RECAP
"Research on European Children and Adults born Preterm"

Grant Agreement number: 733280

Deliverable 9.5

Predicting IQ performance of very low birthweight infants across neonatal eras and countries of birth

Workpackage:	WP 9.5
Task:	T 9.6
Due Date:	30th September 2021 (M57)
Actual Submission Date:	4th October 2021 (M57)
Last Amendment deliverable:	17th September 2021 (M57)
Project Dates:	Project Start Date: January 01, 2017 Project Duration: 51 months
Responsible partner:	UoW
Responsible author:	Dieter Wolke
Email:	D.Wolke@warwick.ac.uk
Contributors:	Robert Eves (UoW), Marina Mendonca (UoW), Yanyan Ni (UoW),

Project funded by the European Commission within H2020-SC1-2016-2017/H2020-SC1-2016-		
Dissemination Level		
PU	Public	X
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission	
CO	Confidential, only for members of the consortium (including the Commission Services)	

Document History:

Version	Date	Changes	From	Review
V0.1	17.09.21	First Draft	UoW	Inserm
V1.0	29.09.21	Version ready for submission	UoW	TNO

Open Issues

No:	Date	Issue	Resolved
1			

SUMMARY

In this deliverable we evaluate how well adult IQ for very low birth weight infants can be predicted by early neonatal factors, early socioenvironmental factors and child IQ assessment. Data from two adult cohorts born in the 1980s and 1990s (BLS and VICS) and one child cohort born in the 2010s (GNN) are used. The analysis splits the adult data into a training dataset, where an initial model predicting adult IQ was developed, and a test dataset, where the accuracy of the model's predictions was validated on "unseen" data. The model was then applied to the GNN data, allowing for the individualised prediction of adult IQ for these VLBW children.

It was initially found that multiple neonatal factors and socioeconomic status significantly predicted adult IQ. However, in the final model, neonatal factors were largely mediated by the most important predictor of adult IQ: child IQ. This model was moderately accurate in predicting continuous IQ scores in the test dataset. In contrast the prediction was highly accurate in classifying those with cognitive disabilities in adulthood. For future outcomes, the model predicts that the GNN cohort will show a small decrease in IQ performance into adulthood.

To conclude, the universally high association between child IQ and adult IQ for VLBW infants underlines the importance of early cognitive screening, with the ability in particular to identify those who will display long term cognitive impairment. Furthermore, the only moderate ability to predict long-term cognitive outcome within the normal IQ range (continuous scores) indicates that future VLBW research should have a stronger focus on identifying protective or further risk factors along childhood in order to further improve the prediction and intervention of cognitive adult outcomes.

TABLE OF CONTENTS

Summary 2
Table of contents 3
1 Introduction 4
1.1 Purpose and Scope 4
1.2 References to other RECAP Documents 5
1.3 Definitions, Abbreviations and Acronyms 6
2 Methods 7
2.1 Included Cohorts 7
2.2 Data Harmonisation 8
2.3 Statistical Analyses 8
3 Main findings 11
3.1 Demographics 11
3.2 Association of Neonatal Factors and Socioeconomic Status with Adult IQ 13
3.3 Association of Child IQ with Adult IQ 14
3.4 Final model of Neonatal Factors, Socioeconomic status and Child IQ predicting Adult IQ 15
3.5 Validation of the Final Model predicting continuous IQ score and cognitive impairment ... 16
3.6 Prediction of Adult IQ for Current VLBW Children 20
4 Meaning of findings 21
5 Literature References 24
6 Appendices 29
6.1 Appendix 1: mediating effect of child IQ on the relation between neonatal factors and adult IQ 29
6.2 Appendix 2: Investigation of factors associated to accuracy of IQ prediction; i.e. the absolute difference between actual and predicted score 32

1 INTRODUCTION

1.1 Purpose and Scope

The aim of deliverable 9.5 was to investigate how accurately adult IQ can be predicted based on neonatal, early childhood factors, and childhood IQ assessments in very low birth weight (VLBW, <1500g) populations. As well as predicting future IQ scores, it is of particular importance to investigate the prediction of those at risk of cognitive impairment, often defined as an IQ score two standard deviations below the mean (<-2 SD, IQ score less than 70). The accurate prediction of cognitive impairment for individuals born VLBW could be particularly beneficial, as then earlier interventions in school could be implemented.

As a result of changing neonatal care in the last 50 years, there has been a dramatic improvement in survival rate for very low birthweight infants (Stoll et al., 2015). With increasing survival rate, long term morbidity has become an increasing focus of outcome monitoring, with the ability to accurately predict morbidity of particular importance (Crilly et al., 2021; de Kleine et al., 2007; Stoll et al., 2015). It has been well established that on average VLBW children have higher rates of major disabilities in the form of blindness (Hirvonen et al., 2018), deafness (Bolisetty et al., 2014) and cerebral palsy (Oskoui et al., 2013) than those born at term. In addition, evidence suggests the IQ scores of VLBW children are on average 12 points lower than those born at term (Twilhaar et al., 2018), with this difference appearing to be stable into adulthood, as shown by both meta-analysis (Eves et al., 2021) and individual studies investigating cognitive trajectories longitudinally (Breeman et al., 2015; Doyle et al., 2015; Linsell et al., 2018).

While the average IQ score for a VLBW individual may be below average, there is a large degree of variability, with many VLBW individuals performing at or above the expected level for their age. Thus, there has been considerable research into the neonatal medical complications or socioenvironmental factors that may allow for more accurate prediction of future cognitive outcomes (Crilly et al., 2021). However, this research has largely been performed in single cohorts (Ambalavanan et al., 2006), in relatively narrow time frames (Farooqi et al., 2011), or has focused largely on predicting developmental quotient (DQ) in the first few years of life, rather than childhood or adulthood IQ (Kalstabakken et al., 2021). This last limitation is of particular importance as early DQ measures are less predictive of adult IQ than child IQ tests (Breeman et al., 2015). Additionally, DQ is thought to measure features more closely related to motor performance rather than core cognitive abilities

(Aylward, 2009; Breeman et al., 2015). Thus, there is a need to determine how accurately early factors universally predict the adult cognitive performance of VLBW infants. Using empirically determined regression functions may then allow for predicting adult outcomes in VLBW children born more recently. Thus, understanding what factors predict long term cognitive outcome into adulthood, and with what degree of accuracy, can provide parents and clinicians with crucial prognostic information.

In order to address the objectives of deliverable 9.5, this study had three aims.

