ERASMUS UNIVERSITY ROTTERDAM
ERASMUS SCHOOL OF ECONOMICS
MSc Economics & Business
Master Specialisation Health Economics

Are birth weight and educational performances causally related?

Using a sensitivity analysis to determine the causal effect of birth weight on school performances

Author:	E. Teichert
Student number:	347322
Thesis supervisor:	dr. J.L.W. van Kippersluis
Finish date:	August 2017

NON-PLAGIARISM STATEMENT

By submitting this thesis the author declares to have written this thesis completely by himself/herself, and not to have used sources or resources other than the ones mentioned. All sources used, quotes and citations that were literally taken from publications, or that were in close accordance with the meaning of those publications, are indicated as such.

ABSTRACT

Background

Politicians and policy makers have tried in many ways to make life more equitable. To do so they need to have insights in which factors influence for example education, and therefore influence wages and lifestyle. Some of the factors policies have been made upon that were considered a causality, have later been proven to be only an association biased by other factors underlying this coherence. One of the factors where opinions of researchers are ambiguous is birth weight. There have been different views, whether birth weight has a positive causal effect on educational performances or being nothing more than a spurious biased association. A lot of the effect attributed to the difference in birth weight has been nullified by controlling for socioeconomic factors, parental education and maternal effects.

Methods

In this paper we try to clarify under which circumstances birth weight has a causal impact on school performances using an informative sensitivity analysis on a GWAS sample of the Avon Longitudinal Study of Parents and Children (ALSPAC). We do so by using the sensitivity analysis proposed by Bowden, Davey Smith & Burgess (2015) to account for possible imperfection of the exclusion restriction, to still be able to draw conclusions about using a polygenic risk score of birth weight as instrument to identify the causal impact of birth weight on school grades and IQ test scores.

Results

We obtain a strong positive significant association of birth weight and school performances, which too a certain extent holds by adjusting for socioeconomic factors, parental education and maternal effects. Some of the outcome measures are identified to be positively causal impacted by birth weight using the polygenic risk score of birth weight as instrument, in case the exclusion restriction would hold. But by performing a sensitivity analysis we can see that a small violation of the exclusion restriction neglect the significant positive results.

Conclusions

Concluding from our data, birth weight is not very likely to have a causal impact on educational performances, since a small violation of the exclusion restriction would neglect the significant results we obtained from the IV regression. Although SNPs and polygenic risk scores are assumed to be valid and strong instruments in most cases we cannot be sure that the exclusion restriction is exactly satisfied. We could identify to what 'imperfection' a significant effect will hold and give a brief insight in the associations birth weight has on different school topics, and how these associations change over time.

Keywords:
Birth weight, education, single-nucleotide polymorphisms (SNPs), instrumental variable analyses, sensitivity analysis.

TABLE OF CONTENTS

1 Introduction

Birth weight has been intensively researched in the last decades. Apart from the determinants of birth weight (Kramer, 1987), and low birth weight specifically (Silvestrin, et al., 2013), the effects that birth weight as determinant can have further in live have also been investigated on different levels. Next to the health-related issues correlated with birth weight (Lawlor, Ebrahim & Davey Smith, 2005; Yu et al., 2011), birth weight also has an economic impact according to Black, Devereux & Salvanes (2005; 2007).

1.1 Economic perspective

From a policy point of view it is interesting to gain insights in the effects of birth weight on for example school performances, so policy makers will be able to introduce programs that lead to more equal opportunities for children, by influencing factors that are a causality rather than just selection. The difference between causality and selection is that in case of selection the fact that the child's parents are successful will make him successful as well later in life, regardless, whereas causality implies that the success of the parents make them raise their child in a different way compared to not so successful parents. So to implement successful policies, a good understanding of causal factors is needed.

Some of the factors policies have been made upon that were considered a causality, have later been proven to be only an association biased by other factors underlying this coherence. One of the factors where opinions of researchers are ambiguous is birth weight.

The interest in the causality between birth weight and academic performance exists at least since 1946 (Asher, 1946). Since surveys and observational studies became more extensive over the years, researchers became more adept in determining if the factor birth weight really has a causal impact on performance indicators.

The question what the relationship between birthweight and academic performance indicators is, has been asked and researched extensively in the last decade. The results, whether the mostly positive association is somehow biased and thus spurious (Chatterji, Kim & Lahiri, 2014; Kirkegaard, Obel & Hedegaard, 2006; Shenkin, Starr & Deary, 2004) or causal (Newcombe et al., 2007), were ambiguous.

1.2 Econometric perspective

The effect of birth weight on performance indicators becomes, too a large extent, nullified after controlling for socioeconomic factors, parental education and maternal effects (Shenkin et al., 2004; Fletcher, 2011; Lawlor et al., 2006; Record, McKeown & Edwards, 1969). To account for possible errors due to environmental and genetic factors, studies within twin pairs have been performed (Tsou et al., 2008; Boomsma et al., 2001; Christensen et al., 2006). Conclusions about the effect of birth weight on academic performances were mixed. Implying hardly any association, an association probably biased by genetic factors (since an effect could only be found in a subsample for dizygotic twins, but not among monozygotic twins), or an association with a small magnitude.

1.3 Contribution of this thesis

Recently, Lin, Leung and Schooling (2017) used instrumental variable analyses to determine if birth weight has an effect on years of schooling or college completion. They performed a Mendelian randomization study with genetic variants as instrumental variables to measure if those with genetically higher birth weight also obtain more years of schooling, using genome-wide association studies (GWASs) data of birth weight in European individuals. Following the study of Lin et al. in this thesis we will investigate the causal impact of birth weight on educational performances using genetics as instrument in an instrumental variable analysis. This thesis differs from the approach of Lin et al. in the regard that we will correct ex post for biases in instruments, whereas they tried to anticipate with an ex ante approach. Besides that we will use scores at different ages and on different topics as dependent variables. By doing so we can obtain if the effect from birth weight diminishes over time or can only be found in alpha or in beta subjects.

1.4 Structure

In chapter two we will explain in depth the key factors that are underlying this research. In chapter three we will dive into the methods used for the analysis In chapter four an overview of the data will be given.. Chapter five contains the results of the research. And in chapter six and seven we will conclude and make some final remarks.

2 Background

2.1 Causality or association

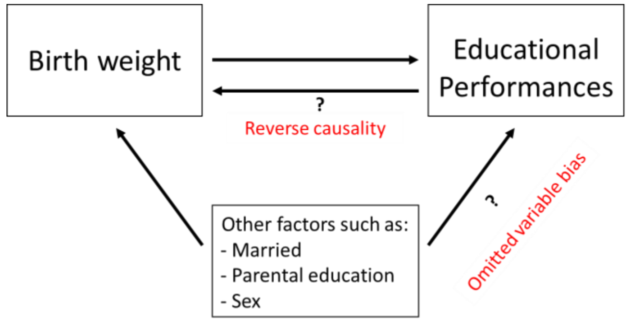
Researchers haven't agreed if the relationship between birth weight and educational outcome measures – years of schooling, the highest obtained degree or grades scored – is causal or just an association biased by omitted factors. In case of a correlation between two factors we can speak of a positive or negative association of those factors, but can only speak of a causal relationship if we are able to determine whether changes in the outcome variable are directly caused by changes in the input factor. In laboratory settings, researchers try to create two identical groups and see what happens to the outcome variable by giving the first group 'treatment A' and the second group 'treatment B', whereas all other factors stay the same as much as possible for both groups. In case of the introduction of a new drug, these kind of experiments are mandatory to show that the new treatment shows better results than the current treatment or a placebo.

But creating such groups is not always possible, due to financial constraints, ethical constraints or the fact that not all other factors can be kept constant, especially not by trying to measure long term effects. Considering this research it is likely that the outcome measure – educational performances – is not only affected by birth weight but by others factors as well. Although we can be certain that educational performances will not influence your birth weight and therefore we will not have any reverse causality bias, we are not able to determine all factors that might impact educational performances. And even if we could determine all those factors we would need data on them as well to control for these differences to isolate the causal effect.

2.2 Instrumental variable analysis

A possible solution to overcome this issue is to use instrumental variable analysis. The name of this method was introduced by Olav Reiersøl in 1945 in his dissertation, although Philip G. Wright is considered to be the first who mentioned usage of this method in his book *The Tariff on Animal and Vegetable Oils* (1928). In 2000 Sander Greenland wrote the paper *An introduction to instrumental variables for epidemiologist* introducing the usage of instrumental variables to become common in other fields than economics as well.

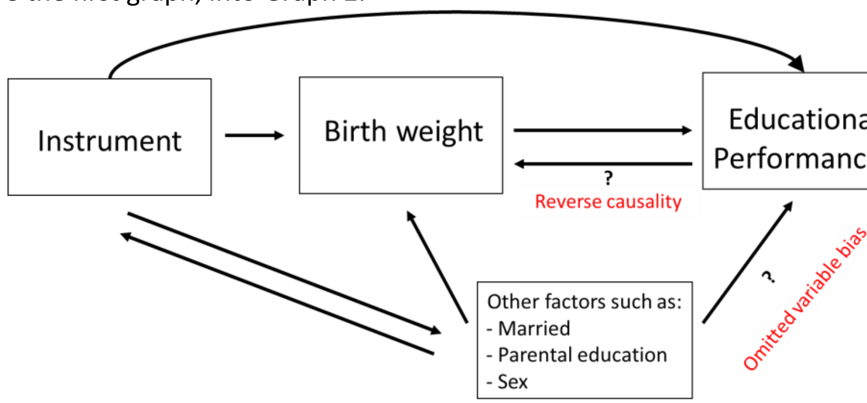
To get a grasp of instrumental variables analyses we will use graphs 1 to 4 to illustrate the idea behind this method, thereafter the mathematical approach of the model will be introduced.



Graph 1. Ordinary least squares analyses to model the effect of birth weight on educational performances corrected for the fact that if the parents from the individual are married, what their highest qualification is and if the individual is a male or a female.

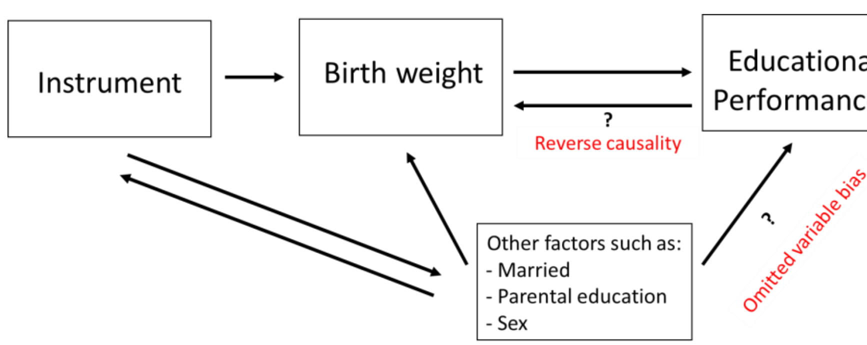
Graph 1 illustrates what we encounter by trying to identify the effect of birth weight on educational performances. Reverse causality, the effect of educational performances on birth weight biasing the effect of birth weight on educational performances, will not be an issue, since your birth weight will be set before you will be able to complete any educational performance. We do have to place a question mark with regard to omitted variable bias. We are able to identify some of the factors influencing educational performances and birth weight, but not all of them. So it is likely that when we establish a linear model, like the ordinary least squares, the error term will be correlated with the explanatory variable – birth weight – and be therefore endogenous. Since one of the underlying assumptions of the OLS model to be consistent is that the regressors are exogenous, the outcome of the model will be biased and will only measure an association of birth weight on educational performances.

We restructure the first graph, into Graph 2.



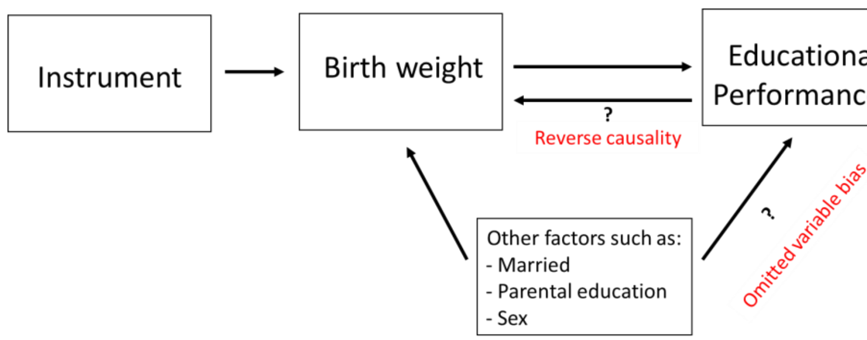
Graph 2. Visualizing the relationship of birth weight on educational performances, whereby we know that an instrument exists that has an effect on birth weight and is correlated with educational performances and might interact with other factors as well.

