Exploiting Genetics to Understand Paediatric Eye Disorders

A thesis submitted to Cardiff University for the degree of Doctor of Philosophy

By

Denis Plotnikov

School of Optometry and Vision Sciences

Cardiff University

JULY 2020

Supervised by Professor

Jeremy A. Guggenheim

AKNOWLEDGEMENTS

I would like to express my gratitude to Professor Jeremy A. Guggenheim for his supervision throughout this project. His support, patience and encouragement made this thesis possible. I would also like to thank my academic advisor Dr Ashley Wood for his feedback on my work.

It has been my great privilege and pleasure to be able to work with Yvonne, Rupal, Alfred and Neema.

Thank you to all in the School of Optometry and Vision Sciences for your support throughout my PhD.

I owe a great gratitude to my wife, Olga and to my daughter, Polina, for their love to me and endless faith in me.

SUMMARY

The primary aim of this series of investigations was to identify *causal* risk factors for paediatric eye disorders (e.g. myopia, hyperopia, strabismus and amblyopia). A series of hypotheses was tested in this project by exploiting state-of-the-art genetic and epidemiological approaches such as genome-wide association studies (GWAS) and Mendelian randomisation (MR).

A GWAS for strabismus was performed in order to identify genetic variant(s) conferring susceptibility to the condition. The genetic variant most strongly associated with the phenotype was rs75078292. This SNP, is situated in an LD block on chromosome 17, containing the genes *TSPAN10, NPLOC4* and *PDE6G*. A non-synonymous variant in *TSPAN10* – in very high LD with rs75078292 – was previously reported to be associated with myopia. These findings were replicated in a cohort of children (ALSPAC).

A Mendelian randomisation approach (both one- and two-sample MR) was used in order to estimate the effect of birth weight within the normal range (2.5 to 4.5kg) on refractive error. The analyses supported the hypothesis that birth weight within normal range plays a causal role in refractive error development.

The influence of education on refractive error development was estimated using a Regression Discontinuity design. In this work in UK Biobank, the Raising Of School Leaving Age (ROSLA) 1972 reform was used as a natural experiment; with use of the genetic data such as Principal Components and Polygenic Risk Scores for educational attainment and for refractive error as covariates. The estimated influence of an additional year of education on refractive error development was statistically significant, supporting existing evidence.

The hypothesis that hyperopia is causal risk factor for lower educational attainment was tested using non-linear Mendelian randomisation analysis. The results of the analyses revealed that the relationship between refractive error and educational attainment in UK Biobank is non-linear, but did not support the main hypothesis.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

25(OH)D	25-hydroxycholecalciferol
7-MX	7-Methylxanthine
95% CI	95% Confidence Interval
ALSPAC	Avon Longitudinal Study of Parents and Children
ATE	Average Treatment Effect
CNV	Copy Number Variant
CREAM	Consortium for Refractive Error and Myopia
CSE	Certificate of Secondary Education
D	Dioptre
DNA	Deoxyribonucleic Acid
EGG	Early Growth Genetics
eQTL	Expression Quantitative Trait Locus
GCSE	General Certificate of Secondary Education
GRM	Genetic Relatedness Matrix
GTEx	Genotype Tissue Expression
GWAS	Genome-wide Association Study
H^2	Broad-sense Heritability
h^2	Narrow-sense Heritability
HNC	Higher National Certificate
HND	Higher National Diploma
HRC	Haplotype Reference Consortium
INFO	Imputation Quality
IV	Instrumental Variable
IVW	Inverse Variance Weighting
kg	Kilogram
LATE	Local Average Treatment Effect
LD	Linkage Disequilibrium
logMAR	Logarithm of the Minimum Angle of Resolution
MAF	Minor Allele Frequency
MR	Mendelian Randomisation
NLMR	Non-linear Mendelian Randomisation
NVQ	National Vocational Qualification
OK	Orthokeratology
OR	Odds Ratio
PAR	Population Attributable Risk
PC	Principal Component
PRS	Polygenic Risk Score
RD	Regression Discontinuity
ROSLA	Raising of School Leaving Age
SNP	Single Nucleotide Polymorphism
SSGAC	Social Science Genetic Association Consortium
UK	United Kingdom
USA	United States of America

V_A	Additive Genetic Effect
V_D	Dominance Genetic Effect
V_E	Environmental Variance
V_G	Total Genetic Variance
V_I	Epistatic Genetic Effect
V_P	Phenotypic Variance

Chapter 1. General Introduction

1.1 Genetics

1.1.1 The Genome, Genes and Polymorphisms

The *genome* refers to the complete set of genetic information stored in an organism's chromosomes. All the essential information necessary for a normally-functioning human organism is coded in the DNA of the human genome, composed of the four nucleotides (DNA bases) adenine (A), cytosine (C), guanine (G), thymine (T). The human genome consists of approximately 3.2 billion nucleotides of DNA, divided into 22 somatic and 2 sex chromosomes.

The term *gene* refers to a small fraction of DNA, typically containing the information required for the synthesis of a specific protein. However, less than 2% of the human genome is believed to provide instructions for building proteins (1). Human genes vary in size from hundreds to hundreds of thousands of DNA bases. Genes are the functional unit of heredity.

The genomes of individual humans are 99.9% identical. The remaining 0.1%, or the genetic variation, partially determines the difference in phenotype between people. Single nucleotide polymorphisms (SNP) are the most common form of human genetic variation (2). A SNP is a change of a single nucleotide at a specific position in the human genome (e.g. when the majority of the population has a thymine nucleotide at a specific position, but in some individuals an adenine nucleotide is found at the same position). Different nucleotides found in that case are called *alleles*. SNP found in a coding region can be categorised into 2 types: synonymous, which do not lead to a change in the protein sequence and non-synonymous, affecting the protein sequence. In the case of a nucleotide substitute that results in a codon that codes for a different amino acid, this type of SNP is called a *missense* variant or a *nonsense* variant when the change results in a stop codon.

Insertion or deletion variants, known as *indels*, are another type of genetic variation. They occurs when a section of DNA (from a few bases to hundreds of nucleotides) is present in some individuals but not in others (3). Copy number variations (CNV) are a class of insertion, deletion or a duplication of a DNA segment with a size more than 1000 bases (1kb) (4).

1.1.2 Complex Traits

The term *trait* in genetics refers to a phenotype that varies between individuals but shows a certain level of stability across time (5). *Mendelian* (or *monogenic*) traits are inherited in concordance with Mendelian principles, i.e. a dominant or recessive mode of inheritance. *Complex* (or *quantitative*) traits result from the actions of (or interaction between) multiple genetic variants and environmental factors (6, 7). Height, refractive error, diabetes, cancer and most common diseases are examples of complex traits.

1.1.3 Heritability

The total amount of variation in a phenotype (variance of a trait) can be partitioned into the following two components: genetic variance and environmental variance (8, 9) (*Equation 1.1*)

$$V_P = V_G + V_E$$

Equation 1.1 Phenotypic variance. V_P = phenotypic variance, V_G = genetic variance, V_E = environmental variance

Genetic variance typically refers to the sum of additive, dominance and epistatic genetic effects (10) (*Equation 1.2*)

$$V_G = V_A + V_D + V_I$$

Equation 1.2 Genetic variance, V_G = total genetic variance, V_A = additive genetic effect, V_D= dominance genetic effect, V_I = epistatic genetic effect

When the phenotypic variance depends on genetic loci (or the number of copies of an allele) linearly, it refers to an additive genetic effect. Dominance genetic effects describe the effect of interaction of alleles at a certain locus; this contrasts with epistatic effects, which refer to the effect of the interaction of alleles at different loci. Heritability in genetic studies refers to the proportion of the variance of the trait that is attributable to the genetic variance (9). Broad-sense heritability refers to the phenotypic variance explained by total genetic variance (*Equation 1.3*).

$$H^2 = \frac{V_G}{V_P}$$

Equation 1.3 Broad-sense heritability; V_P = phenotypic variance, V_G = total genetic variance

Narrow-sense heritability, denoted as h^2, is the proportion of the phenotypic variance occurring due to the additive genetic variance (*Equation 1.4*).

$$h^2 = \frac{V_A}{V_P}$$

Equation 1.4 Narrow-sense heritability. V_P = phenotypic variance, V_A = additive genetic variance

In general, it has been shown that more than 50% of the total genetic variance is explained by the additive component (11).

Previously, heritability was estimated in sets of related individuals in family studies, by studying the phenotypic correlation between parents and offspring, or in twin studies. Nowadays, the use of data from molecular-level association studies has been proposed, which allows estimation of the narrow-sense heritability in groups of unrelated individuals (12, 13).

1.2 Refractive error and Myopia

In the non-accommodated emmetropic eye parallel lights focus on the retina (Figure 1.1).

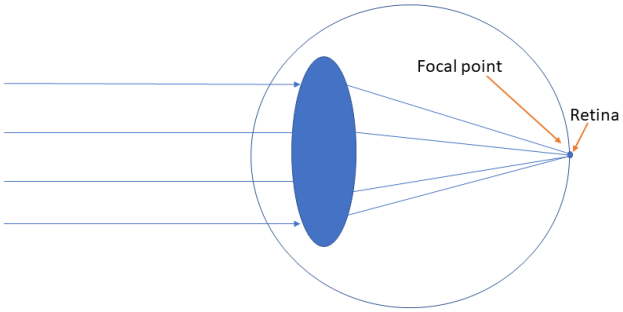

Figure 1.1 Refraction in the emmetropic eye.

Myopia is a refractive error in which parallel rays of a light from a distant object focus in front of the retina, resulting in a blurred image. Close objects are seen clearly, hence myopia is also called ‘near-sightedness’. The condition arises when either the axial length of the eye is too long relative to the corresponding optical power of the eye, or the corneal or lens optical power is too strong relative to the axial eye length.

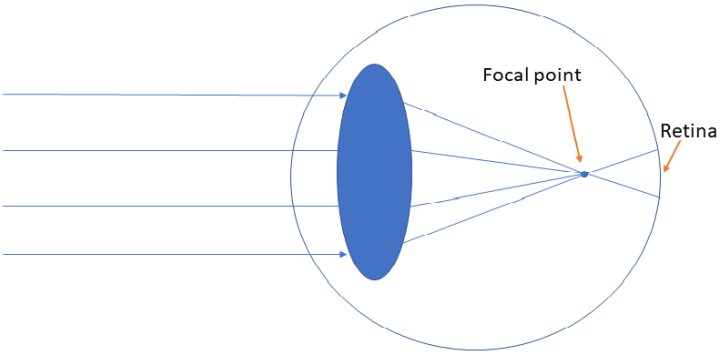

Figure 1.2 Refraction in the myopic eye. Light focuses in front of the retina.

1.2.1 Classification of Myopia

Various criteria have been used to classify myopia based on degree of severity, aetiology, age of onset, progression rate, and its clinical complications. A classification system proposed by Grosvenor (14) is based on the age of onset. It includes the following categories: congenital, youth-onset, early adult-onset and late-onset. A system using degree of myopia for classification was proposed by Fredrick (15); the three categories are: mild myopia (0.00 to -1.50 Dioptres (D)), moderate myopia (-1.50 to -6.00 D) and high myopia (-6.00 D or worse). Pathologic myopia was defined in the latter classification as occurring when refractive error is more than -8.00 D, noting that eye-and-vision threatening complications can develop in individuals with high or even with moderate myopia. More recently (16), other thresholds were proposed to define low (≤ -0.50 D) and high myopia (≤ -6.00 D); also introducing a *pre-myopia* condition with refractive state from +0.75 D to -0.50 D. By clinical entity (17) myopia is classified as simple myopia, night myopia, pseudomyopia, pathologic, and induced myopia. *Simple* myopia refers to the condition in which either the axial length of the eye is too long for its optical power, or the optical power is too strong for its axial length; otherwise this eye is normal. Typically, simple myopia is mild or moderate. *Night* myopia appears due to the accommodative response increase often observed in dim light conditions. *Pseudomyopia* is the consequence of ciliary muscle spasm when the eye is viewing a distant object. *Pathologic* myopia is defined as myopia of a high degree (at least 6.00 D) with degenerative changes in the posterior segment of the eye. These changes include, but are not limited to: diffuse chorioretinal atrophy, choroidal neovascularisation and lacquer cracks. *Induced* myopia occurs as a result of the influence of various drugs (drug-induced transient myopia); other possible causes of induced myopia are nuclear sclerosis (18) and acute hyperinsulinaemia (19)

1.2.2 Emmetropisation

Axial length in newborn infants is relatively short for the corresponding optical power of the cornea and lens; this means that in most cases, humans - like most animals - are born with a hyperopic refraction. During the early years of life the eye elongates, the cornea flattens, the crystalline lens thins (20, 21) and, finally, the eye becomes approximately emmetropic when the axial length matches the optical power of the eye (in reality, the majority of individuals have a small degree of hyperopia rather

than exact emmetropia). The eyeball grows most rapidly in first 1-2 years of human life (22), however there is evidence of adjustment of the axial length to the optical power over a period of up to 15 years (22).

Most dramatically, the axial length, cornea and lens change in the first phase, between 3 and 9 months of age (23). This confirms previous results from animal studies. The elongation of the eye is strongly correlated with the refraction at birth, consistent with a substantial visual dependency in the process of emmetropization (23, 24). Ametropia is a result of failure to reach emmetropia; when elongation of axial length is excessive it leads to the development of myopia. A shift in myopia of -2.00 to -3.00 D occurs as a result of the elongation of the eyeball by 1 mm (25).

There is no single, universally accepted theory regarding the mechanism of emmetropization. However, it has been shown (22, 26) that the ocular growth is regulated by visual experience.

1.2.3 Prevalence of Myopia

Myopia is one of the most concerning global health issues, and is one of the leading causes of blindness in certain population groups (27). Currently, up to 22% of the global population are myopic (28). The condition has an increasingly negative economic impact due to its pathological complications. Furthermore, the economic burden due simply to *uncorrected* myopia was estimated as 200 Billion US Dollars per year (29).

However, it remains difficult to confirm the real prevalence of myopia because different thresholds in refractive error are used to define myopia (30). The most common definition of myopia is a spherical equivalent of -0.50 D or less (16, 31). Numerous studies have reported a significant increase in the prevalence of myopia globally, with some regions (particularly East and South East Asian countries) having considerably higher levels of myopia and high myopia than in the West (32-34).

A US study reported the prevalence of myopia in individuals aged 12 to 54 years as 41.6% (34). An increase in the prevalence of myopia has been shown for all ethnicities and for all levels of myopia severity. Up to 20% of 12-13 years old white European adolescents have been reported as having myopia (35, 36). In adult population, the European Eye Epidemiology (E^3) consortium found an overall

prevalence of myopia as 24.3% (37). More rapid growth in the prevalence and incidence of myopia have been reported in studies from East Asian countries, with myopia affecting up to 80% of children (38, 39). Thus, the prevalence of myopia has increased dramatically since the 1970's, particularly in more recent birth cohorts. The geographical difference in the distribution of myopia, with a significantly higher risk of being myopic among urban population, suggests that environmental factors play a key role in the process of excessive axial elongation and myopia.

1.2.4 Risk Factors for Myopia

1.2.4.1 Environmental Factors Associated with Myopia

1.2.4.1.1 Education

The first articles reporting an association between myopia and higher levels of educational attainment were published in the 1800's and early 1900's (40-42). Since that time, numerous studies have been conducted in order to assess the effect of education on myopia development. Two approaches have been used to investigate and estimate the association between education and myopia. Firstly, time spent in full-time education or years of schooling have been used to assess the effect of education on myopia in adults. Robust association between educational exposure and myopia has been shown in epidemiological studies, with a higher proportion of myopic individuals in those with a university degree in comparison with to those only completing secondary or primary school education (43, 44). Secondly, the possibility of causality in the relationship between years of schooling and myopia has been addressed in two *Mendelian randomisation* (MR) studies (see section 2.4.4), using genetic variants strongly associated with refractive error (45, 46). However, the Mendelian randomisation studies were limited to adults; in epidemiological studies in school children, school performance or additional education load was considered to be factor determining the higher myopia prevalence (38, 47, 48). A trend of highly educated people to have more myopic refraction has been shown in populations of different ethnic background; note that the MR studies were limited to individuals of white European ancestry (49).

Factors potentially mediating a causal effect of education on myopia are near work and insufficient time outdoors (38, 48, 50, 51); however, the exact mechanism(s) remain unclear.

1.2.4.1.2 Near Work

The German astronomer and mathematician Johannes Kepler was one of the first scientists to describe the link between excessive writing and reading with myopia (52). Since that time, this topic has become one of the most controversial in eye-and-vision science (31). The term *near work* in studies investigating myopia is typically used to describe activities performed at a short distance (e.g. reading, writing, watching TV, and playing video games) (38, 50). Near work was reported to be associated with the increased levels of myopia prevalence in cohort studies in schoolchildren (47, 53) with axial elongation as a driver of myopic shift in refraction (54). Higher rates of myopia progression in children reading at a shorter distance were reported in a study conducted in Taiwan (55); however, this association was confounded by a higher myopia level at baseline. In contrast, a weak or even no association of near work and myopia has been reported in some studies (50, 56). A recent meta-analysis confirmed the role of near work as a risk factor of myopia; although with a modest effect size (57).

1.2.4.1.3 Time outdoors

Time spent outdoors or outdoor activity has recently been considered as an important factor for myopia development (58). In a randomized controlled trial (RCT) in China, the effect of 1 additional 40-minute class of outdoor activities on myopia development was tested in 1903 schoolchildren (mean age 6.6 years) (59). A significant difference in cumulative myopia incidence rate between the intervention and control group was reported (30.4% vs. 39.5%, respectively, $p < 0.001$) after a 3-year follow-up period. The difference in spherical equivalent refractive error was marginally significant (a difference of 0.17 D; -1.42 D vs. -1.59 D, $p = 0.04$). A clinical controlled trial in 3051 Chinese schoolchildren has reported a similar effect size resulting from two 20-minute outdoor recess periods between school classes (a difference of 0.17 D; -0.10 D vs. -0.27 D in the intervention and control group, respectively); however, the follow-up period was only 1 year (60). Regardless of physical activity, more time spent outdoors showed a negative correlation with the

incidence of myopia (61). There are several possible mechanisms explaining the protective effect of outdoor activities on myopia development. Vitamin D is one of the factors that may link time outdoors with myopia. Existing evidence of an association of serum 25(OH)D and myopia is inconsistent; moreover, a recent MR study did not support the hypothesis of vitamin D being a causal risk factor for myopia (62). Another possible pathway linking time outdoors with negative refractive error is light exposure; with luminance levels typically lower indoors compared to outdoors (63). Both animal and human studies have revealed a relationship between lower levels of light exposure and an increased risk of myopia development (63-66). Differences in the chromatic spectrum of light have also been reported as a potential risk factor for myopia, including studies in animal models (66-69). Indoor lighting possesses a narrower spectrum than outdoor sunlight, therefore this may also contribute to the protective effects of time outdoors. However, the evidence to date is inconclusive.

1.2.4.1.4 Diet

Diet has also been proposed to be a risk factor for the development of myopia. In some populations, a change in lifestyle from traditional to 'Western' and change to a 'Westernised' diet, rich in carbohydrates, was associated with an increase in myopia prevalence (70). It was suggested by Cordain et al. (70) that a high glycaemic load diet caused a relatively higher level of hyperinsulinaemia, hyperglycaemia and type 2 diabetes. Hyperinsulinaemia starts a cascade of reactions leading to the activation of various growth factors in the human body, including insulin like growth factor-1 that may enhance the scleral tissue growth. Scleral growth, in its turn, causes a myopic shift. A study conducted in Denmark in young adults aged 16-26 with diabetes has reported an association between hyperinsulinaemia and myopia (71). However, even though this Danish study indicated a relationship between metabolism and myopia development, the exact mechanism still remains unclear. Another assumption is that changes in the crystalline lens in diabetic individuals are the primary cause of negative shift in refraction (72, 73), i.e. fluctuations in blood glucose levels lead to swelling of the lens and changes in refraction (73, 74). However, it would be simplistic to limit the consideration of dietary risk factors for myopia only to carbohydrates. A study of Chinese children aged 7-10 years compared the diets of children who developed myopia and those who did not (75). The results indicated that children who became myopic had lower average intake of

fat, protein, vitamins B_1, B_2, C, and such microelements as phosphorus and iron. In summary, there have been few studies of diet and myopia, such that a definitive role for any specific nutrient has yet to be established.

1.2.4.2 Genetics of Myopia

Although lifestyle risk factors play an important role in myopia development, a genetic predisposition to the disease has also been established. Genetic predisposition to myopia has been shown in twin and family studies of different ethnicities, with the reported heritability estimates being as high as 98% (76-79). In a recent meta-analysis of genome-wide association (GWA) studies (see section 2.4.6) for refractive error in 542,934 individuals, independent genetic variants significantly associated with the phenotype explained 18.4% of spherical equivalent heritability (80). The variance in refractive error explained by commonly-occurring SNP, i.e. those with a minor allele frequency (MAF) of at least 5%, has been assessed in the ALSPAC paediatric cohort in children aged from 7 to 15 years old; with the estimated "SNP heritability" of 28% (81). Taking account of measurement error due to non-cycloplegic autorefraction, the authors estimated the SNP heritability in ALSPAC children to be as high as 35%. In an adult UK Biobank cohort, the estimated SNP heritability was 38.7% (82).

1.2.4.2.1 Genetic Linkage Studies

Genetic linkage studies in families with myopia and high myopia have been performed to identify the genetic loci for the disease, mostly under the assumption that inheritance was monogenic.

The first genetic locus for myopia was identified in 1990 (83) in a linkage analysis of a family with a clinical syndrome which included myopia, amblyopia and deuteranopia (Bornholm Eye Syndrome) (83). The genetic locus for this X-linked disease was named MYP1. Guo et al. (84) and other research groups have identified pedigrees affected by this syndrome and found linkage to the MYP1 locus. Ratnamala et al.(85) described X-linked recessive inheritance linked to the MYP1 locus in an Asian Indian pedigree with non-syndromic high myopia (mean refractive error -8.43 D). Recently, an *OPA1LW* (long wavelength opsin gene) mutation was reported to be responsible for syndromic and non-syndromic X-linked high myopia

mapped to MYP1 (86). In 1998, Young et al. (87) identified the MYP2 locus in a linkage study of an 8-generation pedigree with an autosomal dominant pattern of high myopia. MYP3 was identified in a linkage analysis of a large family of Greek/Italian ancestry with autosomal dominant myopia phenotype (88). Decorin (*DCN*) and Lumican (*LUM*) have been suggested as a candidate genes at the MYP3 locus; both genes code for proteoglycans possibly related to the extracellular matrix organisation of the sclera. However, the association of *LUM* gene variants with myopia has not been confirmed by association studies in South Asian and Chinese populations (89, 90). The studies listed above identified genetic loci for rare forms of high myopia inherited following a classical (Mendelian) dominant or recessive pattern. In contrast, low myopia is considered a complex trait caused by a combination of genetic and environmental factors.

A genome-wide scan was performed in the USA in Ashkenazi Jewish families with low myopia, identifying the MYP6 locus (91). Linkage to the locus was successfully replicated in the another USA-based cohort (92). A twin study in 506 dizygotic and monozygotic twin pairs has been conducted to assess the heritability of refractive error as a continuous trait; this analysis was followed by a genome-wide linkage scan in 221 dizygotic twin pairs in order to identify susceptibility loci for the phenotype (93). This study reported 4 novel loci, MYP7-MYP10, for myopia. A regression-based quantitative trait loci (QTL) linkage study in a USA-based cohort of individuals of Ashkenazi Jewish ancestry identified a further novel locus (MYP14) for refractive error on chromosome 1p36 (94). MYP14 locus was replicated in an international study in 2009 (95). The MYP4 locus on chromosome 7q36 was identified by Naiglin et al. (96) in a linkage analysis of 23 families with high myopia; however, no linkage to this loci was found by Paget et al. (97). In the latter study, the myopia locus was mapped to 7p15, also identified in other studies (92, 98); that led to the replacement of MYP4 with MYP17.

Most of the myopia loci showed autosomal dominant inheritance. Exceptions are MYP18, MYP23 and MYP26, which are associated with autosomal recessive high myopia. Xiao et al. (99) identified mutations in the *ARR3* gene at the MYP26 locus linked to female-limited high myopia with early onset. The *LRPAP1* gene on chromosome 4p16 was found to be the causative gene at the MYP23 locus (118). Yang et al. (112) identified an autosomal recessive high myopia locus on chromosome 14q22.1-q24.2 (MYP18); however, the gene responsible has not yet been reported.

To date, 26 myopia loci have been identified in linkage analyses studies (Table 1.1). Very recently, Ouyang et al. (100) have reported mutations in the *CPSF1* gene linked to the early-onset high myopia (MYP27).

MYP gene/locus	Location	Inheritance	Reference	Clinical features
MYP1	Xq28	X-linked	Ratnamala et al. 2011 (85)	High myopia
MYP2	18p11.31	Autosomal Dominant	Young et al. 2001 (101)	High myopia
MYP3	12q21-q23	Autosomal Dominant	Lin et al. 2010 (102), Young et al 1998 (88)	High myopia
MYP5	17q21-q22	Autosomal Dominant	Paluru et al. 2003 (103)	High myopia
MYP6	22ql2	Autosomal Dominant	Klein 2007 (92), Stambolian 2004 (91)	Low myopia
MYP7	11p13	Quantitative Trait Loci	Hammond 2004 (93)	Low myopia
MYP8	3q26	Quantitative Trait Loci	Hammond 2004 (93)	Low myopia
MYP9	4q12	Quantitative Trait Loci	Hammond 2004 (93)	Low myopia
MYP10	8p23	Quantitative Trait Loci	Hammond 2004 (93)	Low myopia
MYP11	4q22-q27	Autosomal Dominant	Zhang 2005 (104)	High myopia
MYP12	2q37.1	Autosomal Dominant	Paluru 2005 (105)	High myopia
MYP13	Xq23-q27.2	X-linked Recessive	Zhang 2006 (106), Zhang 2007 (107)	High myopia
MYP14	1p36	Quantitative Trait Loci	Wojciechowski 2006 (94)	Low myopia
MYP15	10q21.1	Autosomal Dominant	Nallasamy 2007 (108)	High myopia
MYP16	5p15.33-p15.2	Autosomal Dominant	Lam 2008 (109)	High myopia
MYP17 (former MYP4)	7p15	Autosomal Dominant/QTL	Naiglin 2002 (96), Paget 2008(97) ,	High myopia

			Ciner 2008 (98)	
MYP18	14q22.1-q24.2	Autosomal Recessive	Yang 2009 (110)	High myopia
MYP19	5p15.1-p13.3	Autosomal Dominant	Ma 2010 (111)	High myopia
MYP20	13q12.12	Autosomal Dominant	Shi 2011 (112)	High myopia
MYP21	1p22.2	Autosomal Dominant	Shi 2011 (113), Tran-Viet 2012 (114)	Early-onset high myopia
MYP22	4q35.1	Autosomal Dominant	Zhao 2013 (115)	High myopia
MYP23	4p16.3	Autosomal Recessive	Aldahmesh 2013 (116), Jiang 2015 (117)	High myopia
MYP24	12q13.3	Autosomal Dominant	Guo 2014 (118)	Early-onset high myopia
MYP25	5q31.1	Autosomal Dominant	Guo 2015 (119)	Early-onset high myopia
MYP26	Xq13.1	X-linked, Female limited	Xiao 2016 (99)	Early-onset high myopia in females
MYP27	8q24.3	Autosomal Dominant	Ouyang 2019 (100)	Early-onset myopia

Table 1.1 Myopia genetic loci listed on the Online Mendelian Inheritance in Man website (URL:https://omim.org/phenotypicSeries/PS160700?sort=geneSymbols&order=desc)

1.2.4.2.2 Association Analyses

Numerous genome-wide association studies for myopia have been conducted to identify common genetic variants robustly associated with the phenotype. There are two main designs of GWAS for myopia: a case-control study with a binary trait (myopia vs. non-myopia), or analysing refractive error as a continuous trait.

A two-stage design GWAS for pathological myopia was conducted in 2009 in a cohort of 830 cases and 1911 controls (120). In the first stage, a total of 411,777 variants were tested for association with the trait; with further genotyping performed in 537 cases and 980 controls. A locus at 11q24.1 was reported, with two genes, *BLID* and *LOC399959* located in a 200-Kb region tagged by the rs577948 genetic variant showing the strongest association.

In a study of 520 Japanese individuals with high myopia and 520 controls, 39 SNP located on 21q.22.3 (previously reported to be associated with myopia) were tested using the χ^2 test and Fisher's exact test (121). After multiple testing correction only one SNP (rs2839471) located within the *UMOLD1* gene showed significance in the association.

A meta-analysis of 2 GWAS for myopia in cohorts of Chinese ancestry revealed 2 SNP within the *CTNND2* region associated with the trait; one of the SNP, rs6885224 was replicated in an independent cohort (122).

Two separate GWA studies conducted in Europe (123, 124) used refractive error as a continuous trait and found two loci significantly associated with the trait (near the *RASGFR1* gene and near the *GJD2* gene, both in the MYP11 linkage locus). These genetic associations were subsequently replicated (123, 124). In 2010, the Consortium for Refractive Error and Myopia (CREAM) was established and in 2013, the results from genome-wide meta-analysis on refractive error were published (125). Twenty-seven cohorts of European ancestry (n = 37,782) and 5 cohorts of Asian ancestry (n = 12,32) were included in the study, and a total of 24 novel loci were identified. A total of 14 of the loci identified in the CREAM study were confirmed in an independent GWAS study by 23andME (see below). In turn, 16 loci identified in the 23andME study were validated in the CREAM GWAS, despite different phenotypes being used in these studies (126). The GWAS performed by 23andMe Inc., a consumer genomic company, was conducted in 45,771 European individuals (127). A survival analysis on age of the onset of myopia was performed

and a total of 22 genetic variants associated with the phenotype attained genome-wide significance ($P < 5E-08$), with 20 novel loci among them. In an independent cohort of 8,323 individuals classified as having an age of myopia onset earlier or later than 10 years, 10 out of 22 SNP were replicated.

In 2016, a large GWAS for self-reported myopia in 106,086 cases and 85,757 controls was performed and 183 genetic loci associated with myopia were reported (128). The large sample size of the study allowed researchers to identify more than 100 novel loci for myopia.

The most recently published GWAS results are those from a meta-analysis of 542,934 individuals of European ancestry carried out by Hysi et al., which reported a total of 904 independent genetic variants surpassing the genome-wide significance threshold (80).

Several other GWA studies have been performed for phenotypes related to myopia, including GWAS for corneal and refractive astigmatism (82), axial length (129, 130) and macular thickness (131, 132).

1.2.4.3 Mechanisms of myopia

Animal research, along with numerous clinical and epidemiological studies, have helped to generate a better understanding of the possible mechanisms underlying myopia development; e.g. the effect of retinal image focus and/or quality on ocular growth (133-136).

The role of signalling pathways and particularly, dopamine signalling in myopia development has been investigated in numerous studies (137-139). As well as a diurnal variation, dopamine release in the retina is stimulated by light exposure, especially flicker (140). Decreased levels of dopamine and its metabolite 3,4-dihydroxyphenylacetic acid (DOPAC) occur in myopic eyes of chickens and monkeys with form-deprivation myopia (137). The protective effect of bright light has been suggested to involve D_2-dopamine receptors (141, 142). It was also shown that not only brightness of the retinal image but spatial and temporal contrast control the release of dopamine (143).

The insufficient accommodative response during near work induces the hyperopic defocus (50, 144) and animal models show that such a defocus leads to the decrease in retinal or vitreal dopamine levels (145) and to eye growth (133).

The ability of the eye to compensate for the defocus introduced by fitting a lens in front of the eye (so called 'lens-induced defocus') has been shown in a number of animal studies (133, 134, 146). Convex lenses induce myopic defocus, leading to the slowing of ocular growth. Concave lenses induce hyperopic defocus, which results in a thinner choroid and an increase in eye growth. Studies in humans confirmed the ability of the human eye to detect defocus and to act accordingly to its sign as in animal models (136, 147). Lag of accommodation during near work has been proposed as another aetiological factor of excessive myopic ocular growth (144) and a main factor for school myopia development. However, animal studies have not provided compelling support for the involvement of the accommodation system in myopia development; e.g. it has been shown that ciliary nerve section, which blocks accommodation, does not dramatically affect the ability of chicken eyes to emmetropize (148).

Increasing knowledge of the pathogenesis of myopia opens new horizons in diagnosis and treatment of the condition.

1.2.5 Management of Myopia

Optical correction of refractive error is the most common method of myopia management. Until very recently, spectacles or contact lenses have been prescribed purely to provide clear distance vision, i.e. to alleviate the symptoms of the condition. However, in the past few years various types of spectacles have been proposed not only for the correction of visual blur but also to slow the future progression of myopia. Single vision spectacles are still the most widely used means of correction due to convenience, economic reasons, and lack of evidence regarding the long term efficacy of optical treatments for slowing myopia. Undercorrection has been reported to slow the progression of myopia in several studies (149). However, these findings conflict with the results of an RCT showing more rapid myopia progression in children who were undercorrected in comparison to those who were fully corrected (150). Bifocal spectacle lenses with or without base-in prism and progressive addition spectacles are among other types of spectacle correction proposed for slowing myopia progression (151, 152). In

general, the clinical efficacy of these latter treatments in RCT has been insufficient to support widespread adoption (153).

Soft contact lenses with a multifocal design have been reported as a means of slowing myopia progression (154, 155). The best-performing lenses slow myopia progression by approximately 50% (153). Orthokeratology (OK) refers to the use of a gas-permeable rigid contact lenses that are worn overnight to reshape the cornea. Clear vision is obtained without the OK lenses in place during the daytime. OK lenses have been shown to significantly slow axial elongation in myopic children (weighted mean difference, -0.26 mm; 95% CI -0.31 to -0.21; $p < 0.001$) (156). Hence, OK has been recommended to be used in myopia control (156, 157).

Topical application of the muscarinic antagonist atropine has been used as a pharmacological approach to the management of myopia since the 1970's (158). Clinical trials have consistently demonstrated a high level of efficacy (159, 160). The adenosine receptor antagonist 7-methylxanthine (7-MX) has been tested in animal (161) and human studies (162) and has also shown some promise in myopia control.