1. What are the neonatal and childhood factors that consistently predict IQ performance of VLBW adult participants assessed in two adult cohorts, born in different countries with differing healthcare systems?
2. How accurately can adult IQ be predicted based on early childhood and neonatal factors?
3. Using the empirically determined regression function to predict adult IQ – what is the predicted adult IQ for a recently born VLBW cohort?

1.2 References to other RECAP Documents

- WP9: Deliverable 9.1: Report comparing outcomes and survival of children born 2000+ with 1980s/early 90s
- WP9: Deliverable 9.2: Report of statistical analyses identification of universal vs culture specific outcomes promotive and resiliency factors

1.3 Definitions, Abbreviations and Acronyms

Table 1 List of Abbreviations and Acronyms

Abbreviation/ Acronym	DEFINITION
AUC	Area Under the Curve
BPD	Bronchopulmonary dysplasia
GA	GA
IQ	Intelligence quotient
IVH	Intraventricular haemorrhage
KABC	Kaufman Assessment Battery for Children
SES	Socioeconomic Status
SD	Standard deviation
VLBW	Very low birthweight (< 1500g birthweight)
EP/ELBW	Extremely Preterm/ Extremely low birthweight
WPPSI-III	Wechsler Preschool and Primary Scale of Intelligence-III
WISC III	Wechsler Intelligence Scale for Children Third Edition
WAIS II	Wechsler Abbreviated Scale of Intelligence Second Edition
WAIS III	Wechsler Abbreviated Scale of Intelligence Third Edition
NICU	Neonatal Intensive Care Unit
Cohort/Consortium Acronyms	
BLS	Bavarian Longitudinal Study
GNN	German Neonatal Network
VICS	Victorian Infant Collaborative Study
RECAP	Research on European Children and Adults Born Preterm
APIC	Adults Born Preterm International Collaboration

2 METHODS

2.1 Included Cohorts

This study was conducted as part of the Research on European Children and Adults Born Preterm (RECAP) Consortium. Along with two RECAP cohorts (BLS and GNN), one non-European cohort (VICS) from the Adults Born Preterm International Collaboration (APIC) Consortium also took part. All studies had received country-specific ethical reviews, with participants providing written informed consent, and all adhered to the Declaration of Helsinki.

Deliverables 9.1 and 9.2 have both described all the included cohorts in detail. Briefly, the Bavarian Longitudinal Study (BLS) started off as a geographically whole population study of infants admitted to neonatal special care in southern Bavaria in 1985/86 (Riegel et al., 1995). In addition, healthy infants who were cared for at the normal postnatal wards in the obstetric hospitals were recruited as term born controls. Followed repeatedly throughout childhood, including the testing of child IQ at 6 years, the sample was restricted to very preterm or very low birthweight individuals and term born controls at 26 years of age (Eryigit Madzwamuse et al., 2015). For the current analysis, in order to be comparable to the GNN cohort, only those born VLBW were included. This resulted in 162 participants with adult IQ, of which 137 had full neonatal, socioeconomic and child IQ data available. The term born controls were solely used for IQ harmonization purposes, of which there were 197 with adult IQ data available.

The Victorian Infant Collaborative Study (VICS) cohort comprised of infants born below 28 weeks of gestation or less than 1000g (EP/ELBW) in the Australian state of Victoria in 1991 and 1992 (Anderson, 2003). 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. 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). Followed repeatedly throughout childhood, including the testing of child IQ at 8 years, the cohort were assessed again in adulthood at 18 years of age (Doyle et al., 2015). This resulted in 224 EP/ELBW participants completing the adult IQ test, of which 212 had full neonatal, socioeconomic and child IQ data available. The term born

controls were solely used for IQ harmonization purposes, of which there were 146 with adult IQ data available.

The German Neonatal Network (GNN) cohort is a multi-center cohort including infants from 62 level III NICUs throughout Germany (Geisler et al., 2021). Infants born VLBW in 2012 and 2013 were included with their IQ measured at 5 years of age. In contrast to the BLS and VICS, the term born control group for the GNN were recruited at 5 years and were originally participants of the Survey of Neonates in Pomerania (SNiP II), living in a defined geographical region of Pomerania (Geisler et al., 2021). In total, 707 VLBW participants had full neonatal, socioeconomic and child IQ data available. The term born controls were solely used for IQ harmonization purposes, of which there were 196 with child IQ data available.

2.2 Data Harmonisation

Table 1 provides both an overview of the three cohorts and their demographic data. Within each cohort, the full-scale IQ scores were converted to Z scores using the mean and SD of the respective control group. Neonatal data included gestational age at birth, sex, birthweight Z score, presence of bronchopulmonary dysplasia (BPD), and presence of intraventricular haemorrhage (IVH). The definitions of BPD varied, BLS and GNN defining it as oxygen dependency more than 28 days after birth, with VICS defining BPD as oxygen dependency at 36 weeks' postmenstrual age. In past research, the different definitions of BPD have been similarly associated with adult IQ (Eves et al., 2021). Thus, the final harmonised variable was a binary variable, where a score of 1 indicated presence of BPD, regardless of original definition. The definition of IVH was classified according to criteria provided by Papile et al and harmonised into a binary variable (0= no grade/grade 1/ grade 2, 1= grades 3 or 4) (Papile et al., 1978). Birth weight z scores were determined using the Fenton international growth chart for preterm infants (Fenton & Kim, 2013). Socioeconomic status was measured in childhood for all cohorts based on parental occupation. Based on harmonisability, a binary variable (0 = low/medium SES, 1= high SES) was created.

2.3 Statistical Analyses

Participants from all cohorts were only included if they had no missing neonatal, SES or IQ data. Term born control participants were used to harmonise IQ data in each cohort but were subsequently removed, with all further analyses only including VLBW participants. In order to address the first research aim, three linear mixed models were performed in a randomly selected 75% of the combined BLS and VICS adult data (training dataset). Three models

were ran: 1) a prediction of adult IQ with neonatal data and SES as fixed predictors. 2) a prediction of adult IQ with child IQ as the sole fixed predictor. 3) a prediction of adult IQ with neonatal data, SES and child IQ as fixed predictors. In all models, the intercept was modeled as a random effect varying by cohort. The total variance (R^2) accounted for by each model was examined to determine the overall predictive value of the model.