We assume that we can find a variable in our data – from here on forward called 'instrument' – which has an effect on our independent variable birth weight and *is correlated with our dependent variables*, namely the school performances. So we see that this instrument has an impact on educational performances, whereas from this picture it is not clear via which way it will do so. (*Assumption 1*)
If we could find an instrument that has *no causal effect on our outcome variables* we would be able to get rid of one of the routes and therefore Graph 2 would transform into Graph 3. (*Assumption 2*).



Graph 3. Visualizing the relationship of birth weight on educational performances, whereby we know that an instrument exists that has an effect on birth weight and might interact with other factors as well, but has no causal effect on the educational performances.

Now, apart from an effect of the instrument on the educational performances through birth weight there could be an effect through the other factors affecting birth weight. But if we would be able to find an *instrument that is randomly assigned* it would be unrelated to the other factors which would lead to Graph 4. (*Assumption 3*)



Graph 4. Visualizing the relationship of birth weight on educational performances, whereby we assume that an instrument exists that has an effect on birth weight, but is randomly assigned so it will be unrelated to other factors and has no causal effect on the educational performances.

If all three of those conditions would hold, we would be able to obtain an effect of the instrument on our outcome measures – the educational performances – only via our independent variable birth weight. From here on forward we would be able to identify the combined relationship of the instrument and birth weight on educational performances. Since a change in the instrument will lead to a change in the educational performances, that can only occur via birth weight. Since we can find out the correlation between the instrument and birth weight, we can isolate the effect of birth weight on educational performances, which has to be causal.

Following the notation of Bowden et al. (2012), in matrix notation the mathematical approach of the instrumental variable analysis where ε^X and ε^Y are composite error terms including unobserved confounders is composed like this:

$$Y = X\beta + G\alpha + \varepsilon^Y$$
$$X = G\gamma + \varepsilon^X$$

There are three assumptions underlying this model:

1. Relevance: The instrument G is correlated with the exposure X, which means that $\gamma \neq 0$.
2. Independence: The instrument G is uncorrelated with any other variable of the exposure-outcome relationship, which holds if G would be completely random assigned to an individual.
3. Exclusion restriction: The instrument G affects the outcome Y only through the exposure X, which means that $\alpha = 0$.

2.3 *Single-nucleotide polymorphisms*

Not all humans are the same, from the very first start – birth – we do differ from one another. To what extent this comes due to nurture and to what extent due to nature, is a discussion that does not fit into the scope of this paper. Simply the fact that part of the differences between human beings is due to nature – namely genetics – will do for us to be able to use instrumental variables.

The human genome consists of 23 pairs of chromosomes. This genome is composed of around three billion base pairs of nucleotides, each of which can be indexed by its location in the sequence. When the body makes new cells, by copying them, it doesn't make many mistakes. But sometimes a single base pair gets left out, added or substituted. If a single base pair gets substituted it creates a single-nucleotide polymorphism – abbreviated to SNP. Some SNPs account for difference in appearance, others how we develop diseases or respond to drugs, but most of them seem to lead to no observable differences between people at all. The SNPs that are associated with a certain disease for example, allow researchers to evaluate if a person's genome can explain why certain people do develop a certain disease, while others don't. And on the other hand if certain SNPs are associated with a successful trait, then researchers may examine the genes near this SNPs that are responsible for the trait.

There are two possible nucleotide pairs that exists in most of the places on the sequence in case of a SNP, namely AT and GC pairs. One of the two pairs is considered to be the reference pair. Since DNA gets passed on from parents to children, SNPs will be transferred as well. From both of your parents you will get passed on an allele, a type of nucleotide pair. By counting the number of reference alleles you have (0, 1 or 2) your genome can be defined. About 10 million SNPs exist in the human genome, which are not equally divided along the sequence, but seem to cluster at certain locations on the sequence. This is due to the fact that the 'mistakes' in DNA will be passed on from parent to child, so SNPs tend to be correlated with SNPs in the same region of the genome. This is what we call linkage disequilibrium.
Considering SNPs are randomly assigned to an individual at conception– since you will get an allele from you father and one from your mother –, conditional on population stratification variables or family-specific effects and it has been proven that there are SNPs that among other things are correlated with birth weight, SNPs would be ideal to use as instruments in an instrumental variable analysis.

2.4 *Polygenic scores*

Considering that we have three billion base pairs of nucleotides and around 10 million SNPs within them, it is hard to imagine that a single SNP influences an outcome measure rather than a combination of small effects amongst many SNPs. Furthermore to be able to show that a single pair of nucleotides is really different from all others and to be marked as SNP it had to break the 1% barrier – nowadays even a 0.1% barrier is considered - which means that no more than 1% of a population can have the same nucleotide at a certain position in the DNA sequence to be classified as SNP. To be able to discover such a small effect, large sample sizes of more than ten thousand people are needed. And even if the nucleotides break the barrier of 1% and are detected in a so called genome-wide association study, we identified a variable that accounts for less than 1% of the variance of a DNA association. This is where polygenic scores come in handy.

A polygenic score can be composed by adding genotypic values across SNPs (Dudbridge, 2013; Wray et al., 2014; Plomin & Deary, 2015). For example, if in a pair of nucleotides, X, is found that the X_1 allele is associated with higher birth weight, then additive values can be assigned for X. Individuals with X_1X_1 alleles score value two, individuals with X_1X_2 alleles score value one and individuals with X_2X_2 alleles score value zero. Adding these scores found to the scores for alleles Y and Z we can establish a polygenic score that varies between zero and six. Instead of using just three alleles scores, we could use many thousands as well. Since the effect of allele X on birth weight for example might be higher than the effect of allele Y, we can refine the polygenic score by weighting the strength of the association and multiply the score of the allele with the proportion of its strength.

More recently, instead of calculating a polygenic score based on the data sample that is used, genome-wide polygenic scores are calculated including thousands of SNPs or even all SNPs on a DNA sequence weighted by strength of the association between the SNP and the outcome variable. The idea behind this is that more associations between SNPs and outcome variables will be detected.

A polygenic score of birth weight might therefore be an even better instrument since it will have a larger impact on the explanatory variable than a single SNP.

3 Methodology

In this thesis we will describe the effect of birth weight on educational performances from two different angles. First we would like to identify if birth weight has a causal impact on educational performances and if this impact differs across different topics taught at school; Mathematics, Science and English. Thereafter we will compare those results over time, to obtain, if any causal effects exist, if those effects weaken over time, since we expect 'nurture' to kick in and take over at least part of the effect of 'nature'.

Effects of birthweight on educational performances that have been found in the past, have been nullified later on by adding variables on social status and parental intelligence. Those findings have been only associations in the first place, since no causal impact has been found.

To be able to distinguish if the impact of birth weight is causal and not only an association, we will use instrumental variable analysis (IVA). The generic variants will be used as instrumental variables for birth weight in a Mendelian randomization. To be able to use this method some assumptions have to be met:

1. The instrumental variables are relevant by having an effect on the exposure.
2. The instrumental variables are independent by being uncorrelated with confounders of the exposure-outcome relationship.
3. The instrumental variables only influence the outcome through the exposure, which is also known as the exclusion restriction.

3.1 Relevance

The 58 variables included in the dataset containing information on single-nucleotide polymorphisms are not mentioned in the databases[1] to be related to birthweight. Also none of the p-values is smaller than 5×10^{-8} [2]when we look at the individual association with birthweight.

The polygenic risk score that is established does have a strong association with birthweight (p-value < 5×10^{-8}) which makes it a relevant instrument to use to determine if birthweight has a causal effect on educational performances.

Polygenic scores can be useful in the case that individual SNPs do not achieve significance in a large sample. A polygenic risk score can be established by taking the effect sizes from an independent genome-wide association study to weight associated alleles and then calculate the weighted sum of the associated alleles within each subject (Wray, et al., 2014).

[1] SNPedia (https://www.snpedia.com/index.php/SNPedia), the GWAS catalog (https://www.ebi.ac.uk/gwas/) and Ensembl (http://www.ensembl.org/index.html) have been checked until the 1st of August 2018 for updates on the 58 SNPs included in the dataset.

[2] Dudbridge and Gusnanto (2008) estimated that the threshold for a genome of three billion nucleotides should be around 5×10^{-8}.

3.2 Independence

Since genetic variants are randomly assigned to an individual at conception, genetic variants have a strong case regarding being independent, conditional on population stratification variables or family-specific effects. To my knowledge, I do not expect unmeasured confounders of genetic variants and educational performances.

3.3 Exclusion restriction

The exclusion restriction is often the hardest part to be proven in an instrumental variable regression. The same applies in our case, in which we cannot be complete sure that the polygenic score, which contains many SNPs, is only correlated with educational performances through birth weight, meaning we cannot prove that $\alpha = 0$. Some of the SNPs correlated with birth weight might by correlated with educational performances directly as well. The coefficients of our instrumental variable regression therefore will be likely to be biased.

3.4 Sensitivity analysis

Although we cannot overcome that $\alpha \neq 0$, we can perform a sensitivity analysis to get a feeling of the consequences this would have on the direct effect of birth weight on the educational performances (Conley et al., 2012). As Kippersluis and Rietveld (2017) did, we assume that the absolute value of the standardized first stage effect of the polygenic risk score on birthweight, will be larger than standardized direct effect of the polygenic risk score on educational performances, meaning that $0 \leq \delta \leq 1$. Thus we pick μ_α equal to $\delta\hat{\gamma}$ and the variance Ω_α equal to the squared standard error of $\hat{\gamma}$. By applying formula 3 of their paper (Kippersluis & Rietveld, 2017, p.5) we can obtain for what value of δ a causal effect can be found and from what values of δ onward the effect will diminish.

4 Data

4.1 ALSPAC

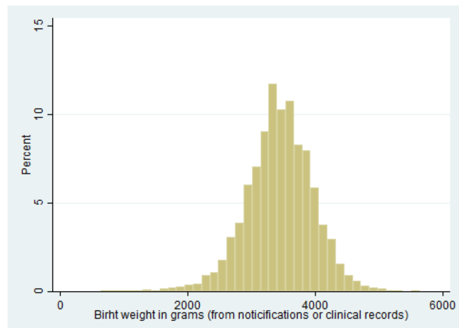
To perform an analyses, data from the Avon Longitudinal Study of Parents and Children (ALSPAC) will be used. The cohort includes all women that lived in a defined area in the South West of England with an expected date of delivery between April 1991 and December 1992 (Fraser et al., 2012).

4.2 Overview

The dataset used contains information on 8106 individuals defined by 195 variables. For 6363 individuals, birth weight and at least one educational performances measure is known.

4.3 Birthweight

In the dataset are five different variables included describing birth weight. We will use the variable with the most entries – from notifications or clinical records - measured in grams and refer to this variable as birth weight throughout the paper. We can obtain a measure on birth weight for 7,700 individuals with mean 3,440 grams and a standard deviation of 532 grams.



Graph 5. Distribution of birth weight

Distinguishing between males (n=3959) and females (n=3741), the average birth weight for males is 3,495 grams with a standard deviation of 557 grams and for females the average birth weight is 3,383 grams with a standard deviation of 498 grams.

In the specific case of twins, the boys (n=49) have an average birth weight of 2532 grams with a standard deviation of 631 grams and the girls (n=37) weigh 2493 grams on average at birth with a standard deviation of 382 grams.

4.4 Educational performance scores

A summary of the educational performance scores is listed in Table 1.

Variable	Obs	Mean	Std. dev.	Min.	Max.
Summary score @ year 2	6454	9.65	3.65	0	15
IQ score @ month 49	745	105.28	14.28	52	154
IQ score @ year 8	5438	104.83	16.41	45	151
IQ score @ year 15.5	3808	92.28	13.09	55	132
English marks @ year 4	6826	48.39	25.89	2	95
English marks @ year 9	5844	38.95	23.20	2	97
English attainment @ year 9	4876	33.71	8.67	0	45
Math marks @ year 4	6903	55.00	30.28	2	99
Math marks @ year 9	5774	52.21	34.72	2	99
Math attainment @ year 9	4876	37.78	10.16	0	51
Math final score @ year 9	4743	6.39	1.30	2.5	8.94
Science marks @ year 4	6904	49.47	24.11	2	79
Science attainment @ year 9	4876	35.15	8.69	0	45
Science final score @ year 9	4760	5.93	1.06	2.5	7.81

Table 1. Summary of educational performances.
Abbreviations: Obs, observations; Std. dev., standard deviaton; Min, minimum; Max, maximum.

