dropout: 79
229 Adult Participants

197 Completed IQ tests

197 Term born participants
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 :

<table>
<thead>
<tr>
<th>Factor</th>
<th>B</th>
<th>SE</th>
<th>Z</th>
<th>p-value</th>
<th>Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>CMM</td>
<td>11.363</td>
<td>0.191</td>
<td>59.465</td>
<td>&lt;0.001</td>
<td>0.828</td>
</tr>
<tr>
<td>AWST</td>
<td>11.137</td>
<td>0.212</td>
<td>52.481</td>
<td>&lt;0.001</td>
<td>0.728</td>
</tr>
<tr>
<td>BEERY</td>
<td>1.707</td>
<td>0.036</td>
<td>47.266</td>
<td>&lt;0.001</td>
<td>0.654</td>
</tr>
</tbody>
</table>
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\textsubscript{N}, right SGA\textsubscript{F})
Appendix 14: Perinatal characteristic of the BLS participants according to SGA status and VP/VLBW (left SGAN, right SGAF)

<table>
<thead>
<tr>
<th>SGA using Neonatal Reference</th>
<th>SGA using Foetal Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGA+ VP/VLBW (n=156)</td>
<td>AGA+ VP/VLBW (n=101)</td>
</tr>
<tr>
<td>AGA+ Term Born (n=183)</td>
<td>AGA+ Term-Born (n=175)</td>
</tr>
<tr>
<td>SGA+ VP/VLBW (n=61)</td>
<td>SGA+ Term-Born (n=22)</td>
</tr>
<tr>
<td>SGA+ Term Born (n=14)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Maternal Age</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean (SD)</td>
<td>29.1 (4.85)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Mother smoked during pregnancy</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Did not smoke</td>
<td>128 (82.1%)</td>
</tr>
<tr>
<td>Did smoke</td>
<td>23 (14.7%)</td>
</tr>
<tr>
<td>Missing</td>
<td>5 (3.2%)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parity</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiparous</td>
<td>70 (44.9%)</td>
</tr>
</tbody>
</table>
## Appendices

### Primiparous

<table>
<thead>
<tr>
<th></th>
<th>86 (55.1%)</th>
<th>44 (72.1%)</th>
<th>105 (57.4%)</th>
<th>7 (50.0%)</th>
<th>51 (50.5%)</th>
<th>79 (68.1%)</th>
<th>96 (54.9%)</th>
<th>16 (72.7%)</th>
</tr>
</thead>
</table>

### Supplemental oxygen duration (days)

<table>
<thead>
<tr>
<th></th>
<th>Mean (SD)</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>6.98 (7.47)</td>
<td>7.41 (8.90)</td>
<td>NA (NA)</td>
<td>NA (NA)</td>
<td>6.34 (7.40)</td>
<td>7.77 (8.25)</td>
<td>NA (NA)</td>
<td>NA (NA)</td>
</tr>
</tbody>
</table>

### Cerebral Palsy Grade

#### 3 or 4

<table>
<thead>
<tr>
<th></th>
<th>133 (85.3%)</th>
<th>49 (80.3%)</th>
<th>183 (100%)</th>
<th>14 (100%)</th>
<th>87 (86.1%)</th>
<th>95 (81.9%)</th>
<th>175 (100%)</th>
<th>22 (100%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>10 (6.4%)</td>
<td>1 (1.6%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>5 (5.0%)</td>
<td>6 (5.2%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>Missing</td>
<td>13 (8.3%)</td>
<td>11 (18.0%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>9 (8.9%)</td>
<td>15 (12.9%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
</tr>
</tbody>
</table>

### Visual Impairment

<table>
<thead>
<tr>
<th></th>
<th>140 (89.7%)</th>
<th>50 (82.0%)</th>
<th>183 (100%)</th>
<th>14 (100%)</th>
<th>90 (89.1%)</th>
<th>100 (86.2%)</th>
<th>175 (100%)</th>
<th>22 (100%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not Blind</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Blind

<table>
<thead>
<tr>
<th></th>
<th>3 (1.9%)</th>
<th>0 (0%)</th>
<th>0 (0%)</th>
<th>0 (0%)</th>
<th>2 (2.0%)</th>
<th>1 (0.9%)</th>
<th>0 (0%)</th>
<th>0 (0%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Missing</td>
<td>13 (8.3%)</td>
<td>11 (18.0%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>9 (8.9%)</td>
<td>15 (12.9%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
</tr>
</tbody>
</table>

### Hearing Impairment

#### Not deaf

<table>
<thead>
<tr>
<th></th>
<th>143 (91.7%)</th>
<th>50 (82.0%)</th>
<th>182 (99.5%)</th>
<th>14 (100%)</th>
<th>92 (91.1%)</th>
<th>101 (87.1%)</th>
<th>174 (99.4%)</th>
<th>22 (100%)</th>
</tr>
</thead>
</table>

#### 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

<table>
<thead>
<tr>
<th></th>
<th>140 (89.7%)</th>
<th>59 (96.7%)</th>
<th>183 (100%)</th>
<th>14 (100%)</th>
<th>94 (93.1%)</th>
<th>105 (90.5%)</th>
<th>175 (100%)</th>
<th>22 (100%)</th>
</tr>
</thead>
</table>

