The role of early life variables in shaping future Incomes in Britain: a longitudinal analysis

Connie Lloyd-Mackay

BSc Economics

School of Economics

University of Kent, July 2025

Abstract

Despite extensive research on intergenerational income inequality, few studies have quantified the cumulative effect of early life variables on later-life income. This paper investigates the role of childhood factors, such as cognitive abilities, socioeconomic background, and early health indicators in determining income at mid-career among British individuals. Drawing on longitudinal data from the 1970 British Cohort Study, the analysis uses robust econometric techniques, including linear regression and hierarchical modelling, to assess the predictive power of multiple early life variables on weekly income measured at age 46 (2016). The results reveal that higher literacy scores and an elevated social class in childhood are strongly associated with increased adult earnings, while some health-related indicators, such as the number of immunisations, demonstrate unexpectedly negative effects. Ultimately, this paper contributes essential insights for policymakers and researchers by suggesting that targeted early interventions in education and social support could be pivotal in mitigating long-term income disparities.

AI Statement

I acknowledge the use of generative AI in literature search and code development in this paper. However, the work reported remains my own.

Acknowledgements

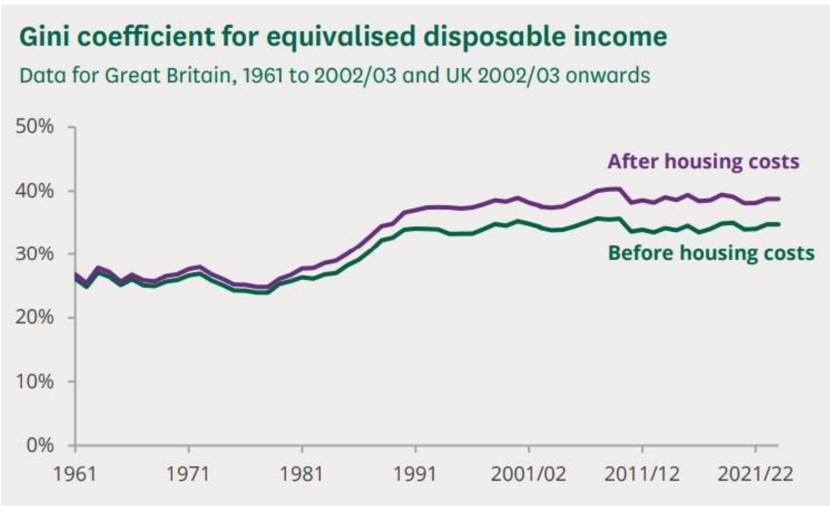
I would like to thank my supervisor Alastair Bailey for his invaluable advice and continual guidance throughout this project and always answering my endless questions/emails; without him this work would not be possible. I would like to thank my parents for getting me to where I am today, also thankyou to James for his support throughout this project.

1: Introduction

Individuals receive income through a variety of sources, such as earnings from employment, state benefits, investment returns, private pensions, and other financial streams. Income is typically measured in two main ways (Francis-Devine and Orme, 2024): gross income refers to the total income received before any deductions, including cash benefits; disposable income represents the amount remaining after the deduction of direct taxes, National Insurance contributions, and council tax (or rates in Northern Ireland) from gross income. According to the Office for National Statistics, median weekly earnings for full-time employees rose to £728 in April 2024, reflecting a 6.0% annual nominal increase. Median gross annual earnings for full-time employees were £37,430, up 6.9% from the previous year of £35,004.

During 2022/23 there were 14.3 million (21%) people living in poverty in the UK, 8.1 million of these were working-age adults (Joseph Rowntree Foundation, 2025). A household is typically classified as being in poverty if its income falls below 60% of the median income in the country (Trust for London 2022). The fact that the majority of those in poverty are working-age adults underscores the pressing need to explore early life variables that shape adult outcomes. Early childhood factors are critical in determining future earning potential (Hill et al., 2013).

The Gini coefficient quantifies income inequality as a single number ranging from 0 to 100, with higher percentages indicating greater inequality. In the UK, it stood at 35% before housing costs and increased to 39% after housing costs, for the 2022/23 financial year (Francis-Devine and Orme, 2024). The United Kingdom has one of the highest income inequality rates in Europe, beaten only by Lithuania with 37% (OECD, 2021), similar countries like France and Germany have 30%. The Netherlands and Ireland have rates of 29%, showing they have more even distribution of income. The below graph shows how the Gini coefficient, which rose sharply during the 1980s and has since remained relatively stable at a higher level, with inequality consistently greater after housing costs than before.



(Institute for Fiscal Studies, 2024)

Reducing income inequality requires interventions to achieve a fairer distribution of resources. However, the creation of effective policies relies on thorough research and evidence to guide decision-making. Kaasa (2005) categorises income inequality factors into five groups: economic development, demographic factors, political factors, cultural and environmental factors, and macroeconomic factors. So, a comprehensive empirical study to identify which of these income inequality factors significantly impact later income is vital.

Intergenerational earning mobility refers to the extent to which a child's earnings are correlated with their parents' earnings (van der Erve et al., 2023), low mobility suggests that life chances are partly determined by factors beyond individual control. Meaning factors like family social class, living situations, parents' qualifications and birthweight all may limit earning potentials before even starting school. Nordic countries, Canada, and Australia show high mobility, with earnings elasticity of under 0.2. However, Italy, the US, and the UK are much lower in terms of mobility, with elasticity levels soaring to 0.4–0.5. This means in countries with high elasticity, children's incomes are strongly tied to their parents' (D'Addio, 2007). This poses the question what (and how) other early life variables influence long term income?

With this study aiming to answer that question by identifying how certain early life variables affect future income, it will highlight the long-term impacts of childhood conditions and the importance of understanding them for regarding future equality. This approach not only addresses poverty at its roots but also promotes long-term economic mobility and resilience. This project's empirical findings offer policymakers a strong foundation for targeted interventions that reduce income inequality. By analysing how key factors shape adult incomes, it helps identify strategies to break the cycle of poverty, ensuring individuals have the necessary skills and opportunities for long-term success. Additionally, these insights guide investment decisions, maximising returns while efficiently allocating limited government resources.

To investigate this the study has formulated the following hypotheses:

$\boldsymbol{H_0}: \boldsymbol{\beta = 1}$ Early life variables have no significant effect on weekly income at age 46.

$\boldsymbol{H_1}: \boldsymbol{\beta \neq 1}$ Certain early life variables significantly affect income, childhood socioeconomic status (SES), education, and cognitive skills do shape long-term economic outcomes.

ORGANISATION

This paper is organised with the following chapter as a literature review summarising relevant studies and identifying gaps in the research. Then chapter 3 goes describes the methodology used going into detail regarding the data source, variables and the model. Next, chapter 4 lists the empirical results and identifies which early life variables significantly influence later life income. After this a discussion section (chapter 5) is used to evaluate the findings in relation to the question and policy implications. The last chapter concludes the findings, restating key points and what these results mean for future research.

2: Literature Review

A review by Rutter, Kim-Cohen, and Maughan (2006) discussed how emotional and conduct problems in childhood often persist into adulthood. For instance, the authors observed how mental health issues during childhood and adolescence can significantly influence outcomes in adulthood. Childhood depression has been linked to lower educational achievement, extended periods of unemployment, early parenthood, as well as increased risks of nicotine dependence and alcohol misuse (Fergusson & Woodward, 2002). The long-term impact of childhood conduct problems includes an increased risk of negative socioeconomic outcomes in adulthood specifically, lower educational levels, unemployment, reliance on welfare, and reduced income. Notably, these associations are less pronounced once the influence of early adversity and intelligence is statistically controlled for (Fergusson et al., 2005; Richards & Abbott, 2009).

Due to the deep research into outcomes of individuals with negative childhoods but little on positive ones Richards and Huppert (2011) decided to investigate the long-term effects of positive early life well-being and mental health. Using data from the 1946 cohort study (NSHD), childhood positive well-being at ages 13 and 15 was evaluated through teacher ratings with the early Rutter A Scale (Rutter, 1967). Then multivariate regression examined associations between childhood positivity, extraversion, and adult outcomes. Children who received two or more positive ratings were 60% less likely to experience emotional problems later in life than those who had no positive ratings. Amongst positive children, there was significantly higher work satisfaction rates, 41.6% of children with two or more positive ratings were “very happy” in their job at 53, this was only 36.3% for children with no positive ratings. This study highlights the significant long-term influence of childhood positive wellbeing on adult mental health and work satisfaction. The use of teacher ratings gives this study a subjective nature that could distort findings, using standardised tests to assess children’s behaviour or family background data might offer stronger validity. This article's exploration of mental health profiles could guide future analysis by suggesting how early life well-being or adversity might indirectly shape income trajectories.

Other research looked at the effect of early life health variables on future economic outcomes. A study (Lambiris et al., 2022) investigated birthweight and adult incomes, particularly focusing on whether low birth weight impacts earnings in adulthood. After conducting a meta-analysis using the Der Simonian and Laird random-effects method to estimate pooled associations. They found that with the increase of each standard deviation in birthweight, annual earnings increased by 2.75%, these results maintained across different sensitivity analyses. Having a low birthweight (<2500g) was associated with a 3.41% reduction in annual earnings (although these results were not statistically significant). Two Canadian studies that explored earnings between individuals with extremely low birth weights ELBW (<1000g) and individuals born >2500g found that males born with ELBW earned 25.3% less annually and females 6.5% less annually (Goddeeris et al., 2010); and that ELBW people earned $17,210 less annually compared to those with birthweights over 2500g (Dobson et al., 2017).