Other methods, including surgical interventions have also been proposed for slowing down the progression of myopia. Of the currently available interventions, atropine eye drops, OK lenses, and multifocal soft contact lenses are the most widely used.

1.3 Strabismus

Strabismus is an oculomotor disorder characterized by constant or intermittent misalignment of the eyes that reduces the ability to achieve optimal binocular vision. It is often associated with amblyopia in the deviated eye and can therefore be associated with childhood-onset visual impairment (163).

1.3.1 Classification of Strabismus

Comitant or concomitant strabismus refers to the condition when an eye deviation is the same in all positions of the gaze; typically, the deviation is congenital or with an early onset. Based on the ocular misalignment, strabismus can be classified into several subcategories. The most common type is convergent strabismus or

esotropia, when the eye deviates inwards; divergent strabismus or exotropia describes an outwards deviation of the eye. Vertical strabismus (hyper- or hypotropia) may develop as a primary disease or in conjunction with convergent or divergent strabismus.

Incomitant or noncomitant strabismus refers to the condition when the deviation of the eye varies in different fields of gaze. Incomitant strabismus results from a limitation of eye movements, and is a secondary condition, relating either to the lack of the innervation of extraocular muscles or muscle weakness, or to an orbital abnormality (164, 165).

1.3.2 Aetiology of Strabismus

1.3.2.1 Environmental factors

Various environmental factor have been reported to increase the risk of strabismus (166-168). Low birth weight and retinopathy of prematurity are among main risk factors for strabismus (169-173). Children with the aforementioned risk factors could suffer a deficit due to malfunction of the geniculostriate pathways during the first months of life, leading to strabismus development (174, 175). Maternal smoking during pregnancy is also considered as a risk factor for strabismus. A positive correlation between the amount of maternal smoking and the increased risk of strabismus development has been reported in several studies; specifically, smoking more than 10 cigarettes per day was found to increase the risk significantly (167, 173, 176). Moreover, mothers who stopped smoking cigarettes during the 1st trimester of pregnancy were at a lower risk of their child having strabismus comparing with those continued smoking (167, 177). This could be explained by the direct toxic effect of nicotine through the placenta (178) during the later stages of the fetal growth, which is crucial for the development and specialization of the visual system.

Paternal age was found to be associated with an increased risk of esotropia (163, 173). There is conflicting evidence regarding the influence of maternal age on strabismus development. Most studies have not reported maternal age as a significant risk factor (163, 179, 180); however, one study found a significant difference in the risk of strabismus in mothers of different ages (173).

A difference in the refractive power between the two eyes (anisometropia) of 1.00 D increases the risk of strabismus development (163, 169). This association could be due to the difference in retinal images between the eyes affecting image fusion. Hyperopia of more than +3.00 D is considered as a risk factor strongly associated with strabismus, with accommodation and convergence as possible mediating factors (181, 182).

The association of myopia and intermittent exotropia has been investigated since the late 1800s; with the hypothesis proposed by Donders that the lower convergence in myopes may lead to an increased risk of divergent strabismus (183). In contrast, Walsh et al. hypothesised that myopia is a consequence of intermittent exotropia and occurs due to the excessive accommodative demand (184). A recent meta-analysis of 7 population based studies in 23,541 individuals reported that myopia was a risk factor for exotropia (OR: 5.23; p = 0.0001) (185).

1.3.2.2 Genetics of Strabismus

Numerous studies have been conducted to understand the genetics of strabismus (186-188). Twin studies have assessed the concordance of the condition (having strabismus in the same direction in both twins) in monozygotic and dizygotic twins (189, 190). Concordance in monozogotic twins varies from 24 to 92%, and from 13 to 60% in dizygotic twins; with higher values of concordance in those having accommodative strabismus (188). A complex inheritance pattern was described in several studies (191-193). However, the hypothesis that there are rare monogenic forms of strabismus was supported, e.g. a locus at 7p22.1 was found in linkage analysis of families with multiple members affected by strabismus (194, 195). In certain rare syndromes, such as Mietens-Weber Syndrome and Lamb-Shaffer Syndrome, strabismus is one of the clinical features, alongside intellectual disability. In the only previous genome-wide association study for non-syndromic strabismus, Shaaban et al. (196) identified a single variant (rs2244352; OR = 1.33, p = 9.58E-11) that was significantly associated with the condition.

1.3.3 Amblyopia

Amblyopia is defined as poor visual acuity in one or both eyes not immediately correctable by glasses and without accompanying ocular pathology. It is one of the

most common clinical features found in conjunction with strabismus. Ocular misalignment in strabismus leads to unequal visual input to the brain from the two eyes. Amblyopia is more frequent in children with a family history of the condition (197) but no studies have directly investigated genetic susceptibility to amblyopia. However, there is a well-recognised variability in the apparent susceptibility to amblyopia amongst individuals that have the same degree of strabismus, and the reasons for this are not clear but could include genetic factors (197).

Bilateral amblyopia is described in cases of equal but severe refractive error, however the clinical course is different from unilateral amblyopia which is much more prevalent. The latter condition is usually defined as a difference in best corrected visual acuity between the two eyes of ≥ 2 lines of a logMAR chart (198, 199).

Strabismic amblyopia typically occurs in unilateral strabismus, when the deviation in one eye is constant. It is less frequent in patients with an alternating deviation (200). Early onset of strabismus and anisometropia associated with the deviation of the eye increase the risk of amblyopia development.

Treatment of amblyopia aims to restore visual functions such as visual acuity, as well as improving accommodation and eye mobility. When it is possible to achieve, the development of a binocular function is the final step in the management of amblyopia. Optical correction of refractive errors associated with amblyopia helps to achieve clear retinal images in both eyes. Full correction by spectacles or contact lenses has been proposed as an initial step in the treatment of amblyopia (201). An alternative therapeutic approach is under-correction of refractive error, as this may have positive effect on emmetropization of the eye (202). Optical correction is typically followed by the stimulation of the amblyopic eye, with only approximately 25% of eye care specialists using the refractive error correction alone (203). Occlusion is an effective method for amblyopia treatment. This well-known intervention was introduced in the 9th century (204). The first type of adhesive eye patch was described in 1927 by C. H. Satler (204). Occlusion of the eye with better visual acuity aims to stimulate the amblyopic eye; it also helps to reduce eccentric fixation. Numerous studies have reported that occlusion therapy is successful in strabismic amblyopia management (205, 206) with higher success rates in younger patients. However, a high level of noncompliance has been reported in children older than 8 years (207). Various patching regimens have been proposed, varying from 2-hours per day to a full-time occlusion (208, 209). A negative correlation

between compliance and efficacy has been reported. Penalization is an alternative approach based on the use of atropine, a muscarinic receptor antagonist. The instillation of atropine sulphate 1% in the eye with a better vision causes a paralysis of the ciliary muscle, leading to blurry vision; hence, reducing the use of better eye in vision processes and thereby stimulating the use of the amblyopic eye. Amblyopia is amenable to treatment if detected early (197, 210). Beyond age 7-10 years treatment outcomes are limited.

1.3.4 Management of Strabismus

There are several aims of strabismus management plans: 1) to achieve good visual acuity in both eyes, 2) to restore appropriate ocular alignment, 3) to restore or obtain binocular fusion, 4) to eliminate double vision and asthenopic symptoms. Treatment options include optical correction, use of pharmacological agents, vision therapy, and/or extraocular muscle surgery. Strabismus surgery is used to correct the misalignment of an eye by altering the physical action or anatomy of one or more of the extraocular muscles. Weakening the muscle or recession to move its attachment site further back from the front of the eye, reduces the action of the muscle when it is stimulated. Conversely, resection surgery involves removal of a section of the muscle, strengthening the muscle function.

1.4 Hyperopia

Hyperopia or far-sightedness is a common refractive error in which light rays focus behind the retina in the non-accommodated eye (Figure 1.3). In most cases it is attributable to a relatively short axial length (211).

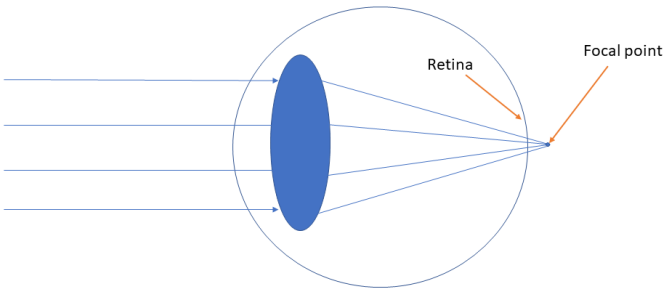

Figure 1.3 Refraction in the hyperopic eye.

1.4.1 Classification of Hyperopia

Two main classifications of hyperopia exist in the current literature. The first one is based on the degree of refractive error, with *low* hyperopia being defined as a refractive error of more than +0.50 D and less than +2.00 D, *moderate* hyperopia as +2.25 D to +5.00 D, and *high* hyperopia as a refractive error more than +5.00 D (212). Different thresholds for defining high hyperopia have been proposed, varying from +4.00 D (213), or +4.50 D (214) to +5.00 D. The second system of classification is based on clinical features, and also consists of three categories (215). *Simple* hyperopia is defined as being due to the normal biological variation in axial length or in the optical components of the eye. *Pathological* hyperopia is defined as being due to a congenital eye malformation or eye trauma that leads to abnormalities in eye anatomy. Finally, paralysis of accommodation causes *functional* hyperopia.

1.4.2 Prevalence of Hyperopia

Hyperopia has been reported as a common refractive error in children (201, 211). However, it is not straightforward to compare various estimates of the prevalence of the condition, as different definitions of hyperopia have been used in different studies (216). Despite this, it is commonly accepted, that hyperopia is age-related (211, 217). Up to 9% of full-term infants 6-9 months old have hyperopia of +3.35 D or more (168, 202). The prevalence drops to 3.5% by the age of 1 year (218).

Children born prematurely are typically less hyperopic (219). A meta-analysis of hyperopia studies reported the prevalence of hyperopia to be 5% in children aged 7 years old, with lower prevalence at ages 9 and 15 years (2-3% and 1%, respectively) (220). However, studies in predominantly Caucasian populations have reported a higher prevalence of hyperopia (211, 217).

A gender difference in the prevalence of hyperopia has been assessed in numerous studies, however the existing evidence is conflicting. Girls were found to be more hyperopic in a meta-analysis of existing studies conducted by Castagno et al (220), whereas Ip et al. (211) reported a higher prevalence in boys in a sample from Australia. No evidence for an association of hyperopia with gender was found in a study conducted in Northern Ireland (217).

1.4.3 Educational Attainment and Hyperopia

Learning at school and academic achievements depend greatly on vision and even minor visual problems may negatively influence the learning process (221, 222). Uncorrected hyperopia, as a common refractive error, is one such vision defect. Increased use of electronic devices and a large amount of near work in class may affect the accommodative-vergence system in hyperopes, leading to development of adverse symptoms such as headache, fatigue and asthenopia (223). Thus, children with uncorrected hyperopia may experience greater difficulties with learning and education. Numerous case-control and cross-sectional studies have examined the effect of uncorrected hyperopia on academic performance in participants with existing (214) or simulated hyperopic refractive error (224). Hyperopia was found to be associated with a deficit in early literacy skills in children aged 4-5 years old (225), confirming results of a study in 8 year-old children conducted in Wales (214). In the latter study, Williams et al. (214) reported that children with hyperopia achieved lower scores in Standardised Assessment Test (SAT) and National Foundation for Education Research (NFER) examinations than their peers.

Two studies have been conducted with the aim of identifying the amount of uncorrected hyperopia that may lead to poorer academic performance. However, only adult participants have been included in these studies (226, 227). Both studies reported that the decrease in test results was restricted to individuals with uncorrected hyperopia of at least +1.50 to +2.00 D.

While the association between uncorrected hyperopia and poorer educational attainment has been assessed in numerous studies, only one study has investigated the causality in the relationship (46). A bidirectional Mendelian randomisation analysis has been performed in a large sample of UK Biobank participants, using genetic variants robustly associated with myopia as instruments to estimate the effect of refractive error on time spent in full-time education. This MR study reported no causality in the association: the estimated reduction in years spent in education per Dioptre of refractive error was 0.008 Years/D (95% CI -0.041 to 0.025, p = 0.6).

More evidence is needed to estimate the causal effect of uncorrected hyperopia on educational attainment due to inconsistency in definition of hyperopia and educational achievement between the studies.

1.4.4 Management of Hyperopia

Optical correction by means of spectacles and contact lenses is the most commonly used method of hyperopia management. Convex spherical or spherocylindrical spectacle lenses is a cost-effective and convenient option (228). Despite the many attempts to standardise the optical correction of hyperopia, many practitioners still rely on their own ad hoc experience when prescribing spectacles. Whereas prescribing the full amount of refractive error found upon non-cycloplegic refraction was proposed by Leat et al. (201), Horwood et al. (229) recommended that prescription of the full cycloplegic refraction is preferable. Soft or rigid contact lenses are also used to correct hyperopia and have some advantages in certain groups of hyperopic patients, e.g. in patients with anisometropia, contact lenses reduce aniseikonia (i.e. the difference in retinal image size between the 2 eyes). Vision therapy has also been suggested as a method to improve the binocular disfunction that may occur due to hyperopia (230). In some patients, optical correction does not completely eliminate the accommodative disfunction; thus, vision therapy may be helpful in such circumstances.

Chapter 2. General Methods

2.1 UK Biobank

UK Biobank recruited 502,633 participants aged 37 to 73 years between February 2006 and July 2010 (231). Ethical approval was obtained from the National Health Service National Research Ethics Service (Ref 11/NW/0382) and all participants provided written informed consent. A comprehensive questionnaire was completed at a baseline visit to one of 22 assessment centres and participants underwent a physical assessment. The information collected included sociodemographic data and medical history. Blood samples were taken in order to perform genome-wide genotyping. Towards the latter stages of the recruitment period, an ophthalmic examination was introduced, and it was completed by approximately 23% of participants.

As an open access large-scale population based study UK Biobank provides researchers with valuable information on genetic, health and lifestyle data (232). UK Biobank data relevant to the current thesis are described in this chapter. However, there are some limitations of the UK Biobank important to consider. The 'healthy volunteer effect' (233) in UK Biobank has been assessed in order to investigate whether the UK Biobank sample differs from the general population (234). It has been shown that UK Biobank individuals live in more socioeconomically advantaged areas, are more educated, and less likely to be obese, to smoke and consume alcohol (234). Thus, while UK Biobank is not representative of the general population, it is suitable for the assessment of the exposure-outcome associations which are, in part, generalizable (234, 235).

The self-report of health conditions that occur in childhood may be affected by recall bias (236). As UK Biobank participant were of aged 37-70 years at the time of data collection, this is an important consideration in the current work. Recall bias may have affected the following variables: age-of-onset of wearing spectacles, birth weight, and presence/absence of strabismus. Where possible, I assessed the reliability of self-reported data in my thesis.

2.1.1 Phenotypes in UK Biobank: Education

498,768 participants completed the questionnaire item #6138: “Which of the following qualifications do you have (you can select more than one)?”; with the options, “(1) College or University degree, (2) A levels/As levels or equivalent, (3) O levels/GCSEs or equivalent, (4) CSEs or equivalent, (5) NVQ or HND or HNC or equivalent, (6) Other professional qualifications, e.g. nursing, teaching, (7) None of the above, (8) Prefer not to answer”. After excluding those without genotype data and withdrawing consent, there were a total of 488,295 individuals with valid Educational qualifications data (*Table 2.1*). Questionnaire item #845: “At what age did you complete your continuous full time education” was asked of all participant except those who indicated having a college or university degree (answers to item #6138). In the current study, a variable *‘EduYears*’ (*Table 2.1*) was derived from the answer to item #845, except that UK Biobank participants with a college or university degree were coded as having finished full time education at age 21 years, and those who reported completing full time education at age 13 or earlier, were assigned a value of 13 years.

Qualfication	**Number of participants**	**(%)**
College or University degree	159,048	32.57
A levels/AS levels or equivalent	35,672	7.31
O levels/GCSEs or equivalent	62,956	12.89
CSEs or equivalent	17,985	3.68
NVQ or HND or HNC or equivalent	52,717	10.80
Other professional qualifications, e.g. nursing, teaching	71,218	14.59
None of the above	82,954	16.99
Prefer not to answer	5,745	1.18
Total	**488,295**	**100**

Table 2.1 Educational qualifications in the sample of 488,295 UK Biobank participants

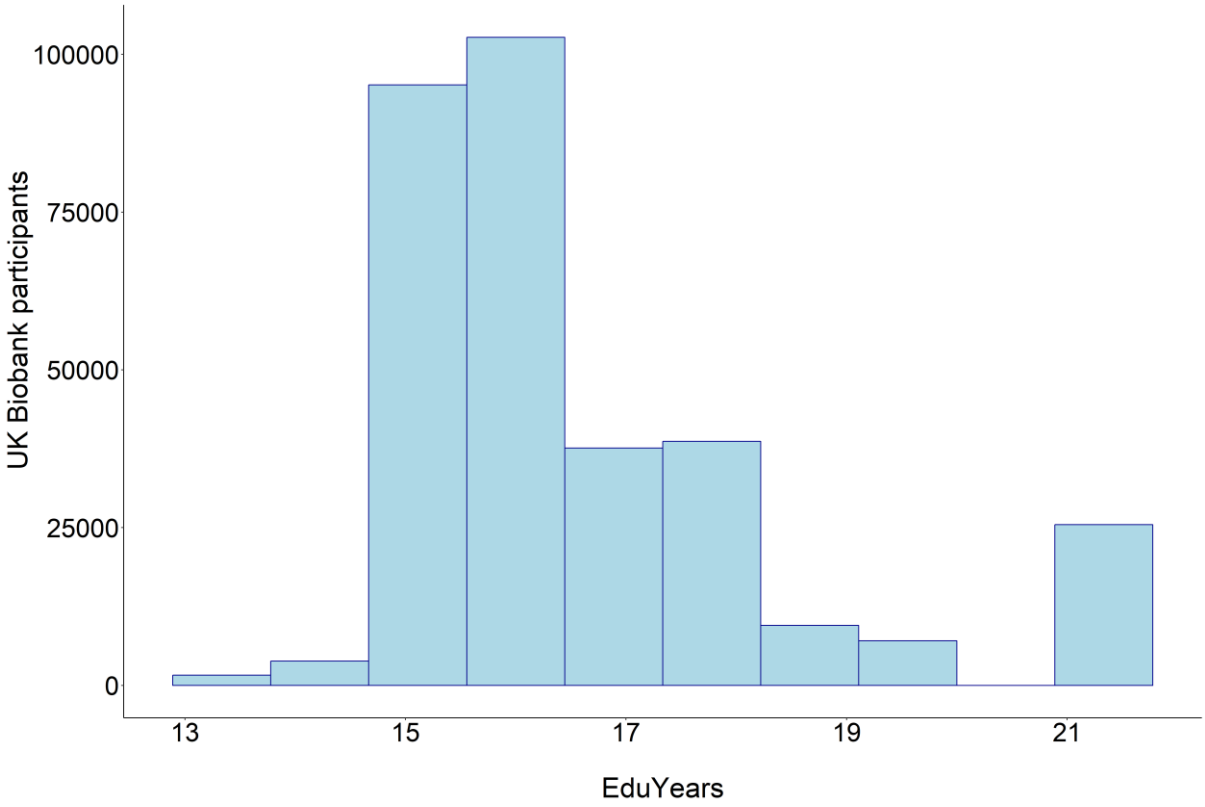

Figure 2.1 Distribution of EduYears variable *in UK Biobank participants.*

A variable *'Educational Year'* was created as analogous to year of birth but starting from September (as for the school year).

2.1.2 Phenotypes in UK Biobank: Eye and Vision-related Data

Measurement of refractive error in UK Biobank was introduced as a component of the ophthalmic assessment, therefore only a subset of participants had refractometry readings (n = 129,739). Non-cycloplegic autorefractometry was performed using the Tomey RC 500 autorefractor-keratometer (Tomey Corp., Nagoya, Japan) after removing habitual glasses or contact lenses. Up to 10 measurements were taken for each recording, with coding all unreliable data as missing. Spherical equivalent was calculated as the spherical power plus half of cylindrical power (237) for each eye, then the values for the two eyes were averaged (average mean spherical equivalent; *avMSE*). The distribution of avMSE in UK Biobank participants is shown on *Figure 2.2*. Following the recommendations of the International Myopia Institute (16), myopic eyes were defined as those with

spherical equivalent refractive error ≤ -0.50 Dioptres (D). Refractive astigmatism was taken as the average cylinder power between the two eyes (238). A binary variable was used to classify individuals with astigmatism ≥1.00 D vs. <1.00 D, as adopted by Shah et al. (82). Anisometropia was calculated as the difference in spherical equivalent between the two eyes. A binary variable was used to classify individuals with anisometropia ≥1.00 D vs. <1.00 D (239).

At the baseline or first follow-up visit participants were asked the reasons for wearing glasses or contact lenses (questionnaire item #6147). They were able to select more than one answer unless they stated myopia or presbyopia as the reason; in that case, no additional choices were allowed. Classification of participants as having strabismus or amblyopia was based on their answer on this item. Note that the "self-reported" strabismus phenotype included all subtypes of strabismus.

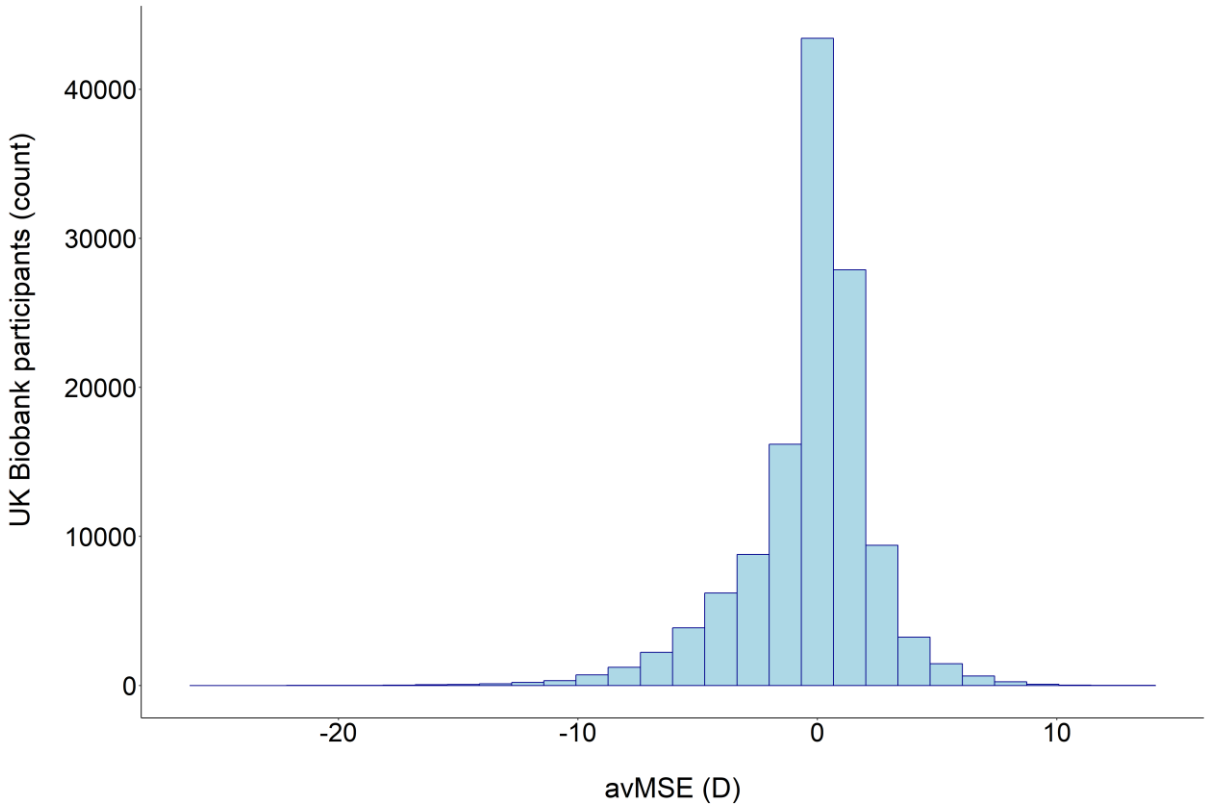

Figure 2.2 Distribution of the avMSE refractive error variable in UK Biobank participants.

2.1.3 Phenotypes in UK Biobank: Birth Weight

All UK Biobank participants were asked to enter their birth weight using either metric or Imperial system units (kg or pounds/ounces, respectively). The Imperial values were converted into kg by the UK Biobank researchers; only metric unit data (field #122) were used in current study. A total of 280,264 individuals recalled their birth weight at the baseline or follow-up visit. The distribution of the phenotype is shown on *Figure 2.3*.

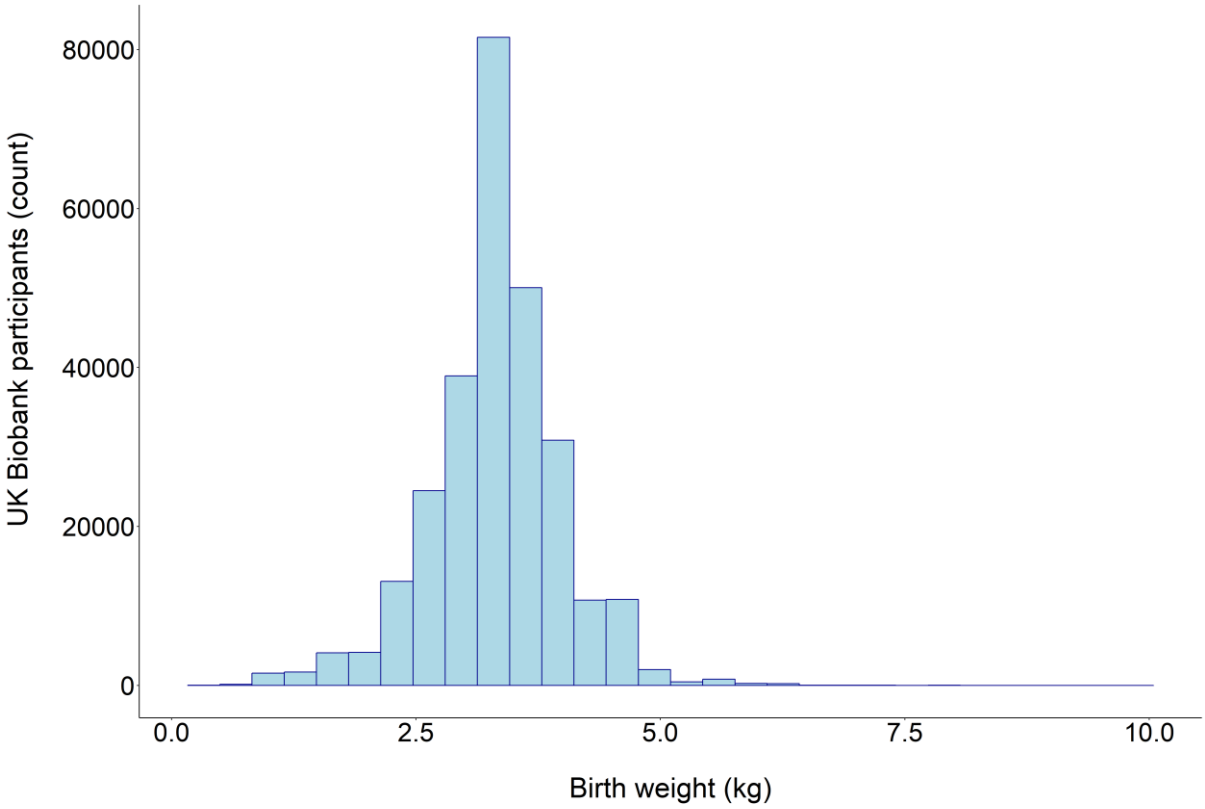

Figure 2.3 Distribution of Birth weight in UK Biobank participants.

2.1.4 Genetic Data in UK Biobank

DNA samples from blood were genotyped by UK Biobank researchers either on the UK BiLEVE genotyping array (n = 49,950) or the UK Biobank Axiom genotyping array (n = 438,427) at approximately 800,000 genetic markers. UK Biobank made two releases of the genetic data: the interim release comprising of data for approximately 150,000 participants was made publicly available in 2015. Imputation was performed using IMPUTE2 (240) at more than 70,000,000 variants using a merged UK10K Project and 1000 Genomes Project Phase 3 reference panel (241, 242). The information metric (INFO) with values between 0 and 1, which

corresponds to the imputation quality (INFO near 1 indicates a high quality of imputation) is reported by IMPUTE2. The second release included genetic data for the full cohort and was made available in July 2017. Prior to release of the genetic data, Bycroft et al. (243) performed extensive quality control of the genotype data and imputed additional genotypes using the HRC reference panel (244) combined with the UK10K Project merged with 1000 Genomes Project Phase 3 reference panel (241, 242). The UK Biobank researchers carried out the imputation using IMPUTE4 software (URL: https://jmarchini.org/software/). This resulted in a dataset with more than 93,000,000 autosomal genetic variants in 487,377 individuals. Due to issues with the imputation of the UK10K and 1000 Genomes variants initially, only variants from the HRC panels were used in current work.

A total of 409,728 UK Biobank participants were in the 'White British ancestry' subset defined by Bycroft et al. (245). A set of well-imputed variants (with IMPUTE4 INFO metric >0.9, minor allele frequency (MAF) > 0.005, missing rate ≤ 0.01, and an 'rs' variant ID prefix) were selected and LD-pruned using the --indep-pairwise 50 5 0.1 command in PLINK 2.0 (246). I created a genetic relationship matrix (GRM) with PLINK 2.0 in order to identify a set of unrelated individuals (command --rel-cutoff 0.025). This left 338,253 unrelated participants of White British ancestry.

2.1.5 Covariates

Age when participants attended the UK Biobank assessment centre, Townsend Deprivation Index, first ten ancestry principal components (PC) and average refractive error (for GWAS for Strabismus) were used as quantitative covariates. Genotyping array (UK BiLEVE or UK Biobank Axiom) and gender were coded as binary variables.

2.2 ALSPAC

In the Avon Longitudinal Study of Parents and Children (ALSPAC) (247, 248) a total of 14,541 pregnant women resident is the southwest of England expecting delivery between 01/04/1990 - 31/12/1992 were recruited. At the age of one year, 13,988 of children born to these women were still alive. Mothers and their partners completed a series of questionnaires. At the age of 7 years, children attended a research clinic

assessment that included an ophthalmic examination carried out by members of the ALSPAC research team (249).

2.2.1 Phenotypes in ALSPAC

Non-cycloplegic refraction was assessed using a Canon R50 (Canon, Tokyo) autorefractor. Refractometry without using cycloplegic agents in children been previously reported as a source of misclassification, with overestimation the proportion of children with myopia (250, 251). To assess the degree of bias, the ALSPAC researchers conducted a nested study comparing non-cycloplegic and cycloplegic refractions in a sample of children with visual acuity through a pinhole worse than 0.2 logMAR (n=345) (252). The sensitivity and specificity of non-cycloplegic refraction were 61% and 99%, respectively, to identify hyperopia of at least 2.00 D. Therefore, lack of cycloplegia will have led to an underestimation of prevalence of hyperopia. Eye movements were assessed by an orthoptist during the ophthalmic assessment. Simultaneous and alternate prism cover tests both at near (33 cm) and distance (6 m) were used to quantify the eye misalignment. Based on the presence in normal viewing with both eyes, strabismus was classified as manifest (> 1 prism Dioptre ($^{\Delta}$) on prism cover testing). With regards to the direction, strabismus was classified as convergent, divergent, vertical or mixed. Horizontal (convergent or divergent) strabismus included manifest and 'large' latent deviations (≥ 10^{Δ} for convergent and ≥ 15^{Δ} for divergent). The ocular phenotype data was available in 5,200 of the ALSPAC children. The following phenotypes were used in the current thesis: Parentally reported history of strabismus (n = 145 cases), manifest strabismus (n = 116 cases), esotropia (n = 143 cases) and exotropia (n = 28 cases). Children with amblyopia were defined (249) as those with a history of patching treatment and/or with an interocular difference in best-corrected visual acuity for each eye of 0.2 logMAR units where the worse-seeing eye had a best-corrected visual acuity of <0.3 logMAR, and the eye looked normal on dilated ophthalmoscopy.

2.2.2 Genetic Data in ALSPAC

ALSPAC children were genotyped using the Illumina HumanHap550 quad chip. Quality control (individual call rate > 0.97, SNP call rate > 0.95, MAF > 0.01, Hardy-Weinberg Equilibrium (HWE) > 1E-7, removal of participants clustered as non-

Europeans), resulted in genetic data available for 9,237 children. ShapeIT (v2.r644) was used to phase haplotypes and IMPUTE2 was used for imputation against all 2,186 reference haplotypes (including non-Europeans) in the December release of the 1000 Genomes reference haplotypes (Version 1 Phase 3). A total of 8,237 children had genotype data available.

2.3 Consortia Data

Summary statistics data for GWAS meta-analyses from the Early Growth Genetics (EGG) consortium, the Consortium for Refractive Error and Myopia (CREAM) and the SOCIAL Science Genetics Association Consortium (SSGAC) have been used in the current thesis.

2.3.1 EGG Consortium

Summary statistics for a GWAS meta-analysis for birth weight (253) were downloaded from the Early Growth Genetics (EGG) consortium website (URL: www.egg-consortium.org). The EGG consortium meta-analysis comprised of 18 European population-based studies (n=26,836). EGG excluded participants from multiple births and those with a gestational age <37 weeks. Genetic variant effect sizes were reported as the change in birth weight (BW) Z-score per copy of the test allele. Studies included in the EGG meta-analysis were: two sub-samples from the 1958 British Birth Cohort (B58C-WTCCC, B58C-T1DGC); the Avon Longitudinal Study of Parents And Children (ALSPAC-Children); the Children's Hospital of Philadelphia (CHOP); the COpenhagen Prospective Study on Asthma in Childhood (COPSAC-2000); the European Prospective Investigation of Cancer (EPIC); the Erasmus Rucphen Family (ERF) study; two sub-samples from the Generation R study (Generation R (Discovery 1), Generation R (Discovery 2)); the Helsinki Birth Cohort Study (HBCS); the Lifestyle – Immune System – Allergy (LISA) study; the Northern Finland 1966 Birth Cohort (NFBC1966); two sub-samples of singleton births from the Netherlands Twin Register (NTR1, NTR2); the Orkney Complex Disease Study (ORCADES); the Prevention and Incidence of Asthma and Mite Allergy (PIAMA) study; the Raine study (RAINE); and the Sorbs study (SORBS).