In order to address the second research aim regarding accuracy of the IQ prediction, the model was then validated using the test dataset (the other 25% of the adult data). Using the model parameters based on the training data, predicted adult IQ scores were then calculated for the test dataset. The accuracy of this prediction was then validated in multiple ways. Firstly, by running a paired samples t-test comparing the predicted and actual IQ score. Second, by calculating the absolute difference between the predicted and actual adult IQ in the test dataset (i.e. absolute difference is calculated by taking the distance between the predicted IQ score and actual score regardless of whether the prediction is an over or under estimation. E.g. Two individuals, both with an IQ prediction of -0.2 but actual scores of 0.3 and -0.7, will both have an absolute difference of 0.5). As accurate prediction in this scenario is somewhat subjective, it was determined that regardless of whether the prediction was an over or underestimation of the actual IQ score, that a prediction within 0.33 Z (5 IQ points) was an accurate prediction and within 1 Z (15 IQ points) was evidence for a clinically useful prediction. In order to further investigate prediction accuracy, a multiple regression was ran with the absolute accuracy of the prediction as the dependent variable. Predictors were the same as those included in model 3, allowing for the investigation of whether certain factors (e.g being male, having a lower gestational age) was associated with particularly poor prediction within the test sample. As well as the prediction of the continuous IQ Z scores, the ability to predict/classify cognitive impairment (actual adult IQ Z score < -2) was investigated. Thus, sensitivity, specificity and area under the curve (AUC) are also reported.

As to address the third research question of future cognitive performance for more recent cohorts, the model's parameters based on the training data were then used to predict the future adult cognitive performance of the GNN child cohort for which child IQ data has only recently been collected. Of particular importance was the change in mean score between the actual child IQ scores and predicted adult IQ scores. Thus whether scores are expected to improve or deteriorate into adulthood and the related changes in cognitive impairment was investigated.

All analyses were performed in R (R Core Team, 2013), linear mixed models were undertaken using the package lme4 (Bates et al., 2015, S. 4), while the calculation and presentation of sensitivity, specificity and area under the curve (AUC) results were performed using the package pROC (Robin et al., 2011).

3 MAIN FINDINGS

3.1 Demographics

In total there were 349 VLBW participants from 2 adult cohorts and 707 VLBW participants from the GNN child cohort, see Table 1. Randomly splitting the initial full adult data with a 3:1 ratio (75% to 25%) resulted in 258 participants in the training dataset and 91 participants in the test dataset.

Table 1: Cohorts included and IQ tests used

	Adult-BLS		**Adult-VICS**		**Child - GNN**
	Test Data (N=39)	**Training Data (N=98)**	**Test Data (N=52)**	**Training Data (N=160)**	**Child Data (N=707)**
Country					
Germany	39 (100%)	98 (100%)	-	-	707 (100%)
Australia	-	-	52 (100%)	160 (100%)	-
Birth_Year					
1985/1986	39 (100%)	98 (100%)	-	-	-
1991/1992	-	-	52 (100%)	160 (100%)	-
2012/2013	-	-	-	-	707 (100%)
Child_IQ_Test					
K-ABC(6 years)	39 (100%)	98 (100%)	-	-	-
WISC III(8 years)	-	-	52 (100%)	160 (100%)	-
WPPSI-III (5 years)	-	-	-	-	707 (100%)
Adult_IQ_Test					
WAIS III (162, 26 years)	39 (100%)	98 (100%)	-	-	NA
WAIS II (18 years)	-	-	52 (100%)	160 (100%)	NA
Gestational Age (weeks)					

	Adult-BLS		Adult-VICS		Child - GNN
	Test Data (N=39)	**Training Data (N=98)**	**Test Data (N=52)**	**Training Data (N=160)**	**Child Data (N=707)**
Mean (SD)	30.0 (2.28)	30.3 (2.36)	26.4 (1.85)	26.7 (2.03)	28.5 (1.76)
Median [Min, Max]	30.0 [25.0, 35.0]	30.0 [26.0, 36.0]	26.0 [23.0, 32.0]	26.0 [23.0, 34.0]	28.0 [26.0, 35.0]
Birthweight (grams)					
Mean (SD)	1170 (239)	1170 (211)	891 (164)	879 (160)	1090 (253)
Median [Min, Max]	1200 [630, 1480]	1190 [680, 1500]	880 [596, 1330]	890 [430, 1280]	1100 [400, 1500]
Birthweight Z score (Fenton reference)					
Mean (SD)	-0.853 (0.921)	-1.03 (1.03)	-0.0259 (0.993)	-0.230 (1.06)	-0.402 (0.888)
Median [Min, Max]	-0.732 [-2.69, 0.587]	-0.822 [-3.58, 1.35]	0.0472 [-2.39, 2.16]	-0.124 [-3.37, 1.94]	-0.313 [-2.95, 2.00]
IVH Grade					
IVH grade 3 or 4	2 (5.1%)	9 (9.2%)	3 (5.8%)	10 (6.3%)	10 (1.4%)
no IVH, or IVH grade 1 or 2	37 (94.9%)	89 (90.8%)	49 (94.2%)	150 (93.8%)	697 (98.6%)
Bronchopulmonary Dysplasia					
BPD	19 (48.7%)	57 (58.2%)	22 (42.3%)	58 (36.3%)	209 (29.6%)
no BPD	20 (51.3%)	41 (41.8%)	30 (57.7%)	102 (63.8%)	498 (70.4%)
Sex					
Male	21 (53.8%)	45 (45.9%)	22 (42.3%)	67 (41.9%)	376 (53.2%)
Female	18 (46.2%)	53 (54.1%)	30 (57.7%)	93 (58.1%)	331 (46.8%)
Socio-Economic Status					

	Adult-BLS		Adult-VICS		Child - GNN
	Test Data (N=39)	**Training Data (N=98)**	**Test Data (N=52)**	**Training Data (N=160)**	**Child Data (N=707)**
Medium/Low SES	20 (51.3%)	43 (43.9%)	22 (42.3%)	65 (40.6%)	341 (48.2%)
High SES	19 (48.7%)	55 (56.1%)	30 (57.7%)	95 (59.4%)	366 (51.8%)
Child IQ Z Score					
Mean (SD)	-0.985 (1.43)	-1.43 (1.57)	-0.816 (1.25)	-0.567 (1.13)	-0.512 (1.10)
Median [Min, Max]	-1.10 [-5.09, 1.26]	-1.10 [-5.09, 1.44]	-0.814 [-3.76, 2.13]	-0.533 [-3.97, 2.20]	-0.534 [-4.23, 2.70]
Adult IQ Z Score					
Mean (SD)	-1.24 (1.44)	-1.47 (1.63)	-0.968 (1.24)	-0.729 (1.16)	NA (NA)
Median [Min, Max]	-0.975 [-5.02, 1.77]	-1.20 [-5.02, 1.31]	-1.07 [-3.68, 1.39]	-0.740 [-3.46, 2.49]	NA [NA, NA]
Missing	0 (0%)	0 (0%)	0 (0%)	0 (0%)	707 (100%)

3.2 Association of Neonatal Factors and Socioeconomic Status with Adult IQ

Table 2 shows the results of the first linear mixed model examining the association of neonatal factors and maternal educational level with adult IQ scores in the training dataset of VLBW participants. Significant associations with adult IQ Z scores were gestational age, birthweight Z score, the presence of neonatal BPD, grade 3/4 IVH and High SES (Table 2). On a traditional IQ test, these findings indicated that among VLBW participants BPD, IVH and low/medium SES would be associated with IQ scores 7.2 (-0.48 Z score), 12.5 (-0.83 Z score) and 11.6 (-0.77 Z score) points lower respectively. In the first analyses, neonatal data and SES in total explained 18% of the variance (Table 2).