Lambiris et al. primary focus on high income countries (Sweden, Norway, Denmark Australia), with only one study from a middle-income country (China) raises concerns with the general applicability of their findings in other more diverse contexts. Additionally, the lack of longitudinal data from low and middle-income regions creates a vital gap, that fails to capture a variety of resources over time. The studies meta-analysis appears to pool studies without accounting for nuanced post hoc analyses that could reveal interaction effects. This is where an own post hoc analysis might excel, particularly in identifying moderating variables like gender or socioeconomic background.

Deshpande and Ramachandran (2022) measured the relationship between childhood stunting and later life outcomes, drawing on longitudinal data from the Young Lives Survey (YLS) in Ethiopia, India, Peru, and Vietnam. They investigated three key areas: the persistence of chronic malnutrition, its effects on human capital and subjective well-being, and the factors linked to early childhood stunting. This study demonstrates the consequences of stunting in early childhood, including poorer educational outcomes, 1.16 fewer grades are achieved by age 22 for those severely stunted at age 8. Reduced cognitive performance at ages 8, 12, and 15 were also observed.

Deshpande and Ramachandran conducted robustness tests in the study assess whether the links between stunting and outcomes like education, cognition, and well-being are valid or influenced by omitted variable bias. Using Oster's (2017) method, researchers found the associations were unlikely due to unobserved factors, strengthening their credibility. However, as the data is observational, there could still be unmeasured factors influencing the observed relationships, and without experimental interventions, causation cannot be definitively proven. So, while the robustness tests enhance the credibility of the findings, the authors caution against if early stunting directly causes the observed negative outcomes. Instead, the results are presented as strong associations that warrant further investigation, ideally with experimental or quasi-experimental methods.

Heckman (2006) summarised literature on the formation of early life skills and the implications for economic and social outcomes. Here a review and synthesis of evidence approach was used, primarily with longitudinal studies, cost-benefit analysis (calculating rates of return on investment in early interventions), and comparative policy analysis which evaluated the effectiveness of early versus later interventions (early childhood education vs. adult training programs). Heckman concluded that early life conditions, especially cognitive and non-cognitive abilities do play a crucial role in moulding future outcomes. With disadvantaged environments creating persistent gaps in skills and abilities, which are difficult and costly to address later in life. Both cognitive and non-cognitive skills were key in providing economic and social success. The Perry Preschool Program (Schweinhart et al., 2005) the Abecedarian Project (Masse and Barnett, 2002) did yield high rates of return and benefit-cost ratios, highlighting how effective investment in young children is. Investment here is associated with better outcomes in earnings, education, welfare dependency and crime rates, compared to later investment. Because of this, policymakers should prioritise early interventions over later ones as it ultimately reduces inequality.

In this article Heckman heavily referred to other studies to find that higher childhood cognitive skills correlate with better academic outcomes which was then only inferred to lead to higher earnings. The effect of early cognitive or non-cognitive skills were not directly measured against income, instead it was assumed that a higher level of education always

translates to greater income. A study involving Heckman and Deshpande and Ramachandran's work could shed light on root cause of reduced cognitive ability, and how it affects income at later life which could substantiate Heckman's claim.
Moyer (2007) noted that early engagement with leisure reading has a profound influence on future outcomes, particularly in adulthood. The study reveals that all interview respondents recognised reading as a vital aspect of their lives since childhood. This underscores the role of early reading experiences in cultivating lifelong reading practices. Additionally, respondents unanimously affirmed the connection between leisure reading and learning, emphasising its educational significance. Their insights suggest that developing a strong reading foundation in childhood contributes to long-term academic benefits. Moyer's study used a survey and follow-up interviews to examine leisure reading. The survey, conducted in two public libraries, gathered responses on recreational and educational reading habits. Interviews with willing participants expanded on these insights, exploring personal reading experiences. Factor analysis validated the survey instrument and assessed the link between educational and recreational reading, providing both quantitative and qualitative findings.

There is a potential self-selection bias here, since survey respondents and interview participants voluntarily chose to engage, they may not fully represent broader reading habits across different demographics. Reliance on self-reported data can introduce recall bias, as participants may overestimate or misremember their childhood reading experiences. The study's library-based sampling also means findings might skew toward individuals who already value reading, limiting generalisability to those with less access or interest in books.

Blanden and Machin (2004) studied the impact of the UK's higher education (HE) expansion from the 1970s to the 1990s. Using longitudinal data from the National Child Development Study, British Cohort Study, and British Household Panel Survey, across three time periods, they examined changes in higher education participation and attainment among children from different parental income groups who attended university in the 1970s, 1980s, and 1990s. They applied econometric models to account for the sequential progression of education choices, examining how income associations vary across different stages of the education pathway. A probit model was used to relate degree acquisition to parental income and control variables. The estimating equation used dummy variables for income quantiles. Nonparametric methods, specifically Nadaraya-Watson kernel regression, explored how the likelihood of earning a degree varied across the family income distribution.

The study found that HE expansion disproportionately benefited wealthier families, with degree attainment at age 23 rising from 14.4% to 37.1% in the top income quintile (1981 to 1999), compared to 6% to 9% in the lowest quintile. This challenges the notion that HE expansion promotes equal opportunity, as it widened participation gaps between income groups. The association between family income and degree attainment grew stronger in more recent periods, persisting even after controlling for test scores. Sequential models revealed declining income effects at later education stages in earlier cohorts, but this trend did not appear in the BHPS cohort, further highlighting rising inequality in education.

However, while this study observes educational inequality, it does not measure the long-term income impact, especially beyond early adulthood. Additionally, for the British Household Panel Survey the sample sizes differ significantly across cohorts, which might affect the reliability of findings, particularly for later cohorts. The focus on HE participation and degree acquisition are narrow, a focus on broader measures like childhood socioeconomic conditions or non-academic factors could capture a fuller reflection on childhood characteristics and future outcomes.

Edwards (1975) illustrated how social class plays a crucial role in shaping children's educational experiences and outcomes, with middle-class students benefiting from a more supportive and value-aligned education than those from lower-income backgrounds. Children from lower-income backgrounds are frequently characterised as "culturally deprived," which can hinder their ability to adjust and thrive in public schools. This perception drives the need for increased educational support, ultimately influencing their long-term income prospects. Edwards also highlighted how institutional racism and poverty have long influenced educational practices, shaping the opportunities accessible to children across social classes. These systemic factors contribute to income disparities that persist into adulthood. Similar, to Blanden and Machin (2004) this research suggested that social class significantly impacts educational experiences but did not empirically look at an effect on later life income.

Research on FSM eligibility (FSME) and GCSE attainment highlights key disparities in educational outcomes influenced by deprivation and school context. Shuttleworth (1995) found that for every 1% in FSM pupils there was a 1% increase in students failing to achieve at least one GCSE grade C or higher, this relationship was weaker than prior findings in England (Kelly, 1993). Despite its significant but small effect on GCSE outcomes, FSME's long-term impact remains unexplored. GCSE grades are linked to lifetime earnings, the Department of Education found that achieving just one grade higher in a single subject was associated with an average increase of £23,000 in lifetime earnings (Busby, 2021) This suggests FSM recipients may face reduced future earnings, though this remains speculative without empirical evidence. Shuttleworth’s multilevel modelling accounts for school-level variation but is limited by cross-sectional data, raising concerns about causality and omitted variable bias. Longitudinal data mitigates these issues, revealing how early interventions influence long-term outcomes and aiding policymakers in designing effective strategies.

In contrast, Taylor (2018) noted FSME's limited reliability in capturing socio-economic disadvantages, with attainment disparities explained better by factors like parental education. Using linear regression models Taylor identified that literacy ability at age 7 showed little to no correlation with being eligible for Free School Meals or living in poverty. Literacy ability at age 11 was strongly linked to all the socio-economic variables analysed (including low income and FSME). Differences between literacy abilities at ages 7 and 11 suggest inequalities deepen as students’ progress, underscoring the value of longitudinal studies for richer insights and bias reduction.

The lack of pre-secondary data confirms the issue with cross-sectional data (the School Leavers Survey was used), as it records information from a single time point, making it difficult to discern causality. This approach increases the risk of omitted variable bias and misreporting, as respondents might inaccurately recall past information. Both Shuttleworth and Taylor focused on individual countries (Northern Ireland and Wales), meaning these results can not represent the UK as a whole and results could vary among regions.

Investigating all 4 nations of the UK will give multiple statistics to compare, while averages can be calculated.

RESEARCH OUTLINE

With many of the discussed studies using cross sectional data, they are limited by not being able to track trends over time. Here a longitudinal data study, particularly a cohort study would be much more reputable as it adds an additional dimension of granularity and depth by following a single cohort longitudinally. This avoids the between-study heterogeneity observed in the meta-analysis used by Lambiris et al. (2022). Additionally, Lambiris (et al.) has underutilised interaction effects, a hierarchical regression could explicitly examine how variables like birthweight interact with other early life variables. Richards and Huppert (2011) regression-based approach provide a foundation for evaluating the reliability and predictive power of linear and hierarchical regression models in this analysis. Despite many articles measuring different outcomes like educational attainment and happiness, which incomes can be surmised from, little have directly measured the impact of multiple different early life variables on later life income. Analysing the effect on income itself is vital as although education is a good predictor of income, it is not a perfect one. Doing so would provide statistical evidence on how these variables influence recent life results (income). As a result, the present study aims to go beyond prior research by investigating how early life variables directly affect income in later life.