2.3.2 CREAM Consortium

Summary statistics for a GWAS meta-analysis for refractive error were provided by the CREAM consortium (238). The CREAM meta-analysis was based on 29 studies of European ancestry (n=44,192 in total), most of which were population based. Spherical equivalent averaged between the two eyes was used as the refractive error phenotype. Studies included in the CREAM meta-analysis were: The 1958 British Birth Cohort (B58C); Antioxydants, Lipides Essentiels, Nutrition et maladies OculaiRes (ALIENOR) study; Avon Longitudinal Study of Parents And Children (ALSPAC Mothers); Australian and New Zealand Registry of Advanced Glaucoma (ANZRAG) study; Age Related Eye Disease Study (AREDS); Brisbane Adolescent Twins Study (BATS); Blue Mountains Eye Study (BMES); Croatia-Korcula; Croatia-Split; Croatia-Vis; Diabetes Control and Complications Trial (DCCT); Estonian Genome Center, University of Tartu (EGCUT); the European Prospective Investigation of Cancer (EPIC-Norfolk); the Erasmus Rucphen Family study (ERF); Fuch's Endothelial Corneal Dystrophy Controls (FECD); Finnish Twin Study of Ageing (FITSA); Framingham Eye Study (FES); Gutenberg Health Study 1 (GHS-1); Gutenberg Health Study 2 (GHS-2); "Kooperative Gesundheitsforschung in der Region Augsburg" (KORA); Ogliastra Genetic Park (OGP) Talana; the Orkney Complex Disease Study (ORCADES); Rotterdam Study I, II, and III (RS I-III), the Twins Eye Study in Tasmania (TEST); the TwinsUK study; Wisconsin Epidemiologic Study of Diabetic Retinopathy (WESDR); and the Young Finns Study (YFS).

2.3.3 SSGAC Consortium

Summary statistics data for an Educational attainment GWAS meta-analysis were provided by the SSGAC Consortium (https://www.thessgac.org/data) (254). Participants from 71 cohorts were included. As the cohorts were heterogeneous in terms of country of birth and birth cohort, the SSGAC researches constructed the *EduYears* (the number of years in schooling completed) variable as a harmonised measure of years-of-education, with levels equivalent to the seven educational categories defined by the 1997 International Standard Classification of Education of the United Nations Educational, Scientific and Cultural Organization.

2.4 Methodology

Various statistical approaches have been used in the current thesis. This section of the thesis describes these methods and the statistical software used.

2.4.1 Causality

The concept of causality is important in epidemiology; however, there is no single definition accepted by the scientific community. Various definitions exist, such as the classification of causes proposed by Parascandola and Weed (2001) (255). These authors described five definitions corresponding to typical epidemiological approaches. The outcome *cannot* be seen without *Necessary Cause*; whereas, it *must* occur if the cause is *Sufficient*. *Sufficient Components* could be seen as a development of the previous definition and refers to the condition when more than one component is, individually, not *sufficient* for the outcome, but combined produce the effect. *Production* refers to the situation when the cause produces the effect or, in other words, affects or changes the outcome. The cause increases the chance (probability) of the outcome occurrence under the *Probabilistic Causation* definition. Under the *Counterfactual* definition, the observed effect given the cause differs from the effect that would have occurred if the effect had been different or even absent.

To distinguish between causal and non-causal association, Sir Austin Bradford Hill proposed nine concepts that should be considered: 1) strength of the association, 2) consistency, 3) specificity, 4) temporality, 5) biological gradient, 6) plausibility and 7) coherence (256).

In epidemiology, randomized controlled studies(RCT) are considered to be the gold standard to establish causality in the relationship between a risk factor and an outcome (257-259). However, sometimes it is not possible or feasible to conduct a RCT (in regard to the current thesis it is impossible to conduct a RCT to estimate the causal effect of birth weight on refractive error and it would be unethical to run a study in schoolchildren in order to assess the causality in the relationship between education and myopia). Hence, other approaches rather than RCT are required; I describe these approaches in the current chapter.

2.4.2 Linear and Logistic Regression

Linear and logistic regression are statistical methods widely used in both observational (260-262) and genetic epidemiology, as well as for genome-wide association studies (GWAS) (263, 264).

Linear regression analysis is used to describe the linear relationship between a continuous outcome Y (dependent variable) and one or more predictors x (independent variables). Predictors in the linear regression model can be continuous, binary or categorical variables. The basic model equation is

$$Y = \beta_0 + \beta_1 x + \varepsilon$$

Equation 2.1 Linear Model Equation. Y = continuous outcome, β_0 = intercept, β_1 = regression coefficient for X, x = observed value of x, ε = error term.

The regression coefficient represents the change in the outcome per unit of measurement change in the predictor variable. In a GWAS, the regression coefficient β_1 is the average change in the trait of interest per copy of the 'risk' allele (the allele that confers a risk of developing disease) carried.

Logistic regression is used when the dependent variable is binary. This model has been utilized in cohort and in case-control studies. Logistic regression answers the same question as linear regression (i.e. Is there an association between the predictor(s) and the outcome?). However, for logistic regression the nature of the outcome variable, the measure of the association, is an odds ratio (OR). OR in epidemiology can be used to assess whether a predictor variable is a risk factor for the outcome. If the predictor variable is associated with higher odds of the outcome, then then OR is greater than 1. Values of the OR less than 1 describe an association of the predictor variable with lower odds of the outcome; and when the predictor variable does not affect the outcome, the OR is equal to 1.

A logit transformation is a critical part of the logistic regression framework, meaning that the estimated association is between the natural logarithm of the odds and predictor(s).

$$\ln\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 x + \varepsilon$$

Equation 2.2 Logistic regression. $ln\left(\frac{\pi}{1-\pi}\right)$ *= natural logarithm of odds, π = is the probability of the binary outcome variable being = 1;* β_0 *= intercept,* β_1 *= natural logarithm of the OR of a one unit increase in* x*,* ε *= error term.*

In a highly unbalanced case-control association study (i.e. where the number of controls is much greater than the number cases) or in studies of rare variants, the standard logistic regression estimate can be biased (265). In that case, a Firth logistic regression that produces a bias-corrected estimate is a more effective approach (266, 267).

Linear mixed models (LMM) extend the traditional linear model framework to allow the inclusion not only of predictor variables as fixed effects but also predictor variables as random effects. In terms of GWAS analysis, LMMs typically model familial relatedness as a random effect. This has the advantage of adjusting for cryptic relatedness and partially accounting for population structure. Allowing the inclusion of relatives in the GWAS sample leads to the increase of the statistical power of the association study. BOLT-LMM v.2.3 (268, 269) is the most commonly-used software for implementing a LMM GWAS analysis.

2.4.3 Instrumental Variable Analysis

As linear regression estimates are susceptible to bias due to confounding (270, 271), they cannot be used to infer the *causality* in the relationship between a predictor and an outcome. Instrumental Variable (IV) analysis is a widely used approach (e.g. econometrics, health sciences, and epidemiology) that seeks to control for confounding. An instrumental variable or instrument is a variable associated with the exposure, but that is not independently associated with the outcome. To be a valid instrument, a variable Z must have following properties:

1) Be associated with the exposure (X)
2) Only affect the outcome via the exposure
3) Have no common cause with the outcome (Y)

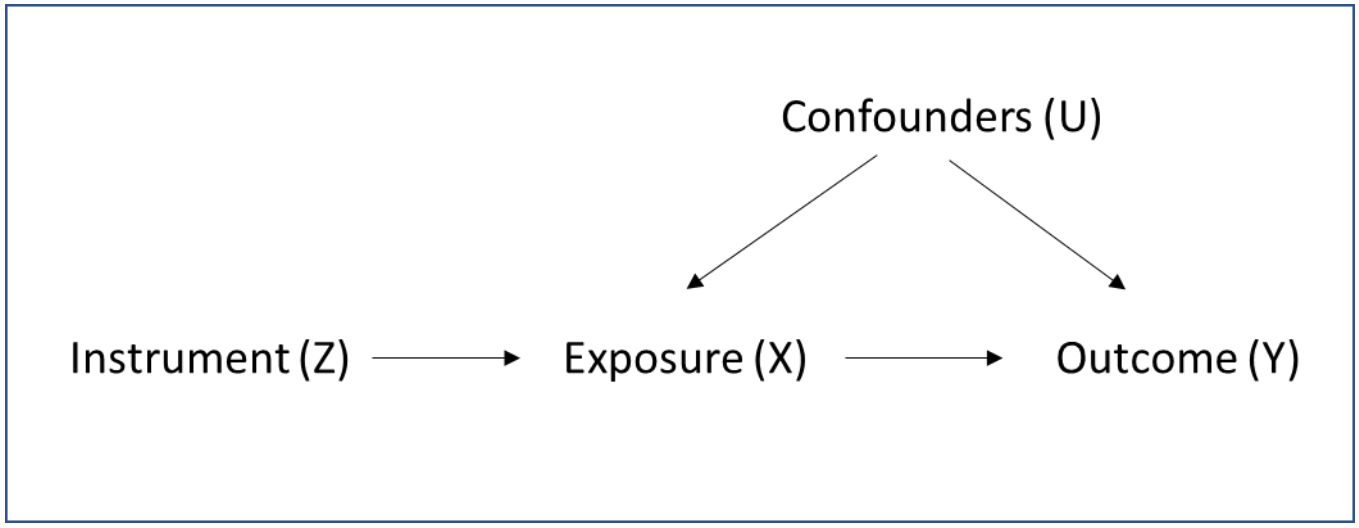

Figure 2.4 Causal Diagram illustrating the IV assumptions for the instrument Z. U is the set of unmeasured confounders of the exposure-outcome association.

The first assumption implies that there is a systematic difference in the average level of exposure in groups defined by the instrument value. The second assumption or the "exclusion restriction" guarantees that unmeasured confounders are distributed equally between these groups. The final assumption means than there is no direct effect of the instrument on the outcome and no other route for the instrument to affect the outcome except via the exposure.

Several methods can be used to estimate the causal effect in IV analysis. The ratio of coefficients method or Wald method (272) is used with a single binary instrument.

$$Wald\ estimate = \frac{\Delta Y}{\Delta \mathrm{X}} = \frac{\Delta\widehat{y_1} - \Delta\widehat{y_0}}{\Delta\widehat{x_1} - \Delta\widehat{x_0}}$$

Equation 2.3 Ratio method (Wald) estimate. ΔY = difference in the outcome, ΔX = difference in exposure.

Mendelian randomisation (MR) analysis is a research method that uses genetic variants as instrumental variables (discussed in detail in Section 2.4.4). In the MR

settings, the causal effect of the exposure on the outcome can be estimated as (259):

$$\widehat{\theta_J} = \frac{\widehat{\beta_{YJ}}}{\widehat{\beta_{XJ}}}$$

Equation 2.4 Wald estimator in MR analysis. $\widehat{\beta Y_J}$ is the estimated effect of jth SNP on the outcome, $\widehat{\beta X_J}$ is the estimated effect of jth SNP on the exposure.

Another method used to calculate the IV estimate is two-stage least squares (2SLS). Under 2SLS, the causal effect is estimated from two regression models.

The first stage (*Equation 2.5*) is a regression of the exposure *X* on the instrumental variable *Z.* In the second stage *Equation 2.6*), the outcome is regressed on the predicted values for the exposure from the first stage.

$$X = \alpha_0 + \alpha_1 Z + \alpha_2 Cov_1 + \ldots + \alpha_{i+1} Cov_i + \omega$$

Equation 2.5 First stage regression model. Cov = set of covariates, ω = error term.

$$Y = \varepsilon_0 + \varepsilon_1 \widehat{X} + \varepsilon_2 Cov_1 + \ldots + \varepsilon_{i+1} Cov_i + \zeta$$

Equation 2.6 Second stage regression model. Cov = set of covariates, ζ = error term.

2.4.4 Inverse-Variance Weighting and Meta-Analysis

Inverse-variance weighting is a statistical method used to average a set of independent effect sizes. In MR analysis settings the variance can be taken as $\frac{se(\widehat{\beta_{YJ}})^2}{\widehat{\beta_{XJ}}^2}$ so, the ratio estimates from each SNP can be combined into one averaged estimate using a formula (273) :

$$\hat{\theta}_{IVW} = \frac{\sum \widehat{\beta_{Y_J}} \widehat{\beta_{X_J}} se(\widehat{\beta_{Y_J}})^{-2}}{\sum \widehat{\beta_{X_J}}^{2} se(\widehat{\beta_{Y_J}})^{-2}}$$

Equation 2.7 IVW MR Estimate. $\widehat{\beta Y_J}$ is the estimated effect of jth SNP on the outcome, $\widehat{\beta X_J}$ is the estimated effect of jth SNP on the exposure.

IVW MR analysis provides a consistent causal effect estimate if all instruments are valid. Methods used when IV assumptions are violated for a proportion of the IVs are described in section 2.4.4.

Meta-analysis is a statistical approach widely used to synthesize evidence from multiple studies in order to increase the precision of the individual study estimate. In genetic epidemiology, meta-analysis is often used to combine the results of individual GWAS (sometimes underpowered due to their small sample size) (274). METAL software (275) was used to perform meta-analyses in studies described in the following chapters. The fixed-effect IVW method is implemented in METAL.

2.4.5 Mendelian Randomisation Analysis

Mendelian randomisation is a research method that exploits the IV framework using genetic variant(s) as instrumental variables to estimate the causal effect of an exposure on an outcome of interest. The concept of using genetic variants in the IV settings was introduced by Katan (276) who hypothesized that the use of genetic predisposition to a risk factor of a disease is a method to overcome issues with reverse causation. He suggested to use genetic variation in the *Apolipoprotein E* gene (*APOE*) associated with lower levels of serum cholesterol in order to assess the causality in the relationship between cholesterol level and risk of cancer. Because of the random assignment of alleles, individuals with the 'cholesterol lowering' allele should not have any systematic difference from those carrying other allele. Hence, comparing the difference in the *APOE* distribution between individuals with and without cancer would allow one to draw a conclusion regarding the *causal* effect of low serum cholesterol level and the increased risk of cancer. The term 'Mendelian randomisation' was first used by Gray and Wheatley (277) and was popularised by Davey Smith and Ebrahim (278).

In genetically informed research methods the counterfactual approach (Section 2.4.1) is used for causal inference (279, 280). Under this scenario, the researcher wants to assess the effect of the exposure on the outcome and compare with the effect of another expose (or, no exposure). In real life it is impossible to observe the outcome under both conditions in a single individual; hence, it is crucial to access the *exchangeability* (281), the balance between exposed and unexposed groups on observed and unobserved confounders. As mentioned previously, in the current chapter the random allocation of SNP alleles to egg/sperm cells provides independence from potential confounders (259).

In support of the validity of causal relationships identified using MR, the results of MR studies have either confirmed the results of previously conducted RCTs (282) or have been supported by subsequent RCTs (283, 284). Specifically, it has been shown that when the effect of a genetic variant is known and effectively mimics the effect of the therapeutic agent used in an RCT, the results of MR study can be used to anticipate the results of clinical trials (285).

MR has found recognition in public health; e.g. the results from MR studies formed a substantial part of the body of the scientific evidence in establishing the guidelines for the management of dyslipidaemias by the European Society of Cardiology and the European Atherosclerosis Society (286).

To be a valid instrument, a genetic variant must meet all the criteria listed in section 2.4.2. To meet the first criteria, a genetic variant must be associated with the risk factor. Currently, the results of numerous GWAS are publicly available, therefore it is straightforward to find variants that satisfy the first assumption. Typically, genetic variants found to be associated with the risk factor at genome wide significance threshold ($p < 5$ E-08) are selected for use as instruments.

It is not possible to test the second assumption that the genetic variant is not associated with any confounders of the exposure-outcome association. Even though the association between known confounders and the genetic variant can be tested, there are still many unknown or unmeasured confounders. A common solution is to test the genetic variant for association with known measured confounders. Genetic variants that have effects on more than one trait are called ‘pleiotropic’; and their inclusion in MR analysis can be problematic as it could violate the second or third IV assumptions. If pleiotropy occurs due to the association of the genetic variant with an intermediary trait that in turn affects the exposure (i.e. the

relationship between the SNP and the exposure is mediated by one or more intermediary traits; known as *'vertical pleiotropy'*), then this variant *can* be used as a valid instrument. However, if the genetic variant affects the exposure via two or more independent pathways (*'horizontal pleiotropy'*), then it does not satisfy the IV criteria and, therefore it is *not* a valid instrument for an MR analysis. When pleiotropic effects are biased in one direction, this is referred to as *'directional pleiotropy'*; in case of *'balanced pleiotropy'* the averaged pleiotropic effect is zero.

An MR study can use either a single instrument (one SNP or a 'polygenic risk score') or multiple instruments (when more than one SNP has been identified in a GWAS). The use of multiple instruments has the advantage of increasing the precision in estimating the causal effect (if all SNP are strongly associated with the exposure and do not suffer from weak instrument bias). The downside of this approach is that including more genetic variants increases the risk that some of the variants will display horizontal pleiotropy and therefore bias the causal effect estimate. Multiple genetic variants can be combined into a single polygenic risk score (287) that protects against weak instrument bias.

MR analyses can be performed in a single sample of participants (one-sample MR) or the variant-exposure and variant-outcome association can be tested in different samples (two-sample MR) (288). The second design has several advantages as it protects against weak instrument bias and allows the researcher to use summary level data (typically made publicly available by large research consortia).

With the increasing number of GWAS conducted, more and more genetic variants have been identified; hence horizontal pleiotropy has become an emerging problem in MR studies. A number of methods have been introduced in order to test for and to correct for horizontal pleiotropy (289-291).

The MR-Egger method exploits the Egger regression meta-analysis approach and requires multiple genetic instruments to be tested (292). The intercept term (the direct effect of variants on the outcome) is included in the inverse variance weighted model used to estimate the effect in MR. An intercept that is shifted from zero indicates the presence of directional pleiotropy; however, the test still provides a valid causal effect estimate.

Median and mode-based MR analyses have been proposed to address the same issue of pleiotropy in MR studies. Median-based analysis (293) produces a causal

effect estimate that is valid when up to one half of the information is from genetic variants that are not valid instruments. A mode-based estimate (294) is consistent even when the majority of variants are not valid instruments.

2.4.6 Regression Discontinuity Analysis

Regression Discontinuity (RD) is a quasi-experimental design widely used in Econometrics research, which posits that when assignment to different levels of an exposure is based on a continuously measured random variable (the 'rating' or 'forcing' variable), for individuals either side of some threshold, the assignment of the exposure is essentially random (295, 296). The RD design has high internal validity (i.e. producing a consistent *causal* estimate) and can be easily performed given all the required conditions are met, yet it has a limitation of low external validity (i.e. it is not straightforward to generalize the RD results) (297, 298). In epidemiology, the RD design can be used in situations where the assignment to experimental or control groups is based on some cut-off point in the forcing variable.

There are two main types of RD design: *'sharp'* and '*fuzzy'* (299). In the sharp design, all the participants with values of the forcing variable above the cut-off value (to the right side of the cut-off point) are assigned to the experimental group, and all the participants with forcing variable values below the cut-off (those to the left site of the cut-off point) are assigned to the control group. In other words, assignment is determined by the value of the forcing variable (300). In the sharp RD design, the discontinuity in the outcome provides a measure of the average causal effect of the treatment (average treatment effect; ATE).

In a fuzzy RD design, not all participants with value of forcing variable more than cut-off value are assigned to the experimental group (the probability of receiving the treatment is less than 1 for those on the right-hand side of the cut-off point). Thus, the researcher must take this into account when estimating the effect in a fuzzy RD study. A local average treatment effect (LATE) is typically used to address this issue (301, 302).

Let Y denote the outcome, X denote the value of the forcing variable, Z is a binary indicator of being on the right side of the cut-off point, T denotes if the individual is assigned to the experimental group (received a treatment), and c is the cut-off point:

$$Z = \begin{cases} 1 \; if \; X \geq c \\ 0 \; if \; X < c \end{cases}$$

$$T = \begin{cases} 1 \; participant \; received \; a \; treatment \\ 0 \; pardicipant \; does \; not \; receive \; a \; treatment \end{cases}$$

Then, the ATE at the cut-off point can be estimated as:

$$ATE = \; E(Yi \mid Z = 1) - E(Yi \mid Z = 0)$$

Equation 2.8 Average treatment effect in sharp RD.

In fuzzy RD, the LATE is estimated as:

$$LATE = \frac{E(Y|Z = 1) - E(Y|Z = 0)}{E(T = 1|Z = 1) - (T = 1|Z = 0)}$$

Equation 2.9 Local average treatment effect in fuzzy RD.

The 2SLS method (section 2.4.2) is the approach typically used for estimating the ATE and the LATE (300, 302). However, other techniques can be used for the estimation of the LATE in a RD study (e.g. maximum likelihood-based or Bayesian estimation methods).

The optimal size for the unit of measurement of the forcing variable, which is known as the 'bin size', can be selected using two main ways: 1) informal, and 2) formal. Using the first approach, the researcher can try different bin sizes and visually compare graphical plots of the bin size vs. outcome relationship to find the most informative one. A bin size that is too small makes the plot noisy and less informative. If the bin size is too big, it is difficult to observe the discontinuity at the cut-off point. The formal method utilises the F-test, which tests the hypothesis that the proposed bin size is too big and a smaller one will provide a better fit to the data.

It is crucial for the validity of an RD study to select the right number of bins either side of the cut-off point (a '*bandwidth*' selection). Optimal bandwidth selection methods utilise the mean square error concept. There is always a trade-off between precision and bias, as using the widest bandwidth increases the number of

observations, hence the estimate is more precise. On the other hand, using a wider bandwidth decreases the accuracy of estimation and increases the risk of bias. Cross-validation (300, 303) or 'plug-in' procedures (304) are typically used to select the optimal bandwidth.

2.4.7 Genome-wide Association Study

In a genome-wide association study (GWAS), a systematic series of single marker tests are performed for genetic variants located across the genome, in order to identify variants associated with the phenotypic trait of interest. A large number of genetic variants are tested in GWAS, which creates a multiple testing problem. Hence, the commonly accepted genome-wide significance threshold is 5.0E-08. This threshold is based on performing a Bonferroni correction for independent SNP across the human genome; that is, it accounts for LD between variants as well as the total number of variants tested (305, 306). Statistical software used to perform GWAS in this thesis include PLINK v1.9/2.0 and BOLT-LMM v2.3 (246, 268, 307). Linear or logistic regression, or linear mixed models are the statistical methods used by these software packages.

A Manhattan plot is typically used for the visualisation of the GWAS results with the negative logarithm of the association p-value for each SNP as a function of genomic position (308). A regional association plot is used to inspect the association within a specific region (typically several hundred kb either side of the SNP of the interest); *LocusZoom* is an online tool that displays such information (https://my.locuszoom.org/) (309).

2.4.8 Population Stratification and Genomic Inflation Factor

Population stratification occurs when a GWAS sample comprises of subgroups of different ancestries with different allele frequencies and is considered to be a source of a bias in association studies (310, 311). If the different ancestries have a different disease prevalence, then any variant whose allele frequency differs between the ancestry groups will be associated with the phenotypic, causing spurious GWAS results.

With regard to the GWAS results, population stratification can be tested by examining the quantile-quantile (Q-Q) plot, a plot of the observed p-values for each SNP from the association study against expected values (typically from the χ^2 distribution) (312). An example of a Q-Q plot is presented in *Figure 2.5*.

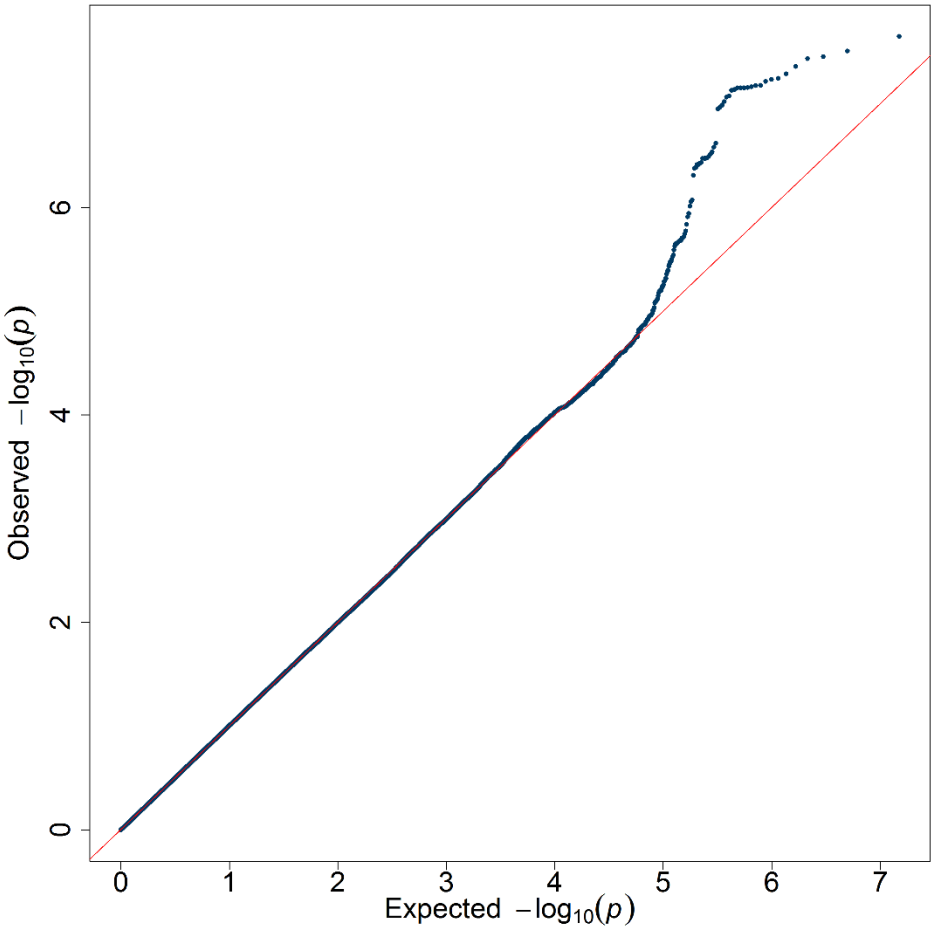

Figure 2.5 Quantile-Quantile (Q-Q) plot for the GWAS for strabismus (Section 4.3.2)

Population stratification occurs when observed association is stronger than expected (i.e. excessive numbers of significant p-values in the observed association compared to the expected distribution). The genomic inflation factor (λ_{GC}) is used to quantify the deviation described above and is calculated as follows (313):

$$\lambda_{GC} = \frac{Median\ of\ the\ observed\ \ \chi^2 statistics}{0.456}$$

Equation 2.10 Genomic inflation factor, 0.456 is the median of the theoretical χ^2 distribution with one degree of freedom

A value of λ_{GC} more than 1.0 indicates that inflation may have occurred due to population stratification; this can be corrected (albeit approximately) by dividing the observed χ^2 statistics by λ_{GC}.

2.4.9 Polygenic Risk Score

To date, GWAS have identified thousands of SNP associated with various phenotypic traits. Typically, the effect of a single SNP on the trait is small. A method of combining the effects of multiple SNP into a polygenic risk score (PRS) was introduced first in animal genetics (314), and later in human genetics by researchers in the psychiatric genetics field (315). PRS can also be used to predict disease risk (316, 317).

Construction of a PRS requires the effect sizes of the SNP–trait association from a 'training' GWAS to be used as weights in an independent replication sample.

$$PRS = \sum_{j=1}^{n} X_j \times W_j$$

Equation 2.11 Polygenic risk score calculation. PRS = polygenic risk score, X_j = allele count (0,1 or 2) for an individual genetic variant j, W_j = weight of individual genetic variant j, n = total number of variants included.

The typical method to construct a PRS is to take the set of independent SNP associated with the trait at genome-wide significance level and to use their effect sizes (regression coefficients from a GWAS) as weights. A Bayesian approach that takes into account the presence of linkage disequilibrium (LD) was proposed to increase the predictive power of the polygenic scores (318); this approach is implemented in the LDpred software package. LDpred version 1.0.6 was used in the current analyses.

2.4.10 Fine-mapping

The aim of fine-mapping is to find the individual genetic variant or variants with the highest likelihood of having a causal effect on the trait of interest among the potentially thousands of SNP in a region identified in a GWAS. One common way to perform a fine mapping analysis is a conditional analysis, when the lead variant with the strongest association with the trait within a certain genomic region is used as a covariate in the association study. This analysis is an iterative one, i.e. it is performed in steps, using the SNP with the lowest p-value from the previous step included as a covariate, until there is no variant attaining the predefined level of significance. This conditional approach can only be performed if the 'raw' genotype data are available. Alternatively, a Bayesian fine-mapping approach can be implemented. FINEMAP v1.3 (319) exploits a shotgun stochastic search algorithm to evaluate the evidence for multiple causal SNP in a region. The maximum number of causal SNP considered is 5 and, given user-selected prior probabilities, the software defines the posterior probability for there being 1-5 causal variants in the region. Methods such as FINEMAP can be performed using GWAS summary statistics rather than requiring access to 'raw' genotypes.

2.4.11 Prediction of the pathogenicity of a variant

Several algorithms can be used for predicting the pathogenicity of a genetic variant; in the current thesis I used the *Ensembl* project website (320, 321) which summarises the results of the various algorithms. The Combined Annotation Dependent Depletion (CADD) tool calculates a 'deleteriousness' score for a SNP or indel in the human genome (322). A machine learning model is used to integrate 63 annotations (conservation, functional prediction, genetic context, epigenetic modification) for a variant of interest into a single score. Two types of scores are calculated for each variant: 'raw' and 'scaled'. The former scores are proposed to be used in comparison of the CADD scores between groups of variants; the latter ones are recommended for fine-mapping or identification of causal variants (322).

SIFT (323) and PolyPhen-2 (324) are two other widely-used algorithms for prediction of the deleteriousness of genetic variants; both are based on sequence homology used to assess whether the amino acid substitution caused by the variant of interest may affect the protein function.

Three databases are available for interpreting the significance of genetic variant identified in GWAS to a range of health outcomes: PhenoScanner (325), ClinVar (326) and gnomAD (327). Each database provides valuable, complimentary information on understanding of the pathogenicity and potential functional significance of variants.

Chapter 3. Assessing Causality in the Relationship Between Birth Weight and Refractive Error: a Mendelian Randomisation Study

3.1 Introduction

Changes in the refractive status in children with low or extremely low birth weight have been reported previously (328). The most common refractive error found in association with low birth weight in preterm-born children is myopia (329, 330), although some studies have also reported an increased prevalence of hyperopia and astigmatism (170, 331). However, the relationship between birth weight within the range of 2.5 to 4.5 kg, which is typically termed the 'normal range' (253, 332), and refractive error has been investigated in only a few observational studies. The normal range was chosen as it encompasses more than 85% of newborns with gestational age between 38-42 weeks (between 2nd and 91st centile by the UK-WHO charts) (333). In a study conducted in Northern Ireland in a cohort of 12-13 year-old children (334), a positive association between birth weight and refractive error was found (OR = 1.31, 95% CI 0.81 to 2.12). In contrast, a study of Singaporean infants aged 3 years (335) reported a decrease of -0.02 D in refractive error per one kg increase in birth weight. However, in both studies there was limited statistical support for the findings. In adults, a trend towards a positive association of birth weight and refractive error has been observed in two population-based studies. The 1958 British Birth Cohort (336) reported the association of myopia with lower birth weight (OR = 0.90, $p < 0.05$); in the Gutenberg Health Study (337) the observed effect was -0.017 D (95% CI -0.011 to 0.023, $p < 0.001$) per SD reduction in birth weight.

The current study aimed to test and to quantify the causal relationship between birth weight (within the range of 2.5 to 4.5 kg) and refractive error, using Mendelian randomisation analysis.

3.2 Methods

3.2.1 Selection of Participants

A GWAS for birth weight (Section 2.1.3) and a GWAS for refractive error (*avMSE;* Section 2.1.2) were performed in UK Biobank participants. The study was restricted to UK Biobank participants with self-reported White ethnicity and with genetically-

inferred European ancestry. From the remaining 444,857 individuals, those with autosomal heterozygosity more than 4 SD from the mean were excluded. 95,504 participants had valid autorefraction data and reported no history of eye surgery and were selected as the sample for the GWAS for refractive error (*Figure 3.1*). The exclusion criteria for the sample for the GWAS for birth weight followed the criteria adopted by Horikoshi et.al (253). These criteria were: birth weight outside the range of 2.5-4.5kg, being part of a multiple birth, a difference > 0.5kg in self-reported birth weight at the baseline and follow-up UK Biobank assessment clinics. A total of 347,701 individuals without *avMSE* data or with a history of eye surgery were included in this sample for further selection. Among them, 151,067 participants had no valid birth weight data and were excluded. Only individuals with a birth weight inside the range of 2.5-4.5kg, being a part of a singleton birth with consistent birthweight data were included, resulting in 162,039 UK Biobank participants being selected as the sample for the GWAS for birth weight; with no overlap with the *avMSE* sample (Figure 3.1).