Table 2: Predicting adult IQ using neonatal and SES data in the training dataset (Model 1)

	Adult IQ Z Score		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	-5.19	-8.25 – -2.12	**0.001**
Gestational Age (weeks)	0.14	0.04 – 0.25	**0.009**
Fenton Birthweight Z (per 1 SD)	0.23	0.01 – 0.44	**0.038**
Female Sex	0.04	-0.26 – 0.35	0.777
IVH Grade 3/4 (No Grade/1/2 as reference)	-0.83	-1.41 – -0.26	**0.005**
BPD (No BPD as reference)	-0.48	-0.81 – -0.15	**0.004**
High SES (Low/Middle SES as reference)	0.77	0.46 – 1.07	**<0.001**
Random Effects			
σ^2	1.48		
$\tau_{00\ cohort}$	0.20		
ICC	0.12		
$N_{\ cohort}$	2		
Observations	258		
Marginal R^2 / Conditional R^2	0.182 / 0.280		

3.3 Association of Child IQ with Adult IQ

The association between child IQ and adult IQ is shown in table 3 and in figure 1. It was found that an increase of child IQ Z score by 1SD was associated with an increase of 0.88 in adult IQ Z score. As the sole fixed predictor of adult IQ in the training dataset, the model explained over 74% of the variance, see table 3.

Table 3: Association between child IQ and adult IQ for VLBW participants in the training dataset (Model 2)

	Adult IQ Z Score		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	-0.23	-0.33 – -0.12	**<0.001**
Child IQ Z Score (per 1 SD)	0.88	0.81 – 0.94	**<0.001**
Random Effects			
σ^2	0.51		
$\tau_{00\ cohort}$	0.00		
$N_{\ cohort}$	2		
Observations	258		
Marginal R^2 / Conditional R^2	0.741 / NA		

3.4 Final model of Neonatal Factors, Socioeconomic status and Child IQ predicting Adult IQ

The final model predicting adult IQ for the training dataset found that SES and child IQ were significant predictors, explaining 76%, of the total variance, see table 4. Thus, following the inclusion of child IQ, it was found that gestational age, birthweight Z score, IVH and BPD all no longer significantly contributed to the prediction of adult IQ. As well as shown here with linear mixed models, evidence that these neonatal factors are significantly "mediated" by child IQ is further demonstrated by a mediation analysis, as shown in appendix 1.

Table 4: Predicting adult IQ using neonatal, SES and child IQ data in the training dataset

	Adult IQ Z Score		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	-1.00	-2.27 – 0.27	0.124
Gestational Age (weeks)	0.02	-0.02 – 0.07	0.290
Fenton Birthweight Z (per 1 SD)	-0.01	-0.13 – 0.11	0.866
Female Sex	-0.12	-0.30 – 0.05	0.162
IVH Grade 3/4 (No Grade/1/2 as reference)	-0.22	-0.55 – 0.11	0.182
BPD (No BPD as reference)	-0.08	-0.26 – 0.10	0.361
High SES (Low/Middle SES as reference)	0.28	0.10 – 0.46	**0.002**
Child IQ Z Score (per 1 SD)	0.85	0.78 – 0.91	**<0.001**
Random Effects			
σ^2	0.47		
$\tau_{00\ cohort}$	0.00		
$N_{\ cohort}$	2		
Observations	258		
Marginal R^2 / Conditional R^2	0.760 / NA		

3.5 Validation of the Final Model predicting continuous IQ score and cognitive impairment

The final model parameters were then used to predict the adult IQ of the test dataset (see Figure 1 and Table 5). The result of the paired samples t-test found that there was no significant difference between the actual IQ scores (M= -1.08, SD= 1.33) and the predicted IQ scores (M= -1.01, SD= 1.19) in the test dataset; t(90)= -0.83, p = 0.411. The absolute difference between the predicted and actual IQ Z score was on average 0.66 (9.9 IQ points). When subcategorizing the accuracy of the prediction, it was found that 79% of predictions were within 15 points, with 31% of predictions within five IQ points. Further investigation using a multiple regression, with the absolute accuracy of the IQ prediction as the outcome of interest, found that those with lower SES parents were likely to have a less accurate prediction, see appendix 2.

In regard to predicting actual cognitive impairment in the test dataset, a predicted IQ Z score of -2 was originally used as the threshold. Using this threshold, it was found that of the 19 participants with cognitive impairments in adulthood, 14 were accurately classified as such, with five incorrectly classified as non-impaired. In addition, two non-impaired participants were incorrectly classified as having cognitive impairments. This resulted in a sensitivity of 74% and a specificity of 97%. When the relationship between predicted IQ Z score and actual cognitive impairment was further investigated, it was found that a threshold of -1.72 on the predicted score was optimal for predicting actual cognitive impairment. Sensitivity was found to be 84% (16/19 impaired participants correctly classified) and a specificity of 94% (69/72 non impaired participants correctly classified) with an area under the curve of 0.92, see figure 2.

Table 5: Accuracy of the IQ prediction- the difference between predicted and actual adult IQ scores in the remaining 25% data (test dataset), differentiated by cohort.