Educational performances have been measured in different ways. An overall score has been measured at four different moments in time. A summary score at age two (n = 6454) has been established. At month 49 (n = 745) a Wechsler preschool and primary scale of intelligence score has been measured. And at eight years (n = 5438) and 15,5 years (n = 3808) of age an IQ score is obtained.

Furthermore, at ages four and nine, different scores have been measured on three topics, which are English, Mathematics and Science. When the respondents were four, a mark in English, Mathematics and Science has been obtained. At an age of nine, the respondents filled out what marks they got in English and Mathematics, what there attainment score was for all three the topics and which final score they obtained in Mathematics and Science.

4.5 Single nucleotide polymorphisms

The dataset contains 60 genetic variables. 58 of those variables are SNPs and the other two are polygenic scores, one for birth weight and one for educational attainment. All of those SNPs have been found to be associated with educational performances by Okbay et al. (2016). Other found associations in these SNPs are BMI, vitiligo, schizophrenia, height, epidermolysis bullosa dystrophica (Horikoshi et al., 2013; Lin et al., 2016; Cahtterji et al., 2014; Rietveld et al., 2013). No associations have been found of those SNPs and birth weight.

4.6 Education and qualification of the parents

37 variables give us insights in the education and qualification the parents of the individuals obtained. In Table 2 we listed the highest educational qualification obtained for both parents.

Mums highest ed qualification	Freq.	Percent	Cum.
cse	1,171	16.07	16.07
Vocational	655	8.99	25.05
O level	2,536	34.79	59.84
A level	1,807	24.79	84.63
Degree	1,120	15.37	100.00
Total	7,289	100.00	

Partners highest ed qualification	Freq.	Percent	Cum.
cse	1,552	21.95	21.95
Vocational	556	7.86	29.81
O level	1,542	21.81	51.62
A level	1,934	27.35	78.97
Degree	1,487	21.03	100.00
Total	7,071	100.00	

Table 2. Summary of the highest educational qualification obtained by the mother and father of the individual.

Abbreviations: ed, education; Freq., frequency; Cum., cumulative; cse, certificate of secondary education; O level, ordinary level; A level, advanced level.

4.7 Consistent subsample

As from Table 1 can be seen, there is a lot of variety in the number of observations for the different performance measures. Since we want to gain insights - apart from the direct effect of birth weight on educational performances - on the development of this effect over the years, we establish a second sample containing only those who filled out the questionnaires consistently to such an extent that we are able to use them in our analysis. By doing so, 1,495 individuals remain in the sample.

The average birthweight in the subsample is 3,449 grams with a standard deviation of 526 grams, which is quite close to the numbers obtained from the total sample. The males (n=699) weigh on average 3491 grams at birth with a standard deviation of 571 grams, whereas the females (n=796) bring 3411 grams on the scale at birth with a standard deviation of 480 grams.

In case of twins we obtain for the boys (n=9) an average birthweight of 2542 grams with a standard deviation of 713 grams and for the girls (n=9) 2607 grams with a standard deviation of 435 grams. Although we would expect boys to be heavier than girls, due to the small number of twins in the subsample these outcomes do not raise suspicion of wrong data entry.

In Table 3 a summary of the educational performances by the subsample can be found.

Variable	Obs	Mean	Std. dev.	Min.	Max.
Summary score @ year 2	1495	10.90	3.02	0	15
IQ score @ month 49	242	107.14	12.75	52	145
IQ score @ year 8	1495	105.44	15.46	60	145
IQ score @ year 15.5	1495	90.65	12.44	55	129
English marks @ year 4	1495	51.48	25.79	2	93
English marks @ year 9	1495	43.11	23.38	2	93
English attainment @ year 9	1495	36.70	5.25	21	45
Math marks @ year 4	1495	58.60	30.50	2	99
Math marks @ year 9	1495	51.91	36.01	2	99
Math attainment @ year 9	1495	41.22	6.64	15	51
Math final score @ year 9	1495	6.79	1.08	2.5	8.94
Science marks @ year 4	1495	52.71	23.91	2	79
Science attainment @ year 9	1495	38.13	5.54	15	45
Science final score @ year 9	1495	6.28	0.84	2.5	7.77

Table 3. Summary of educational performances in the subsample.
Abbreviations: Obs, observations; Std. dev., standard deviaton; Min, minimum; Max, maximum.

We indicated if the values obtained for the subsample are higher (blue) or lower (orange) than the values in the total sample. In most cases the average score is in the subsample higher than in the total sample, indicating that – the children of – those who filled out all the questionnaires score on average higher grades compared to those who dropped out of the sample, due to missing data.
Considering that there are viewer observations in the subsample, one could expect the standard deviations to be higher, but the opposite holds true in most cases. This implies that there are fewer outliers in the subsample and the scores are closer together compared to the total sample.
Table 4 consists of the highest educational qualification of the parents of the individuals. We can see that compared to the total sample there are relatively more mothers with at least an O level qualification in the subsample (80.49%) compared to the total sample (74.95%). The same holds true for the fathers, 73.17% in the subsample compared to 70.19% in the total sample, although this difference is smaller.

Mums highest ed qualification	Freq.	Percent	Cum.
cse	170	11.56	11.56
Vocational	117	7.95	19.51
O level	550	37.39	56.90
A level	412	28.01	84.91
Degree	222	15.09	100.00
Total	1,471	100.00	

Partners highest ed qualification	Freq.	Percent	Cum.
cse	255	17.77	17.77
Vocational	130	9.06	26.83
O level	323	22.51	49.34
A level	433	30.17	79.51
Degree	294	20.49	100.00
Total	1,435	100.00	

Table 4. Summary of the highest educational qualification obtained by the mother and father of the individual.
Abbreviations: ed, education; Freq., frequency; Cum., cumulative; cse, certificate of secondary education; O level, ordinary level; A level, advanced level.

5 Results

5.1 Plain OLS

At first glance we obtain a positive significant association of birth weight and overall performance indicators in Table 5. This positive association seems to be driven by the more exact subjects, like Mathematics and Science.

Effect of Birthweight	n	β^a	95% CI	t-value	P > |t|
Summary score @ year 2	6,129	0.0004	0.0003, 0.0006	5.07	0.000
IQ score @ month 49	744	0.0026	0.0005, 0.0047	2.39	0.017
IQ score @ year 8	5,187	0.0026	0.0017, 0.0035	5.91	0.000
IQ score @ year 15.5	3,600	0.0011	0.0003, 0.0020	2.66	0.008
English marks @ year 4	6,478	0.0010	-0.0002, 0.0021	1.60	0.110
English marks @ year 9	5,550	0.0013	0.0002, 0.0025	2.29	0.022
English attainment @ year 9	4,698	0.0005	-0.0001, 0.0010	1.75	0.080
Math marks @ year 4	6,458	0.0030	0.0016, 0.0044	4.14	0.000
Math marks @ year 9	5,480	-0.0009	-0.0026, 0.0008	-1.03	0.304
Math attainment @ year 9	4,698	0.0010	0.0004, 0.0016	3.36	0.001
Math final score @ year 9	4,568	0.0002	0.0001, 0.0002	4.45	0.000
Science marks @ year 4	6,553	0.0017	0.0006, 0.0028	2.96	0.003
Science attainment @ year 9	4,698	0.0005	0.0000, 0.0010	2.11	0.035
Science final score @ year 9	4,585	0.0001	0.0000, 0.0002	3.01	0.003

Table 5. The effect of Birthweight on performance indicators using OLS regression.
[a] β is the mean difference in the respective score.
Abbreviations: n, number of subjects; CI, confidence interval.

For the IQ score at an age of eight years we obtain a value for a beta of 0.0026, meaning that if the birth weight of an individual increases by one gram on average the IQ score of this eight year old individual increases by 0.0026 points. So a kid at the 75th percentile of birthweight would have on average an IQ score of 88.21 (the value of the constant) + (0.0026 * 3780) = 98.04. For comparison a child at the 25th percentile of birthweight would on average score 88.21 + (0.0026 * 3140) = 96.37, which is almost two points lower.

5.2 Adjusted OLS

Since we assume that educational performances are not only driven by birth weight we will adjust for sex, smoking behavior of the mother, marital status, parental education and if the individual was a singleton or part of multiples to obtain a more precise association of birth weight and school performances.

Effect of Birthweight	n	β[a]	95% CI	t-value	P > |t|
Summary score @ year 2	5,565	0.0004	0.0002, 0.0006	4.28	0.000
IQ score @ month 49	696	0.0027	0.0007, 0.0047	2.67	0.008
IQ score @ year 8	4,827	0.0016	0.0008, 0.0025	3.85	0.000
IQ score @ year 15.5	3,357	0.0007	-0.0001, 0.0016	1.74	0.082
English marks @ year 4	5,902	0.0007	-0.0006, 0.0019	1.05	0.295
English marks @ year 9	5,070	0.0012	0.0001, 0.0024	2.08	0.038
English attainment @ year 9	4,285	0.0003	-0.0002, 0.0008	1.28	0.199
Math marks @ year 4	5,883	0.0025	0.0010, 0.0040	3.29	0.001
Math marks @ year 9	4,981	-0.0008	-0.0027, 0.0010	-0.88	0.379
Math attainment @ year 9	4,285	0.0005	-0.0001, 0.0011	1.76	0.079
Math final score @ year 9	4,170	0.0001	0.0000, 0.0002	3.01	0.003
Science marks @ year 4	5,962	0.0012	-0.0000, 0.0023	1.94	0.053
Science attainment @ year 9	4,285	0.0002	-0.0003, 0.0007	0.74	0.457
Science final score @ year 9	4,191	0.0000	-0.0000, 0.0001	1.50	0.133

Table 6. The effect of Birthweight on overall performance indicators adjusted for sex, smoking behavior of the mother, marital status, parental education and if the individual was a singleton or part of multiples using OLS regression. [a] β is the mean difference in the respective score.
Abbreviations: n, number of subjects; CI, confidence interval.

We can see that the effect of birth weight on educational performances becomes less significant if we correct for other factors, which is in line with the assumption that school performances are influenced by other factors next to birth weight as well. The higher coefficient of beta for the IQ score at the 49th month can be explained by the fact that some individuals did not fill out all of the questionnaire and therefore dropped out of the sample at this stage in the analyses. These were probably individuals who obtained lower scores on average.

Making again a comparison between an individual – female, married mother, no cigarettes smoked by the mother at the eight week of her pregnancy, singleton and both parents have no educational qualification at the 25th and 75th percentile of birthweight and their IQ scores at an age of eight years we obtain respectively;

94.04 (the value of the constant) + (0.0017 * 3140) + (1 * 0.1642) + (1 * -0.3036) + (1 * -0.1886) = **99.05**

94.04 + (0.0017 * 3780) + (1 * 0.1642) + (1 * -0.3036) + (1 * -0.1886) = **100.14**.

Although this difference is smaller, due to the adjusting social economic factors, we still observe a significant difference at a 5% level.

5.3 *Two-stage least squares*

Using a standard two-stage least squares regression with the polygenic risk score of birth weight as instrument for birth weight we obtain the following results listed in Table 7. But as mentioned earlier, the exclusion restriction is likely to be violated, since we cannot proof that α = 0.

Effect of Birthweight	n	β[a]	95% CI	z-value	P > |z|
Summary score @ year 2	6,129	0.0011	0.0003, 0.0019	2.63	0.008
IQ score @ month 49	744	0.0025	-0.0089, 0.0138	0.43	0.667
IQ score @ year 8	5,187	0.0040	0.0007, 0.0073	2.36	0.018
IQ score @ year 15.5	3,600	0.0043	0.0014, 0.0072	2.90	0.004
English marks @ year 4	6,478	0.0030	-0.0024, 0.0085	1.08	0.279
English marks @ year 9	5,550	0.0041	-0.0011, 0.0094	1.56	0.119
English attainment @ year 9	4,698	0.0039	0.0019, 0.0059	3.82	0.000
Math marks @ year 4	6,458	0.0089	0.0022, 0.0155	2.62	0.009
Math marks @ year 9	5,480	-0.0063	-0.0144, 0.0017	-1.54	0.122
Math attainment @ year 9	4,698	0.0043	0.0019, 0.0066	3.51	0.000
Math final score @ year 9	4,568	0.0004	0.0001, 0.0007	2.42	0.015
Science marks @ year 4	6,553	0.0014	-0.0040, 0.0066	0.50	0.616
Science attainment @ year 9	4,698	0.0030	0.0010, 0.0050	2.98	0.003
Science final score @ year 9	4,585	0.0003	0.0000, 0.0005	2.06	0.040

Table 7. The effect of Birthweight on overall performance indicators using 2SLS regression with the polygenic risk score of birth weight as instrument.