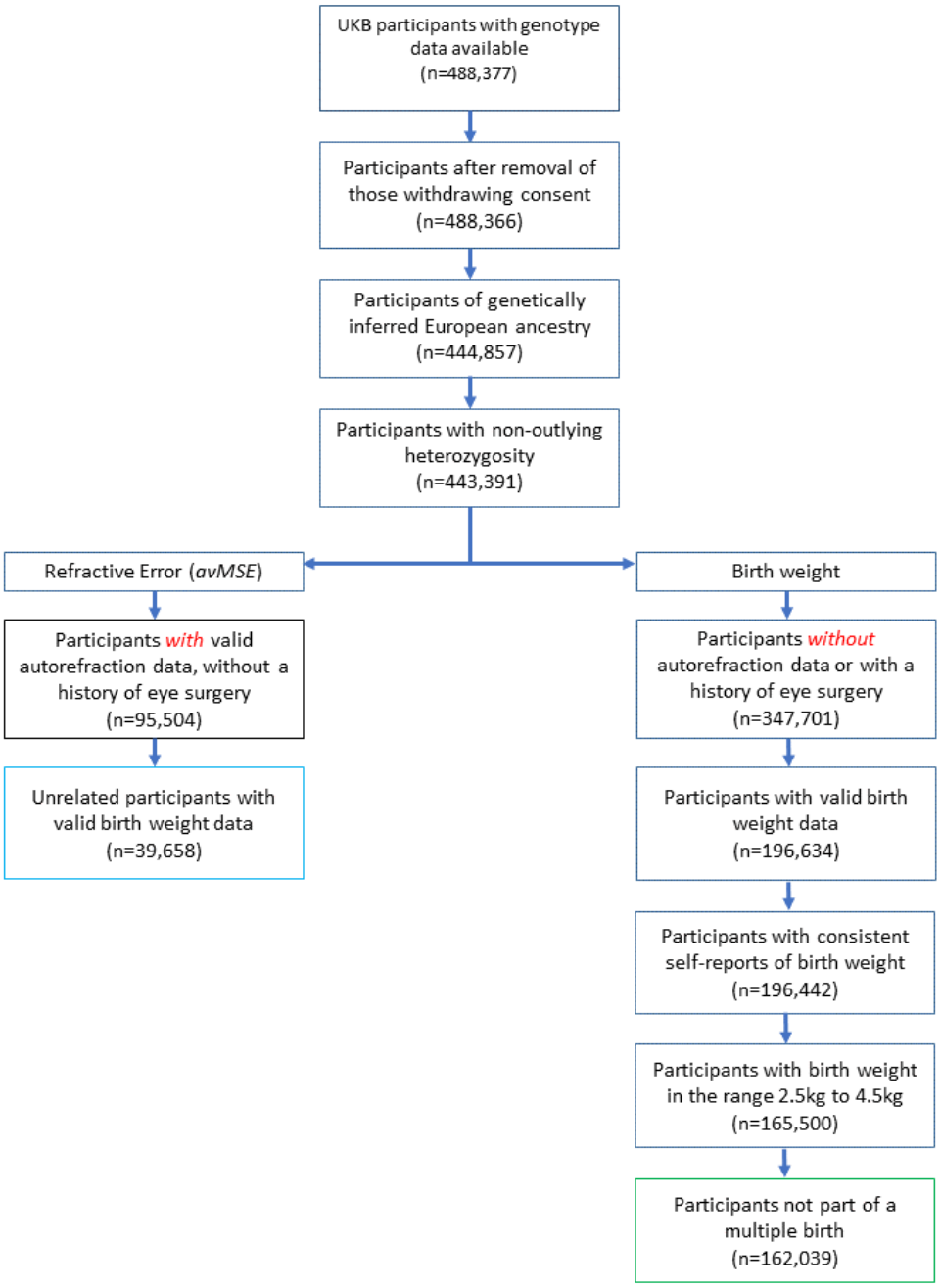

Figure 3.1 Selection of UK Biobank participants for GWAS for refractive error and for GWAS for birth weight

One-sample Mendelian randomisation requires access to the individual level data and valid measurements of both exposure and the outcome. Hence, a set of individuals with both valid birth weight and autorefraction data was selected (n = 39,968) for a one-sample MR analysis. Observational (OLS) association was tested in the same sample of 39,968 UK Biobank participants to keep consistency in the results.

Genotyping is described in Section 2.1.4.

3.2.2 Summary Statistics for GWAS for Refractive Error and for Birth Weight (published by research consortia)

Summary statistics from a GWAS meta-analysis for refractive error was provided by the CREAM Consortium (238). The CREAM Consortium meta-analysis was performed in 44,192 individuals of European ancestry from 29 studies (Section 2.3.2). Summary statistics from a GWAS meta-analysis for birth weight were obtained from the Early Growth Genetics (EGG) consortium, as described in Section 2.3.1. The EGG meta-analysis comprised of 18 studies in a total of 26,836 European individuals; with 7 loci being found genome-wide significantly associated with birth weight.

3.2.3 GWAS for Refractive Error and for Birth Weight in UK Biobank

A genome wide association study for birth weight (Z-score) was performed in the 162,039 UK Biobank participants described above (*Figure 3.1*). Birth weight data were standardized (Z-transformed) as follows: birth weight value minus mean birth weight)/ standard deviation of birth weight. Z-scores in that case are the birth weight values under the assumption of a normal distribution of the data. This Z-score standardisation allowed me to compare the UK Biobank birth weight data with the birth weight data from the EGG consortium (which were also Z-transformed). A total of 10.4 million SNP passed the quality control filters (INFO > 0.9, MAF ≥ 0.05, per-marker missing rate < 0.015 and per-individual missing rate < 0.025; Section 2.1.4) and were tested for the association with birth weight, using BOLT-LMM software (Section 2.4.6). The same set of genetic markers was used to perform a GWAS for refractive error (mean spherical equivalent in Dioptres averaged between the two eyes) in 95,504 individuals (*Figure 3.1*). Both of the traits of interest were considered

as continuous traits and the infinitesimal linear mixed model was used in the analyses (269), with the adjustment for age, gender, genotyping array and the first 10 ancestry principal components. BOLT-LMM allows the researcher to use related individuals in the analysis, accounting for population stratification and relatedness. A genetic relatedness matrix for the analysis was created as described in Section 2.4.6.

Two GWAS meta-analyses were performed. In the first, the summary statistics from the GWAS for refractive error in UK Biobank (n = 95,504) were meta-analysed with the summary statistics from the CREAM Consortium GWAS for refractive error (n = 44,192). In the second, summary statistics from the GWAS for birth weight in UK Biobank (n = 162,039) were meta-analysed with summary statistics from the EGG consortium GWAS for birth weight (n = 26,836). Meta-analyses were performed using METAL software (275) as described in Section 2.4.4.

3.2.4 Instrumental Variable Selection

A total of 75 SNP were associated with birth weight at genome-wide significance level ($p < 5$ E-08) in a GWAS meta-analysis in a total sample of 188,875 UK Biobank and EGG individuals. Among them, two variants were not present in the CREAM consortium summary statistics, leaving 73 SNP. Polygenic risk scores (Section 2.4.8) can be used as an instrument in Mendelian randomisation analyses as they provide increased statistical power and avoid 'weak instrument bias' (338, 339). These 73 SNP were combined into a polygenic risk score for use as an instrumental variable in Mendelian randomisation analysis. The PRS was calculated using PLINK 1.9 ('--score' command).

In order to assess the variance in birth weight explained by the PRS, the increase in R^2 was calculated for a regression of birth weight (Z-score) on the PRS plus age, gender, genotyping array, and the first 10 ancestry principal components compared to a regression of birth weight on age, gender, genotyping array, and the first 10 ancestry principal components alone. Both models were fit in the sample of UK Biobank participants with both refractive error and birth weight data available (n = 39,658).

3.2.5 Relationship Between Birth Weight and Refractive Error Assessed Using Linear Regression and Mendelian Randomisation Analyses

A linear regression analysis was performed in order to estimate the (observational) association between birth weight and refractive error in the sample of unrelated UK Biobank individuals with information on both refractive error and birth weight (n = 39,658).

A 1-sample MR analysis was performed in the same sample of 39,658 participants using a Two-stage Least Squares (2SLS) model implemented in the *AER* R package. The PRS for birth weight was used as the instrumental variable. To test if the causal effect estimate was biased due to the PRS being a 'weak instrument' (340), the F-statistic from the first stage regression was taken. Statistical power was calculated using the *mRnd* online web tool for a type 1 error rate $\alpha = 0.05$ (https://shiny.cnsgenomics.com/mRnd/).

A set of sensitivity analyses was carried out. Estimating a causal effect using a 1-sample MR design with a PRS as a single instrumental variable leaves the potential for bias due to unbalanced pleiotropy. Therefore, weighted median and MR-Egger sensitivity analyses were performed with the 73 genetic variants associated with birth weight used individually as instrumental variables.

Partial overlap between the EGG and CREAM consortia samples may have introduced the risk of bias (259, 341). Hence, as an additional sensitivity analysis, the 2-sample MR analysis was repeated but using GWAS regression (beta) coefficients and standard errors for the second stage (association with refractive error) obtained solely from the GWAS for refractive error in UK Biobank participants, i.e. excluding the results from the CREAM consortium.

3.3 Results

3.3.1 GWAS Results

A GWAS for self-reported birth weight was performed in 162,309 adult UK Biobank participants. This identified 2,087 single nucleotide polymorphisms attaining the genome-wide significance threshold of $p < 5.0e-08$. After excluding variants within ±500 kb from the lead SNP in each region, or those having a pairwise LD of $r^2 > 0.2$

with the lead variant, there were 63 SNP independently associated with birth weight (*Table 3.1*).

A total of 75 SNP associated with birth weight was identified in the meta-analysis of the UK Biobank and EGG consortium GWAS summary statistics. Genetic variants rs1530624 and rs1058026 were not present in the CREAM consortium summary statistics data and were excluded from the downstream analysis (*Table 3.2*).

3.3.2 Observational Analysis

In the sample of UK Biobank participants (n = 39,658) with information about both refractive error and birth weight, those with a lower birth weight were more myopic and those with a higher birth weight were less myopic, on average. Linear regression quantified the relationship as a +0.04D (95% CI 0.02 to 0.07, p = 0.002) shift in refractive error per 1 SD increase in birth weight. Note that the above regression model was adjusted for age and gender.

3.3.3 Mendelian Randomisation Analyses

A 1-sample Mendelian randomisation analysis was used to estimate the causal effect of birth weight on refractive error in the above sample (n = 39,658) of UK Biobank individuals. A PRS for birth weight was used as an instrumental variable in the analysis. The PRS explained 1.2% (p < 2.2E-16) of the variance in birth weight in this sample. The 1-sample MR analysis suggested that a 1 SD increase in birth weight resulted in a refractive error that was +0.28 D (95% CI 0.05 to 0.52, p = 0.02) more hyperopic. With a sample size of 39,658 individuals, there was 62% power to detect the estimated effect of +0.28 D/SD. The current study had 80% power to detect an effect of birth weight on refractive error ≥ +0.35 D/SD (*Figure 3.2*). Note that the power calculation was performed given the observational association of +0.04 D/SD and under the assumption that +0.28 D/SD is the correct estimation of the *causal* effect

The 73 genetic variants robustly associated with birthweight were also used individually as instrumental variables in 2-sample MR sensitivity analyses. MR-Egger and weighted median methods designed to provide a valid estimate even in case of directional pleiotropy and pleiotropy of a proportion of the variants,

respectively, were performed. In the IVW analysis, the causal effect was estimated as 0.15 D/SD (95% CI 0.00 to 0.30, p = 0.044). The MR-Egger intercept was not statistically significantly different from zero, suggesting no evidence of directional genetic pleiotropy (). The MR-Egger causal effect estimate was +0.18 D/SD (95% CI -0.41 to 0.78, p = 0.54). The weighted median MR estimated the causal effect to be +0.18 D/SD (95% CI 0.02 to 0.35, p = 0.03). The estimates obtained by all the methods were consistent in terms of direction and themagnitude (

MR method	Effect size (D/SD)	95% CI		p
Simple median	0.15	-0.02	0.31	0.076
Weighted median	0.18	0.02	0.35	0.029
Penalized weighted median	0.20	0.03	0.36	0.020
IVW	0.15	0.00	0.30	0.044
Penalized IVW	0.21	0.08	0.33	0.002
Robust IVW	0.17	0.02	0.32	0.025
Penalized robust IVW	0.21	0.07	0.34	0.002
MR-Egger	0.18	-0.41	0.78	0.545
(intercept)	0.00	-0.02	0.01	0.918

Table 3.3), supporting the hypothesis of a causal effect of birth weight on refractive error.

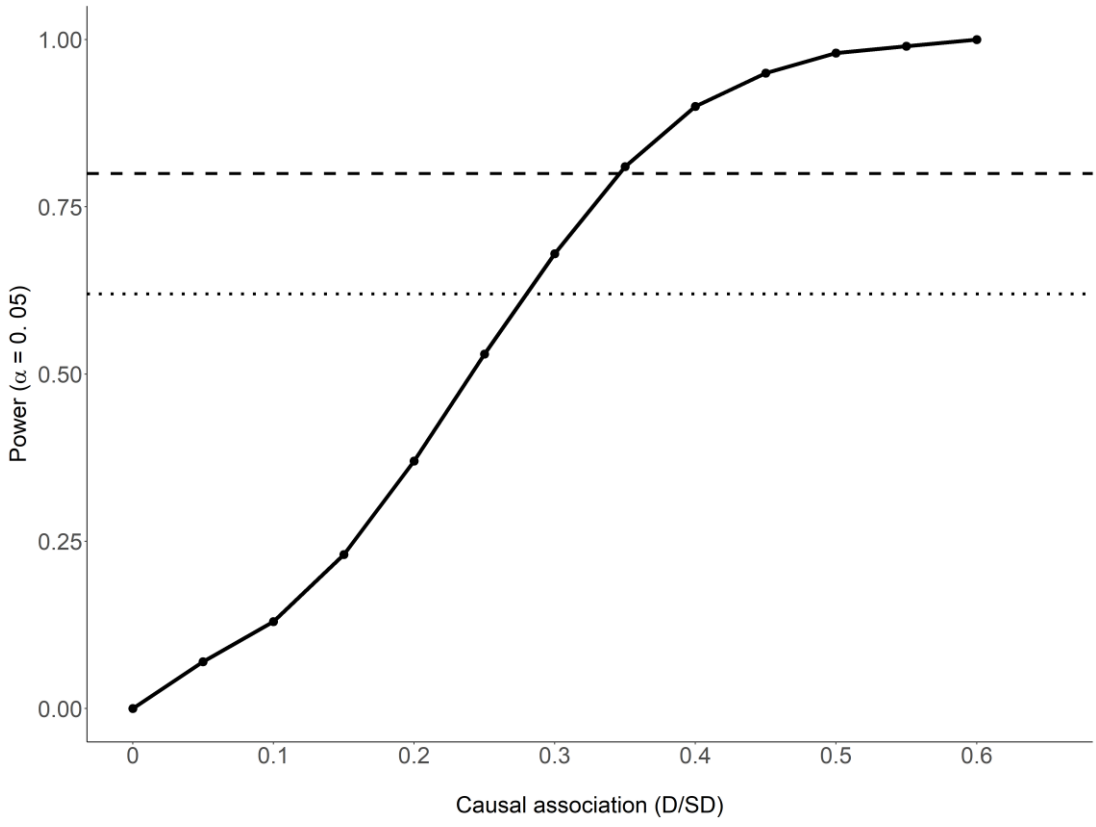

Figure 3.2 Power calculation for 1-sample Mendelian randomisation in 39,658 UK Biobank participants. The dotted line indicates 62% power for the estimated effect size of +0.28D/SD, the dashed line indicates 80% power.

Variant	Chromosome	Position	Effect Allele	Reference Allele	EAF	Effect size	SE	P-value	Nearest Gene
rs13322435	3	156795468	A	G	0.60	0.046	0.004	8.90E-38	CCNL1-LEKR1
rs7306710	12	66376091	T	C	0.48	0.040	0.004	5.50E-30	HMGA2
rs11889583	2	46484882	C	T	0.30	-0.036	0.004	8.50E-21	LOC101926974
rs4144829	4	17903654	C	T	0.26	0.036	0.004	2.90E-20	LCORL
rs10440833	6	20688121	T	A	0.74	0.036	0.004	2.10E-19	CDKAL1
rs11719201	3	123068744	C	T	0.75	-0.034	0.004	1.60E-17	ADCY5
rs12507427	4	145567471	T	A	0.57	-0.029	0.004	9.60E-17	HHIP
rs145965565	9	98273305	T	G	0.91	-0.046	0.006	6.70E-15	PTCH1
rs4766578	12	111904371	T	A	0.49	-0.027	0.003	8.70E-15	ATXN2
rs11187141	10	94467145	A	T	0.62	-0.027	0.004	1.40E-13	HHEX
rs3213221	11	2157044	C	G	0.37	0.026	0.004	3.90E-13	IGF-INS-IGF2
rs11698914	20	31327144	C	G	0.23	0.030	0.004	1.50E-12	COMM7
rs60839038	18	20711927	A	C	0.53	-0.025	0.004	2.40E-12	CABLES1
rs7772579	6	152042502	A	C	0.72	0.027	0.004	3.40E-12	ESR1
rs45560031	7	44149457	G	C	0.94	0.053	0.008	3.40E-12	AEBP1
rs7031933	9	125791823	C	G	0.88	-0.037	0.005	3.70E-12	RABGAP1
rs8034564	15	99190601	G	A	0.42	0.024	0.004	7.30E-12	IGF1R
rs7080472	10	96012950	G	T	0.58	-0.024	0.004	1.30E-11	PLC1
rs6533183	4	106133184	C	T	0.34	0.025	0.004	1.80E-11	TET2
rs1012167	20	39159119	T	C	0.59	-0.024	0.004	2.60E-11	MAFB
rs72656010	8	57122215	T	C	0.87	0.034	0.005	4.30E-11	PLAG1
rs446994	17	7116853	C	A	0.42	-0.023	0.004	4.50E-11	DLG4
rs2934844	6	166142456	A	T	0.33	-0.024	0.004	6.80E-11	PDE10A
rs1870940	1	154984363	G	A	0.73	-0.025	0.004	1.50E-10	ZBTB7B
rs2780226	6	34199092	C	T	0.09	0.039	0.006	1.80E-10	HMCA1
rs1801253	10	115805056	G	C	0.26	-0.025	0.004	2.10E-10	ADRB1
rs55958435	15	96852638	A	G	0.75	0.026	0.004	2.10E-10	NR2F2-AS1
rs854037	5	57091783	A	G	0.81	0.028	0.004	2.90E-10	LOC101928539

rs116794974	4	135009470	G	C	0.98	-0.089	0.014	5.00E-10	PABPC4L
rs10823318	10	70979924	A	T	0.31	-0.023	0.004	5.50E-10	HKDC1 - LOC101928994
rs73354194	17	79905947	T	C	0.98	-0.071	0.012	7.10E-10	MYADML2
rs2779165	19	4915447	G	C	0.19	0.028	0.005	8.60E-10	UHRF1
rs11853312	15	38626475	A	C	0.88	-0.033	0.006	1.20E-09	SPRED1
rs13032333	2	9620701	T	A	0.69	-0.023	0.004	2.00E-09	IAH1
rs11066188	12	112610714	G	A	0.59	0.021	0.004	2.50E-09	HECTD4
rs13397722	2	160515651	G	A	0.63	0.021	0.004	2.70E-09	BAZ2B
rs2901307	10	124128443	C	T	0.53	-0.021	0.003	3.60E-09	PLEKHA1
rs77994518	1	119288847	G	A	0.87	-0.031	0.005	4.60E-09	LOC100421281
rs13266210	8	41533514	A	G	0.79	0.025	0.004	4.80E-09	ANK1
rs13235314	7	23609295	C	T	0.64	0.021	0.004	5.30E-09	LOC442517
rs28378473	9	139245460	T	C	0.27	0.023	0.004	5.70E-09	GPSM1
rs1547550	2	191845725	C	G	0.65	-0.021	0.004	7.60E-09	STAT1
rs12213664	6	109353413	T	C	0.88	-0.031	0.005	8.30E-09	SESN1
rs1051412	20	10654563	A	C	0.51	-0.020	0.004	9.00E-09	JAG1
rs6062527	20	62413838	G	C	0.76	-0.024	0.004	9.20E-09	ZBTB46
rs6930558	6	141878920	G	T	0.25	-0.023	0.004	1.20E-08	RPS3AP23
rs4681464	3	148623443	C	T	0.39	0.020	0.004	1.30E-08	CPA3
rs7090337	10	104775483	G	A	0.59	0.020	0.004	1.50E-08	CNNM2
rs1057412	6	31321752	T	G	0.89	0.032	0.006	1.60E-08	HLA-B
rs138056307	10	96573948	A	T	0.96	-0.048	0.009	1.80E-08	CYP2C19
rs62496903	8	6446938	C	T	0.91	-0.035	0.006	2.00E-08	MCPH1
rs11242219	5	133848760	A	G	0.42	-0.020	0.004	2.10E-08	LOC101927934
rs10269069	7	47276495	A	G	0.91	-0.034	0.006	2.10E-08	TNS3
rs6931945	6	35525751	G	A	0.63	-0.020	0.004	2.20E-08	LOC101929309
rs9526475	13	48964748	T	C	0.27	0.022	0.004	2.40E-08	RB1
rs113934718	17	29214880	C	A	0.73	0.022	0.004	2.90E-08	ATAD5
rs751543	9	119122342	C	T	0.29	-0.021	0.004	3.70E-08	PAPPA

rs11539637	15	91428290	C	T	0.47	-0.019	0.003	4.00E-08	FES
rs2344500	12	26876620	G	A	0.51	0.019	0.003	4.10E-08	ITPR2
rs76895963	12	4384844	T	G	0.98	-0.073	0.013	4.10E-08	CCND2
rs9647618	6	37099123	A	G	0.78	0.023	0.004	4.30E-08	PIM1
rs4350272	10	25056118	A	G	0.27	0.021	0.004	4.40E-08	LOC105376456
rs1649328	2	109392777	C	G	0.76	-0.022	0.004	4.50E-08	RANBP2

Table 3.1 Lead genetic variants associated with birth weight in a GWAS in 162,309 UK Biobank individuals. EAF, Effect Allele frequency; SE, standard error; variants within ±500kb from the variants listed and with pairwise LD $r^2 > 0.8$ are not included.

						UK Biobank			EGG Consortium			UK Biobank + EGG Consortium		
Variant	Chromosome	Position	Effect Allele	Other Allele	EAF	Effect size	SE	P-value	Effect size	SE	P-value	Effect size	SE	P-value
rs905938	1	154991389	T	C	0.26	-0.025	0.004	2.70E-10	-0.035	0.010	2.30E-04	-0.026	0.004	5.00E-13
rs12039075	1	119285701	A	G	0.13	-0.030	0.005	8.30E-09	-0.034	0.012	4.20E-03	-0.031	0.005	1.42E-10
rs1244983	1	215031614	A	G	0.48	-0.019	0.003	5.40E-08	0.027	0.009	2.00E-03	0.020	0.003	4.97E-10
rs17413015	1	161644811	T	C	0.18	-0.020	0.005	7.70E-06	0.044	0.011	9.90E-05	0.024	0.004	1.56E-08
rs12404679	1	43458458	T	C	0.14	0.025	0.005	9.10E-07	-0.032	0.013	1.30E-02	-0.026	0.005	4.47E-08
rs1547550	2	191845725	C	G	0.36	-0.021	0.004	7.60E-09	-0.028	0.010	4.20E-03	-0.022	0.004	1.30E-10
rs13032333	2	9620701	A	T	0.30	-0.023	0.004	2.00E-09	0.017	0.009	5.40E-02	0.022	0.004	3.66E-10
rs4665083	2	160430127	A	G	0.38	0.021	0.004	7.10E-09	0.032	0.018	8.70E-02	0.021	0.004	1.71E-09
rs754868	2	43185532	A	G	0.43	-0.018	0.004	2.70E-07	-0.021	0.008	1.30E-02	-0.019	0.003	1.05E-08
rs1724184	2	109408776	T	C	0.25	-0.022	0.004	4.70E-08	-0.014	0.010	1.60E-01	-0.021	0.004	2.27E-08
rs6713865	2	23899807	A	G	0.18	0.023	0.004	4.40E-07	0.026	0.011	1.90E-02	0.023	0.004	2.56E-08
rs13322435	3	156795468	A	G	0.41	0.046	0.004	8.90E-38	0.080	0.009	5.10E-18	0.050	0.003	1.18E-51
rs11708067	3	123065778	A	G	0.24	-0.034	0.004	1.80E-17	-0.059	0.010	6.20E-09	-0.038	0.004	5.71E-24
rs7647448	3	148623774	T	C	0.40	0.020	0.004	1.30E-08	-0.026	0.008	1.90E-03	-0.021	0.003	1.05E-10
rs2306700	3	142123841	T	C	0.13	-0.026	0.005	3.60E-07	0.030	0.012	1.40E-02	0.027	0.005	1.52E-08
rs2061456	4	17998426	A	C	0.25	-0.036	0.004	6.60E-20	0.045	0.009	1.10E-06	0.038	0.004	4.31E-25

rs17776795	4	145544836	A	G	0.42	-0.029	0.004	1.30E-16	-0.027	0.008	1.00E-03	-0.029	0.003	6.41E-19
rs7674220	4	106148758	C	G	0.36	0.025	0.004	1.90E-11	-0.033	0.009	1.20E-04	-0.026	0.003	1.33E-14
rs854037	5	57091783	A	G	0.20	0.028	0.004	2.90E-10	0.041	0.011	8.70E-05	0.030	0.004	4.45E-13
rs3756668	5	67596088	A	G	0.47	0.019	0.003	5.90E-08	-0.032	0.008	1.30E-04	-0.021	0.003	8.47E-11
rs11242219	5	133848760	A	G	0.41	-0.020	0.004	2.10E-08	-0.027	0.008	1.20E-03	-0.021	0.003	1.43E-10
rs2946164	5	157884706	T	C	0.27	-0.020	0.004	2.70E-07	0.030	0.010	1.80E-03	0.022	0.004	2.76E-09
rs10440833	6	20688121	A	T	0.26	0.036	0.004	2.10E-19	-0.048	0.009	1.30E-07	-0.038	0.004	2.68E-25
rs7772579	6	152042502	A	C	0.28	0.027	0.004	3.40E-12	0.034	0.009	2.00E-04	0.028	0.004	3.63E-15
rs2780226	6	34199092	T	C	0.10	0.039	0.006	1.80E-10	-0.040	0.016	1.10E-02	-0.039	0.006	7.30E-12
rs1202309	6	19919986	A	G	0.30	0.020	0.004	1.30E-07	-0.036	0.009	6.00E-05	-0.023	0.004	1.32E-10
rs1472267	6	166185025	A	G	0.38	0.020	0.004	2.40E-08	0.027	0.009	2.30E-03	0.021	0.003	2.28E-10
rs12197912	6	109338147	T	C	0.12	-0.030	0.005	2.10E-08	0.034	0.012	5.80E-03	0.030	0.005	3.64E-10
rs2064317	6	35477032	A	G	0.39	-0.020	0.004	3.80E-08	-0.021	0.009	1.20E-02	-0.020	0.003	1.68E-09
rs9390019	6	143033056	T	C	0.22	0.023	0.004	5.40E-08	0.026	0.011	1.50E-02	0.023	0.004	3.16E-09
rs6930558	6	141878920	T	G	0.25	-0.023	0.004	1.20E-08	0.016	0.010	9.20E-02	0.022	0.004	3.67E-09
rs12534093	7	23502974	A	T	0.24	0.023	0.004	1.70E-08	-0.031	0.010	1.90E-03	-0.024	0.004	1.67E-10
rs11972595	7	72876445	T	C	0.07	-0.034	0.007	2.30E-07	0.055	0.017	1.20E-03	0.037	0.006	2.05E-09
rs798548	7	2760935	T	C	0.29	0.019	0.004	5.80E-07	0.030	0.009	6.80E-04	0.021	0.004	3.29E-09
rs2301680	7	93116299	A	G	0.50	-0.016	0.003	5.30E-06	-0.035	0.008	2.20E-05	-0.019	0.003	4.70E-09
rs10265057	7	47275737	A	G	0.10	-0.034	0.006	2.30E-08	-0.021	0.015	1.70E-01	-0.032	0.006	1.18E-08
rs2075067	7	44263028	A	C	0.13	0.025	0.005	2.30E-06	-0.040	0.013	2.00E-03	-0.027	0.005	3.10E-08
rs13266210	8	41533514	A	G	0.21	0.025	0.004	4.80E-09	0.017	0.010	8.30E-02	0.024	0.004	1.43E-09
rs12543725	8	142247979	A	G	0.41	0.018	0.004	2.20E-07	-0.027	0.008	1.50E-03	-0.020	0.003	1.71E-09
rs2197131	8	106639923	T	G	0.21	-0.022	0.004	2.90E-07	-0.031	0.010	1.70E-03	-0.023	0.004	2.48E-09
rs8180991	8	126500350	C	G	0.24	0.019	0.004	7.10E-06	0.039	0.010	6.40E-05	0.022	0.004	1.15E-08
rs6984782	8	57135889	T	C	0.14	0.031	0.005	2.40E-09	0.003	0.013	8.40E-01	0.027	0.005	1.91E-08
rs7819593	8	106115172	T	C	0.25	-0.021	0.004	5.40E-07	-0.021	0.009	2.20E-02	-0.021	0.004	3.68E-08
rs10512249	9	98256309	A	G	0.09	-0.046	0.006	1.30E-14	0.031	0.014	2.80E-02	0.044	0.006	1.77E-15
rs587364	9	125760863	T	C	0.16	0.034	0.005	1.10E-11	0.025	0.013	6.50E-02	0.033	0.005	2.02E-12
rs751543	9	119122342	T	C	0.29	-0.021	0.004	3.70E-08	0.022	0.010	2.80E-02	0.021	0.004	3.12E-09

rs2418135	9	113901309	A	G	0.48	-0.017	0.003	1.10E-06	0.025	0.008	2.10E-03	0.018	0.003	1.44E-08
rs3933326	9	123633948	A	G	0.32	-0.019	0.004	8.10E-07	-0.024	0.009	6.70E-03	-0.020	0.004	2.03E-08
rs11187144	10	94469980	T	C	0.39	-0.026	0.004	2.60E-13	-0.025	0.008	3.40E-03	-0.026	0.003	2.92E-15
rs1801253	10	115805056	C	G	0.27	-0.025	0.004	2.10E-10	0.045	0.010	2.50E-06	0.028	0.004	1.84E-14
rs2274224	10	96039597	C	G	0.44	-0.023	0.004	2.80E-11	0.022	0.008	7.50E-03	0.023	0.003	7.83E-13
rs9645500	10	70986723	T	G	0.32	-0.023	0.004	1.20E-09	-0.035	0.009	1.10E-04	-0.025	0.004	1.01E-12
rs6585827	10	124165615	A	G	0.47	-0.020	0.003	1.20E-08	0.014	0.008	9.20E-02	0.019	0.003	3.30E-09
rs1163238	10	104943993	A	G	0.39	0.020	0.004	2.70E-08	-0.014	0.009	1.10E-01	-0.019	0.003	8.17E-09
rs3213225	11	2156536	A	G	0.39	0.026	0.004	7.10E-13	-0.032	0.023	1.70E-01	-0.026	0.004	2.76E-13
rs2923093	11	10360934	A	T	0.40	0.018	0.004	4.00E-07	0.021	0.009	1.70E-02	0.018	0.003	1.96E-08
rs5030317	11	32410337	C	G	0.29	0.020	0.004	5.50E-07	0.020	0.009	3.10E-02	0.020	0.004	4.75E-08
rs7968682	12	66371880	T	G	0.47	0.039	0.003	2.40E-29	-0.049	0.008	2.40E-09	-0.041	0.003	6.80E-37
rs4766578	12	111904371	A	T	0.48	-0.027	0.003	8.70E-15	0.013	0.008	1.10E-01	0.025	0.003	8.47E-15
rs2306547	12	26877885	T	C	0.45	0.019	0.003	5.30E-08	-0.019	0.008	2.30E-02	-0.019	0.003	3.37E-09
rs9535006	13	48849803	A	G	0.25	-0.022	0.004	3.20E-08	-0.017	0.009	6.80E-02	-0.021	0.004	6.78E-09
rs7173576	15	96841524	T	C	0.26	0.025	0.004	4.30E-10	-0.029	0.009	2.00E-03	-0.025	0.004	3.26E-12
rs11853312	15	38626475	A	C	0.11	-0.033	0.006	1.20E-09	-0.017	0.013	2.00E-01	-0.031	0.005	9.71E-10
rs4965425	15	99181663	T	C	0.47	0.019	0.004	3.20E-08	-0.017	0.008	4.20E-02	-0.019	0.003	3.68E-09
rs11539637	15	91428290	T	C	0.47	-0.019	0.003	4.00E-08	0.028	0.024	2.30E-01	0.019	0.003	2.14E-08
rs222852	17	7140606	A	G	0.44	0.023	0.004	9.90E-11	0.036	0.008	1.70E-05	0.025	0.003	2.55E-14
rs11082304	18	20720973	T	G	0.49	-0.024	0.003	1.10E-11	0.001	0.008	9.60E-01	0.020	0.003	2.98E-10
rs2779165	19	4915447	C	G	0.18	0.028	0.005	8.60E-10	-0.025	0.012	3.50E-02	-0.027	0.004	9.55E-11
rs255774	19	54723355	A	G	0.46	0.019	0.004	6.40E-08	0.016	0.009	7.30E-02	0.019	0.003	1.32E-08
rs1129156	19	40719076	T	C	0.26	0.021	0.004	9.90E-08	0.015	0.009	9.20E-02	0.020	0.004	3.03E-08
rs6016373	20	39154095	A	G	0.41	-0.024	0.004	4.40E-11	-0.036	0.008	1.60E-05	-0.026	0.003	9.48E-15
rs2064933	20	31314625	A	G	0.24	0.029	0.004	3.20E-12	0.020	0.009	3.40E-02	0.027	0.004	4.66E-13
rs6062541	20	62449320	T	C	0.23	-0.023	0.004	3.10E-08	0.034	0.010	7.40E-04	0.024	0.004	1.47E-10

Table 3.2 The 73 SNP used for constructing the PRS for birth weight and their association with the trait in UK Biobank, EGG and the combined sample. EAF, Effect Allele frequency; SE, standard error

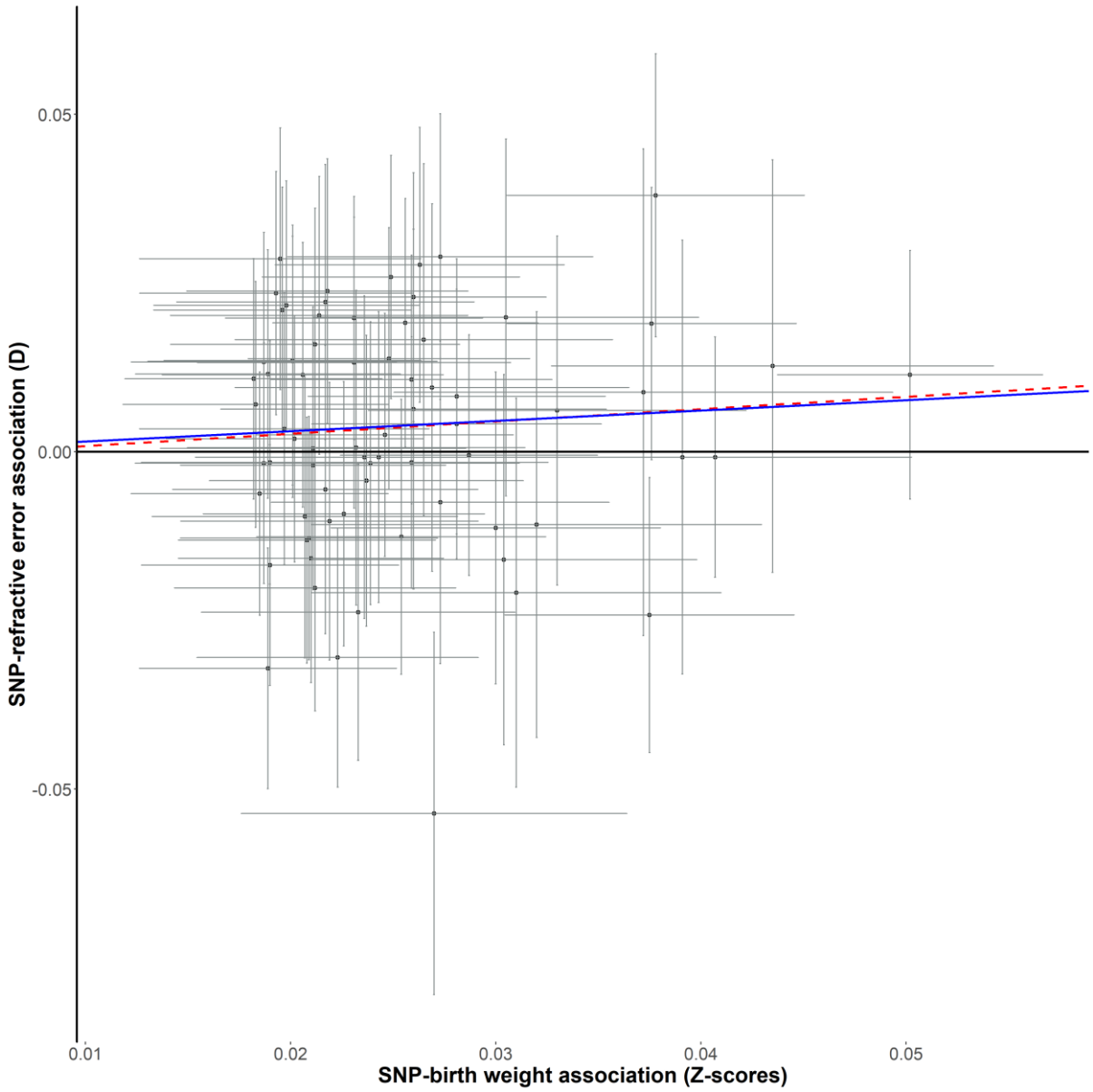

Figure 3.3 IVW and MR-Egger scatter plot. The solid blue line corresponds to the IVW estimate; the dashed red line corresponds to MR-Egger estimate. Error bars correspond to 95% confidence intervals.