3: Methodology

DATA COLLECTION

The first stage to investigate the effects of early life variables on income in later life is to find relevant data. The 1970 British Cohort Study (BCS) was the dataset used. This is a longitudinal study that follows individuals who were born in single week in April 1970 (in Britain) up until today (Butler, 2020). It was designed to investigate the social and biological traits of mothers in relation to neonatal morbidity, and to compare findings with the 1958 National Child Development Study. Over time, the study's focus has expanded beyond its original medical nature to include a broader range of topics, physical health, education, social factors, and economic conditions (Cataloguementalhealth.ac.uk, 2024). There have been 11 sweeps: birth (1970), age 5 (1975), age 10 (1980), age 16 (1986), age 26 (1996), age 30 (2000), age 34 (2004), age 38 (2008), age 42 (2012), age 46 (2016), and age 51 (2021); along with a 22 months, 42 months and, 21 years subsample. Note that the age 51 sweep was only published on 6th March 2025, after I had conducted my statistical research. Participants varied by sex: 53.30% female, 46.70% male; ethnic group and region, distributions as shown in the tables below.

Table 1: Participant Ethnicity overview

Ethnic Group	Percentage (%)
African	0.05
British	92.42
European	1.02%
Indian-Pakistani	1.50%
Other Asian	0.21%
Other	0.30%
West Indian	4.50%

Table 2: Participant Region of Residence overview

Region	Percentage (%)
East Anglia	3.43
East Midlands	7.00
North	6.65
Northwest	12.55
Scotland	9.79
Southeast	27.43
Southwest	7.35
Wales	5.54
West Midlands	10.41
Yorkshire	9.86

(UKDataService, 2024).

Sex, where 1=male and 2=female and height (in centimetres) were taken from the *42-month subsample, 1973.* Cold and cough regularity was also from here, with the question "How often does child get colds and coughs?" (Chamberlain, 2024), with responses reverse coded where 1=frequently, 2=occasionally and 3=rarely. Birthweight (in ounces), Ethnic groupchild, Mothers and Fathers Highest Educ Level and Number of Immunisations were from *the Age 5, Sweep 2, 1975.*

Table 3: Ethnic Group-Child

Value	Label
1	European UK
2	European Other
3	West Indian
4	Indian-Pakistani
5	Other Asian
6	African
7	Other

Table 4: Mothers/Fathers Highest Qualification

Value	Label
1	No Quals
2	Voc Quals
3	O level or Equiv
4	A level or Equiv
5	Srn
6	Cert of Eductn
7	Degree +
8	Other

Gross weekly family income, social Class from fathers' occupation or mothers if missing and estimated reading age at age 10 were extracted from the *Age 10, Sweep 3, 1980.*

Table 5: Gross weekly family income at age 10

Value	Label
0	£250+
1	£200-£249
2	£150-199
3	£100-149
4	£50-£99
5	£35-£49
6	Under £35
8	Refused

Table 6: Social Class from fathers' occupation (or mothers if missing)

Value	Label
1	V Unskilled
2	IV Partly Skilled
3	III Manual
4	III Non-Manual
5	II Managerial and Technical
6	I Professional

Test score variables were from the age 16 follow up otherwise known as the 'Youthscan', which was conducted by the International Centre for Child Studies (Seabrook and Murphy, 2017). Arithmetic test scores were from the *Age 16, Sweep 4 Arithmetic Test, 1986* (Dodgeon, n.d.) included 60 multiple choice questions. Literacy and matrices scores were from *Age 16, Sweep 4 Reading and Matrices Test, 1986* (Hodder and Stoughton,1979).

Free School Meal was from the *Age 16, Sweep 4, 1986* where respondents were asked if they received free school meals last week at lunch, yes=1 and 2=no. Additionally, number of school days lost, average number hours homework per week (1985) and school type were taken from this sweep.

Table 7: School type age 16

Value	Label
1	Comprehensive
2	Grammar
3	Secondary modern/technical
4	Independent private
5	LEA special
6	Independent special
7	Other
8	Scottish LEA

Leisure sport and reading were taken from the *Age 16, Sweep 4 Leisure and Television Diaries, 1986.* Where cohort members were asked to complete a four-day Leisure Diary (Altintas et al., 2022). Sport consisted of 'total time spent active sport, total time spent passive sport and, total time spent walks' added together. Reading involved total time spent 'read books' and 'read paper/magazine'.

Qualification was from the *Age 46, Sweep 10, 2016-2018 where asked* what type of qualification they have. The individuals later life income was from the *Age 46, Sweep 10, 2016-2018* where the variable used was (Derived in CAPI) Weekly take home income.Income at 46 was specifically chose over a younger age because older age is associated with stability as these individuals are mid-career with an established role they have likely been in for many years. This age gives a better reflection of steady earnings, that will not hugely change in nearby years.

For all of the data, negative values ranging from -9 to -1 meant not stated, not applicable, no data etc.

Table 8: Type of qualification obtained (at 46)

Value	Label
1	GCSE grade A-C
2	GCSE grade D or E
3	AS level grade A-C
4	AS level grade D or E
5	A level grade A-C
6	A level grade D or E
7	Access 3 qualification
8	Standard grade qualification
9	Intermediate 1 qualification
10	Intermediate 2 qualification
11	Highers
12	Advanced Highers
13	Certificate of Sixth Year Studies quali
15	Degree
16	Other degree level qualification
17	Higher degree
18	Nursing or paramedical qualification
20	Other teaching qualifications
21	BTEC
22	City and Guilds qualification
23	Level of RSA Qualif
24	Pitman qualification
25	NVQ qualification
26	GNVQ qualification
27	ONC/OND qualification
28	HNC/HND qualification
29	Trade apprenticeship qualification
32	Type of other qualification obtained

METHOD

To investigate the relationship between early life variables and later income cleaning of the data was required. Matrices score; arithmetic score and literacy scores were then turned into percentages to allow for a fairer comparison. For initial visual graphs the categoric data was used and for numerical analysis the numeric file was used. Statistical Analysis was then conducted in Rstudio. These processing steps ensured consistency and reliability throughout the dataset.

The quantitative research uses a linear regression model to test whether early life factors such as educational background, SES, and cognitive scores influence later life outcomes, adult income. This outlines the independent variables as the early life characteristics and the dependent variable as income at age 46. This analysis explores a potential causal link between childhood variables and income at age 46.

To investigate "*The role of early life variables in shaping economic futures*" the linear regression model was used to effectively analyse multiple independent variables, providing coefficients, standard errors, p values and R^2 which facilitates a clear explanation of the models fit.

Linear regression model: $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$

$$
\begin{aligned}
Income_i = \beta_0 &+ \beta_1 Sex + \beta_2 Birthweight + \beta_3 \text{Ethnic Group} + \beta_4 Height \\
&+ \beta_5 Cold\ and\ Cough\ regularity + \beta_6 Mothers\ Highest\ Qualification \\
&+ \beta_7 Fathers\ Highest\ Qualification + \beta_8 Number\ of\ Immunisations \\
&+ \beta_9 Weekly\ Family\ Income + \beta_{10} Social\ Class \\
&+ \beta_{11} Estimated\ Reading\ Age + \beta_{12} Average\ Homework\ Hours \\
&+ \beta_{13} School\ Days\ Lost + \beta_{14} Total\ Leisure\ Reading \\
&+ \beta_{15} Total\ Leisure\ Sports + \beta_{16} Free\ School\ Meals \\
&+ \beta_{17} Arithmetic\ Score + \beta_{18} Literacy\ Score + \beta_{19} Matrices\ Score \\
&+ \beta_{20} School\ Type + \beta_{21} Qualification + \in_i
\end{aligned}
$$

This was the initial model with all the variables of interest.

4: Results

SUMMARY STATISTICS

Table 9: Summary Statistics of all early life variables

Early Life Variable	N	Mean	SD
Arithmetic Score (out of 60)	3677	36.77	11.82
Average Homework Hours (weekly)	11614	3.31	3.13
Birthweight (ounces)	13133	113.28	26.85
Cold and cough regularity	2225	2.23	0.70
Estimated Reading Age (at 10)	14867	6.92	4.46
Ethnic Group	13133	1.17	0.67
Fathers Highest Qualification	13133	2.61	2.12
Free School Meals	11614	1.71	0.45
Height (in cm at 42 months)	2225	93.17	19.73
Literacy Score (out of 75)	3651	46.53	22.71
Matrices Score (out of 11)	3651	7.91	3.01
Mothers Highest Qualification	13133	2.08	1.55
Number of Immunisations	13133	4.19	1.38
Qualification	1847	25.61	7.40
School Days Lost (in last academic year at 16)	11614	5.35	10.99
School Type	18035	1.45	1.31
Sex	1938	1.53	0.50
Social Class (at 10)	14867	3.24	1.44
Total Reading (minutes)	7242	57.78	119.98
Total Sport (minutes)	7242	182.26	256.17
Weekly Family Income (at 10)	14867	2.87	1.60

The summary statistics in Table 9 indicate variation across early life variables in terms of both sample size and distribution. Cognitive test scores (arithmetic, literacy and matrices) are based on smaller subsamples, while broader demographic or socioeconomic indicators (e.g. parental qualifications, family income) are drawn from much larger samples. The means and standard deviations reflect substantial heterogeneity, particularly in measures like ‘Total Sport’ and ‘Total Reading’, where large standard deviations suggest significant differences in leisure time use across individuals. For categorical indicators such as Free School Meals, the mean suggests that majority of the sample did not receive free school meals, pointing toward a less deprived overall sample. Meanwhile, average parental qualifications fall at the lowermiddle end of the education scale. Indicating that most held vocational or O-level equivalent qualifications, but relatively few had degrees. These descriptive results provide a useful baseline for examining how early life conditions may influence later life outcomes, such as income. For income (weekly at 46) itself there were 8580 observations with a mean of £816.81, SD =15.