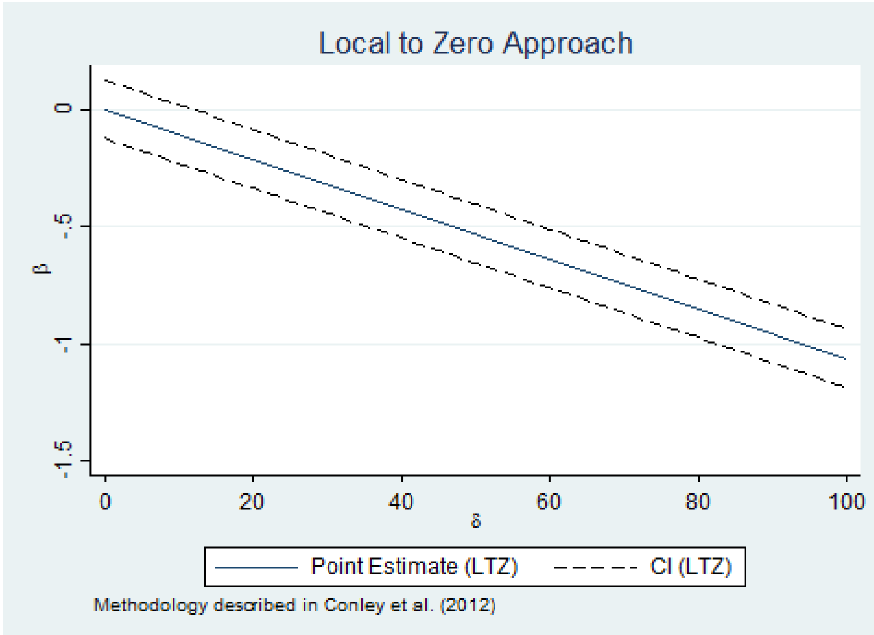
[a] β is the mean difference in the respective score.

Abbreviations: n, number of subjects; CI, confidence interval.

Studying the results, we notice that some of the coefficients are blown up compared to the coefficients of the OLS estimate. So our previous caveat that our instrument violates the exclusion restriction is likely to be true, since we would expect the causal effect of birth weight on educational performances to be smaller than the association measured by the OLS.

5.4 *Sensitivity analysis*

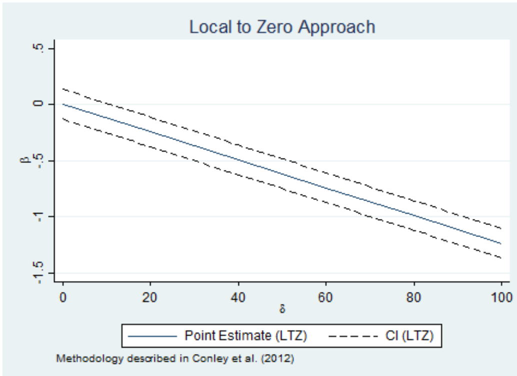
Since we expect the exclusion restriction to be violated, we will use a sensitivity analysis to get a feeling for which consequences this would have for the causal effect of birth weight on the performance measures. The values of β are plotted in graphs 6 to 14 for the different outcome measures that were found significant at a 5% level in the two-stage least squares regressions as function of δ. In case the exclusion restriction would hold, δ = 0, we obtain the same value as for the two-stage least squares estimate.



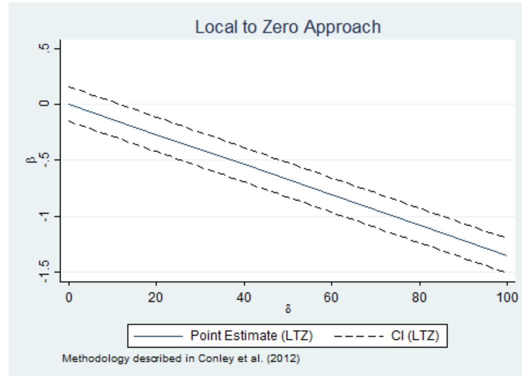
Graph 6. The line shows the point estimate of the direct effect of birth weight on the summary score of an individual at an age of two years. β is the coefficient of the effect, if β would be 1, for every gram of extra birth weight an individual would score on average one point higher on his Summary score at an age of two years. δ indicates by how much the exclusion restriction is harmed. A larger δ means a larger direct effect of the polygenic risk score of birthweight on the performance indicator. The same holds true for graphs 7 through 14, so only the independent variable will be stated.

We can obtain from graph 6 that even with the slightest deviation of δ from 0 the effect of birth weight on the summary score at an age of two will be nullified. So in case the exclusion restriction does not hold, it is not likely that there exists an causal impact of birth weight on this performance indicator.

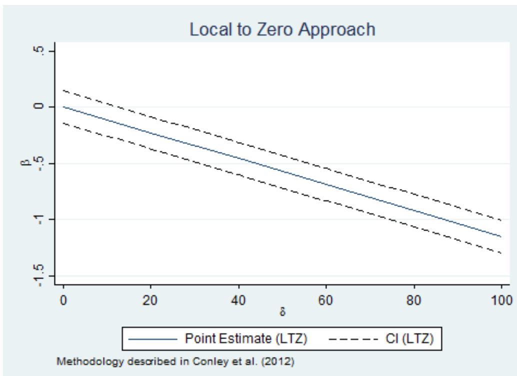
In addition to the exclusion restriction assuming α to be 0, the standard deviation of α is also assumed to be 0. In the graph we see two dotted lines indicating the confidence interval around δ. The fact that the lower bound of the confidence interval at δ = 0 already has a negative value for β indicates that there will be no causal effect of birth weight on the summary score at an age of two years even if on average the exclusion restriction would hold, but not if this would be only the case in 95% of the cases.



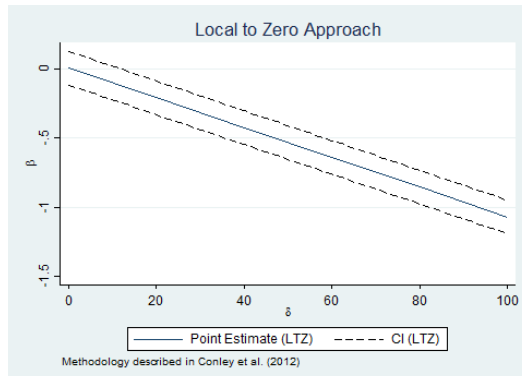
Graph 7. IQ score @ year 8



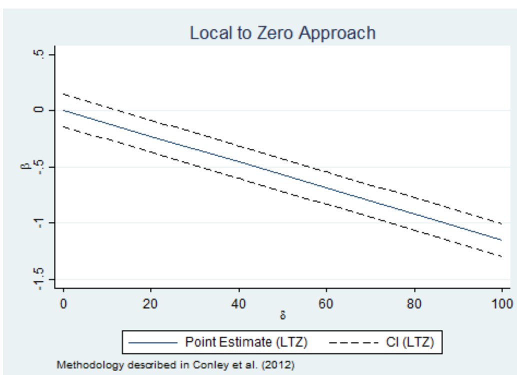
Graph 8. IQ score @ year 15



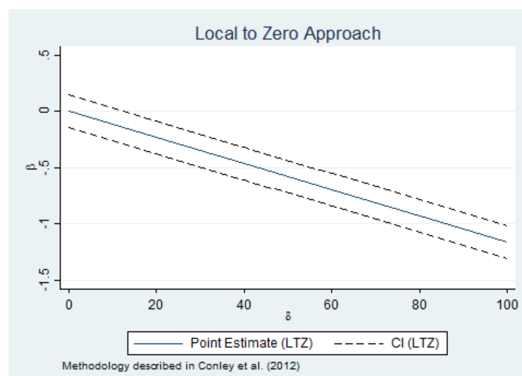
Graph 9. English attainment score @ year 9



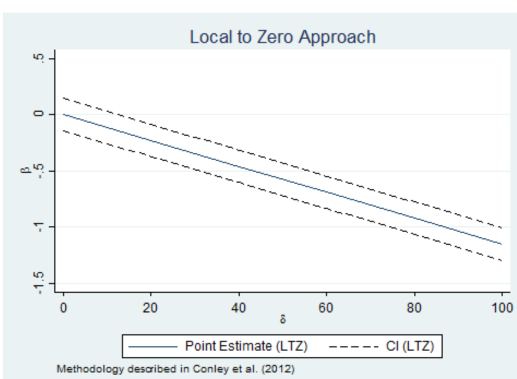
Graph 10. Maths marks @ year 4



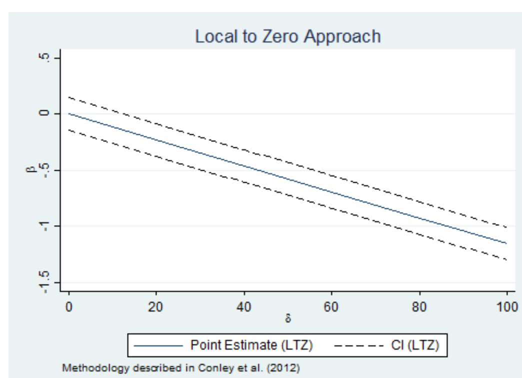
Graph 11. Maths attainment score @ year 9



Graph 12. Maths final score @ year 9



Graph 13. Science attainment score @ year 9



Graph 9. Science final score @ year 9

5.5 Comparison over time

From this point on, analyses will be performed on the sub sample. To be able to not only compare the relationship of birthweight and the educational performance measures over time, but also be able to compare the magnitude of the relationship across the different topics, we standardized the educational performance variables. From this perspective we can obtain which outcome measure gets influenced the most by a change in birth weight compared to the other outcome measures.

5.6 Plain OLS

We can only obtain a significant association at a 5% level of Birthweight with two performance indicators, English and Mathematics marks obtained at age four. Even if we are aware of the fact that the values will be biased, it is odd that the association of birth weight and English marks at the age of four are negative in this subsample.

Effect of Birthweight	n	β^a	95% CI	t-value	P > |t|
English marks @ year 4	1,495	-0.0001	-0.0002, -0.0000	-2.12	0.034
Math marks @ year 4	1,495	0.0001	0.0000, 0.0002	2.40	0.016

Table 8. The effect of Birthweight on performance indicators using OLS regression.
[a] β is the mean difference in the respective score.
Abbreviations: n, number of subjects; CI, confidence interval.

5.7 Adjusted OLS

If we again adjust for sex, smoking behavior of the mother, marital status, parental education and if the individual was a singleton or part of multiples to obtain a more precis association of birth weight and school performances, we obtain that the negative association of English marks at four year of age and birth weight no longer exists. Furthermore, we obtain that next to Maths marks at year four, the overall performance measures at an age of two years, 49 months and eight years are also positive significantly associated with birth weight at a 5% level.

Effect of Birthweight	n	β^a	95% CI	t-value	P > |t|
Summary score @ year 2	1,495	0.0001	0.0000, 0.0002	1.98	0.048
IQ score @ month 49	242	0.0003	0.0001, 0.0005	3.08	0.002
IQ score @ year 8	1,495	0.0001	0.0001, 0.0002	2.58	0.010
Math marks @ year 4	1,495	0.0002	0.0001, 0.0002	3.12	0.002

Table 9. The effect of Birthweight on overall performance indicators adjusted for sex, smoking behavior of the mother, marital status, parental education and if the individual was a singleton or part of multiples using OLS regression. [a] β is the mean difference in the respective score.
Abbreviations: n, number of subjects; CI, confidence interval.

5.8 Two-stage least squares

By trying to identify how much of the effect of birth weight on the performance indicators is not only an association but causal, we use a two-stage least squares regression with the polygenic risk score of birth weight as instrument for birth weight. We can obtain no significant effect of birth weight on the performance indicators using the polygenic risk score of birth weight as instrumental variable for birth weight.