MR method	Effect size (D/SD)	95% CI		p
Simple median	0.15	-0.02	0.31	0.076
Weighted median	0.18	0.02	0.35	0.029
Penalized weighted median	0.20	0.03	0.36	0.020
IVW	0.15	0.00	0.30	0.044
Penalized IVW	0.21	0.08	0.33	0.002
Robust IVW	0.17	0.02	0.32	0.025
Penalized robust IVW	0.21	0.07	0.34	0.002
MR-Egger	0.18	-0.41	0.78	0.545
(intercept)	0.00	-0.02	0.01	0.918

Table 3.3 Causal effect of birth weight on refractive error. Stage 1 summary statistics were taken from a meta-analysis of UK Biobank and EGG Consortium GWAS for birth weight (n = 188,039). Stage 2 summary statistics were taken from a meta-analysis of UK BIOBANK and the CREAM GWAS for refractive error (n = 139,884). IVW, inverse-variance weighted.

The maximum sample overlap in the 2-sample MR analysis was estimated as 11,685 participants, corresponding to 8% of the sample used in the second stage of the analysis. A sensitivity MR analysis was performed using only data from UK Biobank participants in stage 2, which reduced the maximum possible overlap to 6,038 participants (6% of the second stage sample). The results of this additional sensitivity analysis (*Table 3.4*) were consistent with the result from the MR analysis performed in the full sample.

MR method	Effect size (D/SD)	95% CI		p
Simple median	0.28	0.08	0.48	0.007
Weighted median	0.31	0.11	0.51	0.002
Penalized weighted median	0.41	0.21	0.61	0.000
IVW	0.21	0.02	0.40	0.035
Penalized IVW	0.27	0.12	0.42	0.000
Robust IVW	0.24	0.04	0.44	0.018
Penalized robust IVW	0.28	0.12	0.45	0.001
MR-Egger	0.38	-0.39	1.15	0.332
(intercept)	0.00	-0.02	0.01	0.649

Table 3.4 Causal effect of birth weight on refractive error. Stage 1 summary statistics were taken from a meta-analysis of UK Biobank and EGG Consortium GWAS for birth weight (n = 188,039). Stage 2 summary statistics were taken from a UK Biobank GWAS for refractive error (n = 95,504). IVW, inverse-variance weighted.

3.4 Discussion

The Mendelian randomisation study undertaken here is the first study, to date, to assess the causality of the relationship between birth weight and refractive error. The estimated effect of birth weight within the normal range of 2.5-4.5 kg on refractive error was +0.28 D per 1 SD increase in birth weight. Observational analysis estimated the effect size of +0.04 D/SD. The lower value of the linear regression estimate in comparison with the MR estimate (Durbin-Wu-Hausman endogeneity p = 0.04) suggests the presence of bias affecting the observational estimate. Typically, the reason for such bias is confounding. As regards the current study, likely confounders of the refractive error - birth weight relationship include the following. First, maternal age and birth weight have been found to be associated in 5 population-based cohorts, with an inverted U-shape relationship between the phenotypes (342). Association of maternal age with more negative refractive error of a child was reported by Lin et al. (343). Therefore, it is possible that the effect of birthweight on refractive error is mediated by maternal age. Second, parity (birth order) is associated both with birth weight (344) and refractive error (345). The latter study reported a higher risk of being myopic in first-born children. The first-born children were reported to have a higher educational exposure and given the established causal relationship between education and myopia it could be a source of a bias in the observational analysis. Third, a higher maternal educational level is associated with a lower risk of preterm birth (346) and low birth weight in new-borns (347). In addition, maternal education was found to be associated with refractive error in children (348). Maternal educational level is likely to be associated with the reproductive age and higher socioeconomic status, confounding the association of birth weight and refractive error. Maternal weight affects fetal growth. An increased number of fetal macrosomia (newborns with birth weight > 4,500 g) have been reported in obese women (349). A strong positive association between the before-pregnancy maternal BMI and child's weight for age (0.07, 95% CI 0.04 to 0.09) and length for weight (0.06, 95% CI 0.04 to 0.08) has been reported in a cohort study conducted in China; the magnitude of the association was consistent at infants' age from 3 to 24 months (350). There is evidence of an association of fetal macrosomia with various adverse health effects for newborns; e.g. diabetes, obesity, cardiovascular diseases (351-354).

Mendelian randomisation analysis used in the current study is a powerful method to infer the causal effect of a risk factor on an outcome (355). Here, large samples

from Consortia GWAS datasets and UK Biobank data were analysed in order estimate the causal effect.

Sensitivity analyses carried out in the current study suggested an absence of directional pleiotropy and yielded estimates consistent in terms of direction and magnitude, supporting the hypothesis of birth weight being a causal risk factor for refractive error development.

Although the results of MR-Egger and weighted median tests indicated no evidence for unbalanced pleiotropy, it is still not possible to exclude the risk that many of the 73 genetic variants used to create the PRS suffer from horizontal pleiotropy, due to the absence of knowledge of how the variants impact birth weight.

One of the limitations of the current study is the fact that it was performed in UK Biobank individuals of European ethnicity. The rates of myopia differ across different ethnicities, which may lead to differing magnitudes of effect of birth weight on refractive error in other ancestry groups. Hence, it would be of interest in the future to conduct an MR study for individuals living in an East or South-East Asian country with a high prevalence of myopia.

Next, the UK Biobank does not fully represent the UK population; however, the results of analysis carried out in this sample have external validity and can be generalised (234).

Birth weight was ascertained by self-report in UK Biobank. This would have reduced the precision of our instrumental variables for birth weight; however, there was evidence that the large sample size of UK Biobank compensated for the inaccuracy inherent in self-reported birth weight, since the PRS was predictive of birth weight in an independent cohort (r^2 = 1.2%, $p < 2.2e-16$).

The PRS for birth weight explained only 1.2% of the variance in birth weight. Yet, the PRS was robustly associated with birth weight (F-statistic = 496.74 for the first stage of the 1-sample MR), which suggested that the PRS was unlikely to give rise to ‘weak instrument’ bias. However, given the sample size of 39,658, this study had limited power to detect a causal effect less than +0.35 D/SD, suggesting the estimated effect of +0.28 D/SD could be imprecise.

This study supported the hypothesis of a causal role of birth weight on refractive error development. A genetic predisposition to a higher birth weight was associated

with a more hyperopic refraction, and vice versa, although the magnitude of the effect was small (the 95% CI was +0.05 to +0.52 D per SD increase in birth weight). The limited sample size for this analysis meant that the causal effect size could not be estimated with a high degree of precision.

Chapter 4. Genome-wide Association Study for Strabismus

4.1 Introduction

Strabismus is a common condition characterized by constant or intermittent abnormal alignment of the eyes that leads to loss of binocular vision (Chapter 1.3). Strabismus is often associated with amblyopia in the deviated eye and can therefore be associated with childhood-onset visual impairment (163).

Various environmental risk factors, including prematurity, maternal smoking, ill-health during pregnancy, hyperopia and anisometropia increase the risk of strabismus (167, 168, 356). Numerous studies have been conducted to understand the genetics of strabismus (186-188). A complex inheritance pattern was described in several studies (191-193). However, the hypothesis that there are rare monogenic forms of strabismus was also supported, e.g. a locus at 7p22.1 was found in linkage analysis of families with multiple members affected by strabismus. In certain rare syndromes, such as Mietens-Weber Syndrome and Lamb-Shaffer Syndrome, strabismus is one of the clinical features, alongside intellectual disability. In the only previous genome-wide association study for non-syndromic strabismus, Shaaban et al. (196) identified a single variant (rs2244352,; OR = 1.33, p = 9.58E-11) that was significantly associated with the condition.

Here, the hypothesis that specific, commonly-occurring polymorphisms subtly increase the risk of non-syndromic strabismus, was tested by carrying out a genome-wide association study (GWAS) for strabismus. *avMSE* (Section 2.1.2) was included as a covariate in the analysis in order to avoid identification of genetic variants associated with hyperopia which could lie on a causal pathway to strabismus: genetic variant → hyperopia → strabismus. The current study had the advantage of a larger sample size than the previously reported GWAS for strabismus by Shaaban et al. (196).

4.2 Methods

4.2.1 Selection of Participants

The GWAS for strabismus and subsequent analyses in the current study were restricted to unrelated UK Biobank participants of White British genetic ancestry defined by Bycroft et al. (245). Those withdrawing consent were excluded from the analyses, resulting in 338,253 participants; among them 79,727 individuals answered the questionnaire item 'Reason for glasses/contact lenses'. A total of 75,911 individuals had valid *avMSE* data. After excluding those with medical records or self-reported eye trauma or ocular surgery (n = 9,217), the final sample comprised of 66,694 UK Biobank individuals (*Figure 4.1*).

4.2.2 Genome-wide Association Study for Strabismus

A GWAS for the 1,345 cases and 65,349 controls was carried out in the discovery sample. A total of 7,469,170 imputed genetic variants in the Haplotype Reference Consortium (HRC) (244) reference panel were tested for association with strabismus using Firth logistic regression (--glm-firth command in PLINK 2.0) (Section 2.4.1). For the analysis, only SNP with MAF ≥ 1%, INFO > 0.8, missing genotype rate < 1.5%, and HWE p value > 1.0E-06 were used. Participants with a missing genotyping rate > 2.5% were excluded from the study. Age, gender, *avMSE*, genotyping array, and the first 10 ancestry principal components were included as covariates in the analysis.

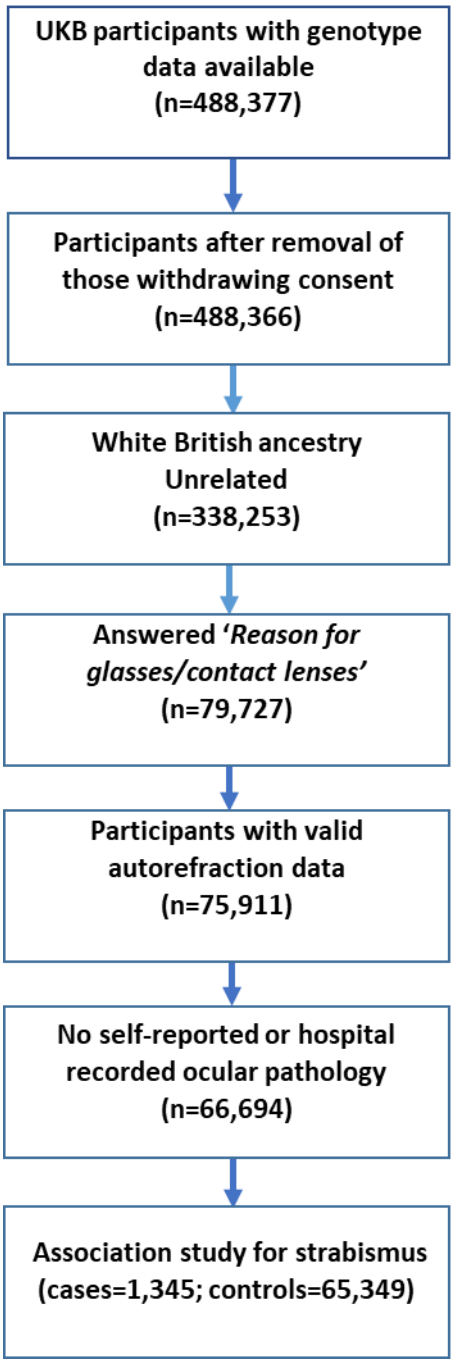

Figure 4.1 Selection of participants for GWAS for strabismus

4.2.3 Post-GWAS Analyses

To search for additional independent signals a conditional analysis (Section 2.4.9) evaluating genetic variants in the region ±250kb from the lead variant was performed using PLINK, followed by visualisation with the *LocusZoom* online tool (309) (http://locuszoom.org/).

Fine-mapping (Section 2.4.9) was performed using FINEMAP v1.3 (319). This analysis included only SNP present on the HRC reference panel within ±250kb of the lead variant or variants in LD (r^2 > 0.1) with the lead variant. A 'causal configuration' with up to 5 SNP being causal in the region was considered. Prior probability of having from 1 to 5 causal genetic variants was defined within the range 0.583 to 0.005. Prior probabilities were calculated following the approach proposed by Benner et al. (319) and were normalized to sum to one.

$$p_k = \Pr(\#\ of\ causal\ SNP = k)\ for\ k = 1, \ldots, K$$

Equation 4.1 Prior probability for the causal configuration of K SNP. Adapted from http://www.christianbenner.com/#input

In order to identify expression quantitative trait loci (eQTL) for genes at the region regulated by the lead variant, the Genotype-Tissue Expression consortium (GTEx) Portal was used (357) and all of the 53 tissue sample analyses were examined (GTEx portal was accessed on 09/18/2018). The *Ensembl* project website was used to assess the CADD scores for the genetic variant strongly associated with strabismus.

4.2.4 Mode of Inheritance and Population Attributable Risk

The GWAS for strabismus results were analysed to identify the most parsimonious genetic model of inheritance. A dominant model was defined as a model with the same risk of strabismus in those carrying one copy of the risk allele as in participants homozygous for the risk allele. The recessive model assumed that the risk of strabismus was higher only in participants carrying two copies of risk allele. For the additive model, the risk of strabismus in risk-allele homozygotes was assumed to be twice that (on a log scale) in heterozygotes. Binary dummy variables for all models (additive, dominant, or recessive) as presented in

Genotype	Dominant	Additive	Recessive	dum_rec	dum_dom	dum_add
AA	0	0	0	0	0	0
AB	1	1	0	0	1	1
BB	1	2	1	1	1	0

Table 4.1.

Genotype	Dominant	Additive	Recessive	dum_rec	dum_dom	dum_add

AA	0	0	0	0	0	0
AB	1	1	0	0	1	1
BB	1	2	1	1	1	0

Table 4.1 Definition of binary dummy variables for dominant, additive, or recessive models. The B allele is the risk allele.

These were then used to form a series of nested models (*Equation 4.2*, *Equation 4.3*). Models were fitted using either standard logistic regression (*glm* function in *R*) or Firth regression (*brglm* function from the *brlrmr R* package). The fit of additive vs. recessive or additive vs. dominant models were compared using a likelihood ratio test (358).

$$Strabismus \sim dum\_rec + cov$$

$$Strabismus \sim dum\_rec + dum\_add + cov$$

Equation 4.2 Nested statistical models to test the mode of inheritance. Recessive vs. additive model, cov = set of covariates used in GWAS for strabismus.

$$Strabismus \sim dum\_dom + cov$$

$$Strabismus \sim dum\_dom + dum\_add + cov$$

Equation 4.3 Nested statistical models to test the mode of inheritance. Dominant vs. additive model, cov = set of covariates used in GWAS for strabismus.

Given the low prevalence of strabismus of 2-4%, the odds ratio (OR) was used as a proxy for the relative risk. Therefore, the population attributable risk (PAR) was estimated using the formula (359): PAR = *p* (OR – 1) / [*p* (OR – 1)] + 1, where *p* is the proportion of controls with the risk genotype. The IMPUTE2 INFO score (Section 2.1.3) for the lead GWAS variants (rs75078292) was 0.999, suggesting that there was no loss in accuracy by using 'hard' genotype calls in the above models rather than dosage accounting for genotype uncertainty.

4.2.5 Replication in an Independent Paediatric Sample

Replication was performed in the ALSPAC paediatric cohort (Section 2.2). At the age of 7 years, children underwent an ophthalmic assessment carried out by

members of the ALSPAC research team (249). A total of 8,237 children had genotype data, and 5,200 of them had ocular phenotype data available. For the replication analyses, the following phenotypes were used: parentally reported history of strabismus (n = 145 cases), manifest strabismus (n = 116 cases), esotropia (n = 143 cases) and exotropia (n = 28 cases). Logistic regression was used to test the association of the genetic variant rs6420484 (available in the ALSPAC genetic data), which is in very high LD ($r^2 = 0.98$) with the lead variant (rs75078292) and predicted to have deleterious effects.

4.3 Results

4.3.1 Validation of Self-reported Strabismus in UK Biobank cohort

As shown in *Figure 4.1*, a total of 66,694 unrelated UK Biobank participants had genetic ancestry principal components that clustered with other White British Europeans, reported their country of birth as England, Wales or Scotland, underwent an ophthalmic assessment, responded to the questionnaire item '*Why were you prescribed glasses/contact lenses?*', and had no history of ocular pathology or ocular surgery. Among these 66,694 participants, 1,345 individuals self-reported strabismus as a reason for wearing spectacles or contact lenses (the answer *'squint or a turn in an eye since childhood').*

Participants with self-reported strabismus had a 11.3-fold greater prevalence of self-reported amblyopia, a 2.5-fold greater prevalence of 1.00 D or more anisometropia, a much more hypermetropic refractive error (median +2.46 vs. +0.21 D), and a much earlier age of starting to wear glasses or contact lenses (median 5 vs. 40 years-old) (*Table 4.2*, *Figure 4.2*). For age and gender, the cases and controls were well-matched (*Table 4.2*), and the difference in Townsend Deprivation Index, a measure of socioeconomic position, in those who did vs. did not report having strabismus was modest (-1.84 vs. -2.12; P=4.40E-03). All of these comparisons between cases and controls supported the validity of the self-reported strabismus phenotype in the majority of participants.

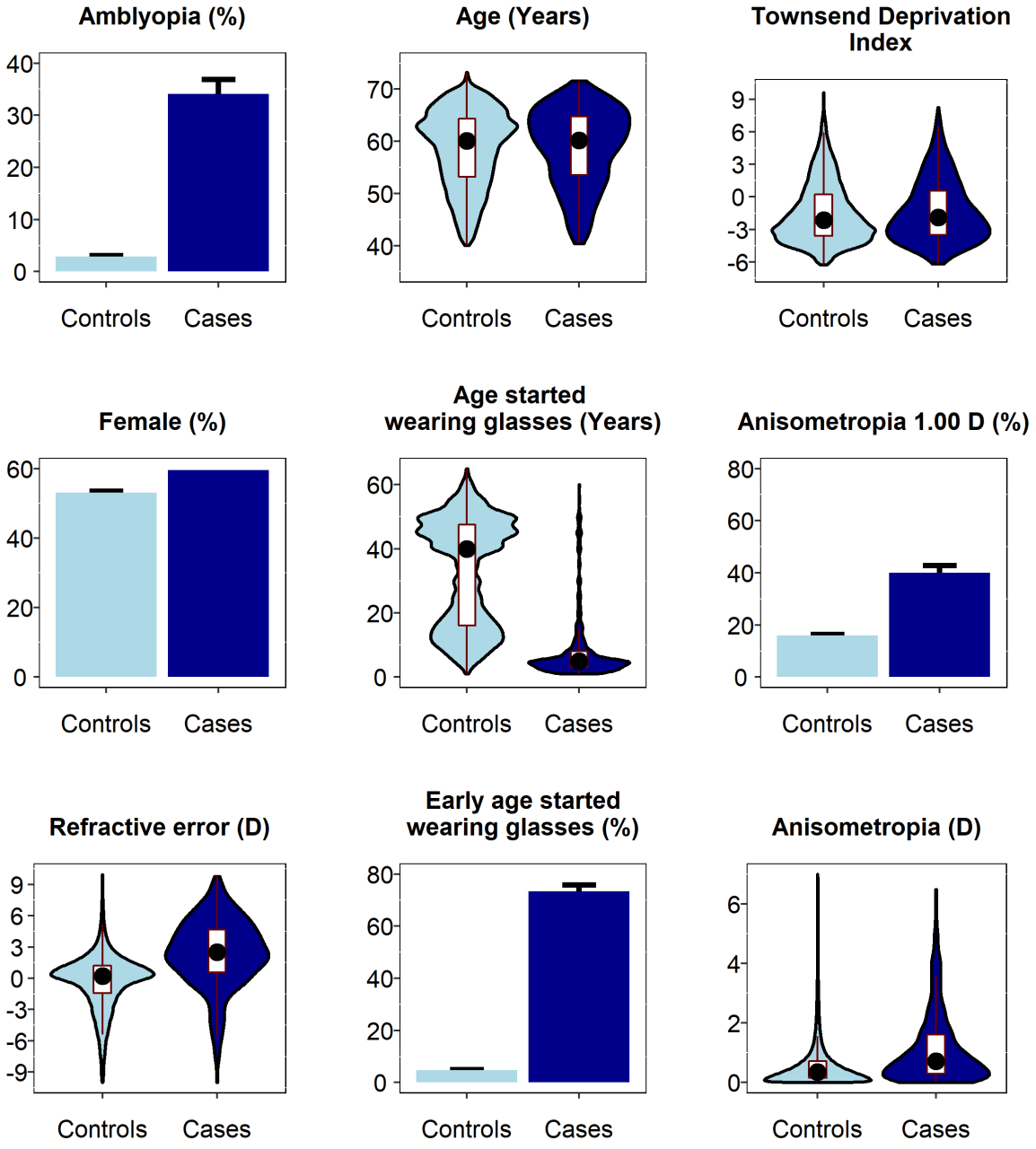

Figure 4.2 Demographic and clinical characteristic of the UK Biobank individuals with self-reported strabismus (n = 1,345) in comparison with the control sample (individuals from the GWAS for strabismus sample without self-reported strabismus; n = 65,349). Early age-of-onset of glasses was defined as ≤7 years-old. Bar chart error bars denote 95% CI. For violin plots, the white rectangle corresponds to the interquartile range and the solid black circle to the median.

Characteristics		**Total (n=66,694)**	**Cases (Self-reported strabismus; n=1,345)**	**Controls (n=65,349)**	***P*-value**
Female	N (%)	35,758 (53.6%)	803 (59.7%)	34,955 (53.5%)	2.20E-01
Self-reported unilateral amblyopia	N (%)	2,465 (3.7%)	471 (35.0%)	1,994 (3.1%)	<1.0E-99
Good VA in both eyes	N (%)	30,098 (45.1%)	266 (19.8%)	29,832 (45.6%)	<1.0E-99
Difference in VA between the two eyes 0.2 logMAR	N (%)	13,135 (19.7%)	624 (46.4%)	12,511 (19.1%)	<1.0E-99
Good VA in one eye and difference > 0.2 log MAR between the two eyes	N (%)	9,123 (13.7%)	454 (33.8%)	8,669 (13.3%)	<1.0E-99
Anisometropia >= 1.0D	N (%)	11,156 (16.7%)	553 (41.1%)	10,603 (16.2%)	<1.0E-99
Anisometropia >= 2.0D	N (%)	3,948 (5.9%)	261 (19.4%)	3,687 (5.6%)	<1.0E-99
Anisometropia (D)	Median (IQR)	0.36 (0.16 to 0.73)	0.75 (0.30 to 1.67)	0.35 (0.15 to 0.72)	8.00E-94
Age (Years)	Median (IQR)	60.17 (53.33 to 64.42)	60.17 (53.67 to 64.83)	60.17 (53.33 to 64.42)	2.30E-01
avMSE (D)	Median (IQR)	0.23 (-1.41 to 1.26)	2.46 (0.52 to 4.65)	0.21 (-1.44 to 1.22)	<1.0E-99
Age started wearing glasses (years)	Median (IQR)	39.00 (15.00 to 47.00)	5.00 (3.00 to 8.00)	40.00 (16.00 to 47.50)	<1.0E-99
Townsend Deprivation Index	Median (IQR)	-2.12 (-3.59 to 0.25)	-1.84 (-3.46 to 0.63)	-2.12 (-3.59 to 0.25)	4.40E-03

Table 4.2 Demographic characteristics of the UK Biobank discovery sample of 66,694 individuals

4.3.2 GWAS for Strabismus results

A GWAS for self-reported strabismus case/control status using 7,469,170 genetic variants identified variants significantly associated (p < 5E-08) with the trait located in a single region on chromosome 17q25.3 (*Figure 4.3*). Of these variants, the most strongly associated SNP was rs75078292 (OR = 1.26, 95% CI 1.16-1.36, p = 2.24E-08). Variants in other regions associated with self-reported strabismus at the suggestive level of genome-wide significance (p < 1E-05) are listed in *Table 4.3*.

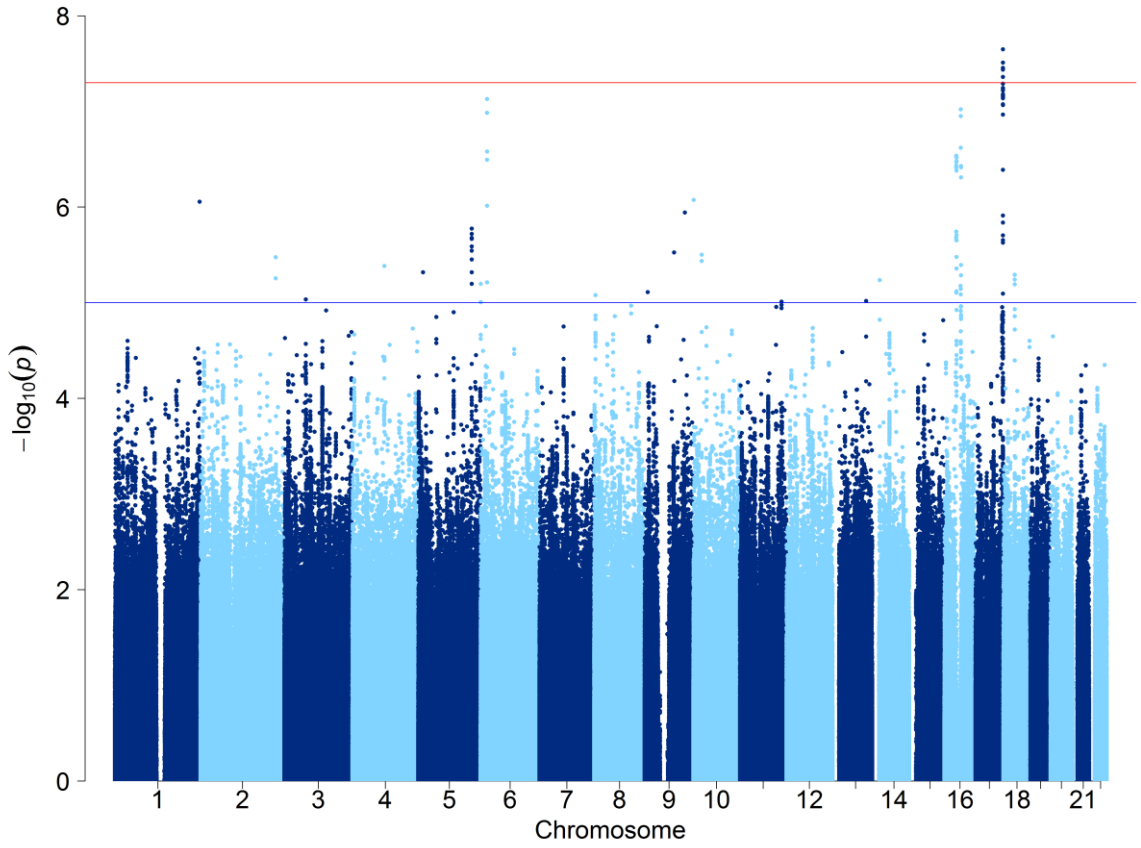

Figure 4.3 Manhattan plot for GWAS for self-reported strabismus in 66,694 UK Biobank participants. The y axis indicates negative log_{10} p values and the x axis indicates genomic position. Each dot represents a single SNP. The red and blue horizontal lines correspond to p values of 5.0E-08 and 1.0E-05, respectively).

Genetic variant	Chromosome	Effect Allele	Reference Allele	MAF	OR	95 %CI	P-value
rs57229473	1	G	T	0.07	1.42	1.24 to 1.63	8.77E-07
rs10932669	2	T	G	0.13	1.29	1.16 to 1.44	3.34E-06
rs75645041	3	G	C	0.15	1.26	1.14 to 1.40	9.19E-06
rs1534560	4	T	C	0.02	1.72	1.37 to 2.17	4.13E-06
rs79576243	5	C	T	0.19	1.26	1.15 to 1.38	1.68E-06
rs116105203	5	A	T	0.01	2.06	1.51 to 2.82	4.81E-06
rs375475939	6	C	T	0.03	1.65	1.38 to 1.98	7.40E-08
rs186974323	6	G	T	0.01	1.94	1.45 to 2.58	6.34E-06
rs141951718	8	A	G	0.01	1.92	1.44 to 2.57	8.31E-06
rs72765677	9	T	C	0.07	0.64	0.64 to 0.77	1.14E-06
rs117517710	9	T	C	0.02	1.69	1.36 to 2.11	2.98E-06
rs143178747	9	A	G	0.02	1.73	1.36 to 2.19	7.75E-06
rs117636134	10	C	G	0.02	1.81	1.43 to 2.30	8.42E-07
rs2274831	10	C	A	0.16	1.27	1.15 to 1.40	3.14E-06
rs58174358	11	T	C	0.04	1.54	1.27 to 1.27	9.76E-06
rs142426391	13	T	C	0.02	1.7	1.34 to 2.14	9.56E-06
rs150685865	14	C	T	0.01	1.82	1.40 to 2.36	5.79E-06
rs116923583	16	C	A	0.03	1.65	1.37 to 1.98	9.51E-08
rs34349606	16	C	T	0.03	1.66	1.37 to 2.02	2.90E-07
rs17744237	16	A	G	0.02	1.7	1.38 to 2.08	3.84E-07
rs34203782	16	A	C	0.03	1.6	1.32 to 1.94	1.96E-06
rs76575122	16	C	A	0.04	1.51	1.26 to 1.81	6.64E-06
rs75078292	17	A	G	0.35	1.25	1.16 to 1.36	2.24E-08
rs117682361	18	T	C	0.02	1.62	1.32 to 2.00	5.10E-06

Table 4.3 Lead variants for regions attaining suggestive genome-wide significance (p < 1.0E-05) in GWAS for self-reported strabismus. Abbreviations: MAF =minor allele frequency; OR = odds ratio; CI = confidence interval.

The genomic inflation factor (λ_{GC}) (Section 2.4.7) was calculated in order to test if there was any systematic bias due to population stratification that could have biased the results of the study. There was no evidence of population stratification (λ_{GC} = 1.004; *Figure 4.4*).