#### stage 3 or stage 4

<p>|            | 16 (10.3%) | 2 (3.3%) | 0 (0%) | 0 (0%) | 7 (6.9%) | 11 (9.5%) | 0 (0%) | 0 (0%) |</p>
<table>
<thead>
<tr>
<th></th>
<th>No</th>
<th>Yes</th>
<th>Missing</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Prenatal infectious disease</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>135 (86.5%)</td>
<td>20 (12.8%)</td>
<td>14 (9.0%)</td>
</tr>
<tr>
<td></td>
<td>52 (85.2%)</td>
<td>9 (14.8%)</td>
<td>8 (13.1%)</td>
</tr>
<tr>
<td></td>
<td>151 (82.5%)</td>
<td>32 (17.5%)</td>
<td>4 (2.2%)</td>
</tr>
<tr>
<td></td>
<td>13 (92.9%)</td>
<td>1 (7.1%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>Yes</td>
<td>86 (85.1%)</td>
<td>14 (13.9%)</td>
<td>11 (10.9%)</td>
</tr>
<tr>
<td></td>
<td>101 (87.1%)</td>
<td>15 (12.9%)</td>
<td>11 (9.5%)</td>
</tr>
<tr>
<td></td>
<td>145 (82.9%)</td>
<td>143 (81.7%)</td>
<td>4 (2.3%)</td>
</tr>
<tr>
<td></td>
<td>19 (86.4%)</td>
<td>18 (81.8%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td><strong>Premature rupture of membrane</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>78 (50.0%)</td>
<td>64 (41.0%)</td>
<td>14 (9.0%)</td>
</tr>
<tr>
<td></td>
<td>38 (62.3%)</td>
<td>15 (24.6%)</td>
<td>8 (13.1%)</td>
</tr>
<tr>
<td></td>
<td>149 (81.4%)</td>
<td>30 (16.4%)</td>
<td>4 (2.2%)</td>
</tr>
<tr>
<td></td>
<td>12 (85.7%)</td>
<td>2 (14.3%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>Yes</td>
<td>47 (46.5%)</td>
<td>43 (42.6%)</td>
<td>11 (10.9%)</td>
</tr>
<tr>
<td></td>
<td>69 (59.5%)</td>
<td>36 (31.0%)</td>
<td>11 (9.5%)</td>
</tr>
<tr>
<td></td>
<td>143 (81.7%)</td>
<td>28 (16.0%)</td>
<td>4 (2.3%)</td>
</tr>
<tr>
<td></td>
<td>18 (81.8%)</td>
<td>4 (18.2%)</td>
<td>0 (0%)</td>
</tr>
</tbody>
</table>
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).

<table>
<thead>
<tr>
<th>SGA/ AGA</th>
<th>Birth Group</th>
<th>n</th>
<th>5M DQ &lt;70</th>
<th>5M DQ &lt;85</th>
<th>20M DQ &lt;70</th>
<th>20M DQ &lt;85</th>
<th>4Y IQ &lt;70</th>
<th>4Y IQ &lt;85</th>
<th>6Y IQ &lt;70</th>
<th>6Y IQ &lt;85</th>
<th>8Y IQ &lt;70</th>
<th>8Y IQ &lt;85</th>
<th>26Y IQ &lt;70</th>
<th>26Y IQ &lt;85</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGA_F</td>
<td>VP/ VLBW</td>
<td>101</td>
<td>17 32 29</td>
<td>56 27 43</td>
<td>24 58 20</td>
<td>44 22 48</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SGA_F</td>
<td>VP/ VLBW</td>
<td>116</td>
<td>36 57 53</td>
<td>75 34 54</td>
<td>33 55 32</td>
<td>60 32 51</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AGA_N</td>
<td>Term</td>
<td>175</td>
<td>3 17 1 21</td>
<td>5 14 3 13</td>
<td>3 15 3</td>
<td>15 3 15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SGA_N</td>
<td>Term</td>
<td>22</td>
<td>0 32 9 23</td>
<td>0 9 9 23</td>
<td>0 14 5</td>
<td>23 5 23</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AGA_N</td>
<td>VP/ VLBW</td>
<td>156</td>
<td>22 40 38</td>
<td>57 30 46</td>
<td>26 54 24</td>
<td>42 27 49</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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

<table>
<thead>
<tr>
<th>Section and topic</th>
<th>Item</th>
<th>Reported on page No</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1: Aim</strong></td>
<td>Report the aim of PPI in the study</td>
<td>The BLS underwent peer review but did not involve public involvement. Page 6.</td>
</tr>
<tr>
<td><strong>2: Methods</strong></td>
<td>Provide a clear description of the methods used for PPI in the study</td>
<td>None</td>
</tr>
<tr>
<td><strong>3: Study results</strong></td>
<td>Outcomes—Report the results of PPI in the study, including both positive and negative outcomes</td>
<td>None</td>
</tr>
<tr>
<td><strong>4: Discussion and conclusions</strong></td>
<td>Outcomes—Comment on the extent to which PPI influenced</td>
<td>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</td>
</tr>
</tbody>
</table>
Appendices

<table>
<thead>
<tr>
<th>5: Reflections/critical perspective</th>
<th>Comment critically on the study, reflecting on the things that went well and those that did not, so others can learn from this experience</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>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</td>
</tr>
</tbody>
</table>

PPI=patient and public involvement

END