	BLS (N=39)	**VICS (N=52)**	**Overall (N=91)**
Child IQ Z Score			
Mean (SD)	-0.985 (1.43)	-0.816 (1.25)	-0.889 (1.32)
Median [Min, Max]	-1.10 [-5.09, 1.26]	-0.814 [-3.76, 2.13]	-0.884 [-5.09, 2.13]
Adult IQ Z Score			
Mean (SD)	-1.24 (1.44)	-0.968 (1.24)	-1.08 (1.33)
Median [Min, Max]	-0.975 [-5.02, 1.77]	-1.07 [-3.68, 1.39]	-1.03 [-5.02, 1.77]
Predicted Adult IQ Z Score			
Mean (SD)	-1.05 (1.27)	-0.988 (1.14)	-1.01 (1.19)
Median [Min, Max]	-1.11 [-4.68, 1.07]	-0.934 [-3.71, 1.71]	-1.03 [-4.68, 1.71]
Absolute Difference between Predicted and Actual IQ Z Score*			
Mean (SD)	0.663 (0.454)	0.660 (0.447)	0.661 (0.447)

	BLS (N=39)	VICS (N=52)	Overall (N=91)
Median [Min, Max]	0.615 [0.0150, 1.68]	0.573 [0.0585, 1.84]	0.590 [0.0150, 1.84]
Accuracy of Predicted Adult IQ Score			
Within 5 IQ Points	11 (28.2%)	17 (32.7%)	28 (30.8%)
5 to 15 IQ Points	20 (51.3%)	24 (46.2%)	44 (48.4%)
15 to 30 IQ Points	8 (20.5%)	11 (21.2%)	19 (20.9%)
More than 30 IQ Points	0 (0%)	0 (0%)	0 (0%)
Predicted Adult Cognitive Impairment (< -2 SD)			
Impaired	8 (20.5%)	8 (15.4%)	16 (17.6%)
Non-Impaired	31 (79.5%)	44 (84.6%)	75 (82.4%)
Actual Adult IQ Impairment (< -2 SD)			
Impaired	10 (25.6%)	9 (17.3%)	19 (20.9%)
Non-Impaired	29 (74.4%)	43 (82.7%)	72 (79.1%)

* Absolute difference is calculated by taking the distance between the predicted IQ score and actual score regardless of whether the prediction is an over or under estimation. E.g. Two individuals, both with an IQ prediction of -0.2 but actual scores of 0.3 and -0.7, will both have an absolute difference of 0.5.

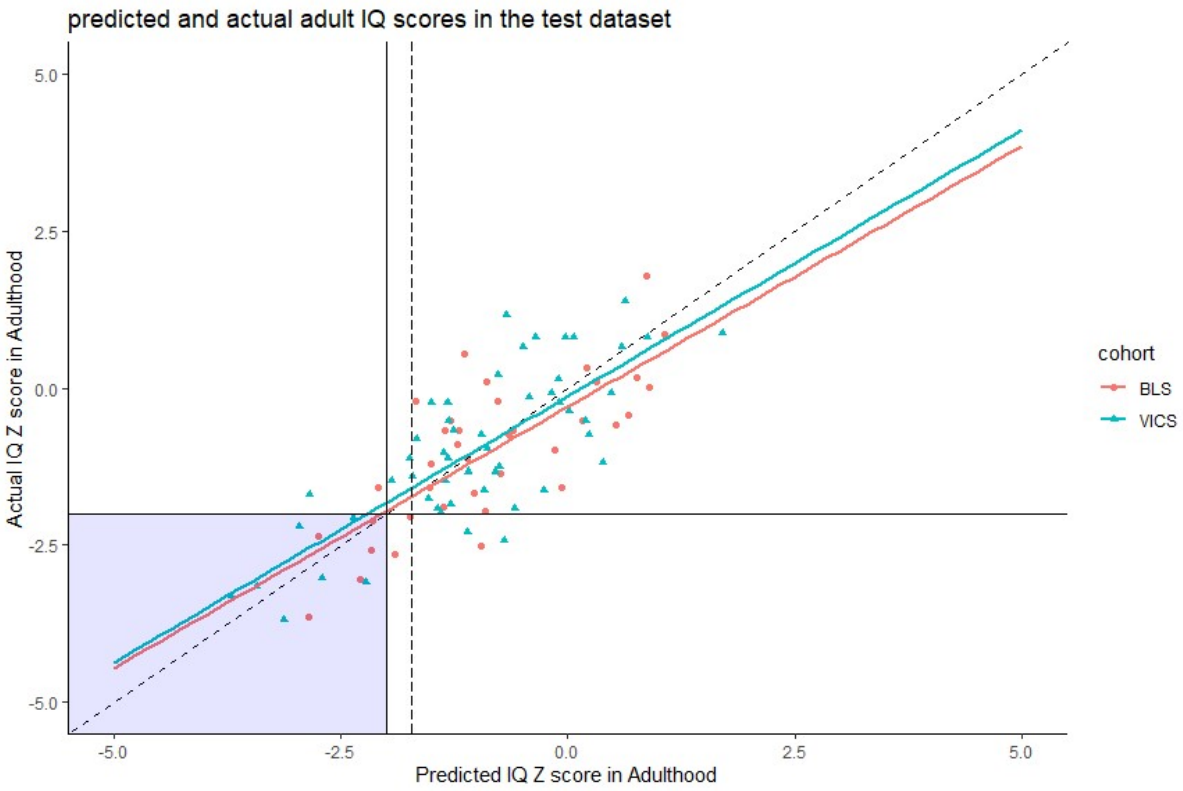

Figure 1: the difference between predicted and actual adult IQ scores in the test dataset, differentiated by cohort. Those within the blue square were successfully classified as having a cognitive impairment (predicted IQ Z score < -2 SD and actual IQ Z Score < -2 SD), the dashed line indicates the "optimal" threshold in the test dataset regarding predicted and actual IQ impairment (predicted IQ Z score <-1.72).