5.9 Sensitivity analysis

Since no significant causal effect of birth weight on the educational performance indicators has been measured by the two-stage least square method, it does not make sense to perform a sensitivity analysis to obtain for what direct impact of the polygenic risk score on to the performance indicators an impact will hold.

6 Conclusion

Since we were not able to pick a valid instrument from our data, we could not determine with instrumental variable analysis what the causal impact of birth weight on educational performances is. By performing a sensitivity analysis, we gained insights in the size of the effect a violation of the exclusion restriction would have on the causal effect of birth weight on educational performances. The association we could obtain from the OLS model and the 'false' instrumental variable regression will be diversified away with even the slightest deviation of the exclusion restriction. So we conclude from our data that birth weight is rather not likely to have a causal impact on educational performances, since a small violation of the exclusion restriction would neglect the significant results we obtained from the IV regression. Although SNPs and polygenic risk scores are assumed to be valid and strong instruments in most cases, we cannot be sure that the exclusion restriction is exactly satisfied. We could identify to what 'imperfection' a significant effect will hold and give a brief insight in the associations birth weight has on different school topics, and how these associations change over time.

These findings are in line with Lin (2016) who used SNP as instruments to determine if birthweight had a causal impact on educational attainment, which was approximated by years of schooling and a binary variable for college completion.

7 Discussion

Since we were not able to identify a single SNP with a strong relationship with the outcome variables, we had to use the polygenic risk score as an instrument, which is likely to be directly related to the outcome measure. So we cannot rule out fully that there is no effect of birth weight on educational performances, since there might exist a SNP that does fulfil the exclusion restriction and is strongly related with educational performances.

In this case, the data collected by Early Growth Genetics Consortium are used to establish the polygenic score for birth weight. Since this is not an independent GWAS there exists a slight chance of overfitting since the data used in this thesis are part of the database that is used to establish the polygenic score.

Although we expect SNPs to be random, studies have shown that population-specific variations in alleles exist. Using principal components normalization of this so called population drift would be applied.

Picking values for Ω_α in the sensitivity analyses is arbitrary and directly influences the value of δ. Following the paper of Kippersluis et al. (2013) we set Ω_α equal to the standard error of $\hat{\gamma}$ in our analyses, but there is no theory to back this value completely.

8 Literature

Asher, C., & Roberts, J.A.F. (1946). A Study on Birthweight and Intelligence. *British Journal of Social Medicine, 3(2): 56 – 68.*

Black, S.E., Devereux, P.G., & Salvanes, K.G. (2005). The more the merrier? The effect of family size and birth order on children's education. *Quarterly Journal of Economics, v120, 669-700.* https://doi.org/10.3386/w10720.

Black, S.E., Devereux, P.G., & Salvanes, K.G. (2007). From the cradle to the labor market? The effect of birth weight on adult outcomes. *Quarterly Journal of Economics, 122(1):409 – 439.* http://www.nber.org/papers/w11796.pdf.

Boomsma, D.I., Van Beijsterveldt, C.E., Rietveld, M.J., Bartels, M., & Van Baal, G.C. (2001). Genetics mediate relation of birth weight to childhood IQ. *British Medical Journal, 323(7326):1426–1427*. https://doi.org/10.1136/bmj.323.7326.1426a. Chatterji P, Kim D, & Lahiri K. (2014) Birth weight and academic achievement in childhood. *Health Economics, 23(9):1013–1035.* https://doi.org/ 10.1002/hec.3074

Bowden, J., Davey Smith, G., & Burgess, S. (2015). Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *International Journal of Epidemiology, 44:512-525.* https://doi.org/10.1093/ije/dyv080.

Christensen, K., Petersen, I., Skytthe, A., Herskind, A.M., McGue, M., & Bingley, P. (2006). Comparison of academic performance of twins and singletons in adolescence: follow-up study. *British Medical Journal 333(7578): 1095–1097*. https://doi.org/10.1136/bmj.38959.650903.7C.

Dudbrigde, F., & Gusnanto, A. (2008). Estimation of significance thresholds for genomewide association scans. *Genetic Epidemiology, 32(3):227 – 34.* https://doi.org/10.1002/gepi.20297.

Dudbridge, F. (2013). Power and Predictive Accuracy of Polygenic Risk Scores. *PLOS genetics, 9(4):10.* https://doi.org/10.1371/journal.pgen.1003348.

Fletcher, J.M. (2011). The medium term schooling and health effects of low birth weight: evidence from siblings. *Economcis of Education Review, 30(3):517–527.* https://doi.org/10.1016/j.econedurev.2010.12.012.

Fraser, H.B., Levy, S., Chavan, A., Shah, H.B., Perez, J.C., Zhou, Y., Siegal, M.L., & Sinha, H. (2012). Polygenic cis-regulatory adaptation in the evolution of yeast pathogenicity. *Genome Research, 22(10):1930 – 1939*. https://doi.org/10.1101/gr.134080.11.

Greenland, S. (2000). An introduction to instrumental variables for epidemiologists. *International Journal of Epidemiology, 29:722-729.* https://doi.org/10.1093/ije/29.4.722.

Horikoshi M., YaghootkarH., Mook-Kanamori, D.O., Sovio, U., Taal, H.R., Hennig, B.J., Bradfield, J.P., St Pourcain, B., & et al. (2014). New loci associated with birth weight identify genetic links between intrauterine growth and adult height and metabolism. *Nature Genetics, 45(1):76–82*. https://doi.org/10.1038/ng.2477.

Van Kippersluis, H., & Rietveld, C.A. (2017). Pleiotropy-robust Mendelian randomization. *International Journal of Epidemiology*. https://doi.org/10.1093/ije/dyx002.

Kirkegaard, I., Obel, C., Hedegaard, M., & Brink Henriksen, T. (2006). Gestational age and birth weight in relation to school performance of 10-year-old children: a follow-up study of children born after 32 completed weeks. *Pediatrics,118(4):1600–1606.* https://doi.org/10.1542/peds.2005-2700.

Lawlor D.A., Ebrahim S., Davey Smith G. (2005). Association of birth weight with adult lung function: finding from the British Women's Heart and Health Study and a meta-analysis. *Thorax. 2005; 60: 851–858.* https://doi.org/10.1136/thx.2005.042408.

Lawlor, D.A., Clark, H., Smith, G.D., & Leon, D.A. (2006). Intrauterine growth and intelligence within sibling pairs: findings from the Aberdeen children of the 1950s cohort. *Pediatrics, 117(5):e894–e902.* https://doi.org/10.1542/peds.2005-2412

Lin, S.L., Leung, G.M., & Schooling, C.M. (2017). The effect of birth weight on academic performance: instrumental variable analysis. *American Journal of Epidemiology, 1;185(9):853-859.* https://doi.org/10.1093/aje/kwx034.

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

Newcombe, R., Milne, B.J., Caspi, A., Poulton, R., & Moffitt, T.E. (2007). Birthweight predicts IQ: fact or artefact. *Twin Research and Human Genetics, 10(4):581–586.* https://doi.org/10.1375/twin.10.4.581.

Okbay, A., Beauchamp, J.P., Fontana, M.A., Lee, J.J., Pers, T.H., Rietveld, C.A., Turley, P., & et al. Genome-wide association study identifies 74 loci associated with educational attainment. *Nature, 533: 539–542.* http://dx.doi.org/10.1038/nature17671.

Plomin, R., & Deary, I.J. (2015). Genetics and intelligence differences: five special findings. *Molecular Psychiatry, 20(1):98-108.* https://doi.org/10.1038/mp.2014.105.

Record, R.G., McKeown, T., & Edwards, J.H. (1969). The relation of measured intelligence to birth weight and duration of gestation. *Annals of Human Genetics, 33(1):71–79.* https://doi.org/10.1111/j.1469-1809.1969.tb01630.x.

Reiersøl, O. (1945). *Confluence analysis by means of instrumental sets of variables*. Arkiv for Mathematik, Astronomie och Fysik, 32.

Rietveld, C.A., Medland, S.E., Derringer, J., Yang, J., Esko, T., Martin, N.W., Westra, H., Shakhbazo, K., & et al. (2013). GWAS of 126,559 Individuals Identifies Genetic Variants Associated with Educational Attainment. *Science, 340(6139):1467 – 1471.* https://doi.org/10.1126/science.1235488.

Shenkin, S.D., Starr, J.M., & Deary, I.J. (2004). Birth weight and cognitive ability in childhood: a systematic review. *Psychol Bull, 130(6):989–1013.* https://doi.org/10.1037/0033-2909.130.6.989.

Silvestrin, S., Silva C.H., Hirakata V.N., Goldani A.A., Silveira P.P., & Goldani M.Z. (2013). Maternal education level and low birth weight: a meta-analysis. *Jornal de Pediatria (Versão em Português), Volume 89, Issue 4, Pages 339-345.* https://doi.org/10.1016/j.jped.2013.01.003.

Tsou, M.T., Tsou, M.W., Wu, M.P, & Liu, J.T. (2008). Academic achievement of twins and singletons in early adulthood Taiwanese cohort study. *British Medical Journal, 337:a438.* https://doi.org/10.1136/bmj.a438.

Wray, N.R., Lee, S.H., Mehta, D., Vinkhuyzen, A.A., Dudbridge, F., & Middeldorp, C.M. (2014). Research review: Polygenic methods and their application to psychiatric traits. *Journal of Child Psychology and Psychiatry, 55(10):1068-1087.* https://doi.org/10.1111/jcpp.12295.

Wright, P.G. (1928). *The Tariff on Animal and Vegetable Oils*. New York: Macmillan.

Yu, Y. Yoon, S. Poulogiannis, G. Yang, Q. Max Ma, X. Villén, J. Kubica, N. Hoffman, G.R., Lewis, C.C., & Gygi, S.P. (2011) Quantitative Phosphoproteomic Analysis Identifies the Adaptor Protein Grb10 as an mTORC1 Substrate that Negatively Regulates Insulin Signaling. *Science 332(6035): 1322–1326*. https://doi.org/10.1126/science.1199484.