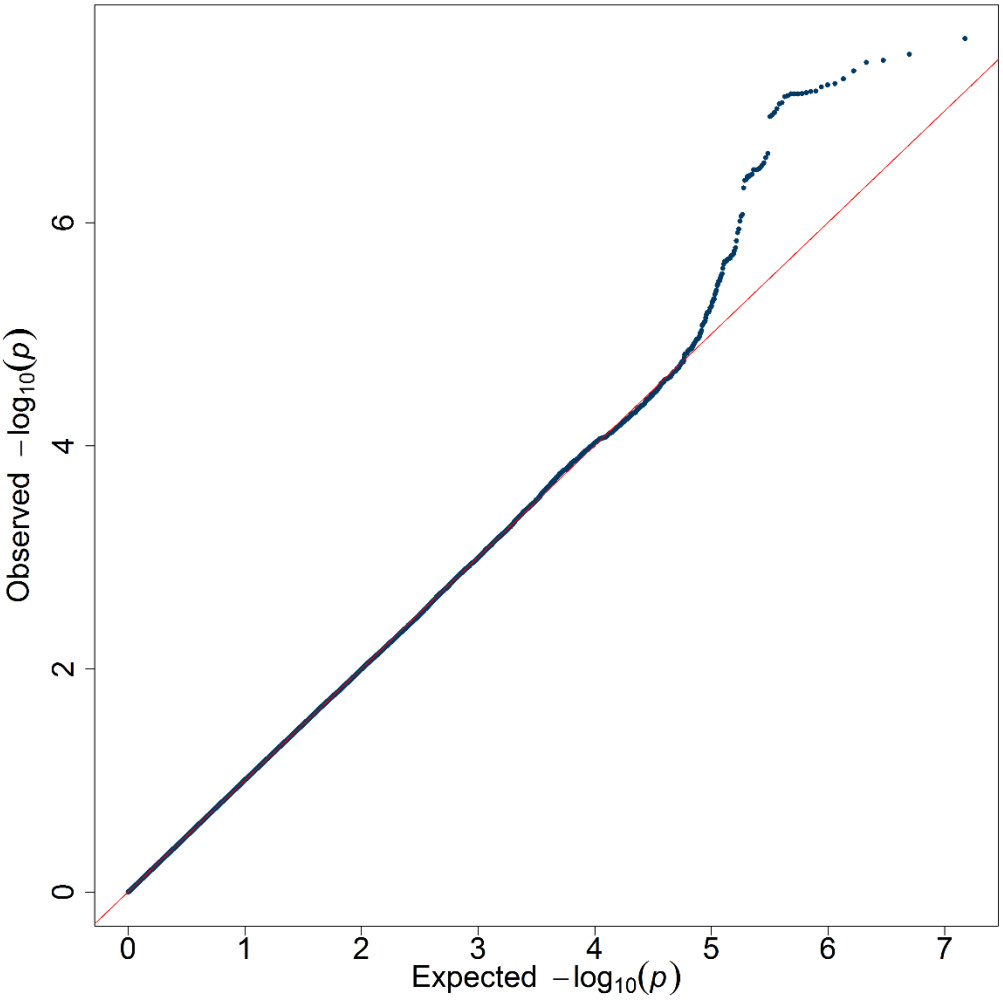

Figure 4.4 Quantile-quantile plot for GWAS for strabismus. The x shows expected negative log_{10} p values and y axis shows observed negative log_{10} p values for association with self-reported strabismus.

Shaaban et.al (2018) reported genetic variant rs2244352 to be associated with non-accommodative esotropia. In the current study, the reported SNP was not significantly associated with self-reported strabismus (OR = 1.01, 95% CI 0.93-1.10, p = 0.83).

4.3.3 Fine-mapping of the 17q25.3 Locus

A conditional analysis was performed and visualised using the *LocusZoom* online tool with a flanking region of 250Kb either side from the lead variant. This region included 23 genes, and was centred on the genes *TSPAN10*, *NPLOC4* and *PDE6G*. This analysis did not identify any secondary signals, suggesting that the association was driven by one causal variant (*Figure 4.5*).

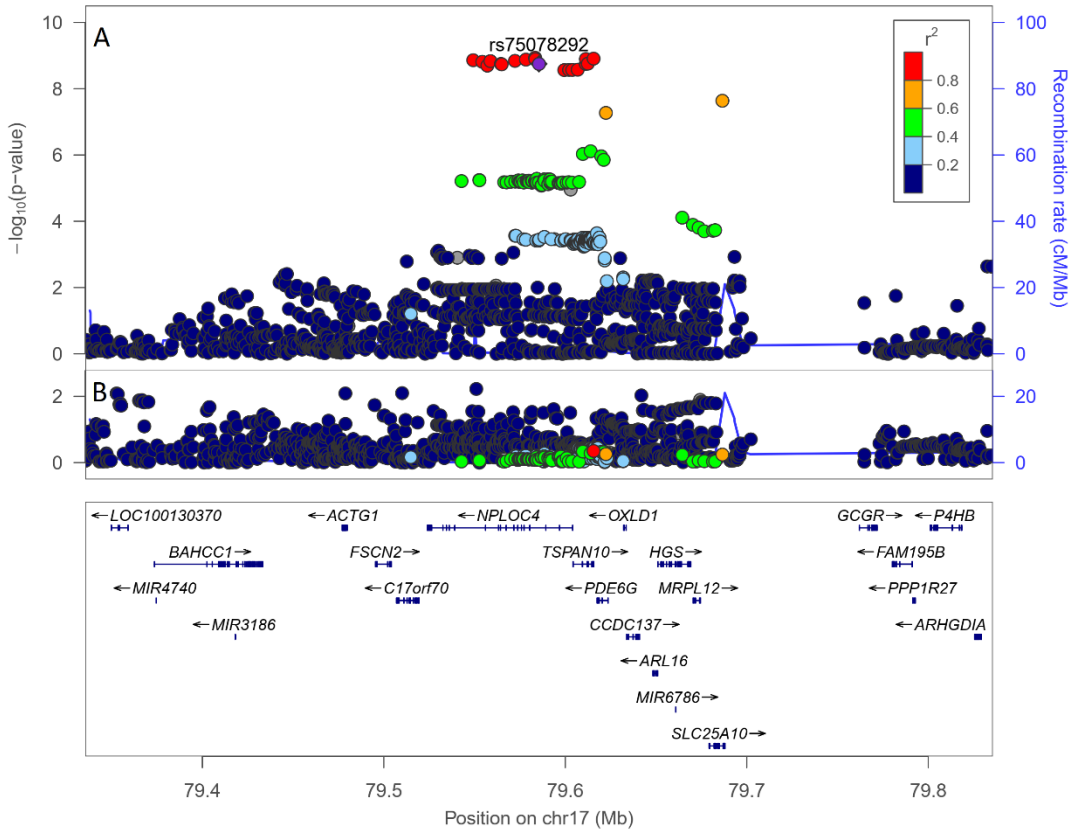

Figure 4.5. Regional association plot for self-reported strabismus in the region ±250kb from the lead variant rs75078292, before (a) and after (b) conditioning on the lead variant.

Further analyses using FINEMAP were not able to identify the precise causal variant(s) in the genomic region, possibly because more than 20 highly-associated variants were in perfect or almost perfect LD. (Notably, the pattern of LD across this region varies markedly across ancestry group, therefore it could be fruitful to fine-map this region in non-European ancestry groups; the MAF of the lead variant, rs75078292, in diverse ancestry groups is shown in *Figure 4.6*).

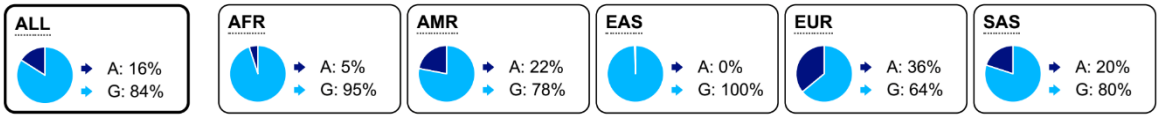

Figure 4.6 Allele frequency spectrum of lead GWAS variant (rs75078292) across ancestry groups. Abbreviations: AFR, Africans; AMR, Native Americans; EAS, East Asians; EUR, Europeans; SAS, South Asians). The results were obtained from Ensembl (320) for participants of the 1000 Genomes project (360).

4.3.4 Functional Evaluation of the 17q25.3 Locus

The lead strabismus variant rs75078202 is in very tight LD ($r^2 = 0.98$) with rs6420484. The minor allele of rs6420484 produces a C177Y substitution (amino acid cysteine is replaced with tyrosine) in *TSPAN10*, the gene encoding the protein tetraspanin-10, which would be expected to have deleterious consequences. Another genetic variant, indel rs397693108, which is also predicted to have deleterious effects, is in almost perfect LD with both the lead variant rs75078292 and rs6420484. The CADD scores (322) were obtained from *Ensembl (320)* to assess the pathogenicity of the variants of interest. CADD scores for highly-associated markers in the region are presented in *Table 4.4*. The rs397693108 indel introduces a 4-bp deletion in the mRNA for 2 isoforms of *TSPAN-10*, which would be predicted to result in a frameshift in the coding region. For a third isoform, the indel is predicted to result in nonsense-mediated decay (NMD) of the mRNA. Given the same MAF of the risk alleles of rs75078292, rs6420484 and rs397693108 and the fact that they are in tight LD, it suggests that they occur on the same haplotype.

Genetic variant	Position	Effect allele	Reference allele	MAF	GWAS OR	GWAS P-value	Scaled CADD score	eQTL effect size	eQTL P-value
rs6420484	79612397	A	G	0.35	1.21	8.24E-10	16.29	0.84	3.30E-21
rs397693108	79614932	TTAAC	T	0.35	1.21	6.90E-10	16.08	0.84	3.30E-21
rs9895741	79603831	A	G	0.35	1.21	1.43E-09	10.84	0.84	3.30E-21
rs9747347	79606820	T	C	0.35	1.21	1.27E-09	7.09	0.84	1.80E-21
rs7405453	79615572	A	G	0.35	1.21	5.02E-10	6.92	0.84	3.30E-21
rs62075722	79611271	A	G	0.35	1.21	8.48E-10	6.29	0.84	3.30E-21
rs67050149	79557043	C	A	0.35	0.83	1.25E-09	5.98	-0.84	3.30E-21
rs11656126	79564542	G	A	0.35	0.83	7.96E-10	5.98	-0.84	3.00E-21
rs8081701	79599441	T	C	0.35	1.21	1.47E-09	2.41	0.86	5.40E-21
rs11650127	79572253	G	A	0.35	0.83	6.00E-10	1.85	-0.84	1.70E-20
rs71373084	79564930	C	G	0.35	0.83	8.03E-10	1.65	-0.84	3.40E-21
rs9905786	79602063	G	T	0.35	1.21	1.44E-09	1.08	0.86	7.20E-22
rs112364254	79578287	G	A	0.35	0.82	4.93E-10	1.04	-0.82	2.70E-20
rs7503894	79583473	T	C	0.35	1.21	5.94E-10	0.64	0.84	3.30E-21
rs12953229	79554271	G	A	0.35	0.83	9.50E-10	0.48	-0.84	3.30E-21
rs75078292	79585492	A	G	0.35	1.21	7.95E-10	0.38	0.84	3.20E-21
rs62075723	79611326	G	A	0.35	1.21	5.91E-10	0.15	0.84	3.30E-21
rs34635363	79549250	G	A	0.35	0.83	9.34E-10	0.13	-0.84	3.30E-21
rs12948708	79558741	G	A	0.35	0.83	9.34E-10	0.02	-0.84	3.30E-21

Table 4.4 CADD scores and eQTL effects for TSPAN10 gene expression in cerebellum for variants surpassing the genome-wide significance threshold and in very high LD ($r^2 > 0.95$) with the lead variant at the locus.

TSPAN10 gene expression was examined in tissue samples analyses by the GTEx consortium (357) in order to identify the further effect of the indel rs397693108. The risk allele of rs397693108 was associated with reduced expression of *TSPAN10* in human cerebellum and cerebellar hemisphere (p = 1.30E-10 and p = 8.10E-12, respectively; **Error! Reference source not found.**A, B). These findings were consistent with degradation of an mRNA isoform via NMD. The risk allele of rs397693108 was found to have *cis*-eQTL effects at other genes. An association with reduced phosphodiesterase 6G (*PDE6G*) gene expression was found in testis (p = 5.82E-06; *Figure 4.7*C), and with the expression of ADP ribosylation factor like GTPase 16 (*ARL16*) mRNA levels in thyroid (p = 2.10E-06; *Figure 4.7*D). In testis, rs397693108 was the 47th-ranked eSNP associated with *PDE6G* expression; however it had an almost 1.6-fold lower effect than the top-ranked variant located 15 kb distant (rs11150804, P=1.80-41). Like *PDE6G*, the *ARL16* gene encoding ADP ribosylation factor like GTPase 16 is located nearby to *NPLOC4* and *TSPAN10* (*Figure 4.5*). In thyroid, rs397693108 was the 232nd ranked eSNP associated with *ARL16* expression, and had a more than 4-fold lower effect than the top-ranked variant situated 34.5 kb away (rs7503637, P=5.50E-96). By contrast, rs397693108 was the 15th highest-ranked *TSPAN10* eQTL in cerebellum and had an effect size equal to the top-ranked variant (rs9905786, P=7.20 10E-122). In cerebellar hemisphere tissue, rs397693108 was the 63rd-ranked *TSPAN10* eQTL. Gene expression in human tissues can be regulated by structural variations (Section 1.1.1), e.g. Chiang et al. (361) found evidence for co-regulation of *TSPAN10* gene expression by structural variants and SNP in 17q25.3 region. Thus, in summary, there was strong evidence that the causal variant underlying the association with strabismus acts as a *cis*-eQTL for *TSPAN10* in neural tissue, and for *PDE6G* and *ARL16* in certain other tissues. Of these 3 eGenes, the evidence suggested *TSPAN10* as the most likely causal eGene.

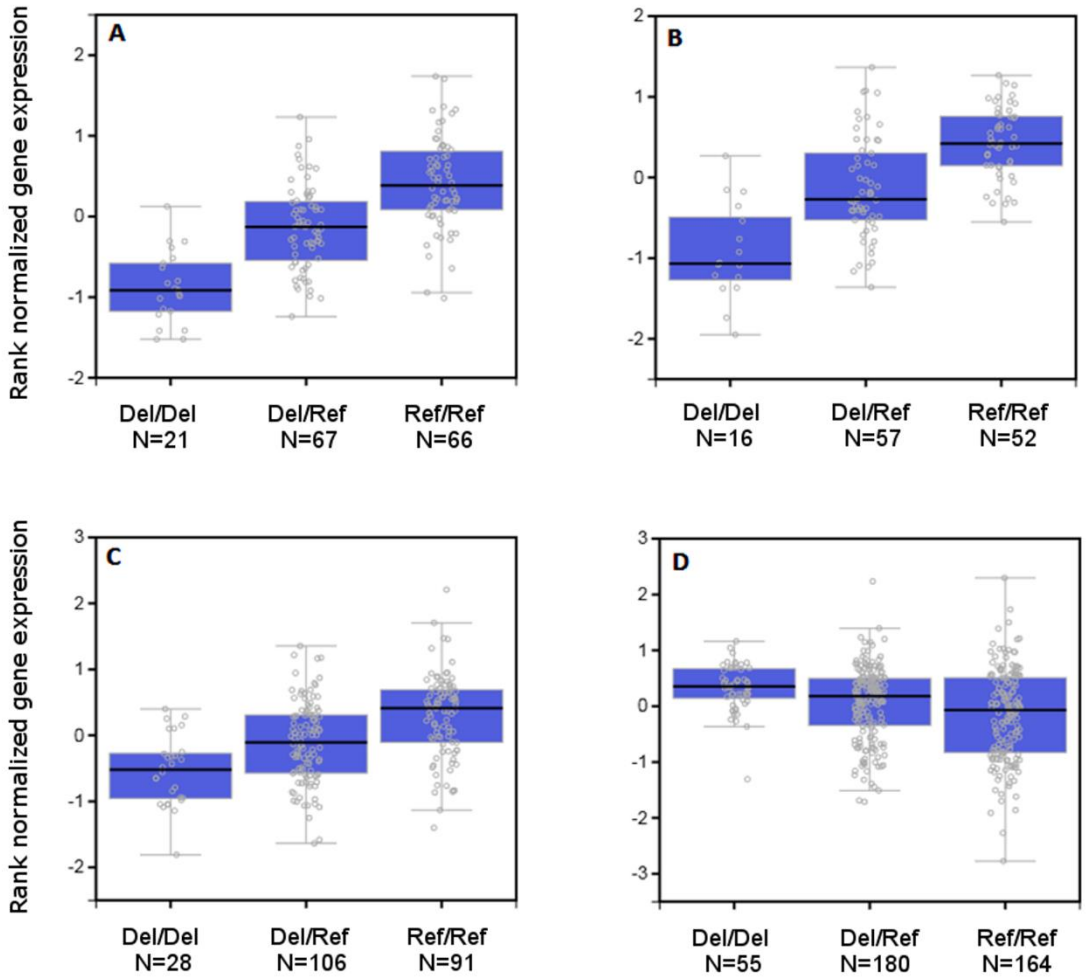

Figure 4.7 Strabismus associated indel rs397693108 is an eQTL. Gene expression levels of tetraspanin-10 mRNA in cerebellum (A), human cerebellar hemisphere (B), and of phosphodiesterase 6G in testis (C), and of ADP ribosylation factor like GTPase 16 in thyroid (D) for individuals with the indicated rs397693108 genotype (Ref allele = TTAAC, Del allele = T). Numbers indicate the number of donors for each genotype category. Boxes depict the first and third quartile; whiskers extend to the furthest data point up to a maximum of 1.5 times the height of the box. Reproduced from Plotnikov et al. (362).

4.3.5 Co-localisation of Association Signals for Strabismus and Refractive Error

Pickrell et al. (128) carried out a GWAS for self-reported myopia in individuals of European-ancestry using the data from the personal genomics company 23andMe. They identified the *NPLOC4-TSPAN10-PDE6G* region as strongly associated with the phenotype. Genetic variant rs9747347 was reported as the SNP with strongest association in the aforementioned region (OR = 1.07, 95% CI 1.06 to 1.09, p = 1.6E-18). rs9747347 is in perfect LD with the lead strabismus variant, rs75078292,

(r^2=1.00 in Europeans). The PhenoScanner (325) web tool was used in order to cross-reference the lead variant associated with strabismus against a wide range of other phenotypes. The rs75078292 variant was found to be associated with the following traits: myopia (p = 1.20E-07) (127); age of onset of myopia (p = 9.80E-07) (127); and age-related macular degeneration (5.04E-11) (363). In addition, a recent GWAS for morphological retinal phenotypes in more than 30,000 UK Biobank individuals reported a variant, rs7503894, in perfect LD with rs75078292 (r^2 = 1.00 in Europeans) that is strongly associated with the retinal nerve fibre layer thickness (p = 2.49E-29) (364). Notably, the lead strabismus risk variant is associated with a *more negative* refractive error and with strabismus, which is counter-intuitive, since strabismus is typically associated with hyperopia (181, 182). The association of rs75078292 with refractive error and myopia were estimated in the sample of 66,694 UK Biobank individuals; the analyses were adjusted for age, gender, genotyping array, and the first 10 ancestry principal components. The lead strabismus variant rs75078292 was found to be associated with myopia (OR = 1.06, p = 5.78E-07) and refractive error (β = -0.09 D, 95% CI -0.12 to -0.06 per copy of the risk allele, p = 2.77E-08). After adjusting for amblyopia or strabismus, the estimates remained consistent in magnitude and direction.

The above results suggested that the risk allele of rs75078292 was independently associated with strabismus and myopia. Since, on average, strabismus is more commonly associated with hyperopia than with myopia, this makes it highly unlikely that the association with strabismus is driven by an association with refractive error, i.e. genetic variant → hyperopia → strabismus. However, in order to further confirm that the association between the lead variant rs75078292 and strabismus was not driven by refractive error, the discovery sample was stratified into myopic, emmetropic and hyperopic groups (*Table 4.5*). In the myopic group, there was no evidence for an association between strabismus and rs75078292 (OR = 1.11, p = 6.19E-01). In hyperopic group, the association was strong (OR = 1.64, p = 8.30E-07). In emmetropes, the association was statistically significant (OR = 1.88, p = 7.24E-03) but weaker than in the hyperopic subsample. The results of the stratified analysis supported the hypothesis that the association between the lead genetic variant and strabismus was not driven by an association with refractive error.

Group	Model	Cases/controls	Additive model			Recessive model		
			OR	95% CI	P-value	OR	95% CI	P-value
Myopic	Baseline	224/23,190	1.13	0.94 to 1.37	1.90E-01	1.21	0.85 to 1.73	2.86E-01
Emmetropic	Baseline	178/22,710	1.23	0.99 to 1.51	5.69E-02	1.79	1.24 to 2.57	1.69E-03
Hyperopic	Baseline	943/19,449	1.30	1.18 to 1.30	4.66E-08	1.71	1.43 to 2.04	3.28E-09
Myopic	Adjusted[a]	224/23,190	1.13	0.94 to 1.37	1.98E-01	1.20	0.84 to 1.73	3.19E-01
Emmetropic	Adjusted[a]	178/22,710	1.19	0.96 to 1.46	1.15E-01	1.60	1.10 to 2.32	1.43E-02
Hyperopic	Adjusted[a]	943/19,449	1.20	1.09 to 1.32	2.80E-04	1.51	1.25 to 1.81	1.14E-05

Table 4.5 Association of lead variant for strabismus at the NPLOC4-TSPAN10-PDE6G locus with self-reported strabismus in groups stratified by refractive status. Participants were grouped into refractive error categories based on their spherical equivalent refractive error in each eye meeting the following criteria: for myopic group: ≤-0.5 D; emmetropic: > -0.5 D and ≤ +1.0 D; hyperopic: > +1.0 D. Participants were excluded when two eyes were not classified into the same group (e.g. one eye myopic and one eye emmetropic). The baseline covariates age, sex, genotyping array, and the first 10 PCs were included in all models.

[a]Adjusted for baseline covariates plus presence/absence of amblyopia, and anisometropia (D)

4.3.6 Mode of Inheritance

The fit of strabismus associations to dominant, recessive or additive models of inheritance was examined in a baseline model and in an adjusted model. In the baseline setting, with age, sex, first 10 ancestry principal components and genotyping array included as covariates, an additive model provided a better fit in a comparison with a dominant model (p = 4.76E-06). However, a recessive model provided a better fit than the additive model (p = 8.10E-05). After including additional covariates (presence of amblyopia, refractive error, astigmatism, or anisometropia) in the model, the pattern remained the same (*Table 4.6*).

In the 14.8% of the risk-allele homozygotes, the estimated risk of strabismus was OR = 1.62 (95% CI 1.41 to 1.86, p = 1.01E-11). Given the OR = 1.62, the estimated population attributable risk (PAR) was approximately 8.4%.

Model	Additive model			Dominant model			Recessive model		
	OR	95% CI	P-value	OR	95% CI	P-value	OR	95% CI	P-value
Baseline	1.24	1.14 to 1.34	8.92E-08	1.16	1.04 to 1.29	1.03E-02	1.62	1.41 to 1.86	1.01E-11
Baseline + Amblyopia binary	1.16	1.08 to 1.26	1.70E-04	1.11	0.99 to 1.24	7.66E-02	1.42	1.23 to 1.64	2.16E-06
Baseline + *avMSE*	1.18	1.09 to 1.28	5.84E-05	1.13	1.00 to 1.27	4.24E-02	1.45	1.25 to 1.68	9.90E-07
Baseline + Anisometropia	1.23	1.13 to 1.33	2.97E-07	1.16	1.03 to 1.29	1.13E-02	1.58	1.37 to 1.82	1.65E-10
Baseline + Astigmatism	1.21	1.12 to 1.31	2.01E-06	1.13	1.01 to 1.26	3.45E-02	1.57	1.36 to 1.81	4.47E-10
Baseline + Myopia binary	1.25	1.16 to 1.35	1.71E-08	1.17	1.05 to 1.31	5.39E-03	1.66	1.44 to 1.90	1.25E-12
Baseline + Anisometropia binary	1.22	1.13 to 1.32	7.28E-07	1.15	1.03 to 1.29	1.38E-02	1.55	1.35 to 1.78	8.50E-10
Baseline + Astigmatism binary	1.22	1.12 to 1.32	1.07E-06	1.14	1.02 to 1.28	2.23E-02	1.57	1.36 to 1.81	3.77E-10

Table 4.6 Comparison of modes of inheritance of rs75078292 for association with self-reported strabismus in UK Biobank participants (n = 66,694).

Covariate definitions. avMSE=mean spherical equivalent refractive error in Dioptres averaged between the two eyes; Anisometropia=anisometropia in Dioptres coded as a continuous variable; Astigmatism=Refractive astigmatism in Dioptres averaged between the two eyes; Myopia binary=binary variable coded as zero unless refractive error averaged between the two eyes <= -0.5 D; Anisometropia status=binary variable coded as zero unless anisometropia >= 1.00 D; Astigmatism binary=binary variable coded as zero unless refractive astigmatism averaged between the two eyes >= 1.00 D.

4.3.7 Replication in the ALSPAC Paediatric Sample

Independent replication was performed in a sample of children aged 7 years from the ALSPAC cohort (n = 5,200), clinically examined by an orthoptist. The results are presented in *Table 4.7*. The lead GWAS variant showed independent replication of association in the clinician-diagnosed sample of children. For manifest strabismus, the OR = 1.44 (95% CI 1.11 to 1.88, p = 0.007) for an additive model and OR = 1.85 (95% CI 1.16 to 2.95, p = 0.009) for a recessive model. The results were consistent with those obtained in the UK Biobank sample in terms of effect size and likely mode of inheritance. After adjusting the model for amblyopia, the results were similar to the baseline model. The association was statistically significant for exotropia ($p < 0.05$ for both additive and recessive models), but not for esotropia. However, one possible reason for such association is that some children may have had secondary exotropia (as a result of previous surgery correction of esotropia). Given the small number of children with exotropia (n = 28), the interpretation of this result is not straightforward.

Phenotypes for replication	Cases/Controls	Mode of inheritance	Baseline model		Adjusted model	
			OR (95% CI)	P-value	OR (95% CI)	P-value
Amblyopia	189/5011	Additive	1.19 (0.97 to 1.48)	1.00E-01	0.97 (0.77 to 1.23)	7.89E-01
		Recessive	1.36 (0.91 to 2.02)	1.35E-01	1.10 (0.69 to 1.74)	6.90E-01
History of strabismus	145/5055	Additive	1.36 (1.07 to 1.73)	1.10E-02	1.34 (1.04 to 1.73)	2.60E-02
		Recessive	1.72 (1.12 to 2.63	1.30E-02	1.58 (0.98 to 2.55)	6.10E-02
Esotropia	143/5057	Additive	1.08 (0.85 to 1.38)	5.42E-01	1.02 (0.78 to 1.32)	9.10E-01
		Recessive	1.20 (0.74 to 1.95)	4.50E-01	1.00 (0.59 to 1.72)	9.89E-01
Manifest strabismus	116/5084	Additive	1.44 (1.11 to 1.88)	7.00E-03	1.43 (1.07 to 1.92)	1.60E-02
		Recessive	1.85 (1.16 to 2.95)	9.00E-03	1.72 (1.00 to 2.95)	5.00E-02
Exotropia	28/5172	Additive	1.76 (1.04 to 2.99)	3.50E-02	1.73 (1.02 to 2.93)	4.00E-02
		Recessive	2.47 (1.05 to 5.83)	3.90E-02	2.37 (1.00 to 5.63)	5.10E-02

Table 4.7 Association of strabismus and amblyopia with lead GWAS variant in ALSPAC replication sample (n = 5,200).

Baseline model was adjusted for gender; Association analyses for strabismus were adjusted for gender and amblyopia; Association analysis for amblyopia was adjusted for gender and strabismus.

4.4 Discussion

This study identified a commonly occurring genetic variant located on chromosome 17q25.3 that was associated with strabismus. The conditional analysis suggested that the association was driven by a single SNP. The presence of approximately 20 SNP in high LD in this region made it difficult to pinpoint the exact causal SNP. Further fine-mapping analyses in non-European populations may be useful. Genetic variants in this locus were found to be associated with myopia previously; however, the results of analyses after adjustment for myopia, refractive error, or anisometropia provided evidence that the association between the lead variant and strabismus was independent of the association with other traits. The influence of genes in *the TSPAN10-NPLOC4-PDE6G* region on the visual system development remains unclear. A previous GWAS for esotropia by Shaaban et al. (196) reported a single genetic variant associated with the disorder. In the current study, the lead SNP (rs2244352) from the GWAS conducted by Shaaban et al. was tested for association with strabismus in UK Biobank sample. However, no association of rs2244352 with self-reported strabismus was found, which suggests that the genetic architecture of non-accommodative esotropia differs from that of self-reported strabismus generally, or that the association reported by Shaaban et al. was a false-positive.

Three criteria need to be met in order for a case-control GWAS to detect genetic variants associated with a trait. First, the trait must have a genetic component. Second, the sample of 'controls' should contain only a small proportion of individuals truly affected by the disorder (misclassified 'cases'). Third, the participants truly affected by the disorder should constitute the bulk of the 'cases' group. Regarding the first criterion, recent progress in identifying mutations causing strabismus (187, 188, 192) supports the hypothesis that susceptibility to strabismus has a genetic component. In the study conducted by the Wellcome Trust Case Controls Consortium (365), the statistical power of case-control GWAS was shown to be minimally reduced if the 'control' group includes less than 5% of misclassified 'cases'. In the current study, controls were selected as those not self-reporting a history of strabismus. Given the prevalence of strabismus in general population of 2-4%, this approach was considered as an effective method with regards to the second criterion. A comprehensive validation of the self-reported strabismus phenotype was performed to meet the third criterion. Thus, the level of comorbidities (i.e. self-reported amblyopia, anisometropia and asymmetric visual acuity) was compared in participants with and without self-reported strabismus (*Table 4.2*; *Figure 4.2*). The association with strabismus was replicated in an independent cohort of children with clinician-diagnosed strabismus (OR = 1.85, 95% CI 1.16 to 2.95).

In the current study the number of controls was more than 48-fold greater than the number of cases. This unbalanced case-control ratio is a feature of the association studies for disorders with low prevalence when the total sample size is large (366). The use of standard logistic regression in that case can lead to the bias in the estimation of the trait-variant association. Thus, to avoid this source of bias, Firth regression was used in the discovery GWAS in 66,694 UK Biobank individuals. The discovery association analysis was restricted to unrelated individuals to control for relatedness, one of the main confounding factors in GWAS analyses (366, 367).

The findings that the association of the lead GWAS variants with strabismus, amblyopia and myopia were mutually independent implies a complex relationship between the causal variant(s) at the locus and visual development. The suggestive evidence that the association with strabismus was restricted to non-myopic individuals may reflect a different mechanism of strabismus development in myopic vs. non-myopic individuals, perhaps related to the higher prevalence of strabismus in hypermetropes and anisometropes compared to myopes. To further dissect the inter-relationships between the locus and these ocular traits will require the collection and genotyping of large cohorts of individuals with clinician-diagnosed strabismus.

In summary, this genome-wide association study in a large sample of UK Biobank participants identified a single locus associated with self-reported strabismus. The association was replicated in an independent sample (OR = 1.85, p = 0.009). Among the associated variants, 2 are predicted to have deleterious effects: rs6420484, which substitutes tyrosine for a conserved cysteine (C177Y) in the *TSPAN10* gene, and 4-bp deletion variant, rs397693108, predicted to cause a frameshift in *TSPAN10*.

Therefore, this study suggests gentic variants in *TSPAN10* confer susceptibility to strabismus.

Chapter 5. The Effect of Education on Refractive Error: Quasi-experimental evidence from the 1972 Reform of the Raising of the School Age (ROSLA).

5.1 Introduction

A plethora of observational (cross-sectional and longitudinal) studies has been performed in order to identify and to estimate the association between education and refractive error (particularly myopia). Despite the large number of such studies, the evidence of education being a causal risk factor for refractive error development has only recently been examined in two MR studies (45, 46).

The 1944 Education Act (Butler Act) in England and Wales and the Education (Scotland) Act 1946 legislated for free universal schooling in the United Kingdom, as well as the raising of the minimum school-leaving age first from 14 to 15 years-old in April 1947 and then from 15 to 16 years-old in September 1972 (following the recommendations of the Robbins Report) (368). Children born in September 1957 were the first to be affected by the raising of school-leaving age (ROSLA) in 1972; those who would have left school aged 15 were required to remain at school for up to one additional academic year (369). This led to the sharp increase in the educational attainment in the aforementioned school children. The fact that the educational reform was implemented nation-wide at the same date has made it possible to use the regression discontinuity (RD) design (Section 2.4.5) to assess the effect of schooling on various health outcomes, including refractive error. In that scenario, September the 1st 1972 is considered as a cut-off point dividing the study cohort into two subgroups: those affected by the reform (born after September 1957 and therefore 15 years of age or older in September 1972), and those unaffected (born before September 1957).

The availability of educational attainment and refractive error data in UK Biobank; and the large number of UK Biobank individuals affected by the ROSLA 1972 reform has made it possible to estimate the *causal* effect of education on myopia using regression discontinuity analysis in the current study.

5.2 Methods

5.2.1 Analysis samples

Analyses in the current study were restricted to UK Biobank participants of White British genetic ancestry (245), who reported England or Wales as their country of birth and who had data available for *EduYears* (Section 2.1.1) and autorefraction-measured refractive error (*avMSE;* Section 2.1.2). Those participants who reported having a history of eye surgery

(specifically, cataract, laser refractive surgery or corneal graft surgery) were excluded from the analyses. Only participants born between September 1945 and September 1968 (*Educational Year*; Section 2.1.1) were selected for inclusion in the final sample (n = 62,812 individuals). Instrumental variable analysis was performed in a sub-sample of participants with *Educational Year* between 1956-1958 (n = 4,761; '*IV sample*'). Regression discontinuity analysis was carried out in a sub-sample (n = 21,548; '*RD sample*') selected using the optimal bandwidth selection procedure described by Calonico, Cattaneo and Titiunik (370). Participants selection is presented in *Figure 5.1*.