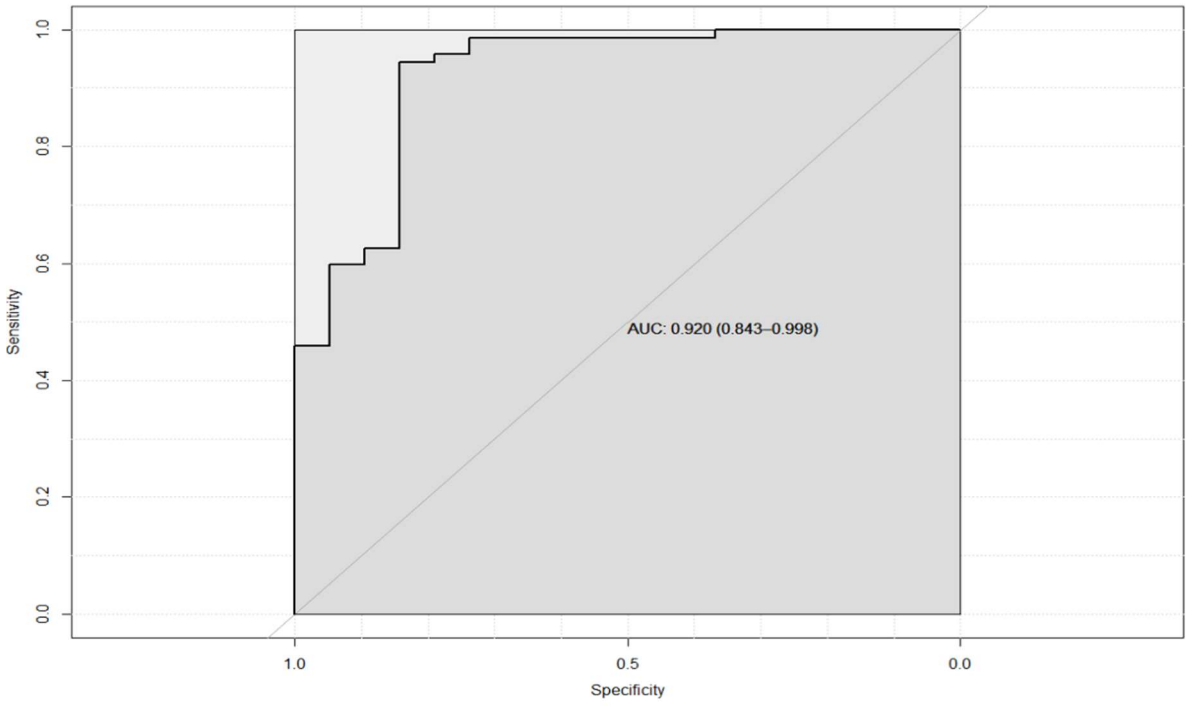

Figure 2: Sensitivity and specificity between for adult IQ Z score and actual adult cognitive impairment. Highest specificity (94%) and sensitivity (84%) at Predicted score of SD -1.72 to detect actual adult cognitive impairment.

3.6 Prediction of Adult IQ for Current VLBW Children

The final objective was to predict the future adult IQ of the more recently born GNN cohort. From the -0.51 IQ Z score difference to term born controls at age 5, the model predicts that this difference will slightly increase to -0.66 into adulthood, see figure 3. In addition, the model predicts that 8.8% of participants will have cognitive impairment, defined as scoring 2 SD below their peers. If the potentially more *optimal threshold of -1.72 is instead used, it is found that 13.4% of participants are predicted to be classified* as impaired in adulthood.

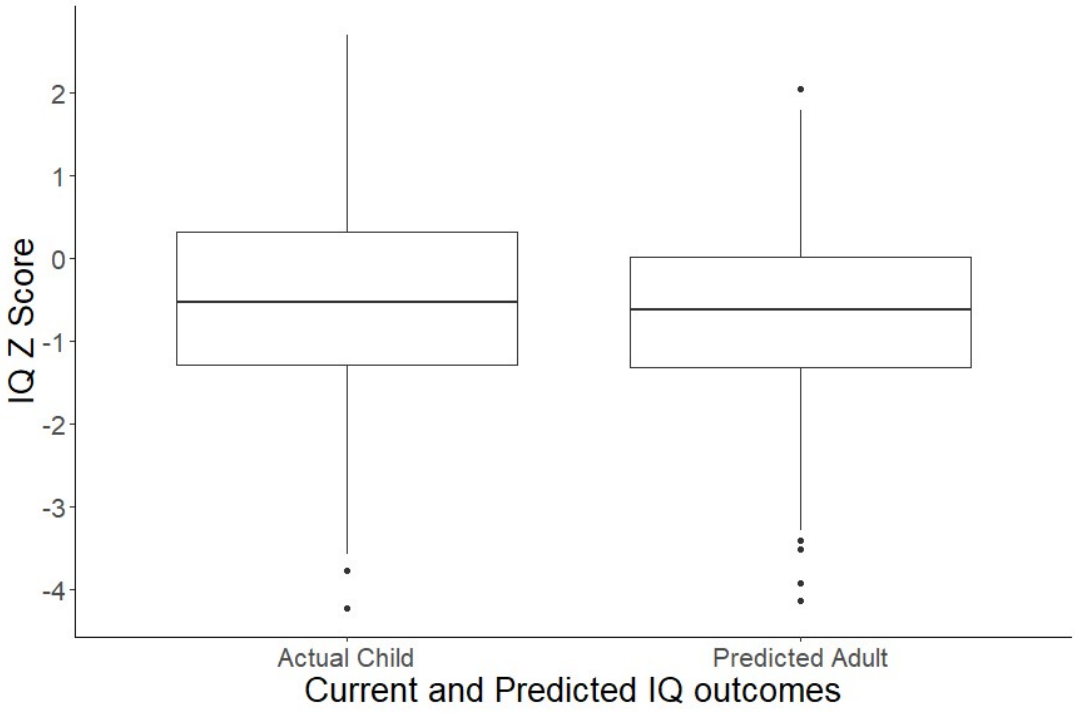

Figure 3: Predicted IQ outcomes for the GNN cohort

4 MEANING OF FINDINGS

In this analysis of 3 different VLBW cohorts, born in 2 different countries and from 1985 to 2012, the prediction of adult IQ using neonatal factors, parental socioeconomic status and child IQ was investigated. It was found that while early neonatal factors are significantly associated with long term IQ, their ability to improve prediction of adult IQ is limited once child IQ is also known. Childhood IQ appears a common pathway of how neonatal complications affect the brain and cognitive function into adulthood.

It was found that the accuracy of individual IQ point prediction in the test dataset was only moderately successful, within 15 IQ points for 79% of participants. On the one hand, with 79% of participants having their prediction within 15 IQ points, the prediction is going to be useful for a sizable number of participants. However, this also means that 21% of VLBW participants had predictions off by more than 1 SD. This result therefore indicates that even once neonatal data, parental SES and most importantly child IQ is known, there are still a sizable number of individuals who do not follow a predicted pattern, either significantly under or over performing on adult IQ. However, in contrast to the prediction of exact scores within the whole range, *the prediction of cognitive impairment (IQ<-2SD) in adulthood was highly specific and with good sensitivity*. The AUC was .92; an excellent value considering that prediction spans 2 decades in the BLS and a decade in VICS.

One factor that was important in addition to child IQ was parental SES, but this was a relatively simplistic binary variable. As this was the only parental/socio-environmental predictor available in the GNN sample and thus harmonisable across cohorts, it was the only socio-environmental factor that could be included in the analysis. However, based on the current findings, more emphasis should be placed on collection of more fine-grained parental/socio-environmental factors in more recent cohort studies, as this may improve the prediction of adult IQ. For example, prior evidence from the BLS found maternal sensitivity and the parent-infant relationship to both be significantly associated with education and cognitive outcomes (Eves et al., 2020; Jaekel et al., 2015). Future research should look to investigate more intensely the social and family environment where protective factors may be identified while growing up (Wolke, 2019).