9 Do file

```
clear all
use "\\campus.eur.nl\users\home\347322et\Desktop\Alspac.dta", clear
log using "C:\Users\Erik_\Documents\Master\Thesis\Master thesis.log", replace
set more off
*having a look at the variables included in the dataset:
des
destring k2_tote, generate(k2_tote_destring) ignore(`"-10"', illegal) force
destring k2_totm, generate(k2_totm_destring) ignore(`"-10"', illegal) force
destring k2_tots, generate(k2_tots_destring) ignore(`"-10"', illegal) force
destring k3_tote, generate(k3_tote_destring) ignore(`"-10"', illegal) force
destring k3_totm, generate(k3_totm_destring) ignore(`"-10"', illegal) force
recode kz021 (2= 0)
*Label Male=1, Female=0.
label define KZ021 0 "Female" 1 "Male", replace
rename kz021 Male
*Have a look at the different variables for birthweight.
sum kz030 kz030b kz030c kz030d
*Variable kz030d has the most entries and seems to be the best fit.
rename kz030d Birthweight
rename kz032c Birth_length
rename cf813 IQ_month49
rename cf058 Height_month49
rename cf048 Weight_month49
rename f8ws112 IQ_year8
rename f8lf020 Height_year8
rename f8lf021 Weight_year8
rename fh3000 Height_year15,5
rename fh3010 Weight_year15,5
*Have a look at the different variables for IQ at the age of 15,5 years.
sum fh6280 fh6281, detail
*There is no difference between the variables, so we can use either of them.
rename fh6280 IQ_year15,5
rename sat190a Score_year2
rename k2_tote_destring Englishmarks_year4
rename k2_totm_destring Mathsmarks_year4
rename k2_tots_destring Sciencemarks_year4
rename k3_tote_destring Englishmarks_year9
rename k3_epts Englishattainment_year9
rename k3_totm_destring Mathsmarks_year9
rename k3_mpts Mathsattainment_year9
rename k3_mfine Mathsfinal_year9
rename k3_spts Scienceattainment_year9
rename k3_sfine Sciencefinal_year9
recode mz010a (1 = 0)
recode mz010a (2 = 1)
label define MZ010A 0 "Singleton" 1 "Multiple", replace
rename mz010a Twins
*All those who died before the age of 1, will/should have no entries for educational performance scores.
drop if mz014 == 1
rename a200 Cigs_week8
gen Smoker = Cigs_week8
replace Smoker =1 if Cigs_week8>0
rename a214 Change_cigs
rename b023 Mum_age_1st
gen Married = a525
replace Married = 0 if a525 < 5
replace Married = 1 if a525 > 4
rename dw032 Mum_birthweight
rename dw042 Mum_BMI_week12
```

```
rename c645a Mum_educ_highest
rename c666a Dad_educ_highest
label define K6280 0 "No", modify
rename k6280 Mum_educ_non
recode Mum_educ_non (missing = 0)
label define K6281 0 "No", modify
rename k6281 Mum_educ_CSE
recode Mum_educ_CSE (missing = 0)
label define K6282 0 "No", modify
rename k6282 Mum_educ_Olvl
recode Mum_educ_Olvl (missing = 0)
label define K6283 0 "No", modify
rename k6283 Mum_educ_Alvl
recode Mum_educ_Alvl (missing = 0)
label define K6284 0 "No", modify
rename k6284 Mum_educ_voc
recode Mum_educ_voc (missing = 0)
label define K6292 0 "No", modify
rename k6292 Mum_educ_uni
recode Mum_educ_uni (missing = 0)
label define K6300 0 "No", modify
rename k6300 Dad_educ_non
recode Dad_educ_non (missing = 0)
label define K6301 0 "No", modify
rename k6301 Dad_educ_CSE
recode Dad_educ_CSE (missing = 0)
label define K6302 0 "No", modify
rename k6302 Dad_educ_Olvl
recode Dad_educ_Olvl (missing = 0)
label define K6303 0 "No", modify
rename k6303 Dad_educ_Alvl
recode Dad_educ_Alvl (missing = 0)
label define K6304 0 "No", modify
rename k6304 Dad_educ_voc
recode Dad_educ_voc (missing = 0)
label define K6312 0 "No", modify
rename k6312 Dad_educ_uni
recode Dad_educ_uni (missing = 0)
rename bestgest Pregnancy_length

*Descriptive statistics
sum Birthweight, detail
histogram Birthweight, percent xtitle(Birht weight in grams (from noticifications or clinical records))
sum Birthweight if Male == 1
sum Birthweight if Male == 0
sum Birthweight if Male == 1 & Twins == 1
sum Birthweight if Male == 0 & Twins == 1

tab Married
tab Smoker
tab Mum_educ_highest
tab Dad_educ_highest

sum Score_year2 IQ_month49 IQ_year8 IQ_year15
sum Englishmarks_year4 Englishmarks_year9 Englishattainment_year9
sum Mathsmarks_year4 Mathsmarks_year9 Mathsattainment_year9 Mathsfinal_year9
sum Sciencemarks_year4 Scienceattainment_year9 Sciencefinal_year9
sum Birth_length Height_month49 Height_year8 Height_year15
sum Birthweight Weight_month49 Weight_year8 Weight_year15
sum Cigs_week8 Pregnancy_length
```

```
corr Score_year2 IQ_month49 IQ_year8 IQ_year15 Birthweight
corr Englishmarks_year4 Englishmarks_year9 Englishattainment_year9 Birthweight
corr Mathsmarks_year4 Mathsmarks_year9 Mathsattainment_year9 Mathsfinal_year9 Birthweight
corr Sciencemarks_year4 Scienceattainment_year9 Sciencefinal_year9 Birthweight

reg Score_year2 Birthweight, robust
reg IQ_month49 Birthweigh, robust
reg IQ_year8 Birthweight, robust
reg IQ_year15 Birthweight, robust

reg Englishmarks_year4 Birthweight, robust
reg Englishmarks_year9 Birthweight, robust
reg Englishattainment_year9 Birthweight, robust

reg Mathsmarks_year4 Birthweight, robust
reg Mathsmarks_year9 Birthweight, robust
reg Mathsattainment_year9 Birthweight, robust
reg Mathsfinal_year9 Birthweight, robust

reg Sciencemarks_year4 Birthweight, robust
reg Scienceattainment_year9 Birthweight, robust
reg Sciencefinal_year9 Birthweight, robust

reg Score_year2 Birthweight Male Cigs_week8 Married Twins Mum_educ_non Mum_educ_CSE Mum_educ_Olvl
Mum_educ_Alvl Mum_educ_voc Mum_educ_uni Dad_educ_non Dad_educ_CSE Dad_educ_Olvl Dad_educ_Alvl Dad_educ_voc
Dad_educ_uni, robust
reg IQ_month49 Birthweigh Male Cigs_week8 Married Twins Mum_educ_non Mum_educ_CSE Mum_educ_Olvl
Mum_educ_Alvl Mum_educ_voc Mum_educ_uni Dad_educ_non Dad_educ_CSE Dad_educ_Olvl Dad_educ_Alvl Dad_educ_voc
Dad_educ_uni, robust
reg IQ_year8 Birthweight Male Cigs_week8 Married Twins Mum_educ_non Mum_educ_CSE Mum_educ_Olvl Mum_educ_Alvl
Mum_educ_voc Mum_educ_uni Dad_educ_non Dad_educ_CSE Dad_educ_Olvl Dad_educ_Alvl Dad_educ_voc Dad_educ_uni,
robust
reg IQ_year15 Birthweight Male Cigs_week8 Married Twins Mum_educ_non Mum_educ_CSE Mum_educ_Olvl Mum_educ_Alvl
Mum_educ_voc Mum_educ_uni Dad_educ_non Dad_educ_CSE Dad_educ_Olvl Dad_educ_Alvl Dad_educ_voc Dad_educ_uni,
robust

reg Englishmarks_year4 Birthweight Male Cigs_week8 Married Twins Mum_educ_non Mum_educ_CSE Mum_educ_Olvl
Mum_educ_Alvl Mum_educ_voc Mum_educ_uni Dad_educ_non Dad_educ_CSE Dad_educ_Olvl Dad_educ_Alvl Dad_educ_voc
Dad_educ_uni, robust
reg Englishmarks_year9 Birthweight Male Cigs_week8 Married Twins Mum_educ_non Mum_educ_CSE Mum_educ_Olvl
Mum_educ_Alvl Mum_educ_voc Mum_educ_uni Dad_educ_non Dad_educ_CSE Dad_educ_Olvl Dad_educ_Alvl Dad_educ_voc
Dad_educ_uni, robust
reg Englishattainment_year9 Birthweight Male Cigs_week8 Married Twins Mum_educ_non Mum_educ_CSE Mum_educ_Olvl
Mum_educ_Alvl Mum_educ_voc Mum_educ_uni Dad_educ_non Dad_educ_CSE Dad_educ_Olvl Dad_educ_Alvl Dad_educ_voc
Dad_educ_uni, robust

reg Mathsmarks_year4 Birthweight Male Cigs_week8 Married Twins Mum_educ_non Mum_educ_CSE Mum_educ_Olvl
Mum_educ_Alvl Mum_educ_voc Mum_educ_uni Dad_educ_non Dad_educ_CSE Dad_educ_Olvl Dad_educ_Alvl Dad_educ_voc
Dad_educ_uni, robust
reg Mathsmarks_year9 Birthweight Male Cigs_week8 Married Twins Mum_educ_non Mum_educ_CSE Mum_educ_Olvl
Mum_educ_Alvl Mum_educ_voc Mum_educ_uni Dad_educ_non Dad_educ_CSE Dad_educ_Olvl Dad_educ_Alvl Dad_educ_voc
Dad_educ_uni, robust
reg Mathsattainment_year9 Birthweight Male Cigs_week8 Married Twins Mum_educ_non Mum_educ_CSE Mum_educ_Olvl
Mum_educ_Alvl Mum_educ_voc Mum_educ_uni Dad_educ_non Dad_educ_CSE Dad_educ_Olvl Dad_educ_Alvl Dad_educ_voc
Dad_educ_uni, robust
reg Mathsfinal_year9 Birthweight Male Cigs_week8 Married Twins Mum_educ_non Mum_educ_CSE Mum_educ_Olvl
Mum_educ_Alvl Mum_educ_voc Mum_educ_uni Dad_educ_non  Dad_educ_CSE Dad_educ_Olvl Dad_educ_Alvl Dad_educ_voc
Dad_educ_uni, robust
```

reg Sciencemarks_year4 Birthweight Male Cigs_week8 Married Twins Mum_educ_non Mum_educ_CSE Mum_educ_Olvl Mum_educ_Alvl Mum_educ_voc Mum_educ_uni Dad_educ_non Dad_educ_CSE Dad_educ_Olvl Dad_educ_Alvl Dad_educ_voc Dad_educ_uni, robust
reg Scienceattainment_year9 Birthweight Male Cigs_week8 Married Twins Mum_educ_non Mum_educ_CSE Mum_educ_Olvl Mum_educ_Alvl Mum_educ_voc Mum_educ_uni Dad_educ_non Dad_educ_CSE Dad_educ_Olvl Dad_educ_Alvl Dad_educ_voc Dad_educ_uni, robust
reg Sciencefinal_year9 Birthweight Male Cigs_week8 Married Twins Mum_educ_non Mum_educ_CSE Mum_educ_Olvl Mum_educ_Alvl Mum_educ_voc Mum_educ_uni Dad_educ_non Dad_educ_CSE Dad_educ_Olvl Dad_educ_Alvl Dad_educ_voc Dad_educ_uni, robust