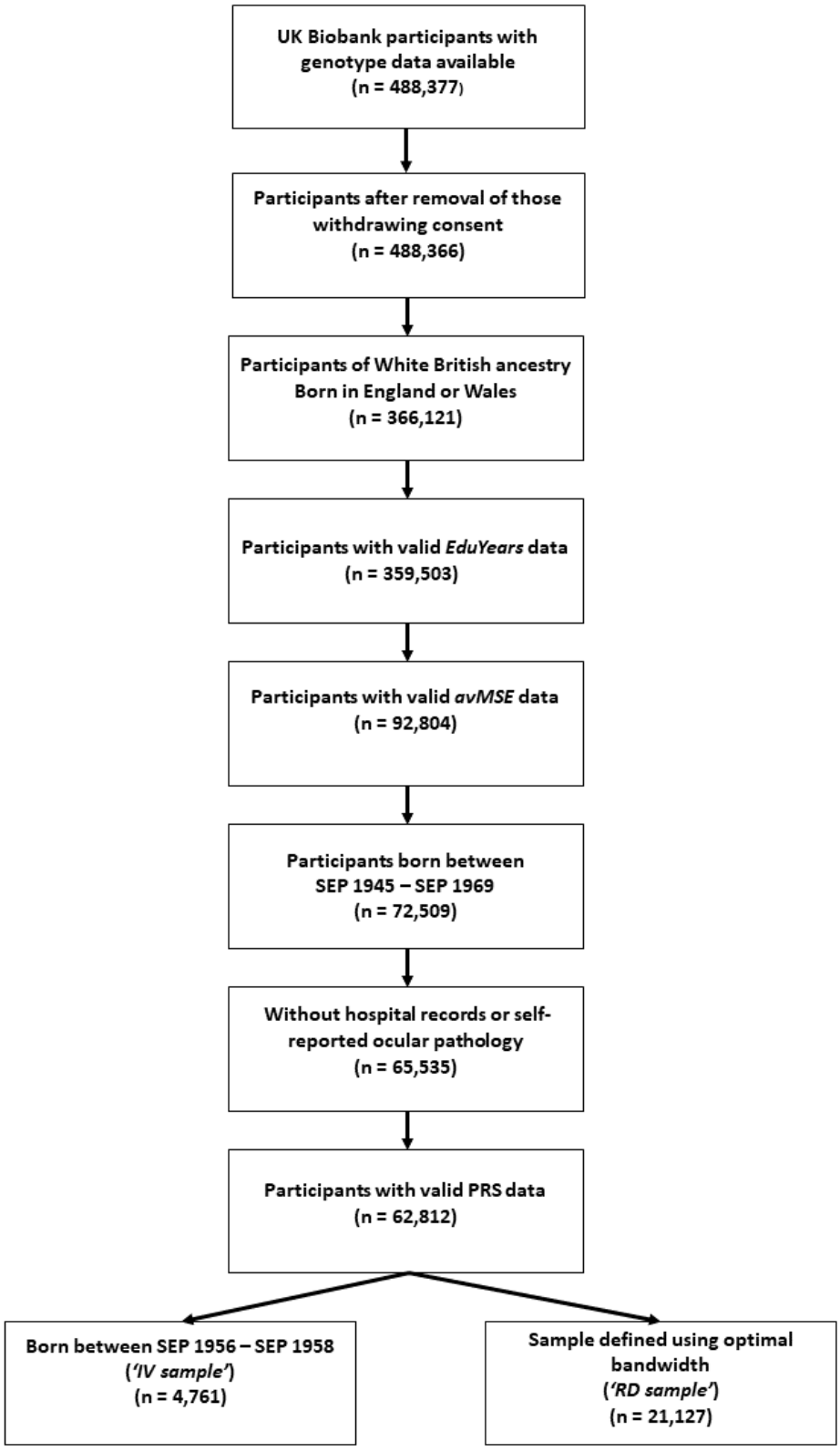

Figure 5.1 Selection of UK Biobank participants for the RD and IV samples

5.2.2 Forcing Variable and Bin Size Selection

The discontinuity in educational attainment induced by ROSLA 1972 was determined by the participant's birth date, hence an additional ('forcing variable') term was included in the RD analysis corresponding to the birth cohort of each participant. The birth date of participants (in months) is typically used as a forcing variable in regression discontinuity design studies investigating the effect of ROSLA on health or social-economic outcomes (371-373). Therefore, the forcing variable '*run1*' was created and coded as the number of months before or after the cut-off date (September 1957) that the participant was born; negative values were used for those born before the cut-off date and positive values otherwise. In this scenario, the bin size was 1 month; however, the optimal bin size can also be selected using either informal or formal procedures (Section 2.4.5). For the formal approach, the use of an F-test has been recommended (374), selecting the widest bin size such that a one-step wider bin size would provide a worse fit to the data (i.e. F-test, $p < 0.05$). In order to perform this test, K dichotomous variables for each bin were created (e.g. for bin size 12 months, there were 24 dummy indicators corresponding to 12 bins before and 12 bins after the cut-off point). Next, each bin was divided into two smaller bins and 2K indicators were created; refractive error was regressed on the new set of indicators. An F-statistic was calculated to compare the fit of the two regression equations, and then continued with further division of bin size.

5.2.3 Polygenic Risk Scores for Refractive Error and Educational Attainment

Polygenic risk scores (PRS; Section 2.4.8) were created for educational attainment and for refractive error. The PRS for educational attainment was constructed using summary statistics for a published educational attainment GWAS meta-analysis (approximate n = 324,162; Section 2.3.3) (254). UK Biobank participants were excluded from this educational attainment GWAS dataset, which guarded against 'overfitting' that may have occurred if the same participants were used to determine the weights of the PRS regression coefficients and to explore the effects of the PRS. The PRS for refractive error was constructed using summary statistics from a GWAS for Age of Onset of Spectacle Wear (AOSW)-inferred spherical equivalent refractive error (n = 287,448 participants of UK Biobank) (317). Note that there was no overlap between the AOSW-inferred refractive error GWAS sample and either the *IV sample* or the *RD sample.* For variants in LD, PRS weights were adjusted using LDpred v1.0.6. In total, 1,175,465 HapMap3 (375) SNP were used for constructing the PRS. PRS were standardized to have a mean of zero and variance of one. Next, the PRS for each

trait was converted from a continuous standardised variable to a binary variable. For the educational attainment binary PRS, the score was assigned a value of 1 if the standardised PRS score was more than 0; and 0 otherwise. For the binary refractive error PRS, the score was assigned a value of 1 if the standardised PRS was *less* than zero (meaning that participants with higher genetic risk of developing a negative refractive error, i.e. myopia, had a value of 1); and 0 otherwise.

5.2.4 Effects of ROSLA 1972 on Educational Attainment

Two binary variables were created in order to assess the effect of the ROSLA 1972 on educational attainment and refractive error. The first variable, '*leave16*' was coded '1' if the individual reported completing full time education at age 16 years or more; and '0' otherwise. The second variable, '*ROSLA*' was assigned a value of '1' if the individual was born during or after September 1957; and '0' otherwise. Logistic regression was used to investigate the association between *EduYears* and *ROSLA.* The '*glm.cluster*' command from the *miceadds* R package was used, and standard errors were clustered by month and year of birth. In addition, the proportion of UK Biobank individuals who completed full time education at age 15 years was plotted against the *Educational Year* to visually assess the effect of the reform. These analyses used the full sample of 62,812 UK Biobank participants born 12 years before and after September 1957.

5.2.5 Effect of the Reform on Refractive Error

The binary variables *leave16* and *ROSLA* were used as predictors in linear regression models to investigate the observational ordinary least squares (OLS) association of the reform with refractive error. Standard errors were clustered by month and year of birth. These OLS estimates could potentially be used to gauge the causal effect of the reform on refractive error, however such an approach is not generally recommended since OLS estimates from observational studies typically suffer from bias due to confounding (271, 376). Therefore, as an alternative to OLS regression, an IV analysis (Section 2.4.2) was performed. If the so-called 'instrumental variable assumptions' hold (Section 2.4.2), an IV analysis will estimate the causal effect free from bias from confounders. In the current study, a 2SLS IV analysis (Section2.4.2) was carried out, using the '*ivreg*' command from the *AER* R package. In the first stage (Equation 5.1), variable *leave16* was regressed on the instrument *ROSLA*. In the second stage (Equation 5.2) refractive error was regressed on the

model-estimated values from stage one. The analysis was performed with adjustment for age, gender, first 5 PC and the binary PRS for educational attainment.

$$leave16 = \beta_0 + \beta_1 ROSLA + \beta_2 Cov_1 + \ldots + \beta_{i+1} Cov_i + \omega$$

Equation 5.1 First stage regression model. Cov = set of covariates; ω = error term

$$RE = \mu_0 + \mu_1 \widehat{leave16} + \mu_2 Cov_1 + \ldots + \mu_{i+1} Cov_i + \zeta$$

Equation 5.2 Second stage regression model Cov = set of covariates; ζ = error term; RE = refractive error

The analyses were carried out in the *IV Sample*, which comprised of 4,761 UK Biobank individuals born in the interval from 12 months before and 12 months after the cut-off date. Selection of individuals from the last cohort before the reform and the first cohort after the reform (one full school year either side from the cut-off) was performed in order to minimize the secular variation either side of the cut-off date.

After the implementation of ROSLA, some children still left school at age 15 or earlier, i.e. not all school pupils were affected by the reform. Specifically, the probability of leaving school at or above age 16 years was approximately 80% after the implementation of ROSLA 1972; i.e. approximately 20% of individuals left full-time education at an earlier age in the year after ROSLA 1972 was introduced. Accordingly, a 'fuzzy' regression discontinuity design (Section 2.4.5) was used in the current study. The Local Average Treatment Effect (LATE; Section 2.4.5) was used to estimate the effect of an additional year of schooling introduced by the reform on refractive error in the sub-population of participants who left full-time education at age 16 or more. Specifically, the LATE was estimated with refractive error as the dependent variable, the *leave16* variable as the independent variable, and *ROSLA* as the instrument. In this scenario, the LATE can be estimated as (301, 302):

$$LATE = \frac{\big(E(RE|ROSLA = 1) - E(RE|ROSLA = 0)\big)}{(E(leave16 = 1 \,\big|\, ROSLA = 1) - (leave16 = 1|ROSLA = 0))}$$

Equation 5.3 RE denotes the refractive error; ROSLA and leave16 denote binary indicators of threshold attainment and intervention.

The optimal bandwidth (Section 2.4.5) was chosen using the method of Calonico, Cattaneo and Titiunik (370) with the '*rdbwselect*' command in the '*rdrobust*' R package. Following the suggestions of Imbens and Kalyanaraman (304), a triangular kernel was used for weighting the observations. The model was adjusted for gender, month of birth, first 5 ancestry principal components and for the education PRS binary variable.

5.2.6 Inverse Probability Weighting

In the UK Biobank sample (n = 62,812) only 17.5% of participants reported completing full-time education at age 15 or less compared with 33% of the Health Survey for England and the General Household Survey (reported in Clark and Royer (371)). Non-random sampling can lead to biased regression estimates (377); therefore it is suggested to use an inverse probability weighting for correcting for endogenous sampling (378). In the current study, the weighting factor (33/17.5) was used in main analyses for individuals with reported age left school of 15 years or less.

5.2.7 Heterogenous Effects of the Reform on Refractive Error

In order to assess if genetic predisposition to refractive error modified the causal effect of ROSLA 1972 on refractive error, the *IV Sample* (n = 4,761) was stratified into 2 groups: those with either a relatively high (n = 2,360) or low genetic predisposition to myopia (n = 2,401), using the binary PRS for refractive error for assignment into groups. Then, IV analyses were carried out in both sub-samples.

To test if the effects of the ROSLA 1972 reform on educational attainment were heterogeneous across the education distribution, the total sample of 62,812 UK Biobank participants was stratified by highest level of educational qualification based on the response for the questionnaire item (Section 2.1.1). Linear regression analysis was performed in each stratum (Equation 5.4)

$$EduYears = \nu_0 + \nu_1 run + \nu_2 ROSLA + \nu_3 run{:}\, ROSLA + \chi$$

Equation 5.4 Estimating the effect of the reform within each educational stratum. run = running variable; run:ROSLA = an interaction term; χ = error term

5.2.8 Sensitivity Analyses

According to the null hypothesis, there should be no shift in refractive error before vs. after cut-off points that are selected randomly rather than the cut-off point corresponding to an educational reform. To test for the absence of an effect at points where no effect was expected (e.g. other than the true cut-off point) (300), the association with refractive error was investigated at 'false' cut-off points. Dates of 2 years before and 2 years after the true cut-off date were selected for such testing. For each of these 2 points, the binary variable for the 'artificial reform' was created and used in the analyses.

To test the no-manipulation assumption of the forcing variable (the density of the forcing variable is continuous at the cut-off point), a McCrary test was used (379).

As further sensitivity analyses, the regression discontinuity analyses were performed with bin sizes of 1, 2, 3, 6 and, also of 12 months at different bandwidths between 1 and 12 years to test the robustness of the results to differences in bin width and bandwidth. Note that for the bin sizes of 6 and 12 months the estimated effect cannot be calculated with bandwidth of 1 and 2 years.

5.3 Results

5.3.1 Polygenic Risk Scores for Refractive Error and for Educational Attainment

The binary PRS for refractive error explained an additional 6.0% of the variance in refractive error ($p<2.2E-16$) compared to a baseline model that included the predictor variables age, sex, binary PRS for education, and the first 5 genetic PC. The binary PRS for education explained an additional 4.2% ($p<2.2E-16$) of the variance in *EduYears* compared to a baseline model that included the predictor variables age, sex, binary PRS for refractive error, and the first 5 genetic PC.

5.3.2 Effect of the ROSLA 1972 Reform on Educational Attainment

During the 12 years before the ROSLA 1972 reform there was a gradual decline in the proportion of UK Biobank participants from England and Wales who reported completing their education at age 15 or younger. However, the reform coincided with a much sharper decrease in the proportion of participants who left school at age of 15 or earlier (Figure 5.2).

The proportion of UK Biobank participants reporting an age completing full time education of 15 years or less was reduced by 22.1% (95% CI 21.6 to 22.7%) after the introduction of ROSLA 1972, from approximately 25% to 2.9%. This is consistent with the reduction of 26.1% observed by Clark and Royer (371) in the Health Survey for England and the General Household Survey. In UK Biobank participants born 12 month before or after the cut-off point, the estimated effect was of a similar magnitude, namely a 12.8% (95% CI 11.2 to 14.4%) reduction in the proportion of those leaving school at age 15 or less (from 15.2% before the cut-off to 2.4% after).

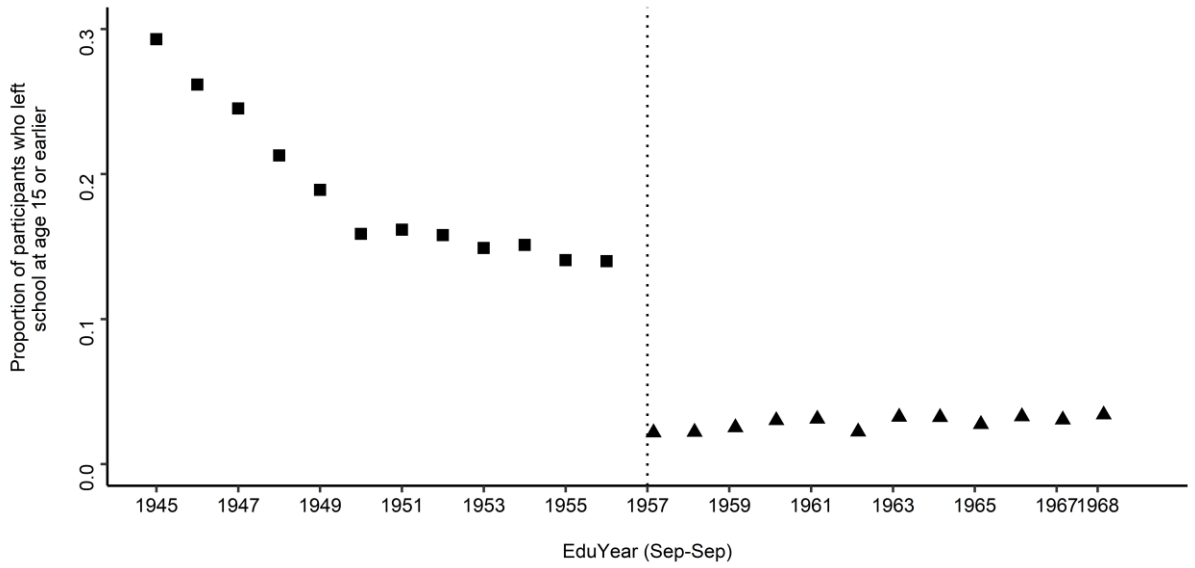

Figure 5.2 Proportion of UK Biobank participants who left school at age 15 years or earlier. The vertical dotted line represents the cut-off point; Each data point represents the proportion of participants who left school before the age of 16 per educational year (running from September to September).

5.3.3 Bin Size Selection

Both informal and formal approaches were used to select the optimal bin size. For the informal approach, plots of various bin sizes were visually compared (*Figure 5.3-Figure 5.5*).

Figure 5.3-5.5 show the graphs for the mean average spherical equivalent Three sets of graphs with bandwidths varying from 4 to 12 years are reported using bin sizes of 1, 2 ,3, 6 and 12 months. All graphs provide us with the information about a discontinuity at the cut-off

points; however, the ones with the bin size of 1 and 2 months contain an excessive number of data points (288 and 144, respectively) for visual evaluation.

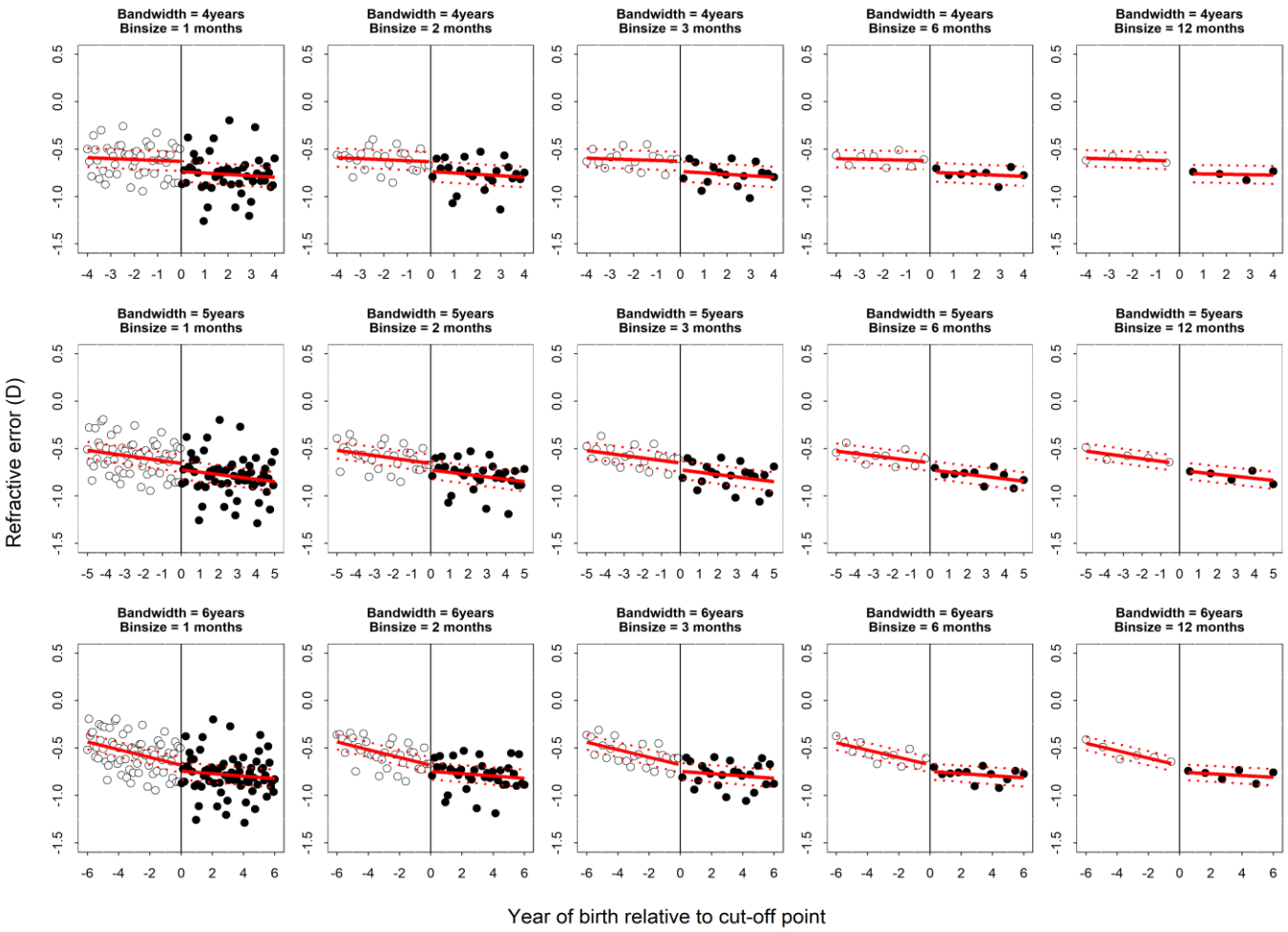

Figure 5.3 Refractive error of UK Biobank participants (n = 62,812) in relation to year of birth. Each data point represents the mean refractive error. Data are plotted across a grid of bin sizes and bandwidths (4-6 years).

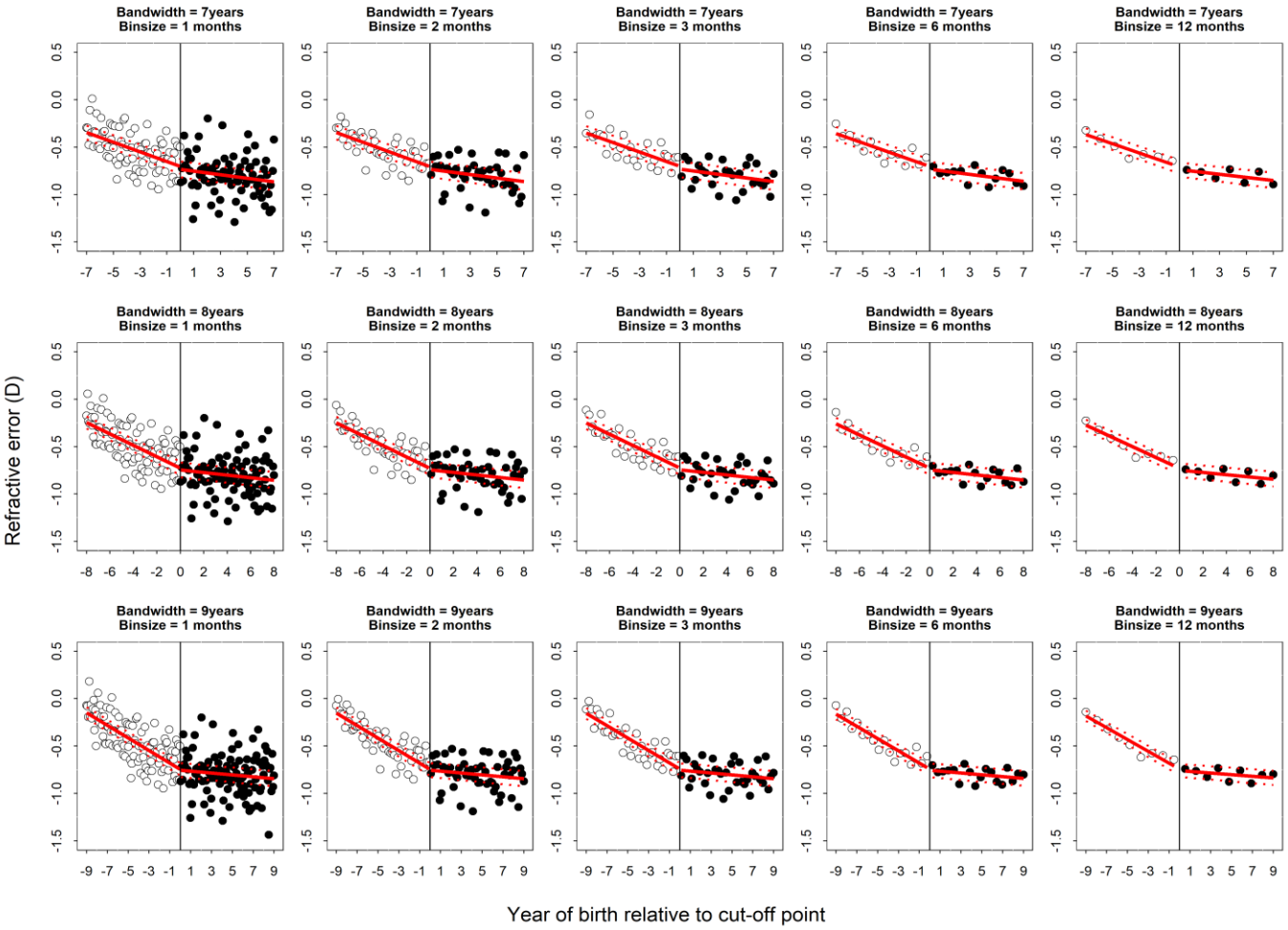

Figure 5.4 Refractive error of UK Biobank participants (n = 62,812) in relation to year of birth. Each data point represents the mean refractive error. Data are plotted across a grid of bin sizes and bandwidths (7-9 years)

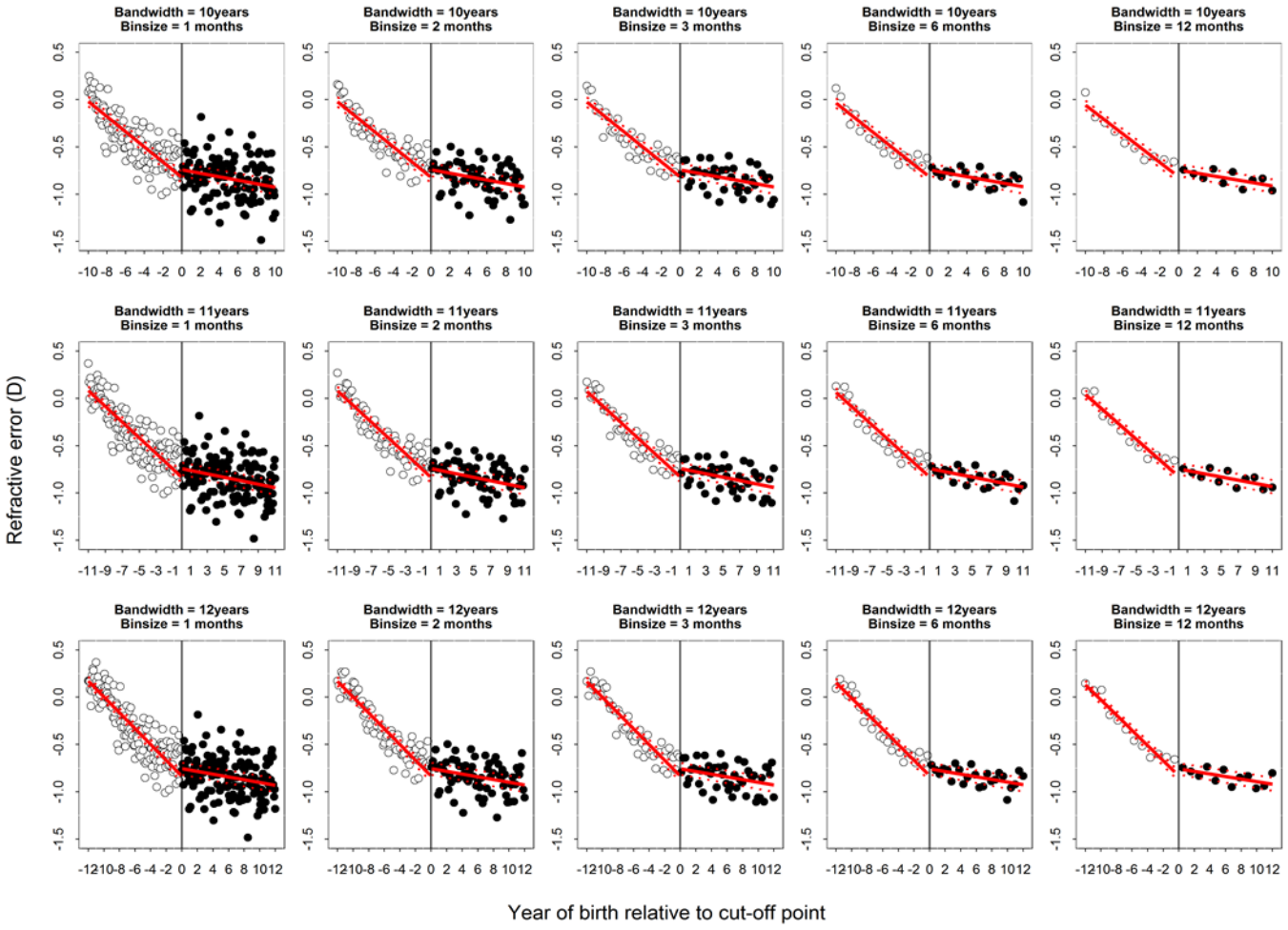

Figure 5.5 Refractive error of UK Biobank participants (n = 62,812) in relation to year of birth. Each data point represents the mean refractive error. Data are plotted across a grid of bin sizes and bandwidths (10-12 years).

For the formal method, an ANOVA was performed to select the optimal bin size (see Methods). Comparing bin sizes of 1, 2, 3, 6 or 12 months, the F-statistics and the corresponding p-value for the ANOVA suggested that any of the bin sizes could be used in the analyses. However, using the 12-month bin would have the advantage that month-of-birth effects acting between August and September would not contribute to any discontinuity at the cut-off point. Therefore, the main analyses in the current study were performed using a bin size of 1 month and with bin sizes of 2, 3, 6 or 12 months as sensitivity analyses to test the results for robustness.

Bin size	F-statistics	P-value
12 months	1.20	0.23
6 months	1.34	0.06
3 months	0.85	0.76
2 months	0.94	0.70
1 month	Reference	Reference

Table 5.1 ANOVA results for selecting the optimal bin size.

5.3.4 Effect of the ROSLA 1972 Reform on Refractive Error

An OLS linear regression analysis was performed in the *IV sample* of UK Biobank participants born one year before to one year after the reform (n= 4,761) in order to test the null hypothesis that remaining at school after age 16 had no effect on refractive error. The null hypothesis was rejected: the estimated effect was a -0.75 D (95% CI -0.96 to -0.54 D, p = 3.18E-12) difference in refractive error associated with completing 16 or more vs. 15 or fewer years of education.

Likewise, the effect on refractive error associated with the ROSLA 1972 reform obtained by an OLS linear regression in the above sample of 4,761 participants was -0.18 D (95% CI -0.32 to -0.04 D, p= 0.01).

These two results are consistent with the theory that additional years spent in education increase the risk of myopia, i.e. education is associated with a more negative refractive error. Importantly, however, conventional OLS analyses such as those above are susceptible to bias from confounding factors such as socioeconomic position and level of parental education (271, 376). Therefore, in order to provide a more accurate estimate of the causal effect of the ROSLA reform on refractive error a 2SLS instrumental variable analysis was performed, using the ROSLA reform as the instrumental variable. Applying the IV approach in the same sample of UK Biobank participants born one year before to one year after the reform (n = 4,761) the causal effect of educational attainment on refractive error was estimated to be -0.91 D (95% CI -1.68 to -0.14 D, p = 0.02). The F-statistic from the first-stage regression was 245.4, confirming that the instrumental variable was not a 'weak instrument'(380). A regression discontinuity analysis was also carried out. For the RD analysis, an optimal bandwidth of 54.8 months (for bin size of 1 month) was selected, which resulted in a sample size of n = 21,548. The estimated causal

effect using RD was -0.77 D (95% CI: -1.53 to -0.02, p = 0.04). The 2SLS and the regression discontinuity estimated showed no significant difference between them (independent samples *t*-test p = 0.80).

In order to test the robustness of the results to the choice of bandwidth, the RD analysis was repeated at different bandwidths (for a bin width of 1 month). The causal effect estimates were consistent in terms of direction and magnitude (*Table 5.2* , *Figure 5.6*).

Bin size	Bandwidth (years)	Observations	Estimate (D)	95% CI	P value
1 month	1	4370	-1.17	-2.43 to 0.09	0.07
1 month	2	9129	-0.60	-1.62 to 0.41	0.24
1 month	3	13805	-0.72	-1.63 to 0.18	0.12
1 month	4	18642	-0.79	-1.60 to 0.01	0.05
1 month	5	23530	-0.76	-1.48 to -0.03	0.04
1 month	6	28359	-0.76	-1.43 to -0.09	0.03
1 month	7	33271	-0.71	-1.33 to -0.08	0.03
1 month	8	38542	-0.66	-1.25 to -0.07	0.03
1 month	9	44330	-0.63	-1.19 to -0.06	0.03
1 month	10	50517	-0.56	-1.11 to -0.00	0.05
1 month	11	57176	-0.45	-1.00 to 0.09	0.10
1 month	12	62468	-0.39	-0.94 to 0.16	0.16

Table 5.2 Regression discontinuity analyses estimated effect of ROSLA 1972 on refractive error for bin size of 1 month at different bandwidths.

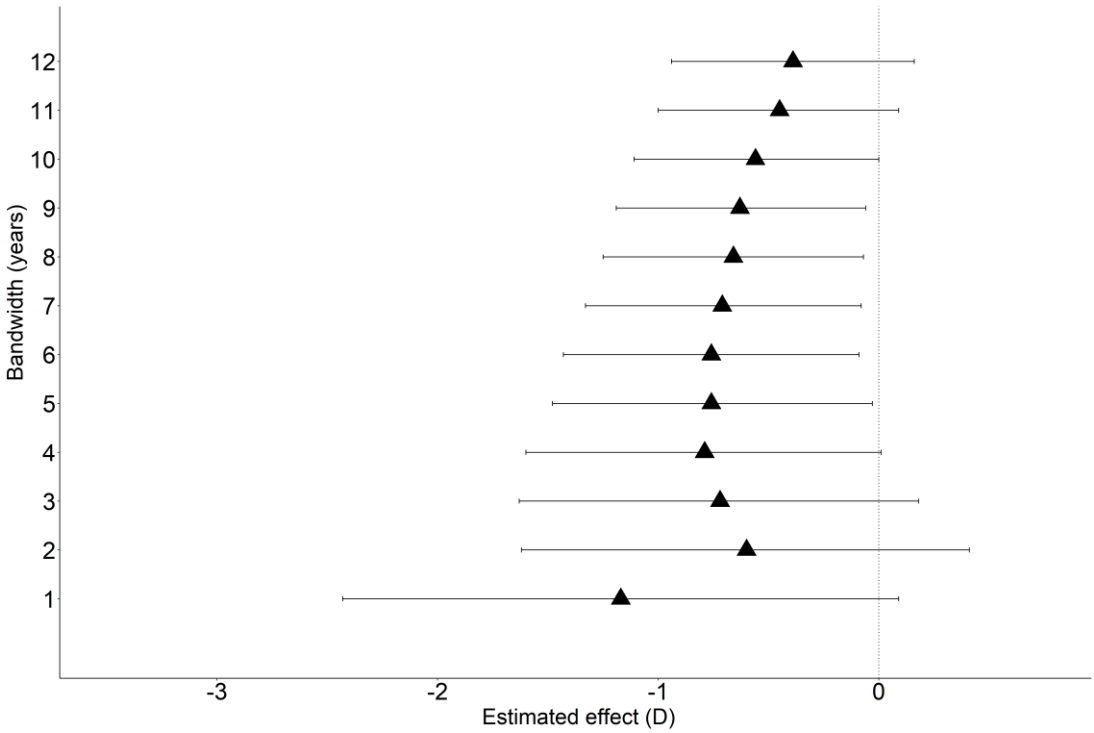

Figure 5.6 Causal effect of ROSLA 1972 reform on refractive error for bin size 1 month at different bandwidths. Vertical dotted line represents no-effect; error bars represent 95% confidence interval.