The current study has a number of strengths. It includes the use of testing and training datasets in order to test how accurate predictions from the model are, rather than simply developing a model without subsequent verification. In addition, the use of multiple cohorts ranging in

birth year and country vastly increases the applicability to other groups of VLBW individuals from other countries, and provides a crucial insight into what future outcomes may be expected for VLBW participants born in the 21st century. As well as this, the use of adult IQ as the outcome of interest is also a strength of this study over past research. As DQ may only have limited association with adult IQ, it suggests using DQ as an end point to determine the outcomes of VLBW individuals is limited (Aylward, 2009; Breeman et al., 2015). Instead, both the current study and past research has indicated that child IQ and adult IQ are strongly correlated, for both the VLBW and general population (Aylward, 2009). The current study therefore suggests that for those interested in long term outcomes of VLBW individuals, an early child IQ test at the time of primary school entry may offer more reliable prediction of long-term outcome, in particular for those at risk of cognitive impairment. Due to the fact that childhood IQ assessed at early primary school age is highly sensitive and specific in predicting long-term cognitive impairment, assessment at this time may help to utilize more educational resources and support to maximize potential of these children with cognitive impairment.

The limitations of the current study were the differences between cohorts regarding IQ tests used and differences in recruiting of controls, in particular for the GNN. IQ scores were always harmonised according to the mean and standard deviation of each cohort's control group. This means the relative difference and comparisons across cohorts is reliant upon the fact that each control group accurately reflects the healthy term born population, in particular for cognitive performance (Wolke et al., 1994). As the GNN controls were not a nationally representative sample, their average IQ score may not reflect the average IQ score for the term born population of Germany. If for whatever reason the GNN control group is particularly low performing, this may artificially lower the estimated IQ difference between VLBW and healthy term born populations. Another limitation is that only few perinatal factors could be considered as predictors for adult IQ, largely due to the difference in which variables were collected in each cohort, and how that varies depending on their respective neonatal era. Conversely, as the included neonatal factors appeared to be largely mediated once child IQ was included, it is possible that the addition of further neonatal factors would also be similarly mediated and thus add little in predicting adult IQ as their common pathway is via child IQ.

Overall, this research will aid to search for potential further predictors of adult cognitive functioning. Factors across the lifespan, whether further risks or protective factors may

explain how some VLBW individuals improve or show deterioration in IQ between childhood and adulthood. Any personalized predictive models do require the inclusion of socio-environmental factors for more accurate prediction of long-term adult cognitive performance. In particular, factors that are amenable to modification such as parenting or educational interventions should be considered in addition to child IQ.

To conclude, the universally high association between child IQ and adult IQ for VLBW infants emphasizes the importance of early cognitive screening, with the ability in particular to identify those who will display long term cognitive impairment. In particular, assessment of child IQ at school entry/early school age allows for highly accurate determination of those with long-term cognitive impairment and thus coordination of potential educational and psychosocial intervention. Further VLBW research needs to investigate how socioenvironmental factors may act as protective factors in order to both improve the prediction of long-term outcomes and potentially aid interventions for VLBW children.

5 LITERATURE REFERENCES

Ambalavanan, N., Baibergenova, A., Carlo, W. A., Saigal, S., Schmidt, B., & Thorpe, K. E. (2006). Early prediction of poor outcome in extremely low birth weight infants by classification tree analysis. *The Journal of Pediatrics*, *148*(4), 438-444.e1. https://doi.org/10.1016/j.jpeds.2005.11.042

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

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

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

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

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

Crilly, C. J., Haneuse, S., & Litt, J. S. (2021). Predicting the outcomes of preterm neonates beyond the neonatal intensive care unit: What are we missing? *Pediatric Research*, *89*(3), 426–445. https://doi.org/10.1038/s41390-020-0968-5

de Kleine, M. J. K., den Ouden, A. L., Kollée, L. A. A., Ilsen, A., van Wassenaer, A. G., Brand, R., & Verloove-Vanhorick, S. P. (2007). Lower mortality but higher neonatal morbidity

over a decade in very preterm infants. *Paediatric and Perinatal Epidemiology*, *21*(1), 15–25. https://doi.org/10.1111/j.1365-3016.2007.00780.x

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

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

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

Eves, R., Mendonça, M., Baumann, N., Ni, Y., Darlow, B. A., Horwood, J., Woodward, L. J., Doyle, L. W., Cheong, J., Anderson, P. J., Bartmann, P., Marlow, N., Johnson, S., Kajantie, E., Hovi, P., Nosarti, C., Indredavik, M. S., Evensen, K.-A. I., Räikkönen, K., … Wolke, D. (2021). Association of Very Preterm Birth or Very Low Birth Weight With Intelligence in Adulthood: An Individual Participant Data Meta-analysis. *JAMA Pediatrics*. https://doi.org/10.1001/jamapediatrics.2021.1058

Farooqi, A., Hägglöf, B., Sedin, G., & Serenius, F. (2011). Impact at age 11 years of major neonatal morbidities in children born extremely preterm. *Pediatrics*, *127*(5), e1247-1257. https://doi.org/10.1542/peds.2010-0806

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

Geisler, I., Rausch, T. K., Göpel, W., & Spiegler, J. (2021). Extremely and very preterm-born children <1500 g show different weight development in childhood compared to their peers. *Acta Paediatrica*, *110*(7), 2093–2099. https://doi.org/10.1111/apa.15785

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

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

Kalstabakken, A. W., Molitor, S. J., Gross, A. C., Georgieff, M. K., & Boys, C. J. (2021). Predictive Value of Developmental Assessment in a Neonatal Intensive Care Unit (NICU) Follow-Up Clinic. *Journal of Pediatric Psychology*, *jsab048*. https://doi.org/10.1093/jpepsy/jsab048

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

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

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

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

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

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C., & Müller, M. (2011). PROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC bioinformatics*, *12*(1), 1–8.

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

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

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

Wolke, D., Ratschinski, G., Ohrt, B., & Riegel, K. (1994). The cognitive outcome of very preterm infants may be poorer than often reported: An empirical investigation of how methodological issues make a big difference. *European Journal of Pediatrics*, *153*(12), 906–915. https://doi.org/10.1007/BF01954744

6 APPENDICES

6.1 Appendix 1: mediating effect of child IQ on the relation between neonatal factors and adult IQ

The DV (Y) was Adult IQ Z Score. The IV (X) was Gestational Age (weeks), Fenton Birthweight Z (per 1 SD), Male Sex, High SES (Low/Middle SES as reference), IVH Grade 3 or 4, BPD, and Cohort-VICS. The mediating variable was Child IQ Z Score.