*Having a look at the individual SNPs
reg Birthweight rs11191193_g rs113520408_a rs11588857_a rs11712056_c rs12531458_c rs12987662_a rs13294439_c rs148734725_a rs1606974_a rs16845580_c rs17167170_g rs17824247_c rs192818565_g rs301800_t rs34305371_a rs4493682_c rs4500960_t rs4851251_t rs4863692_t rs61160187_g rs62263923_g rs62379838_c rs7306755_a rs76076331_t rs9320913_a rs2992632_t rs2456973_c rs1043209_g rs10496091_a rs11210860_a rs11689269_c rs11690172_g rs11768238_a rs12969294_a rs13402908_t rs1402025_c rs17119973_a rs1777827_g rs1871109_g rs2245901_a rs2431108_c rs2615691_a rs2837992_t rs2964197_t rs324886_t rs34072092_c rs35761247_a rs55830725_a rs572016_a rs6739979_c rs6799130_g rs7131944_t rs7767938_c rs7854982_t rs7955289_t rs8005528_c rs895606_a rs9537821_g, robust
test rs11191193_g rs113520408_a rs11588857_a rs11712056_c rs12531458_c rs12987662_a rs13294439_c rs148734725_a rs1606974_a rs16845580_c rs17167170_g rs17824247_c rs192818565_g rs301800_t rs34305371_a rs4493682_c rs4500960_t rs4851251_t rs4863692_t rs61160187_g rs62263923_g rs62379838_c rs7306755_a rs76076331_t rs9320913_a rs2992632_t rs2456973_c rs1043209_g rs10496091_a rs11210860_a rs11689269_c rs11690172_g rs11768238_a rs12969294_a rs13402908_t rs1402025_c rs17119973_a rs1777827_g rs1871109_g rs2245901_a rs2431108_c rs2615691_a rs2837992_t rs2964197_t rs324886_t rs34072092_c rs35761247_a rs55830725_a rs572016_a rs6739979_c rs6799130_g rs7131944_t rs7767938_c rs7854982_t rs7955289_t rs8005528_c rs895606_a rs9537821_g
reg Birthweight rs11191193_g, robust
reg Birthweight rs113520408_a, robust
reg Birthweight rs11588857_a, robust
reg Birthweight rs11712056_c, robust
reg Birthweight rs12531458_c, robust
reg Birthweight rs12987662_a, robust
reg Birthweight rs13294439_c, robust
reg Birthweight rs148734725_a, robust
reg Birthweight rs1606974_a, robust
reg Birthweight rs16845580_c, robust
reg Birthweight rs17167170_g, robust
reg Birthweight rs17824247_c, robust
reg Birthweight rs192818565_g, robust
reg Birthweight rs301800_t, robust
reg Birthweight rs34305371_a, robust
reg Birthweight rs4493682_c, robust
reg Birthweight rs4500960_t, robust
reg Birthweight rs4851251_t, robust
reg Birthweight rs4863692_t, robust
reg Birthweight rs61160187_g, robust
reg Birthweight rs62263923_g, robust
reg Birthweight rs62379838_c, robust
reg Birthweight rs7306755_a, robust
reg Birthweight rs76076331_t, robust
reg Birthweight rs9320913_a, robust
reg Birthweight rs2992632_t, robust
reg Birthweight rs2456973_c, robust
reg Birthweight rs1043209_g, robust
reg Birthweight rs10496091_a, robust
reg Birthweight rs11210860_a, robust
reg Birthweight rs11689269_c, robust
reg Birthweight rs11690172_g, robust
reg Birthweight rs11768238_a, robust
reg Birthweight rs12969294_a, robust
reg Birthweight rs13402908_t, robust
reg Birthweight rs1402025_c, robust

```
reg Birthweight rs17119973_a, robust
reg Birthweight rs1777827_g, robust
reg Birthweight rs1871109_g, robust
reg Birthweight rs2245901_a, robust
reg Birthweight rs2431108_c, robust
reg Birthweight rs2615691_a, robust
reg Birthweight rs2837992_t, robust
reg Birthweight rs2964197_t, robust
reg Birthweight rs324886_t, robust
reg Birthweight rs34072092_c, robust
reg Birthweight rs35761247_a, robust
reg Birthweight rs55830725_a, robust
reg Birthweight rs572016_a, robust
reg Birthweight rs6739979_c, robust
reg Birthweight rs6799130_g, robust
reg Birthweight rs7131944_t, robust
reg Birthweight rs7767938_c, robust
reg Birthweight rs7854982_t, robust
reg Birthweight rs7955289_t, robust
reg Birthweight rs8005528_c, robust
reg Birthweight rs895606_a, robust
reg Birthweight rs9537821_g, robust

*Having a look at the polygenic risk score for birthweight
reg Birthweight bw_pgs, robust
display ttail(7698,abs(_b[bw_pgs]/_se[bw_pgs]))*2

corr Score_year2 IQ_month49 IQ_year8 IQ_year15 bw_pgs
corr Englishmarks_year4 Englishmarks_year9 Englishattainment_year9 bw_pgs
corr Mathsmarks_year4 Mathsmarks_year9 Mathsattainment_year9 Mathsfinal_year9 bw_pgs
corr Sciencemarks_year4 Scienceattainment_year9 Sciencefinal_year9 bw_pgs

*Two stage least squares
ivregress 2sls Score_year2 (Birthweight=bw_pgs), robust first
ivregress 2sls IQ_month49 (Birthweight=bw_pgs), robust first
ivregress 2sls IQ_year8 (Birthweight=bw_pgs), robust first
ivregress 2sls IQ_year15 (Birthweight=bw_pgs), robust first

ivregress 2sls Englishmarks_year4 (Birthweight=bw_pgs), robust first
ivregress 2sls Englishmarks_year9 (Birthweight=bw_pgs), robust first
ivregress 2sls Englishattainment_year9 (Birthweight=bw_pgs), robust first

ivregress 2sls Mathsmarks_year4 (Birthweight=bw_pgs), robust first
ivregress 2sls Mathsmarks_year9 (Birthweight=bw_pgs), robust first
ivregress 2sls Mathsattainment_year9 (Birthweight=bw_pgs), robust first
ivregress 2sls Mathsfinal_year9 (Birthweight=bw_pgs), robust first

ivregress 2sls Sciencemarks_year4 (Birthweight=bw_pgs), robust first
ivregress 2sls Scienceattainment_year9 (Birthweight=bw_pgs), robust first
ivregress 2sls Sciencefinal_year9 (Birthweight=bw_pgs), robust first

*Do not forget to install the package; via findit plausexog
plausexog ltz Score_year2 (Birthweight = bw_pgs), omega(0) mu(0) level(.95) vce(robust) graph(Birthweight) graphmu(0
1324606.8	2649213.6	3973820.4	5298427.2	6623034.0) graphomega(157933913281.0
	157933913281.0	157933913281.0	157933913281.0	157933913281.0	157933913281.0) graphdelta(0 20 40
60 80 100)
plausexog ltz IQ_month49 (Birthweight = bw_pgs), omega(0) mu(0) level(.95) vce(robust) graph(Birthweight) graphmu(0
1128106.2	2256212.4	3384318.6	4512424.8	5640531.0) graphomega(1260237495609.0
	1260237495609.0	1260237495609.0	1260237495609.0	1260237495609.0	1260237495609.0) graphdelta(0 20 40
60 80 100)
```

```
plausexog ltz IQ_year8(Birthweight = bw_pgs), omega(0) mu(0) level(.95) vce(robust) graph(Birthweight) graphmu(0 1542255.2
	3084510.4	4626765.6	6169020.8	7711276.0) graphomega(175218090408.4
	175218090408.4	175218090408.4	175218090408.4	175218090408.4	175218090408.4) graphdelta(0 20 40
60 80 100)
plausexog ltz IQ_year15(Birthweight = bw_pgs), omega(0) mu(0) level(.95) vce(robust) graph(Birthweight) graphmu(0
1690766.8	3381533.6	5072300.4	6763067.2	8453834.0) graphomega(239343546756.3
	239343546756.3	239343546756.3	239343546756.3	239343546756.3	239343546756.3) graphdelta(0 20 40
60 80 100)

plausexog ltz Englishmarks_year4 (Birthweight = bw_pgs), omega(0) mu(0) level(.95) vce(robust) graph(Birthweight) graphmu(0
1354933.0	2709866.0	4064799.0	5419732.0	6774665.0) graphomega(146809439911.8
	146809439911.8	146809439911.8	146809439911.8	146809439911.8	146809439911.8) graphdelta(0 20 40
60 80 100)
plausexog ltz Englishmarks_year9 (Birthweight = bw_pgs), omega(0) mu(0) level(.95) vce(robust) graph(Birthweight) graphmu(0
1387647.2	2775294.4	4162941.6	5550588.8	6938236.0) graphomega(171947725689.6
	171947725689.6	171947725689.6	171947725689.6	171947725689.6	171947725689.6) graphdelta(0 20 40
60 80 100)
plausexog ltz Englishattainment_year9 (Birthweight = bw_pgs), omega(0) mu(0) level(.95) vce(robust) graph(Birthweight)
graphmu(0 1435943.6	2871887.2	4307830.8	5743774.4	7179718.0)
graphomega(208667519042.0	208667519042.0	208667519042.0	208667519042.0	208667519042.0
	208667519042.0) graphdelta(0 20 40 60 80 100)

plausexog ltz Mathsmarks_year4 (Birthweight = bw_pgs), omega(0) mu(0) level(.95) vce(robust) graph(Birthweight) graphmu(0
1343153.2	2686306.4	4029459.6	5372612.8	6715766.0) graphomega(149209920729.0
	149209920729.0	149209920729.0	149209920729.0	149209920729.0	149209920729.0) graphdelta(0 20 40
60 80 100)
plausexog ltz Mathsmarks_year9 (Birthweight = bw_pgs), omega(0) mu(0) level(.95) vce(robust) graph(Birthweight) graphmu(0
1372139.4	2744278.8	4116418.2	5488557.6	6860697.0) graphomega(177340285700.4
	177340285700.4	177340285700.4	177340285700.4	177340285700.4	177340285700.4) graphdelta(0 20 40
60 80 100)
plausexog ltz Mathsattainment_year9 (Birthweight = bw_pgs), omega(0) mu(0) level(.95) vce(robust) graph(Birthweight)
graphmu(0 1435943.6	2871887.2	4307830.8	5743774.4	7179718.0)
graphomega(208667519041.96	208667519041.96	208667519041.96	208667519041.96	208667519041.96
	208667519041.96) graphdelta(0 20 40 60 80 100)
plausexog ltz Mathsfinal_year9 (Birthweight = bw_pgs), omega(0) mu(0) level(.95) vce(robust) graph(Birthweight) graphmu(0
1444026.8	2888053.6	4332080.4	5776107.2	7220134.0) graphomega(210822854870.25
	210822854870.25	210822854870.25	210822854870.25	210822854870.25	210822854870.25) graphdelta(0 20 40
60 80 100)

plausexog ltz Sciencemarks_year4 (Birthweight = bw_pgs), omega(0) mu(0) level(.95) vce(robust) graph(Birthweight) graphmu(0
1339389.2	2678778.4	4018167.6	5357556.8	6696946.0) graphomega(147256924836.49
	147256924836.49	147256924836.49	147256924836.49	147256924836.49	147256924836.49) graphdelta(0 20 40
60 80 100)
plausexog ltz Scienceattainment_year9 (Birthweight = bw_pgs), omega(0) mu(0) level(.95) vce(robust) graph(Birthweight)
graphmu(0 1435943.6	2871887.2	4307830.8	5743774.4	7179718.0)
graphomega(208667519041.96	208667519041.96	208667519041.96	208667519041.96	208667519041.96
	208667519041.96) graphdelta(0 20 40 60 80 100)
plausexog ltz Sciencefinal_year9 (Birthweight = bw_pgs), omega(0) mu(0) level(.95) vce(robust) graph(Birthweight) graphmu(0
1438508.2	2877016.4	4315524.6	5754032.8	7192541.0) graphomega(211043490388.09
	211043490388.09	211043490388.09	211043490388.09	211043490388.09	211043490388.09) graphdelta(0 20 40
60 80 100)

drop if Score_year2 == . | IQ_year8 == . | IQ_year15 == .
drop if Englishmarks_year4 == . | Englishmarks_year9 ==. |  Englishattainment_year9 == .
drop if Mathsmarks_year4 == . | Mathsmarks_year9 == . | Mathsattainment_year9 == . | Mathsfinal_year9 == .
drop if Sciencemarks_year4 == . | Scienceattainment_year9 == . | Sciencefinal_year9 == .
drop if Married == . | Twins == . | Cigs_week8 == .

sum Birthweight, detail
sum Birthweight if Male == 1
sum Birthweight if Male == 0
```

```
sum Birthweight if Male == 1 & Twins == 1
sum Birthweight if Male == 0 & Twins == 1

tab Married
tab Smoker
tab Mum_educ_highest
tab Dad_educ_highest