INITIAL VISUALS

Initial visual graphs were created to aid the understanding of the data. This can allow for spotting trends, for example if income at 46 is directly proportional to family income at 10.

Figure 1: Relationship between Free School Meals and Income

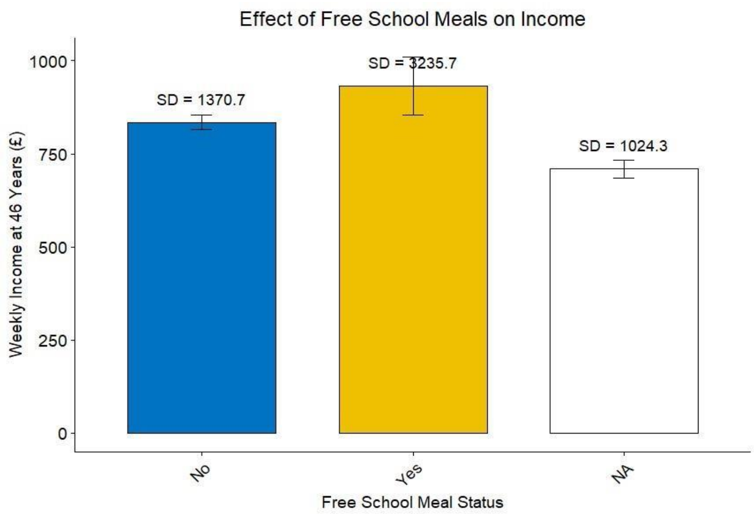


Figure 1 presents the average weekly income at age 46 based on FSM status during childhood. Interestingly, individuals who received FSM report slightly higher average incomes compared to those who did not. Although this pattern is visually notable, further statistical analysis is required to determine whether these differences are significant.

Figure 2: Relationship between family Income at age 10 on income at 46

Figure 2 illustrates the relationship between weekly family income at age 10 and weekly income at age 46. A clear upward trend is visible: individuals from higher-income households during childhood tend to earn more in adulthood. Those from families in the top bracket (£250+) have the highest average adult income. In contrast, those from poorest households (<£35) show the lowest adult income. This suggests a positive association between childhood family income and later-life earnings, though further statistical analysis is needed to confirm the strength and significance of this relationship.

Figure 3: Relationship between childhood social class and later Income

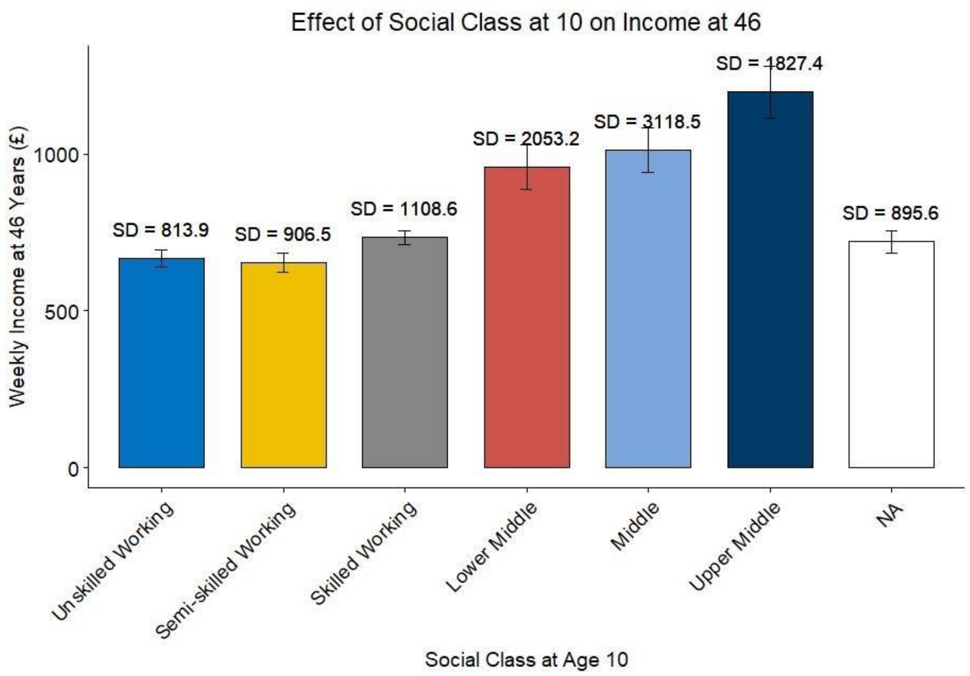


Figure 3 displays the relationship between childhood social class at age 10 and weekly income at age 46. The graph shows a clear positive trend: individuals from higher childhood social classes tend to earn more in adulthood. Those from upper middle-class backgrounds have the highest adult income, while those from unskilled or semi-skilled working-class backgrounds earn considerably less. This highlights the long-term economic impact of early social class positioning, though statistical analysis is needed to determine the significance of these differences.

Figure 4: Literacy scores at 16 and Income at 46

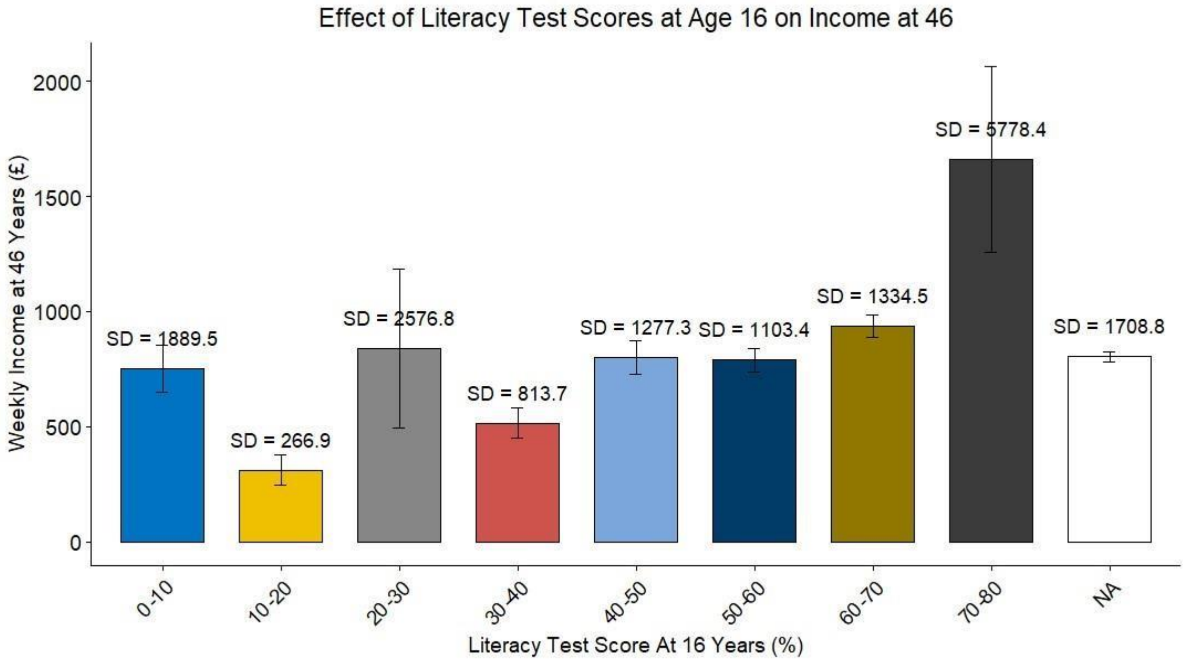


Figure 4 demonstrates the effect of literacy test scores on Income at 46. Whilst no clear trend is visible, the top scores (70-80%) have considerably higher earnings in later life. Scores of 20-30% surprisingly have the lowest incomes. This underscores the long-term economic significance of literacy skills acquired during teenage years. Again, further analysis is essential for discovering the significance of this correlation.

REVISING THE MODEL

Starting with the full model, individual coefficients and the complete model's statistics were compared to get a revised model that reflects a more accurate and reliable relationship between the dependent and independent variables. To do so beta and Spearman coefficients of each independent variable were compared. In this model β_1 shows how much income at 46 changes with a 1 unit increase in the independent variables. Comparing these 2 coefficients can identify which variables are suppressors in the model, they occur when the Spearman is greater than beta, or when the two coefficients have negative signs. These variables suppress irrelevant variance in those predictors, thus enhancing their regression weights, improving the overall fit of the model. (Pandey and Elliott, 2010) Number of immunisations, sex, leisure sport and social class were a few suppressors. Next a hierarchal regression was computed, where predictors were added one at a time to find the contributors of the model. This was carried out by examining the R^2, watching which variables make the model a better fit. Certain variables like Mothers and Fathers highest qualification didn't change the R^2 at all, so they were removed to protect the simplicity of the model. All three test scores (arithmetic, literacy, matrices) were highly correlated, which could cause multicollinearity issues. To address this, they could be combined or two of them removed, as arithmetic and matrices scores didn't contribute any variance to the model they were removed.

The revised model was:

$$Income = \beta_0 + \beta_1 School\ Type + \beta_2 Leisure\ Reading + \beta_3 Arithmetic\ Score +$$

$$\beta_4 Literacy\ Score + \beta_5 Free\ School\ Meals + \beta_6 Average\ Homework\ Hours +$$

$$\beta_7 Social\ Class + \beta_8 Cold\ and\ coughs + \beta_9 Height + \beta_{10} Number\ of\ Immunisations +$$

$$\epsilon_i$$

The statistics of this model were: $F(10,$

$157) = 6.493$, $p < 0.001$

$R^2 = 0.292$, adjusted $R^2 = 0.248$

These results show that the new overall model is statistically significant, so the set of predictors significantly explain variation in income at 46. Approximately 29.2% of the variance in income can be explained by the early life variables included in the revised model. After adjusting for the number of predictors a more modest value is displayed with 24.8% of the variation in income being explained by the model, accounting for the degrees of freedom.