Total effect(c) of Gestational Age (weeks) on Adult IQ Z Score = 0.16 S.E. = 0.06 t = 2.84 df= 250 with p = 0.005. Direct effect (c') of Gestational Age (weeks) on Adult IQ Z Score removing Child IQ Z Score = 0.06 S.E. = 0.03 t = 1.7 df= 249 with p = 0.091

Indirect effect (ab) of Gestational Age (weeks) on Adult IQ Z Score through Child IQ Z Score = 0.11. Mean bootstrapped indirect effect = 0.11 with standard error = 0.05 Lower CI = 0.01 Upper CI = 0.21.

Total effect(c) of Fenton Birthweight Z (per 1 SD) on Adult IQ Z Score = 0.24 S.E. = 0.11 t = 2.19 df= 250 with p = 0.029. Direct effect (c') of Fenton Birthweight Z (per 1 SD) on Adult IQ Z Score removing Child IQ Z Score = 0.02 S.E. = 0.06 t = 0.31 df= 249 with p = 0.76. Indirect effect (ab) of Fenton Birthweight Z (per 1 SD) on Adult IQ Z Score through Child IQ Z Score = 0.22. Mean bootstrapped indirect effect = 0.22 with standard error = 0.1 Lower CI = 0.02 Upper CI = 0.41.

Total effect(c) of Male Sex on Adult IQ Z Score = -0.04 S.E. = 0.16 t = -0.24 df= 250 with p = 0.81. Direct effect (c') of Male Sex on Adult IQ Z Score removing Child IQ Z Score = 0.13 S.E. = 0.09 t = 1.46 df= 249 with p = 0.15. Indirect effect (ab) of Male Sex on Adult IQ Z Score through Child IQ Z Score = -0.17. Mean bootstrapped indirect effect = -0.17 with standard error = 0.13 Lower CI = -0.43 Upper CI = 0.09.

Total effect(c) of High SES (Low/Middle SES as reference) on Adult IQ Z Score = 0.76 S.E. = 0.16 t = 4.79 df= 250 with p = 2.9e-06. Direct effect (c') of High SES (Low/Middle SES as reference) on Adult IQ Z Score removing Child IQ Z Score = 0.27 S.E. = 0.09 t = 2.95 df= 249 with p = 0.0035. Indirect effect (ab) of High SES (Low/Middle SES as reference) on Adult IQ Z Score through Child IQ Z Score = 0.49. Mean bootstrapped indirect effect = 0.49 with standard error = 0.14 Lower CI = 0.23 Upper CI = 0.76-

Total effect(c) of IVH Grade 3 or 4 on Adult IQ Z Score = -0.82 S.E. = 0.3 t = -2.75 df= 250 with p = 0.0065

Direct effect (c') of IVH Grade 3 or 4 on Adult IQ Z Score removing Child IQ Z Score = -0.21 S.E. = 0.17 t = -1.24 df= 249 with p = 0.22

Indirect effect (ab) of IVH Grade 3 or 4 on Adult IQ Z Score through Child IQ Z Score = -0.61

Mean bootstrapped indirect effect = -0.6 with standard error = 0.3 Lower CI = -1.2 Upper CI = -0.01

Total effect(c) of BPD on Adult IQ Z Score = -0.45 S.E. = 0.17 t = -2.61 df= 250 with p = 0.0097

Direct effect (c') of BPD on Adult IQ Z Score removing Child IQ Z Score = -0.04 S.E. = 0.1 t = -0.4 df= 249 with p = 0.69

Indirect effect (ab) of BPD on Adult IQ Z Score through Child IQ Z Score = -0.41

Mean bootstrapped indirect effect = -0.4 with standard error = 0.14 Lower CI = -0.68 Upper CI = -0.13.

Total effect(c) of Cohort-VICS on Adult IQ Z Score = 0.99 S.E. = 0.23 t = 4.21 df= 250 with p = 3.6e-05. Direct effect (c') of Cohort-VICS on Adult IQ Z Score removing Child IQ Z Score = 0.19 S.E. = 0.14 t = 1.37 df= 249 with p = 0.17. Indirect effect (ab) of Cohort-VICS on Adult IQ Z Score through Child IQ Z Score = 0.8. Mean bootstrapped indirect effect = 0.8 with standard error = 0.22 Lower CI = 0.38 Upper CI = 1.24.

Summary of Mediation: Total Covariates effects (C) on Adult IQ, and then the direct effect after the mediator of Child IQ is removed (C‘)

Variable	Beta (C)	P value	Beta (C‘)	P value
Intercept	-6.19	3.05e-04	-2.11	3.15e-02
Gestational Age (weeks)	0.16	4.95e-03	0.06	9.07e-02
Fenton Birthweight Z (per 1 SD)	0.24	2.93e-02	0.02	7.57e-01
Male Sex	-0.04	8.11e-01	0.13	1.47e-01
High SES (Low/Middle SES as reference)	0.76	2.92e-06	0.27	3.46e-03
IVH Grade 3 or 4	-0.82	6.48e-03	-0.21	2.17e-01
BPD	-0.45	9.67e-03	-0.04	6.86e-01
Cohort:VICS	0.99	3.61e-05	0.19	1.71e-01

6.2 Appendix 2: Investigation of factors associated to accuracy of IQ prediction; i.e. the absolute difference between actual and predicted score

	Absolute Difference between Actual and Predicted IQ		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	1.01	-0.86 – 2.88	0.285
Gestational Age (weeks)	-0.00	-0.07 – 0.06	0.882
Fenton Birthweight Z (per 1 SD)	0.00	-0.13 – 0.14	0.950
Female Sex	-0.15	-0.34 – 0.05	0.138
High SES (Low/Middle SES as reference)	-0.21	-0.40 – -0.01	**0.042**
IVH Grade 3/4 (No Grade/1/2 as reference)	-0.25	-0.67 – 0.17	0.232
BPD (No BPD as reference)	0.08	-0.12 – 0.28	0.427
Child IQ Z Score (per 1 SD)	0.06	-0.02 – 0.14	0.119
VICS-cohort (BLS as reference)	0.01	-0.25 – 0.27	0.950
Observations	91		
R^2 / R^2 adjusted	0.111 / 0.024		