sum Score_year2 IQ_month49 IQ_year8 IQ_year15
sum Englishmarks_year4 Englishmarks_year9 Englishattainment_year9
sum Mathsmarks_year4 Mathsmarks_year9 Mathsattainment_year9 Mathsfinal_year9
sum Sciencemarks_year4 Scienceattainment_year9 Sciencefinal_year9
sum Birth_length Height_month49 Height_year8 Height_year15
sum Birthweight Weight_month49 Weight_year8 Weight_year15
sum Cigs_week8 Pregnancy_length

egen float z_Score_year2 = std(Score_year2), mean(0) std(1)
egen float z_IQ_month49= std(IQ_month49), mean(0) std(1)
egen float z_IQ_year8 = std(IQ_year8), mean(0) std(1)
egen float z_IQ_year15 = std(IQ_year15), mean(0) std(1)
egen float z_Englishmarks_year4 = std(Englishmarks_year4), mean(0) std(1)
egen float z_Mathsmarks_year4 = std(Mathsmarks_year4), mean(0) std(1)
egen float z_Sciencemarks_year4 = std(Sciencemarks_year4), mean(0) std(1)
egen float z_Englishmarks_year9 = std(Englishmarks_year9), mean(0) std(1)
egen float z_Mathsmarks_year9 = std(Mathsmarks_year9), mean(0) std(1)
egen float z_Englishattainment_year9 = std(Englishattainment_year9), mean(0) std(1)
egen float z_Mathsattainment_year9 = std(Mathsattainment_year9), mean(0) std(1)
egen float z_Scienceattainment_year9 = std(Scienceattainment_year9), mean(0) std(1)
egen float z_Mathsfinal_year9 = std(Mathsfinal_year9), mean(0) std(1)
egen float z_Sciencefinal_year9 = std(Sciencefinal_year9), mean(0) std(1)

reg Score_year2 Birthweight, robust
reg IQ_month49 Birthweight, robust
reg IQ_year8 Birthweight, robust
reg IQ_year15 Birthweight, robust

*reg Englishmarks_year4 Birthweight, robust
*reg Mathsmarks_year4 Birthweight, robust
*reg Sciencemarks_year4 Birthweight, robust

*reg Englishmarks_year9 Birthweight, robust
*reg Mathsmarks_year9 Birthweight, robust

*reg Englishattainment_year9 Birthweight, robust
*reg Mathsattainment_year9 Birthweight, robust
*reg Scienceattainment_year9 Birthweight, robust

*reg Mathsfinal_year9 Birthweight, robust
*reg Sciencefinal_year9 Birthweight, robust

reg z_Score_year2 Birthweight, robust
reg z_IQ_month49 Birthweight, robust
reg z_IQ_year8 Birthweight, robust
reg z_IQ_year15 Birthweight, robust

reg z_Englishmarks_year4 Birthweight, robust
reg z_Mathsmarks_year4 Birthweight, robust
reg z_Sciencemarks_year4 Birthweight, robust

reg z_Englishmarks_year9 Birthweight, robust
```

```
reg z_Mathsmarks_year9 Birthweight, robust

reg z_Englishattainment_year9 Birthweight, robust
reg z_Mathsattainment_year9 Birthweight, robust
reg z_Scienceattainment_year9 Birthweight, robust

reg z_Mathsfinal_year9 Birthweight, robust
reg z_Sciencefinal_year9 Birthweight, robust

*reg Score_year2 Birthweight Male Cigs_week8 Married Twins Mum_educ_non Mum_educ_CSE Mum_educ_Olvl Mum_educ_Alvl Mum_educ_voc Mum_educ_uni Dad_educ_non Dad_educ_CSE Dad_educ_Olvl Dad_educ_Alvl Dad_educ_voc Dad_educ_uni, robust
*reg IQ_month49 Birthweigh Male Cigs_week8 Married Twins Mum_educ_non Mum_educ_CSE Mum_educ_Olvl Mum_educ_Alvl Mum_educ_voc Mum_educ_uni Dad_educ_non Dad_educ_CSE Dad_educ_Olvl Dad_educ_Alvl Dad_educ_voc Dad_educ_uni, robust
*reg IQ_year8 Birthweight Male Cigs_week8 Married Twins Mum_educ_non Mum_educ_CSE Mum_educ_Olvl Mum_educ_Alvl Mum_educ_voc Mum_educ_uni Dad_educ_non Dad_educ_CSE Dad_educ_Olvl Dad_educ_Alvl Dad_educ_voc Dad_educ_uni, robust
*reg IQ_year15 Birthweight Male Cigs_week8 Married Twins Mum_educ_non Mum_educ_CSE Mum_educ_Olvl Mum_educ_Alvl Mum_educ_voc Mum_educ_uni Dad_educ_non Dad_educ_CSE Dad_educ_Olvl Dad_educ_Alvl Dad_educ_voc Dad_educ_uni, robust

*reg Englishmarks_year4 Birthweight Male Cigs_week8 Married Twins Mum_educ_non Mum_educ_CSE Mum_educ_Olvl Mum_educ_Alvl Mum_educ_voc Mum_educ_uni Dad_educ_non Dad_educ_CSE Dad_educ_Olvl Dad_educ_Alvl Dad_educ_voc Dad_educ_uni, robust
*reg Mathsmarks_year4 Birthweight Male Cigs_week8 Married Twins Mum_educ_non Mum_educ_CSE Mum_educ_Olvl Mum_educ_Alvl Mum_educ_voc Mum_educ_uni Dad_educ_non Dad_educ_CSE Dad_educ_Olvl Dad_educ_Alvl Dad_educ_voc Dad_educ_uni, robust
*reg Sciencemarks_year4 Birthweight Male Cigs_week8 Married Twins Mum_educ_non Mum_educ_CSE Mum_educ_Olvl Mum_educ_Alvl Mum_educ_voc Mum_educ_uni Dad_educ_non Dad_educ_CSE Dad_educ_Olvl Dad_educ_Alvl Dad_educ_voc Dad_educ_uni, robust

*reg Englishmarks_year9 Birthweight Male Cigs_week8 Married Twins Mum_educ_non Mum_educ_CSE Mum_educ_Olvl Mum_educ_Alvl Mum_educ_voc Mum_educ_uni Dad_educ_non Dad_educ_CSE Dad_educ_Olvl Dad_educ_Alvl Dad_educ_voc Dad_educ_uni, robust
*reg Mathsmarks_year9 Birthweight Male Cigs_week8 Married Twins Mum_educ_non Mum_educ_CSE Mum_educ_Olvl Mum_educ_Alvl Mum_educ_voc Mum_educ_uni Dad_educ_non Dad_educ_CSE Dad_educ_Olvl Dad_educ_Alvl Dad_educ_voc Dad_educ_uni, robust

*reg Englishattainment_year9 Birthweight Male Cigs_week8 Married Twins Mum_educ_non Mum_educ_CSE Mum_educ_Olvl Mum_educ_Alvl Mum_educ_voc Mum_educ_uni Dad_educ_non Dad_educ_CSE Dad_educ_Olvl Dad_educ_Alvl Dad_educ_voc Dad_educ_uni, robust
*reg Mathsattainment_year9 Birthweight Male Cigs_week8 Married Twins Mum_educ_non Mum_educ_CSE Mum_educ_Olvl Mum_educ_Alvl Mum_educ_voc Mum_educ_uni Dad_educ_non Dad_educ_CSE Dad_educ_Olvl Dad_educ_Alvl Dad_educ_voc Dad_educ_uni, robust
*reg Scienceattainment_year9 Birthweight Male Cigs_week8 Married Twins Mum_educ_non Mum_educ_CSE Mum_educ_Olvl Mum_educ_Alvl Mum_educ_voc Mum_educ_uni Dad_educ_non Dad_educ_CSE Dad_educ_Olvl Dad_educ_Alvl Dad_educ_voc Dad_educ_uni, robust

*reg Mathsfinal_year9 Birthweight Male Cigs_week8 Married Twins Mum_educ_non Mum_educ_CSE Mum_educ_Olvl Mum_educ_Alvl Mum_educ_voc Mum_educ_uni Dad_educ_non Dad_educ_CSE Dad_educ_Olvl Dad_educ_Alvl Dad_educ_voc Dad_educ_uni, robust
*reg Sciencefinal_year9 Birthweight Male Cigs_week8 Married Twins Mum_educ_non Mum_educ_CSE Mum_educ_Olvl Mum_educ_Alvl Mum_educ_voc Mum_educ_uni Dad_educ_non Dad_educ_CSE Dad_educ_Olvl Dad_educ_Alvl Dad_educ_voc Dad_educ_uni, robust

reg z_Score_year2 Birthweight Male Cigs_week8 Married Twins Mum_educ_non Mum_educ_CSE Mum_educ_Olvl Mum_educ_Alvl Mum_educ_voc Mum_educ_uni Dad_educ_non Dad_educ_CSE Dad_educ_Olvl Dad_educ_Alvl Dad_educ_voc Dad_educ_uni, robust
```

```
reg z_IQ_month49 Birthweigh Male Cigs_week8 Married Twins Mum_educ_non Mum_educ_CSE Mum_educ_Olvl
Mum_educ_Alvl Mum_educ_voc Mum_educ_uni Dad_educ_non Dad_educ_CSE Dad_educ_Olvl Dad_educ_Alvl Dad_educ_voc
Dad_educ_uni, robust
reg z_IQ_year8 Birthweight Male Cigs_week8 Married Twins Mum_educ_non Mum_educ_CSE Mum_educ_Olvl
Mum_educ_Alvl Mum_educ_voc Mum_educ_uni Dad_educ_non Dad_educ_CSE Dad_educ_Olvl Dad_educ_Alvl Dad_educ_voc
Dad_educ_uni, robust
reg z_IQ_year15 Birthweight Male Cigs_week8 Married Twins Mum_educ_non Mum_educ_CSE Mum_educ_Olvl
Mum_educ_Alvl Mum_educ_voc Mum_educ_uni Dad_educ_non Dad_educ_CSE Dad_educ_Olvl Dad_educ_Alvl Dad_educ_voc
Dad_educ_uni, robust

reg z_Englishmarks_year4 Birthweight Male Cigs_week8 Married Twins Mum_educ_non Mum_educ_CSE Mum_educ_Olvl
Mum_educ_Alvl Mum_educ_voc Mum_educ_uni Dad_educ_non Dad_educ_CSE Dad_educ_Olvl Dad_educ_Alvl Dad_educ_voc
Dad_educ_uni, robust
reg z_Mathsmarks_year4 Birthweight Male Cigs_week8 Married Twins Mum_educ_non Mum_educ_CSE Mum_educ_Olvl
Mum_educ_Alvl Mum_educ_voc Mum_educ_uni Dad_educ_non Dad_educ_CSE Dad_educ_Olvl Dad_educ_Alvl Dad_educ_voc
Dad_educ_uni, robust
reg z_Sciencemarks_year4 Birthweight Male Cigs_week8 Married Twins Mum_educ_non Mum_educ_CSE Mum_educ_Olvl
Mum_educ_Alvl Mum_educ_voc Mum_educ_uni Dad_educ_non Dad_educ_CSE Dad_educ_Olvl Dad_educ_Alvl Dad_educ_voc
Dad_educ_uni, robust

reg z_Englishmarks_year9 Birthweight Male Cigs_week8 Married Twins Mum_educ_non Mum_educ_CSE Mum_educ_Olvl
Mum_educ_Alvl Mum_educ_voc Mum_educ_uni Dad_educ_non Dad_educ_CSE Dad_educ_Olvl Dad_educ_Alvl Dad_educ_voc
Dad_educ_uni, robust
reg z_Mathsmarks_year9 Birthweight Male Cigs_week8 Married Twins Mum_educ_non Mum_educ_CSE Mum_educ_Olvl
Mum_educ_Alvl Mum_educ_voc Mum_educ_uni Dad_educ_non Dad_educ_CSE Dad_educ_Olvl Dad_educ_Alvl Dad_educ_voc
Dad_educ_uni, robust

reg z_Englishattainment_year9 Birthweight Male Cigs_week8 Married Twins Mum_educ_non Mum_educ_CSE Mum_educ_Olvl
Mum_educ_Alvl Mum_educ_voc Mum_educ_uni Dad_educ_non Dad_educ_CSE Dad_educ_Olvl Dad_educ_Alvl Dad_educ_voc
Dad_educ_uni, robust
reg z_Mathsattainment_year9 Birthweight Male Cigs_week8 Married Twins Mum_educ_non Mum_educ_CSE Mum_educ_Olvl
Mum_educ_Alvl Mum_educ_voc Mum_educ_uni Dad_educ_non Dad_educ_CSE Dad_educ_Olvl Dad_educ_Alvl Dad_educ_voc
Dad_educ_uni, robust
reg z_Scienceattainment_year9 Birthweight Male Cigs_week8 Married Twins Mum_educ_non Mum_educ_CSE Mum_educ_Olvl
Mum_educ_Alvl Mum_educ_voc Mum_educ_uni Dad_educ_non Dad_educ_CSE Dad_educ_Olvl Dad_educ_Alvl Dad_educ_voc
Dad_educ_uni, robust

reg z_Mathsfinal_year9 Birthweight Male Cigs_week8 Married Twins Mum_educ_non Mum_educ_CSE Mum_educ_Olvl
Mum_educ_Alvl Mum_educ_voc Mum_educ_uni Dad_educ_non Dad_educ_CSE Dad_educ_Olvl Dad_educ_Alvl Dad_educ_voc
Dad_educ_uni, robust
reg z_Sciencefinal_year9 Birthweight Male Cigs_week8 Married Twins Mum_educ_non Mum_educ_CSE Mum_educ_Olvl
Mum_educ_Alvl Mum_educ_voc Mum_educ_uni Dad_educ_non Dad_educ_CSE Dad_educ_Olvl Dad_educ_Alvl Dad_educ_voc
Dad_educ_uni, robust

ivregress 2sls Score_year2 (Birthweight=bw_pgs), robust first
ivregress 2sls IQ_month49 (Birthweight=bw_pgs), robust first
ivregress 2sls IQ_year8 (Birthweight=bw_pgs), robust first
ivregress 2sls IQ_year15 (Birthweight=bw_pgs), robust first
ivregress 2sls Englishmarks_year4 (Birthweight=bw_pgs), robust first
ivregress 2sls Mathsmarks_year4 (Birthweight=bw_pgs), robust first
ivregress 2sls Sciencemarks_year4 (Birthweight=bw_pgs), robust first
ivregress 2sls Englishmarks_year9 (Birthweight=bw_pgs), robust first
ivregress 2sls Mathsmarks_year9 (Birthweight=bw_pgs), robust first
ivregress 2sls Englishattainment_year9 (Birthweight=bw_pgs), robust first
ivregress 2sls Mathsattainment_year9 (Birthweight=bw_pgs), robust first
ivregress 2sls Scienceattainment_year9 (Birthweight=bw_pgs), robust first

ivregress 2sls Mathsfinal_year9 (Birthweight=bw_pgs), robust first
ivregress 2sls Sciencefinal_year9 (Birthweight=bw_pgs), robust first

log close
```