This level of explained variance is moderate, especially considering the intricate nature of income determination in later life, which can be influenced by a wide range of additional factors (e.g. adult education, labour market experience, location, etc.) not captured in this model. Thus, this model is a moderate fit.

FINAL RESULTS

Table 10: Results of the revised model

Early Life Variable	Estimate	*p*
Arithmetic Score	6.38	0.132
Average Homework Hours	20.48	0.100
Cold & Coughs	73.53	0.173
Free School Meals	277.63	0.086*
Height (at 42 months)	0.71	0.659
Literacy Score	6.30	0.002***
No. of Immunisations	-56.01	0.045**
School Type	26.93	0.323
Social Class	67.72	0.015**
Total Reading (Minutes)	-0.52	0.092*

Note. *** indicates significance at the 1% level, ** is at the 5% level, * is at the 10% level.

Table 10 provides strong evidence to reject the null hypothesis (H_0) at the 5% level, which states that early life variables have no significant effect on income in later life. Literacy score is the most significant variable in predicting income at 46, significant at the 1% level. Social class (of fathers occupation at age 10) was also a significant predictor of income. Number of immunisations at age 5 is significant, but counterintuitive. Not receiving free school meals and total leisure reading were significant at the 10% level, only providing moderate evidence and can be considered ‘marginally’ significant. Arithmetic scores, average homework hours, height, school type and having cold and coughs more rarely (reverse coded data) have expected coefficient estimates but did not yield significant results.

POST HOC ANALYSIS

To ensure the robustness and validity of the model, a series of diagnostic tests were conducted following the initial statistical tests. These provide a deeper understanding of the underlying trends in the dataset, which were essential for verifying the assumptions underlying the model and ensuring that the results were reliable and accurate.

Table 11: Post Hoc Results

Analysis	Result
Multicollinearity	All VIF <2
Misspecification: Ramsay Reset	P=0.353
Normality	Histogram of std. residuals normally distributed
Linearity	No non-linear shapes in residuals scatter plot
Outliers: Cook's Distance	CD <1
Heteroscedasticity: Breusch-Pagan	P = 0.364

Table 11 observes the outcomes of the further diagnostic tests. Variance Inflation Factor (VIF) checked for multicollinearity. All the VIFs were less than 2 so not much variance inflation was present and there was no cause for concern. The Ramsay Reset test had a p value greater than 0.05, so there was no sign of misspecification. Both the histogram of standard residuals and QQ plot were normally distributed, as shown in Figure 5 and Figure 6 respectively.

Figure 5: Normality Test (Histogram of standard residuals)

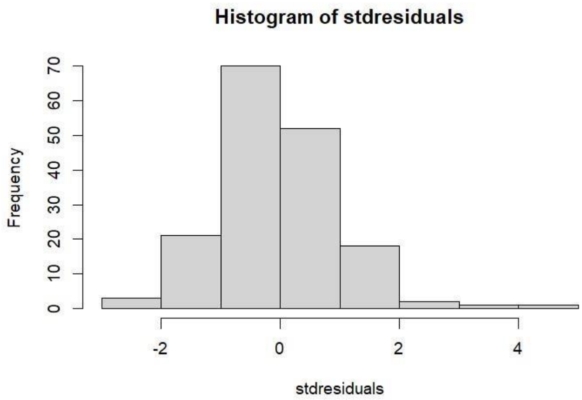
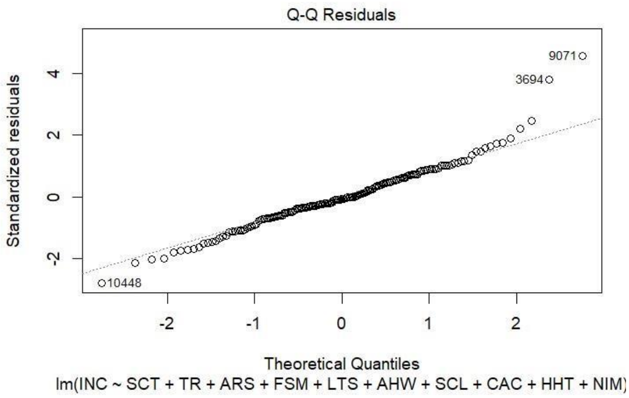


Figure 6: Normality Test (QQ Plot)



There were no non-linear shapes in the residuals scatter plot, confirming that the relationship between the independent and dependent variables was linear. Based on the Cook's Distance test, no data points seem to be significantly influencing the regression model, as all the values were less than 1. This indicates that the models fit, and parameter estimates are not being disproportionately affected by any one data point. A Breusch-Pagan test was conducted to check for heteroscedasticity. Since the p-value was greater than 0.05 (or 0.10 for a 10% significance level) there is no significant evidence of heteroscedasticity, and the model likely satisfies the homoscedasticity assumption of linear regression.

5: Discussion

INTERPRETATION OF RESULTS

With literacy scores being the most significant variable, an additional single point in the literacy test very likely increases weekly income (at age 46) by £6.30 on average. This result supports Heckman's (2006) argument that cognitive abilities do play a vital role in determining future outcomes. This could be down to individuals with a higher cognitive ability (and performing better on the literacy test) being able to process complex information more efficiently (Gladstone and Barrett, 2023). Additionally, cognitive abilities act as barriers to well-paid careers, with a large proportion of top earners in management/tech roles demonstrating above average literacy scores (OECD, 2013). Thus, investing in literacy early on can help promote better future earnings, and policy initiatives such as enhanced curricula and early screening for literacy challenges, to identify who need help are essential. Furthermore, policymakers could provide additional training for teachers of early years on literacy instruction and the most effective practices based on proven methods. Similar patterns in arithmetic scores further emphasise the importance of early cognitive development for future earnings.

The significance of childhood social class supports the notion that childhood SES has a lasting effect on adult earnings. With an increase of each social class as listed in Table 6, weekly income at 46 increases by £67.72. As Blanden and Machin (2004) found that degree attainment from students in the top family income quintile were much higher (37.1%) than in the lowest quintile (9%) in 1999, higher social classes are clearly correlated with higher education levels, which is known to significantly increase income with university graduates earning substantially more than high school graduates (Card, 1999). Apart from education level, children in higher social classes have access to greater learning materials (Lurie et al., 2021). Cognitive stimulation, such as language exposure and caregiver involvement is greater in high SES households and directly links SES to academic achievement (Ali, Bashir and Ahmad, 2021). Wealthier parents can also afford homes in desirable school catchment areas or pay for private education and extracurricular activities (Manstead, 2018), providing their child with more opportunities and experiences.

The negative coefficient (-56.01) displays that an additional immunisation is associated with a £56.01 decrease in weekly earnings. This could be explained by immunisations being inversely proportional to social class in the 1960/70s. Working class/poorer parents were more reliant on NHS services, who pushed mass immunisation programmes such as vaccines against measles in 1968 (Hand, 2016). The NHS provided immunisations free of charge, which particularly appealed to the poor. Additionally, public health campaigns in the 1960/70s were particularly targeted at poorer urban areas because of the Urban Aid Programme under the Local Government Grants (Social Need) Act 1969. The Community Development Projects (1969) also encouraged residents in deprived urban areas to participate in local services to address health inequalities (Bertrand, 2017). During the MMR vaccine controversy from 1990 to 2010 socio-economic factors influenced vaccine resistance significantly amongst the middle class (Holmberg, Blume and Paul Robert Greenough, 2017). Middle-class parents, despite being well-educated and informed, often questioned medical authorities and pharmaceutical companies, because of their mistrust and their exposure to media fabrication highlighting alleged vaccine risks. Wealthier groups also shared similar scepticism regarding government motives, which heightened their hesitancy. Meanwhile, lower-income groups were less prominently featured in discussions of the MMR controversy and less resistant to the vaccine. This evidence discussed displays how poorer families in the UK were more likely to get their children vaccinated than wealthier ones, especially during vaccine scares. As wealthier families were more likely to express their scepticism or completely refuse immunisations because of safety concerns or disbelief in the government.

The above results both stress the importance of government intervention at an early age to reduce social class disparities, by ensuring children from lower social classes have equal access to resources and opportunities so they can go on to perform just as well as their peers of greater SES.

Some would consider FSM ‘marginally’ significant at the 10% level, it is not significant at the 5% level. Marginally significant variables can give useful insights into the relationship. With a larger sample or one with less variability and missing data. For example, if FSM was measured on a continuous scale of need instead of binary the true relationship between FSM and income might stand out more clearly, resulting in a significant one. The coefficient estimate was large stating that not receiving free school meals (1=yes, 2=no) is associated with a £277.63 weekly income increase, £14,436.76 a year. This estimate builds on the relationship between FSM status and academic achievement established by Shuttleworth (1995), essentially stating that FSM recipients achieved fewer GCSEs by showing how these recipients earn £277.63 less than those who didn’t receive them (who also obtained more

GCSEs). If this was significant it would hugely emphasise the importance of reducing inequality for children, promoting more investment in poorer households with children. The initial visual graph (figure 1) contradicts these results stating that FSM recipients have higher weekly income. This can likely be explained by the graph showing raw averages of income by FSM status, before any adjustments for confounding variables (family income, social class) through a hierarchical regression. The regressions more complex treatments reveal that once confounding variables are dealt with not receiving FSM is linked with a substantial income boost. This discrepancy underscores the importance of controlling for confounding factors; what appears in a bivariate analysis may noticeably differ from the true effect.

Average weekly homework hours also trended towards significance (p-value 0.100), with a single additional hour of homework per week relating to a £20.48 weekly income increase. This reinforces ideas about how student effort is highly correlated with success, particularly for students from lower-income backgrounds (Jin, 2023).

Leisure reading with a p-value of 0.092, explains how an additional minute of leisure reading is connected to a 52p decrease in weekly earnings, somewhat opposing Moyer's (2007) idea of reading habits increasing learning potential. Although this cannot be said for certain as the present study did not directly measure leisure reading on education attainment. This result could highlight non-academic sources being read like cartoons or magazines, evidently proving that this type of literature does not enrich earning potentials. Otherwise, this negative effect could reflect the opportunity cost of other activities. Additionally, leisure reading could be a proxy for time spent alone, and anti-social behaviour. Suggesting that social interaction is key for higher earnings. As these variables are not actually significant, they do not give robust evidence, but useful reflections for future studies.

Despite the following being non-significant, understanding what doesn't contribute to later life income is important for a comprehensive review of the results. Non-significant variables can be extremely helpful for future studies with the same goal.

Cold and cough regularity stated that rarer frequencies of illness at 42 months, are correlated with an additional £73.53 per week. Although school type was positively associated with income at age 46, this effect was not statistically significant after accounting for early life SES, cognitive ability, and health indicators. This may suggest that the type of school attended plays a less critical role than the broader socioeconomic and individual factors shaping later-life income. Alternatively, it may reflect overlapping influence between school type as private schools are primarily attended by children from higher economic backgrounds so they're so their effect may be absorbed by those variables.

There is no strong evidence that early height (in cm) is predictive of later income once other early life factors are controlled for, different to Deshpande and Ramachandran's (2022) notion that stunting influences educational outcomes. Early height can be a good indicator of early health and nutrition because it can be correlated with early life adversity and household food security. The lack of association may reflect the mediating effects of other variables like social class or education. Height does not directly predict income in this model, but it reflects early health conditions, signalling the importance of health equity, prenatal care, and nutritional support in early life.

STRENGHTS OF THIS STUDY

This study has a longitudinal data advantage from using the 1970 British cohort study, which gives a fuller view than cross sectional data where you cannot observe any trends. The choice of the 1970 BCS over another similar study like the millennium cohort study (MCS) was because the 1970 study has older participants now in there mid 50s with well-established careers and earnings, which allows for a more long-term analysis. Unlike the MCS where participants are in there mid 20s and share a vast level of career uncertainty, using the MCS would give an unrepresentative result.

This study could act as a strong longitudinal foundation for future studies that would compare the 1970 BCS to other cohort studies like the 1958 BCS. Cross-national cohort studies could also be useful to provide how early life variables effect income in different countries, with different education systems and work attitudes. This comparison would identify trends between countries and provide stronger evidence for the most influential variables, if similar results were noted multiple times.

Including variables that are likely related like school type and social class reduces the chance of omitted variable bias. Omitted variable bias arises in linear regression when key independent variables are excluded from the model. This omission can distort the estimated coefficients if the missing variable has a correlation with the dependent variable or multiple included independent variables (Nikolopoulou, 2022). So, by using related variables the model is strengthened by making it more accurate, less biased and more credible for producing a real relationship.

The use of further statistical analysis that was earlier discussed and shown in Table 12 enhanced the reliability and validity of this model by confirming that there were no major flaws in the model that could jeopardise the credibility of the results.

LIMITATIONS OF THIS STUDY

This study did have a relatively large amount of missing data values, with their being only 4 fully comprehensive data participants for the 22 initial variables collected. Despite this, there were still approximately 2000 useful pieces of data. Because of this amount of missing data, the model could falsely reject the null hypothesis or have biased parameter estimates (Kang, 2013). These distortions can threaten the validity of the model, leading to invalid results and conclusions. To address this multiple imputation techniques could be used to estimate missing values. Multiple imputation involves two main steps, imputation where multiple replacement values are generated for missing data based on characteristics (relationships, patterns, distributions) of the dataset (Li, Stuart and Allison, 2015). Then analysing each imputed data

set independently as if no data were missing and combining the results for a comprehensive understanding. Ultimately using multiple imputation could minimise bias by preserving relationships between variables (Lall, 2016), which can lead to more reliable *SE* and *P* values (Pedersen et al., 2017). However, it is not advised to use multiple imputation techniques on large amounts of missing data (Hyuk Lee and Huber Jr., 2021), which this study had. Variables with high missingness yield larger standard errors, limiting the ability to detect significant effects (Hyuk Lee and Huber Jr., 2021). Moreover, with extensive missing data, it becomes harder to include all variables predictive of missingness in the imputation model, increasing the risk of violating the Missing-at-Random assumption (Chaput-Langlois et al., 2024). If future studies have a relatively small amount of missing data, they could use multiple imputation techniques to restore the sample size and reduce bias.
Additionally, there may still be multicollinearity issues as FSM, family income and social class are likely to be highly correlated. Despite all the VIFs being low, showing there was no serious multicollinearity in the model, these 2 variables could be theoretically related because what they measure overlap (family background, SES). VIF measures overall multicollinearity across the whole model (Hayes, 2024) but fails to always detect strong pairwise correlations between two variables, unless they are extremely noticeable. To check for this, sample-based trimming could be used (Zhu et al., 2021). This involves identifying non-overlapping regions and excluding subjects in these non-overlapping groups. Here, the maximum values of weekly family income for those receiving FSM could be identified, add a 5% buffer then use that value to remove high income values from the sample. This strategy resembles Crump et al.'s trimming method of removing extreme values and addresses potential covariate imbalances for a more robust analysis (Zhu et al., 2021).

However, as family income was recorded in groups (displayed in Table 5) the individual maximum value could not be identified. Alternatively, data from the lowest FSM recipients' group and the group above could just be used; this would create a poor versus almost poor sample, removing middle class and richer children who might distort the result. Yet this would lead to a less accurate comparison as income bands don't report the exact income of each individual and the distribution of the grouped income data is not known. Some children in higher family income bands may be well above the FSM threshold, but others just made it into that group, so generalising would rule out individuals that could be useful. Lastly, this would involve making too many assumptions that may not be true which could cause selection bias, small sample bias and an outcome that cannot be generalised to the entire sample and rest of the population.

Although this limitation is about a non-significant variable it is worth mentioning that the coding of the school type responses (shown in Table 7) was not strictly ordinal, but were treated as numerical in the model, meaning that the interpretation can be distorted. However, this variable could not have been treated as continuous because the model would then assume that moving up through the categories was a linear progression which it was not. This is because the responses where 1=comprehensive, 2=grammar, 3=secondary modern etc are not strict improvements on one another.

Despite many other deeper analysis tests being conducted, an endogeneity test was not. Endogeneity describes a situation where the independent variable is correlated with unexplained variation, or the error term, in the dependent variable (Kovac, 2023).
In regression analysis, this occurs when the predictor variable influences or is influenced by the error term, leading to biased estimates. This issue is particularly significant because it can compromise the validity of statistical inferences, potentially resulting in misleading conclusions.

This study had a standard cross-sectional regression model because it was modelling at a single outcome (income at 46); and an imbalanced data panel meaning that there were different variables recorded at different times, not a standard set of data collected at each point in time. This makes an endogeneity test very difficult. To perform an endogeneity test on panel data a Hausman test can be carried out, which compares the Ordinary Least Squares (OLS) estimator with an alternative estimator (Baltagi, 2014). Due to the high amount of missing data, and missing data levels being inconsistent for each variable, a Hausman test cannot compare the standard OLS to a two-stage least squares (2SLS). Furthermore, potentially endogenous variables do not have their valid instruments in the dataset, so alternative variables would have to be used, for example, using mother's and father's highest qualification as an instrument for social class of free school meals.

The alternative here would be to do a manual calculation of the Hausman value, comparing the variance-covariance matrices of coefficients. Doing this compares the difference in coefficients between your OLS and 2SLS (IV) models, and tests whether that difference is statistically significant, given the variance of the difference (Wooldridge, 2010).

However, the variance-covariance matrices would not have been aligned because many have missing values, and some sweeps are missing data completely. To fix this it would involve removing variables from one model so both models include the same set or adding variables to align both models. Manually forcing the models to match is risky because it can increase the chance of errors in the model. Removing variables just to make OLS and IV match can lead to omitted variable bias if those variables are important to the outcome. Adding variables to the IV model without have strong, valid instruments for them creates a risk of weak instrument bias (Greene, 2012). As there were 18,000 respondents doing this would be extremely time consuming and increase the chance of inaccurate results by human error. Ultimately there are too many risks involved to get a result that may be invalid.

Nevertheless, this model was tested with several other strong measures confirming the reliability of the model, so the model's results are still presented with confidence. Because of this, future studies should aim to collect a balanced panel, with the same set of variables collected over all time periods. This would facilitate a smoothly run Hausman (endogeneity) test.

Lastly, even further long-term analysis could be performed by using the Age 51 Sweep to measure the effect of early life variables on income at 51, instead of 46. This sweep was not used in the current study as it was only released in March of this year (CLS, 2025), after this statistical analysis had taken place.

6: Conclusion

This paper has investigated the role of early life variables in determining later life income (at 46), using the 1970 British cohort study. The evidence discussed shows how higher social class at age 10 and, literacy scores at 16 increase income. On the contrary, number of immunisations reflecting underlying socioeconomic differences, was associated with lower earnings. This suggests that early life variables and parental SES may interact in ways more than meets the eye, warranting further investigation. The statistically significant outcomes from this analysis confirm the notion that economic advantages can be conferred early in life, setting individuals on paths toward higher future earnings.

Nevertheless, this study acknowledges that the factors that determine adult income go far beyond early life variables. Future research should adopt a longitudinal approach that tracks these variables over time using more detailed and balanced data to pinpoint when interventions have the greatest impact. To summarise, while the role of early life variables in shaping future income is just one aspect of a very detailed framework. The evidence presented provides a compelling start as to why policies that broaden childhood opportunities could promote economic well-being and reduce deep-rooted inequality.

Bibliography

Ali, R., Bashir, F. and Ahmad, R. (2021). Imprints of Lower Socioeconomic Class in English Speaking Anxieties and Academic Performance of Rural and Urban Students. *iRASD journal of economics*, [online] 3(3). doi:https://doi.org/10.52131/joe.2021.0303.0055.

Altintas, E., Campbell, S., Dennison, K., Gershuny, J., Peters, A., Shepherd, P., (2022) 1970 British Cohort Study Age 16 Survey Leisure and Television Diaries User Guide. London: UCL Centre for Longitudinal Studies.

Baltagi, B.H. (2014). Panel Data and Difference-in-Differences Estimation. *Elsevier eBooks*, pp.425–433. doi:https://doi.org/10.1016/b978-0-12-375678-7.00720-3.

Bertrand, M. (2017) The National Community Development Projects in the United Kingdom, 1969-78. 1970-79: Community in the UK. Intercalaires: Agrégation d'anglais, 1, pp.11 - 31, ffhal-01714925f

Blanden, J. and Machin, S. (2004). Educational Inequality and the Expansion of UK Higher Education. *Scottish Journal of Political Economy*, 51(2), pp.230–249

Butler, K. (2020). *The origins of the 1970 British Cohort Study | CLS*. [online] CLS. Available at: https://cls.ucl.ac.uk/the-origins-of-the-1970-british-cohort-study/ [Accessed 2 Apr. 2025].

Busby, E. (2021). *Pupils with higher GCSE grades can expect lifetime earnings boost report*. Retrieved from The Standard: https://www.standard.co.uk/news/uk/gcse-departmentforeducation-geoff-barton-nick-gibb-grade-b944871.html

Bynner, J., Butler, N., University of London, Institute of Education, Centre for Longitudinal Studies, 2024, *1970 British Cohort Study: Age 16, Sweep 4, 1986*, [data collection], UK Data Service, *9th Edition,* Accessed 4 April 2025. SN: 3535, DOI: http://doi.org/10.5255/UKDASN-3535-6

Cataloguementalhealth.ac.uk. (2024). *Catalogue of Mental Health Measures - 1970 British Cohort Study (BCS70)*. [online] Available at: https://www.cataloguementalhealth.ac.uk/?content=study&studyid=BCS [Accessed 5 Apr. 2025].

Card, D. (1999). The Causal Effect of Education on Earnings. *Handbook of Labor Economics*, [online] 3(A), pp.1801–1863. doi:https://doi.org/10.1016/s1573-4463(99)03011-4.

Centre for Longitudinal Studies. (2024). *MCS Age 23 Sweep*. Retrieved from https://cls.ucl.ac.uk/cls-studies/millennium-cohort-study/mcs-age-23-sweep/

Chaput-Langlois, S., Stickley, Z.L., Little, T.D. and Rioux, C. (2024). Multiple Imputation When Variables Exceed Observations: An Overview of Challenges and Solutions. *Collabra. Psychology*, 10(1). doi:https://doi.org/10.1525/collabra.92993.

Chamberlain, R., 2024, *1970 British Cohort Study: 42-Month Subsample, 1973*, [data collection], UK Data Service, *2nd Edition,* Accessed 4 April 2025. SN: 2690, DOI: http://doi.org/10.5255/UKDA-SN-2690-1

Chamberlain, S. a. (2001). *Improving the take up of free school meals.* Department for Education and Employment.

CLS. (2025). *CLS | BCS70 Age 51 Sweep*. [online] Available at: https://cls.ucl.ac.uk/clsstudies/1970-british-cohort-study/bcs70-age-51-sweep/ [Accessed 27 Apr. 2025].

Department for Education. (2024, 6 6). *Schools, pupils and their characteristics.* Retrieved from GOV.UK: https://explore-education-statistics.service.gov.uk/find-statistics/schoolpupilsand-their-characteristics

Deshpande, A. and Ramachandran, R. (2022). Early childhood stunting and later life outcomes: A longitudinal analysis. *Economics & Human Biology*, 44, p.101099. doi:https://doi.org/10.1016/j.ehb.2021.101099.

Dobson, K.G., Ferro, M.A., Boyle, M.H., Schmidt, L.A., Saigal, S. and Van Lieshout, R.J. (2017). Socioeconomic Attainment of Extremely Low Birth Weight Survivors: The Role of Early Cognition. *Pediatrics*, 139(3). doi:https://doi.org/10.1542/peds.2016-2545.

D'Addio, A. (2007), "Intergenerational Transmission of Disadvantage: Mobility or Immobility Across Generations?", *OECD Social, Employment and Migration Working Papers*, No. 52, OECD Publishing, Paris, https://doi.org/10.1787/217730505550.

Dodgeon, B. (n.d.). *Centre for Longitudinal Studies CLS Guide to the Dataset BCS70 16year follow-up APU Arithmetic Test CLS Cohort Studies*. [online] Available at: https://doc.ukdataservice.ac.uk/doc/6095/mrdoc/pdf/bcs70_16year_arithmetic_test_guide_to_the_dataset.pdf [Accessed 4 Apr. 2025].

EDWARDS, R.M. (1975). Race and Class in Early Childhood Education. *Young Children*, [online] 30(6), pp.401–411. doi:https://doi.org/10.2307/42658416.

Fergusson, D.M. and Woodward, L.J. (2002). Mental Health, Educational, and Social Role Outcomes of Adolescents With Depression. *Archives of General Psychiatry*, [online] 59(3), p.225. doi:https://doi.org/10.1001/archpsyc.59.3.225.

Fergusson, D.M., John Horwood, L. and Ridder, E.M. (2005). Show me the child at seven: the consequences of conduct problems in childhood for psychosocial functioning in adulthood. *Journal of Child Psychology and Psychiatry*, 46(8), pp.837–849. doi:https://doi.org/10.1111/j.1469-7610.2004.00387.x.

Francis-Devine, B. and Orme, S. (2024). *Income inequality in the UK*. [online] Available at: https://researchbriefings.files.parliament.uk/documents/CBP-7484/CBP-7484.pdf.

Gladstone, J. and Barrett, J.D. (2023). Understanding the functional form of the relationship between childhood cognitive ability and adult financial well-being. *PLOS ONE*, 18(6), pp.e0285199–e0285199. doi:https://doi.org/10.1371/journal.pone.0285199.

Greene, W.H. (2012). *Econometric analysis*. 7th ed. Boston, Mass.: Pearson.

Goddeeris, J.H., Saigal, S., Boyle, M.H., Paneth, N., Streiner, D.L. and Stoskopf, B. (2010). Economic Outcomes in Young Adulthood for Extremely Low Birth Weight Survivors. *PEDIATRICS*, [online] 126(5), pp.e1102–e1108. doi:https://doi.org/10.1542/peds.2010-1242.

Gorard, S. (2013). *Who is eligible for free school meals? Characterising free school meals as a measure of disadvantage in England.* British Educational Research Journal.

GOV.UK. (1980). *Education Act 1980.* Retrieved from legislation.go.uk: https://www.legislation.gov.uk/ukpga/1980/20/section/38/enacted?view=extent#extent%3Dengland

Hand, J. (2016). *Childhood Vaccination and the NHS*. [online] People's History of the NHS. Available at: https://peopleshistorynhs.org/encyclopaedia/childhood-vaccination-andthenhs/.

Hayes, A. (2024). *Multicollinearity*. [online] Investopedia. Available at: https://www.investopedia.com/terms/m/multicollinearity.asp.

Heckman, J.J. (2006). Skill Formation and the Economics of Investing in Disadvantaged Children. *Science*, 312(5782), pp.1900–1902

Hill, H. D., Morris, P., Gennetian, L. A., Wolf, S., & Tubbs, C. (2013). The Consequences of Income Instability for Children's Well-Being. *Child development perspectives*, *7*(2), 85–90. https://doi.org/10.1111/cdep.12018

Hodder and Stoughton (1979) The Shortened Edinburgh Reading Test. Kent: Hodder and Stoughton Educational Ltd

Holmberg, C., Blume, S., & Greenough, P. (Eds.). (2017). *The Politics of Vaccination: A Global History*. Manchester University Press. http://www.jstor.org/stable/j.ctt1wn0s1m

Hyuk Lee, J. and Huber Jr., J.C. (2021). Evaluation of Multiple Imputation with Large Proportions of Missing Data: How Much Is Too Much? *Iranian Journal of Public Health*, 50(7). doi:https://doi.org/10.18502/ijph.v50i7.6626.

Institute for Fiscal Studies. (2024). *Living standards, poverty and inequality in the UK*. [online] Available at: https://ifs.org.uk/living-standards-poverty-and-inequality-uk.

Jin, X. (2023). The role of effort in understanding academic achievements: empirical evidence from China. *European Journal of Psychology of Education*, [online] 39, pp.389–409. doi:https://doi.org/10.1007/s10212-023-00694-5.

Joseph Rowntree Foundation (2025). *UK Poverty 2025: The essential guide to understanding poverty in the UK*. [online] Joseph Rowntree Foundation. Available at:

https://www.jrf.org.uk/uk-poverty-2025-the-essential-guide-to-understanding-poverty-intheuk.

Kaasa, A. (2005). *Factors of Income Inequality and Their Influence Mechanisms: A Theoretical Overview*. [online] papers.ssrn.com. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=868491.

Kang, H. (2013). The Prevention and Handling of the Missing Data. *Korean Journal of Anesthesiology*, 64(5), pp.402–406. doi:https://doi.org/10.4097/kjae.2013.64.5.402.

Kovac, D. (2023). *What is Endogeneity?* [online] Statistics Solutions. Available at: https://www.statisticssolutions.com/what-is-endogeneity/.

Lall, R. (2016). How Multiple Imputation Makes a Difference. *Political Analysis*, 24(4), pp.414–433. doi:https://doi.org/10.1093/pan/mpw020.

Li, P., Stuart, E.A. and Allison, D.B. (2015). Multiple Imputation: A Flexible Tool for Handling Missing Data. *JAMA*, 314(18), p.1966. doi:https://doi.org/10.1001/jama.2015.15281.

Lurie, L.A., Hagen, M.P., McLaughlin, K.A., Sheridan, M.A., Meltzoff, A.N. and Rosen, M.L. (2021). Mechanisms Linking Socioeconomic Status and Academic Achievement in Early childhood: Cognitive Stimulation and Language. *Cognitive Development*, 58(101045), p.101045. doi:https://doi.org/10.1016/j.cogdev.2021.101045.

Legislation.gov.uk. (2025). *Local Government Grants (Social Need) Act 1969*. [online] Available at: https://www.legislation.gov.uk/ukpga/1969/2/enacted [Accessed 13 Apr. 2025].

L. J. Schweinhart et al., Lifetime Effects: The High/ScopePerry Preschool Study Through Age 40 (High/Scope,Ypsilanti, MI, 2005)

L. N. Masse, W. S. Barnett, A Benefit Cost Analysis of theAbecedarian Early Childhood Intervention (RutgersUniversity, National Institute for Early EducationResearch, New Brunswick, NJ, 2002

Manstead, A.S.R. (2018). The psychology of social class: How socioeconomic status impacts thought, feelings, and behaviour. *British Journal of Social Psychology*, [online] 57(2), pp.267–291. doi:https://doi.org/10.1111/bjso.12251.

Moyer, J.E. (2007). Learning From Leisure Reading: A Study of Adult Public Library Patrons. *Reference & User Services Quarterly*, [online] 46(4), pp.66–79. doi:https://doi.org/10.2307/20864750.

Nikolopoulou, K. (2022). *What Is Omitted Variable Bias? | Definition & Examples*. [online] Scribbr. Available at: https://www.scribbr.com/research-bias/omitted-variable-bias/.

North, A. (2022). *The long, strange history of anti-vaccination movements*. [online] Vox. Available at: https://www.vox.com/the-goods/22958419/covid-vaccine-mandatepandemichistory.

OECD (2013) Skills Matter: Further Results from the Survey of Adult Skills. Paris: OECD Publishing. Available at: http://www.oecd.org/skills/piaac/SkillsMatter-Further-Resultsfromthe-Survey-of-Adult-Skills.pdf

OECD (2021). *Income inequality*. [online] OECD. Available at: https://www.oecd.org/en/data/indicators/income-inequality.html.

ONS (2024). *Employee Earnings in the UK*. [online] Ons.gov.uk. Available at: https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/earningsandworkinghours/bulletins/annualsurveyofhoursandearnings/2024.

Oster, E. (2017). Unobservable Selection and Coefficient Stability: Theory and Evidence. *Journal of Business & Economic Statistics*, 37(2), pp.187–204. doi:https://doi.org/10.1080/07350015.2016.1227711.

Pandey, S. and Elliott, W. (2010). Suppressor Variables in Social Work Research: Ways to Identify in Multiple Regression Models. *Journal of the Society for Social Work and Research*, 1(1), pp.28–40. doi:https://doi.org/10.5243/jsswr.2010.2.

Pedersen, A., Mikkelsen, E., Cronin-Fenton, D., Kristensen, N., Pham, T.M., Pedersen, L. and Petersen, I. (2017). Missing Data and Multiple Imputation in Clinical Epidemiological Research. *Clinical Epidemiology*, [online] 9, pp.157–166. doi:https://doi.org/10.2147/clep.s129785.

Richards, M. and Abbott, R., 2009. Childhood mental health and adult life chances in postwar Britain: insights from three national birth cohort studies.

Richards, M. and Huppert, F.A. (2011). Do positive children become positive adults? Evidence from a longitudinal birth cohort study. *The Journal of Positive Psychology*, 6(1), pp.75–87. doi:https://doi.org/10.1080/17439760.2011.536655.

Rutter, M. (1967). A CHILDREN'S BEHAVIOUR QUESTIONNAIRE FOR COMPLETION BY TEACHERS: PRELIMINARY FINDINGS. *Journal of Child Psychology and Psychiatry*, 8(1), pp.1–11. doi:https://doi.org/10.1111/j.14697610.1967.tb02175.x.

Rutter, M., Kim-Cohen, J. and Maughan, B. (2006). Continuities and discontinuities in psychopathology between childhood and adult life. *Journal of Child Psychology and Psychiatry*, 47(3-4), pp.276–295. doi:https://doi.org/10.1111/j.1469-7610.2006.01614.x.

Seabrook, G. and Murphy, T. (2017). *Reading and Matrices Tests*. [online] Available at: https://doc.ukdataservice.ac.uk/doc/8288/mrdoc/pdf/bcs70_1986_reading_and_matrices_tests data_note.pdf [Accessed 4 Apr. 2025].

Shuttleworth, I. (1995). The Relationship between Social Deprivation, as Measured by Individual Free School Meal Eligibility, and Educational Attainment at GCSE in Northern Ireland: A Preliminary Investigation. *British Education Research Journal*.
Taylor, C. (2018). The Reliability of Free School Meal Eligibility as a Measure of SocioEconomic Disadvantage: Evidence from the Millenium Cohort Study in Wales. *British Journal of Educational Studies*.

Trust for London (2022). *Poverty definitions and thresholds*. [online] Trust for London. Available at: https://trustforlondon.org.uk/data/poverty-thresholds/.

van der Erve, L., Krutikova , S., Macmillan , L. and Sturrock, D. (2023). *Intergenerational mobility in the UK*. [online] ifs.org.uk. Available at: https://ifs.org.uk/inequality/intergenerational-mobility-in-the-uk/.

UKDataService (2024). *End User Licence Agreement*. [online] Available at: https://ukdataservice.ac.uk/app/uploads/cd137-enduserlicence.pdf.

University College London, UCL Institute of Education, Centre for Longitudinal Studies, 2024, *1970 British Cohort Study: Age 16, Sweep 4 Reading and Matrices Tests, 1986*, [data collection], UK Data Service, Accessed 4 April 2025. SN: 8288, DOI: http://doi.org/10.5255/UKDA-SN-8288-1

University of London, Institute of Education, Centre for Longitudinal Studies, Osborn, A., Dowling, S., Butler, N., 2024, *1970 British Cohort Study: Age 5, Sweep 2 1975*, [data collection], UK Data Service, *5th Edition,* Accessed 4 April 2025. SN: 2699, DOI: http://doi.org/10.5255/UKDA-SN-2699-4

University of London, Institute of Education, Centre for Longitudinal Studies, Butler, N., Bynner, J., 2024, *1970 British Cohort Study: Age 10, Sweep 3, 1980*, [data collection], UK Data Service, *7th Edition,* Accessed 4 April 2025. SN: 3723, DOI: http://doi.org/10.5255/UKDA-SN-3723-8

University of London, Institute of Education, Centre for Longitudinal Studies, 2024, *1970 British Cohort Study: Age 16, Sweep 4 Arithmetic Test, 1986*, [data collection], UK Data Service, *2nd Edition,* Accessed 4 April 2025. SN: 6095, DOI: http://doi.org/10.5255/UKDASN-6095-2

University of London, Institute of Education, Centre for Longitudinal Studies, 2024, *1970 British Cohort Study: Age 46, Sweep 10, 2016-2018*, [data collection], UK Data Service, *2nd Edition,* Accessed 4 April 2025. SN: 8547, DOI: http://doi.org/10.5255/UKDA-SN-8547-2

University College London, Institute of Education, Centre for Longitudinal Studies, 2024, *1970 British Cohort Study: Age 16, Sweep 4 Leisure and Television Diaries, 1986*, [data collection], UK Data Service, Accessed 4 April 2025. SN: 8949, DOI: http://doi.org/10.5255/UKDA-SN-8949-1

Wooldridge, J.M. (2010). *Econometric Analysis of Cross Section and Panel Data*. [online] *JSTOR*. The MIT Press. Available at: https://www.jstor.org/stable/j.ctt5hhcfr [Accessed 16 Apr. 2025].

Zhu, Y., Hubbard, R.A., Chubak, J., Roy, J. and Mitra, N. (2021). Core concepts in pharmacoepidemiology: Violations of the positivity assumption in the causal analysis of observational data: Consequences and statistical approaches. *Pharmacoepidemiology and Drug Safety*, 30(11), pp.1471–1485. doi:https://doi.org/10.1002/pds.5338.