5.3.5 Heterogeneity in the Effect of the ROSLA 1972 Reform on Genetic Predisposition to Myopia

In order to assess whether genetic predisposition to myopia had an impact on the effects of ROSLA 1972, UK Biobank participants in the *IV Sample* were stratified into sub-samples with either high (n = 2,360) or low (n = 2.401) genetic predisposition for myopia development, respectively. Both the conventional OLS linear regression analysis and the instrumental variable analysis showed suggestive evidence of heterogeneity in the effect of ROSLA 1972 on refractive error across strata. Specifically, in the stratum of UK Biobank individuals with a high genetic predisposition to myopia, the IV analysis estimated effect was -0.43 D (95% CI: -1.68 to 0.81, p = 0.56), whereas in the subsample with low genetic predisposition to myopia the causal effect estimate was -1.22 D (95% CI: -2.11 to -0.34, p = 0.007). OLS analysis estimated the association in the subsample with high genetic predisposition to myopia as -0.08 D (95% CI: -0.29 to 0.14, p = 0.07), while a larger effect size was apparent for the subsample with low genetic predisposition to myopia: -0.25 D (95% CI: -0.41 to -0.09, p = 0.005). However, an

independent samples *t*-test did not support the hypothesis that there was a meaningful level of heterogeneity in the effect of the reform across strata (p = 0.28 and p = 0.30 for instrumental variable and OLS analyses, respectively).

5.3.6 Sensitivity Analyses

There was little evidence of 'forcing variable manipulation' around the cut-off point (McCrary density test for bin size 1 month, p = 0.21. Regarding the robustness of the results to the exact choice of bandwidth and bin size, the LATE estimates showed consistency in the magnitude and the direction of effect (*Table 5.2*, *Figure 5.7*).

To test for the presence of a discontinuity in the refractive error-education relationship at dates other than the true ROSLA cut-off point, the regression discontinuity analysis was carried out using 'artificial' cut-off dates, namely 2 years before or after September 1957. Both estimates were non-significant (p = 0.57 and p = 0.62, respectively).

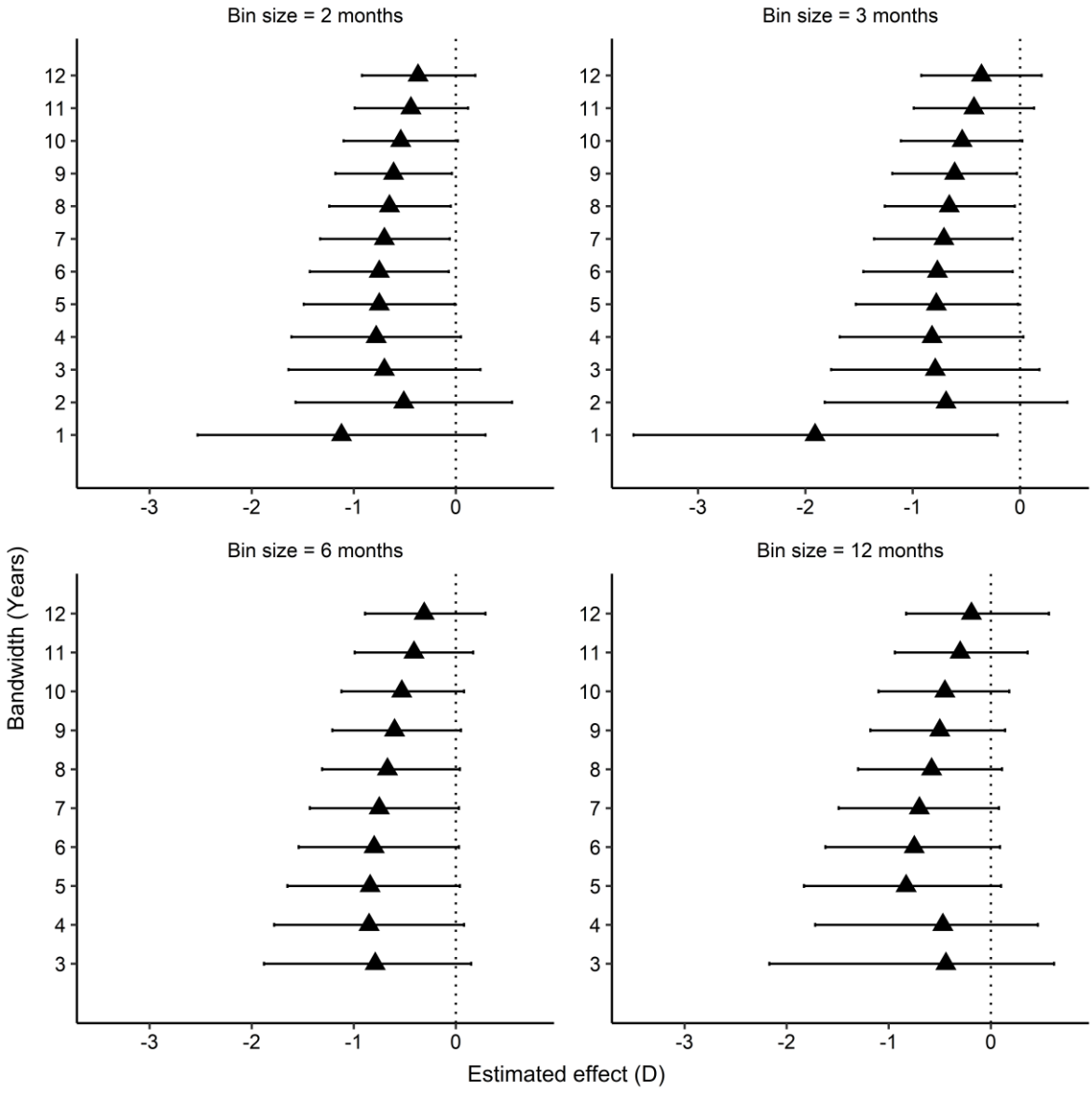

Figure 5.7 Causal effect of ROSLA 1972 reform on refractive error for different bin sizes at different bandwidths. Vertical dotted line represent no-effect; error bars represent 95% confidence intervals.

5.4 Discussion

In the current study, the ROSLA 1972 educational reform was used as a natural experiment to assess the causality of the association between educational attainment and refractive error. The discontinuity in educational exposure generated by ROSLA 1972 for those born after September 1957 made it possible to estimate the causal effect of the reform on refractive error with less bias from confounding factors than would be the case with conventional OLS analysis. UK Biobank

individuals affected by the reform had a shift in refractive error towards myopia. Estimates from both regression discontinuity and instrumental variable analyses were consistent in terms of magnitude and direction; however, the sample size of the current study was limited to the UK Biobank individuals born close to the cut-off date. This resulted in the low precision of the estimate (namely, within the range of -1.68 to -0.02 D), meaning that the true effect could be modest. A larger sample size would be required in order to increase the precision of the estimated effect of education on refractive error.

The heterogeneity in the effects of ROSLA 1972 across strata of highest educational qualification *Figure 5.8*) confirms previous suggestions that people in the lower tail of the educational attainment distribution were more affected by the reform than those who may have stayed on at school irrespective of the reform (381).

Suggestive evidence of heterogeneity in the effect of reform was found after stratifying the sample using the binary PRS for refractive error. The data suggested that individuals with a low genetic risk for myopia responded more to the additional year of education rather those at high risk of myopia.

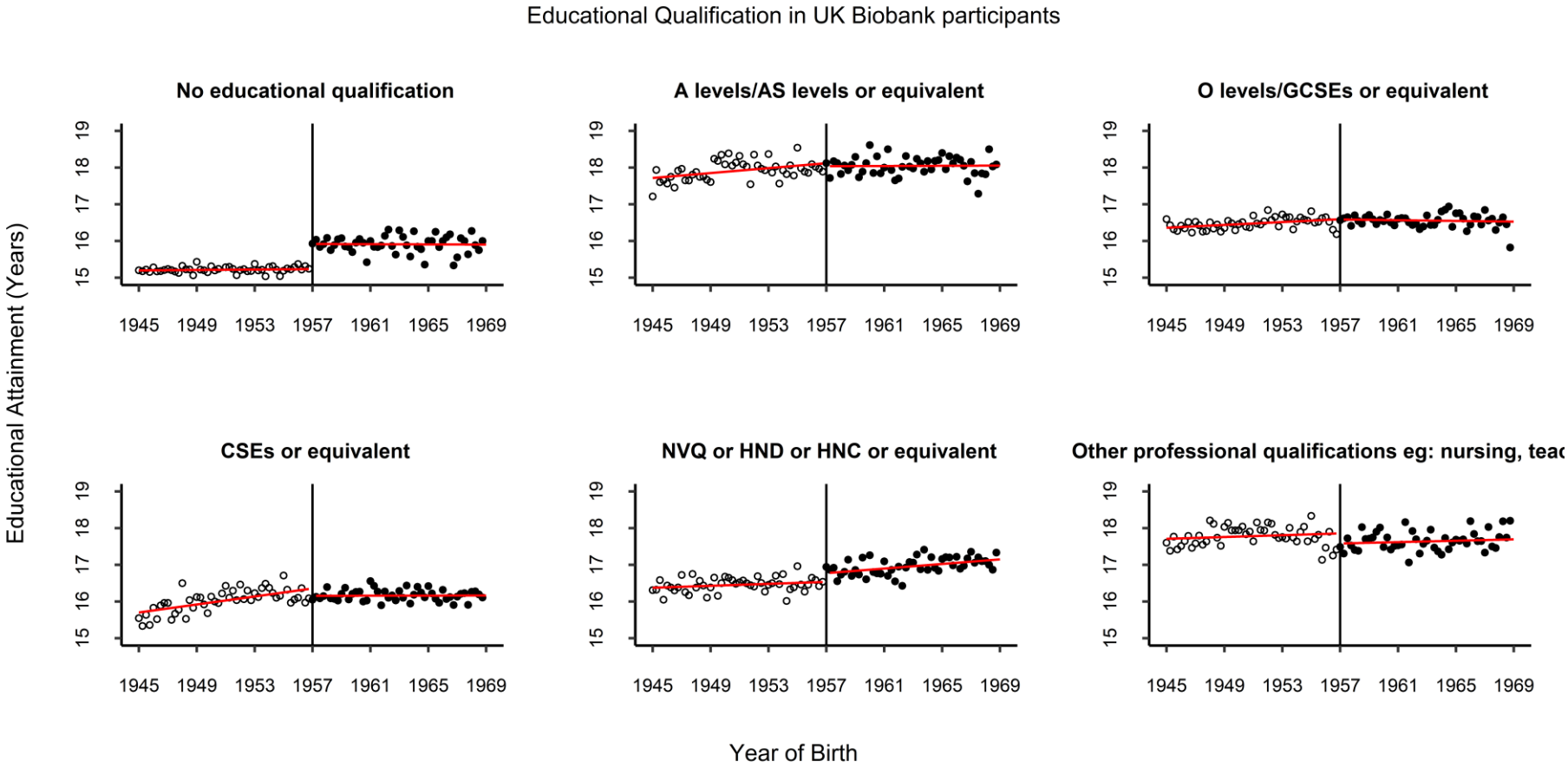

Figure 5.8 Educational discontinuity in UK Biobank participants stratified by educational qualification. Vertical line represents the cut-off date; data points represent mean educational attainment (age completed full time education) for the bin size of 3 months.

The regression discontinuity design allows to estimate the causal effect of an epidemiological exposure when it is not possible to conduct a randomized controlled trial. The major strength of the regression discontinuity design is its ability to balance confounders in the exposed and unexposed groups, as the participants in the groups just before and just after the cut-off point differ only in treatment assignment but – in theory – not in measured or unmeasured confounders (382).

Strengths of this study were that refractive error in UK Biobank participants was measured using a reliable objective method (autorefraction) and that both the month and as well as the year of birth were known. In order to avoid the effect of population stratification, the sample was restricted to individuals with white British genetic ancestry. Weaknesses of the study included the fact that well-educated people are overrepresented in UK Biobank, i.e. selection bias. To address this, inverse probability weighting was used. Variation of refractive error and educational attainment has been found in relation to season or month of birth (383, 384). To limit bias from month of birth, sensitivity analyses were carried out using a 12-month bin size allowed to exclude the impact of month of birth related effects.

The current study provided support for the theory that education is a causal risk factor for myopia. However, the study design did not allow me to explore the biological mechanism of the effect. Notably, the estimated effect reflects the influence of the educational reform per se and is not necessarily equal to the average effect of an additional year of education. Instead, it encompasses the effect of potentially dramatic changes in educational exposure and attainment for individuals born just before vs. after the cut-off point.

Chapter 6. Hyperopia and Educational Attainment: Assessing Causality by Non-linear Mendelian Randomisation

6.1 Introduction

It has been suggested that children require good vision in order to benefit fully from schooling in terms of their educational attainment and academic achievement (221). Children with hyperopia need to make more accommodative effort to perform near viewing tasks such as reading (223). A plethora of epidemiological studies have investigated the effect of uncorrected hyperopic refractive error on academic achievements (214, 224, 385, 386). A US-based study in 492 pre-schoolers aged 4-5 years showed an association of uncorrected hyperopia ≥ +4.00 D with reduced scores in a test of early literacy (225). A pilot study in 41 children aged 4 to 7 years reported a significant difference in early literacy skills ($p< 0.05$) in children with hyperopia ≥ +2.00 D in comparison with those with emmetropia (386). In a study conducted in Wales, a group of 1,298 8-years-old children underwent a screening for hyperopia; those who failed the fogging test were referred to an optometrist to assess the refractive error. A subsequent analysis of the school test results (SATs and NFERs) was performed in order to examine the association between hyperopic refractive error and the test results (214). The study reported lower SAT test scores in the children with hyperopia > +3.00 D in comparison with the non-referred group (SAT results: proportion of attaining CSI level 2; 79.5% vs. 85.9%) with a similar trend in the NFER results.

Even the existing evidence from observational studies suggests that children with uncorrected hyperopia have poorer educational outcomes (214, 224, 226). However, such a relationship might instead reflect the causal relationship between education and myopia. In other words, children who spend the least time in education are less likely to develop myopia and therefore they are more likely to be emmetropic (Section 1.2) or hyperopic (Section 1.4) instead.

Causality in the relationship between refractive error and educational attainment has been tested in a recently published bi-directional Mendelian randomisation (MR) study (46). This existing MR study found no evidence to suggest a non-zero causal effect of refractive error on time spent in full time education (the number of years of schooling associated with a +1.00 D increase in refractive error was -0.008 years/D, 95% CI -0.041 to 0.025, $p = 0.60$). However, the aforementioned MR study was performed under the assumption that the association between the exposure and the outcome is linear. It has been shown that in some cases the true association is non-

linear, e.g. the association between the body mass index and all-cause mortality (387, 388); and it is important to assess the shape of the true association and to estimate the effect when the assumption of linearity is not fulfilled.

The aim here was to test for evidence of a causal relationship between hyperopia and educational attainment, using non-linear MR (NLMR).

6.2 Methods

6.2.1 Selection of Participants

Analyses in the current study were restricted to unrelated UK Biobank participants of White British genetic ancestry as defined by Bycroft et al. (245). Those withdrawing consent were excluded from the analyses as were those without *EduYears* data (as defined in Section 2.1.1). This resulted in 272,588 participants; among them 61,158 individuals had valid data on average mean spherical equivalent refractive error (Section 2.1.2) and no medical record-reported or self-reported eye trauma or ocular surgery. Only participants with valid data on polygenic risk scores for refractive error were selected for the final sample (*NLMR sample*), which comprised of 60,988 individuals.

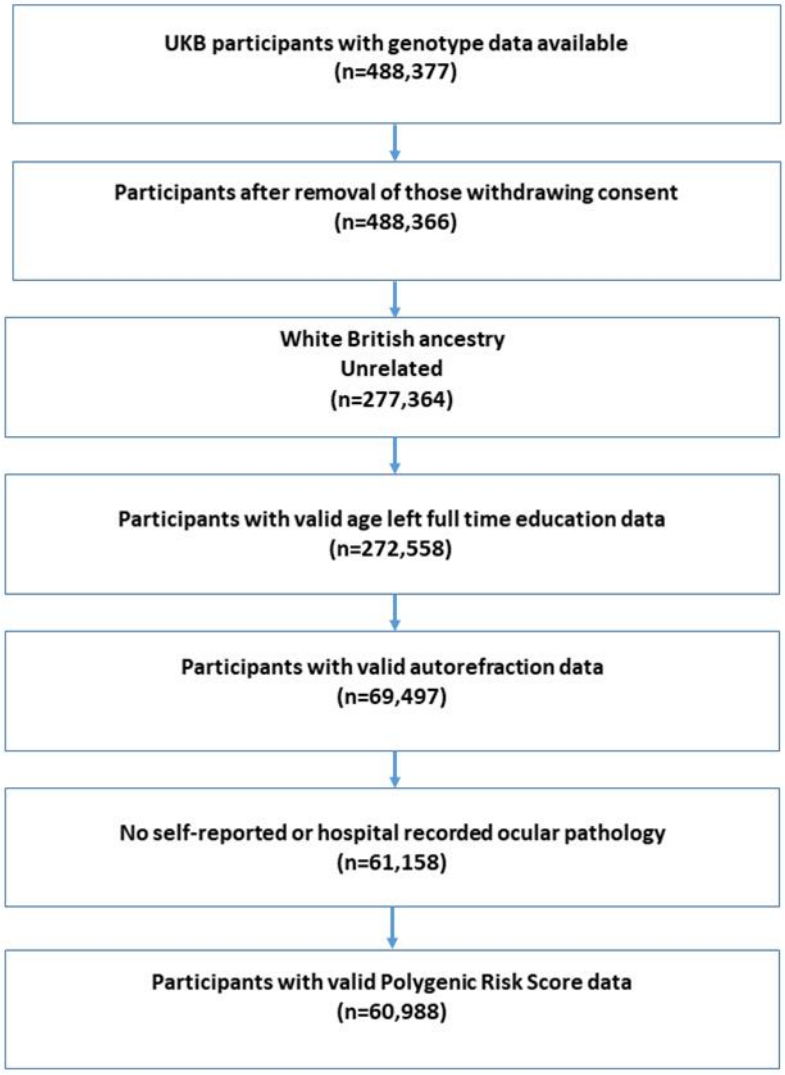

Figure 6.1 Selection of participants for linear and non-linear Mendelian randomisation analysis

6.2.2 Instrumental Variables for Mendelian Randomisation

A total of 161 genetic variants robustly associated with refractive error ($p < 5 \times 10E-08$) in the large GWAS meta-analysis of studies carried out by the CREAM Consortium and the 23andME personal genomics company (Section 2.3.2) (238) were used for instrumental variable selection. Further selection was based on the criteria adopted by Wood and Guggenheim (389). Of the 161 variants, 149 successfully replicated in an independent sample of 95,505 UK Biobank individuals

(390) and after exclusion of 18 variants in linkage disequilibrium with another SNP in the set, a total of 131 SNP was used in the main analysis (*Table 6.1*).

SNP	Effect Allele	Reference Allele	Effect size	SE	P-value
rs10122788	G	A	0.103	0.016	1.82E-10
rs10187371	T	C	0.066	0.015	1.25E-05
rs10458138	A	G	-0.095	0.018	1.36E-07
rs10500355	A	T	-0.099	0.018	4.53E-08
rs10511652	A	G	-0.123	0.018	4.14E-12
rs1064583	G	A	-0.089	0.015	8.22E-09
rs10760673	A	G	-0.114	0.016	3.13E-12
rs10853531	A	A	-0.101	0.017	1.60E-09
rs10880855	T	C	-0.099	0.015	7.72E-11
rs10887262	C	T	-0.072	0.015	2.12E-06
rs11088317	T	C	0.117	0.028	2.03E-05
rs11101263	T	C	0.075	0.020	1.34E-04
rs11118367	T	C	-0.126	0.022	7.86E-09
rs11145465	A	C	-0.051	0.020	1.09E-02
rs11178469	C	T	0.094	0.015	7.29E-10
rs11202736	T	A	-0.088	0.017	1.66E-07
rs11210537	A	G	0.066	0.018	1.59E-04
rs1150687	C	T	-0.135	0.016	3.61E-17
rs115152181	A	T	0.174	0.026	2.90E-11
rs11589487	A	G	-0.194	0.019	5.01E-24
rs11602008	T	A	-0.043	0.016	8.80E-03
rs11654644	T	C	-0.085	0.017	3.07E-07
rs117735470	A	G	-0.047	0.015	1.83E-03
rs11802995	C	A	-0.048	0.016	2.29E-03
rs11952819	T	C	0.069	0.016	2.26E-05
rs12193446	G	A	0.099	0.015	6.68E-11
rs1237670	G	A	0.035	0.015	2.09E-02
rs12451582	A	G	-0.094	0.017	4.92E-08
rs12526735	A	T	0.133	0.019	1.73E-12
rs12883788	T	C	-0.111	0.019	8.93E-09
rs12898755	A	G	-0.350	0.047	1.31E-13
rs12965607	G	T	-0.120	0.015	3.26E-15
rs1313240	C	T	-0.031	0.017	6.77E-02
rs1358684	C	T	-0.084	0.018	4.65E-06
rs1359543	G	A	0.034	0.016	2.65E-02
rs1454776	G	T	-0.075	0.018	3.82E-05
rs1532278	T	C	0.071	0.016	1.18E-05
rs1555075	T	C	0.072	0.015	2.73E-06
rs1556867	T	C	-0.059	0.015	1.32E-04
rs1649068	A	C	0.073	0.016	5.09E-06
rs17032696	C	A	0.220	0.016	2.31E-45
rs17125093	A	G	0.085	0.015	2.13E-08
rs17382981	T	C	-0.082	0.015	8.94E-08

rs17428076	G	C	0.415	0.026	2.62E-59
rs1790165	A	C	0.052	0.016	1.39E-03
rs1858001	G	C	-0.062	0.017	2.49E-04
rs1928175	G	A	-0.281	0.069	5.15E-05
rs1954761	T	C	0.062	0.015	5.99E-05
rs1969091	A	C	-0.037	0.015	1.56E-02
rs1983554	A	T	0.131	0.019	1.67E-12
rs2116093	C	G	-0.097	0.022	1.33E-05
rs2143964	G	C	0.070	0.016	2.06E-05
rs2150458	A	G	0.068	0.019	2.76E-04
rs2155413	A	C	0.081	0.016	3.22E-07
rs2166181	G	A	-0.055	0.016	5.14E-04
rs2225986	A	T	0.105	0.016	3.17E-11
rs2229742	C	G	-0.081	0.018	1.06E-05
rs235770	T	C	-0.113	0.018	4.29E-10
rs2573210	G	A	-0.078	0.019	3.24E-05
rs2573232	C	C	-0.048	0.015	1.61E-03
rs2622646	A	C	-0.095	0.017	2.04E-08
rs2745953	T	A	-0.071	0.015	3.01E-06
rs28471081	G	A	-0.061	0.016	1.23E-04
rs284818	T	C	0.071	0.015	2.97E-06
rs2855530	C	G	-0.136	0.016	3.26E-18
rs28658452	G	A	-0.112	0.017	1.55E-11
rs2908972	A	T	0.050	0.017	2.47E-03
rs297593	T	C	-0.052	0.015	6.25E-04
rs3110134	A	G	0.079	0.016	6.69E-07
rs3138137	A	C	-0.099	0.019	1.43E-07
rs34539187	G	C	0.131	0.017	4.24E-14
rs35337422	C	A	0.048	0.016	2.38E-03
rs36024104	G	A	-0.238	0.020	7.99E-32
rs41393947	A	G	0.044	0.015	4.03E-03
rs4237285	T	C	-0.092	0.015	2.58E-09
rs4260345	C	T	-0.102	0.016	7.67E-11
rs4687586	G	C	0.054	0.015	4.01E-04
rs4764038	T	G	-0.048	0.017	3.90E-03
rs4793501	C	T	0.074	0.015	1.34E-06
rs4795364	G	A	-0.260	0.029	1.05E-18
rs4808962	G	A	-0.069	0.016	2.49E-05
rs4894529	A	G	-0.091	0.017	1.69E-07
rs511217	T	A	-0.041	0.017	1.38E-02
rs524952	A	T	-0.069	0.026	7.61E-03
rs5442	A	G	-0.068	0.015	7.75E-06
rs55885222	A	C	0.118	0.015	1.34E-14
rs56014528	T	G	0.056	0.018	2.02E-03
rs56055503	G	A	-0.073	0.016	3.45E-06
rs56075542	G	T	0.058	0.016	2.33E-04
rs62070229	G	A	-0.092	0.017	9.41E-08
rs6420484	A	G	0.050	0.015	9.44E-04

rs6433704	G	T	-0.116	0.015	4.25E-14
rs6495367	A	G	0.104	0.015	4.37E-12
rs6753137	T	C	0.084	0.015	2.77E-08
rs7042950	G	A	-0.093	0.019	1.31E-06
rs7107014	A	A	-0.106	0.015	1.90E-12
rs7122817	A	G	-0.111	0.017	1.25E-10
rs7207217	A	G	-0.080	0.017	1.63E-06
rs72655575	A	C	-0.073	0.019	1.08E-04
rs72826094	T	A	0.106	0.021	4.22E-07
rs7337610	T	C	-0.065	0.021	2.45E-03
rs73730144	C	A	-0.264	0.015	9.69E-69
rs7449443	G	T	0.124	0.022	2.63E-08
rs74764079	A	T	0.135	0.019	4.38E-13
rs7624084	C	T	-0.149	0.015	1.55E-22
rs7662551	G	A	0.075	0.017	6.69E-06
rs7667446	C	T	0.115	0.019	9.15E-10
rs7737179	A	G	-0.158	0.016	6.70E-23
rs7744813	C	A	0.047	0.018	9.40E-03
rs7747	T	C	0.038	0.016	1.38E-02
rs7829127	G	A	-0.123	0.016	6.07E-15
rs7895108	T	G	-0.075	0.015	1.07E-06
rs7933504	G	A	-0.125	0.019	1.46E-10
rs7941828	T	C	-0.082	0.017	2.32E-06
rs7968679	G	A	-0.078	0.020	6.78E-05
rs7971334	T	G	0.093	0.016	3.57E-09
rs807037	G	C	-0.115	0.020	7.26E-09
rs8073754	T	C	0.073	0.015	2.32E-06
rs8075280	A	T	-0.044	0.016	4.81E-03
rs837323	T	C	-0.080	0.016	4.55E-07
rs9295499	A	C	0.086	0.019	3.75E-06
rs931302	T	C	-0.130	0.021	4.51E-10
rs9388766	T	C	-0.057	0.020	4.76E-03
rs9395623	A	T	-0.066	0.015	2.11E-05
rs9416017	T	C	0.067	0.016	2.28E-05
rs9516194	A	G	-0.157	0.024	1.57E-10
rs9517964	C	T	-0.049	0.017	2.90E-03
rs9547035	G	T	-0.046	0.042	2.70E-01
rs9606967	C	G	0.086	0.015	1.74E-08
rs9680365	A	G	-0.026	0.019	1.62E-01
rs9681162	C	T	-0.066	0.016	4.96E-05

Table 6.1 The association of the 131 SNP selected for the main analysis with refractive error in the sample of 60,988 UK Biobank participants.

The degree of the association of each of 131 aforementioned individual genetic variants with refractive error was assessed one-by-one using a series of 131 linear regression analyses (*Equation 6.1*) in the sample of 60,988 UK Biobank individuals.

$$RE = \beta_0 + \beta_1 SNP + \beta_2 \, cov_1 + \cdots + \beta_i cov_{i-1} + \varepsilon$$

Equation 6.1 Linear model equation. RE = refractive error, SNP = genotype of test SNP (coded as 0,1 or 2), Cov_1 …Cov_{i-1} = set of covariates, β_1 = regression coefficient for PRS, ε = error term.

A polygenic risk score (PRS) for refractive error was calculated (*Equation 6.2*), using beta coefficients from the SNP-refractive error regression as weights in the sample of 60,988 UK Biobank individuals.

$$PRS = \sum_{m=1}^{n} \beta_m \times X_m$$

Equation 6.2 PRS = polygenic risk score for refractive error, X_m = number of effect alleles (0, 1 or 2) of individual genetic variant m, β_m = weight of individual genetic variant, n = number of SNP included (n=131).

The variance in refractive error explained by the PRS was assessed in the *NLMR sample* by performing two linear regressions : (1) baseline model with refractive error as the response variable and age, sex, genotyping array, Townsend Deprivation Index, and first five ancestry principal components (PC1-5) as explanatory variables; (2) full model with the PRS included as an additional explanatory variable. The variance in refractive error explained by PRS was calculated as the difference in adjusted R^2 of the full and baseline models (*Equation 6.3*).

$$VarRE = adjR^2 \, Full - adjR^2 \, Baseline$$

Equation 6.3 Variance in refractive error explained by PRS; VarRE = variance in refractive error, $adjR^2 \, Full$ = adjusted coefficient of determination of the full model, $adjR^2 \, Baseline$ = adjusted coefficient of determination of the baseline model.

6.2.3 Observational Analysis, Linear and Non-linear Mendelian Randomisation

A linear regression was performed in order to estimate the observational association between refractive error and *EduYears* in the final sample of 60,988 UK Biobank individuals. Then, the regression was repeated with adjustment for age, gender, Townsend Deprivation Index, genotyping array and the first 5 ancestry principal components. Inclusion of ancestry principal components was done to provide additional protection from potential bias due to population stratification.

A one-sample linear Mendelian randomisation analysis with refractive error as the exposure variable and with *EduYears* as the outcome was carried out in the same sample of 60,988 individuals using the *AER* R package. The two stage least squares model (Section 2.4.2) was used, with the PRS for refractive error (Equation 6.2) as the instrumental variable. The first-stage F-statistics was assessed to determine if the PRS was a 'weak instrument' (340). A statistical approach proposed by Brion et al. (391) and implemented in the *mRnd* online tool (https://shiny.cnsgenomics.com/mRnd/) was used to assess the statistical power of the analysis for a type-I error rate α = 0.05.

A non-linear Mendelian randomisation does not rely on the assumption that the effect of the exposure on the outcome is linear. In order to perform a non-linear Mendelian randomisation analyses, the *NLMR* sample was first stratified into 10 deciles using the refractive error distribution. The 10 refractive error deciles were selected based on the *residual refractive error* ('genetic risk-free'). Residual refractive error can be explained as an IV-free exposure; or the exposure, free of genetic risk. Genetic risk in that case was calculated as the effect of all SNP in the PRS for refractive error (*Equation 6.4*).

$$RE = \beta_0 + \beta_1 PRS + \chi$$

Equation 6.4 Calculation of Residual Refractive Error. RE = refractive error, PRS = polygenic risk score for refractive error, β_1 = regression coefficient for PRS, χ = residual refractive error.

Residual refractive error was calculated using the *lm* function in R-3.5.0 statistical software in two steps. In the first step, refractive error was regressed on the PRS in the *NLMR sample*. Next, the residuals from the regression (i.e. the IV-free exposure) were obtained. Stratification of the sample on the IV-free exposure allows

to avoid a violation of one of the core IV assumption (Section 2.4.2) that could be induced when stratifying the sample into quantiles of the exposure directly (392).

Two approaches were applied to implement the non-linear Mendelian randomisation analysis: (1) a 2SLS analysis was performed in each of the 10 residual refractive error deciles using the PRS for refractive error as an instrument; (2) an inverse variance weighted Mendelian randomisation (IVW; Section 2.4.3) was carried out in each of the 10 deciles using all 131 SNP as discrete instrumental variables. Note that for the latter analysis using 131 SNP as instrumental variables, the beta coefficients quantifying the SNP-exposure association (Equation 6.1) were calculated within each decile.

Next, Cochran's Q test was used to test for heterogeneity in the estimates obtained in the 10 refractive error deciles using the *rma* command in the R package *metafor.*

6.2.4 Sensitivity Analysis Using SNP for PRS from an Independent GWAS

In the NLMR analyses described above, the SNP weights for constructing the PRS for *avMSE* were derived using the same dataset in which the NLMR analysis was carried out (due to the absence of published summary statistics of a very large GWAS for refractive error). Non-independency of the discovery and validation samples can lead to the over-estimation of the accuracy of prediction (393); hence, the results of the main analysis could have been biased. Therefore, a sensitivity analysis was performed using summary statistics from an independent GWAS for Age of Onset of Spectacle Wear (AOSW)-inferred mean spherical equivalent in 287,448 UK Biobank participants (317) (note that there was no overlap between the AOSW-inferred refractive error sample and the *NLMR sample*). A total of 15,735 SNP were genome-wide significantly associated (p< 5E-08) with AOSW-inferred refractive error. After excluding variants within ±500 kb from the lead SNP in each region, or those having a pairwise LD of $r^2 > 0.2$ with the lead variant, there were 170 SNP independently associated with the trait. The strength of the association of each of the 170 SNP (GWAS β coefficient) was used as weight for PRS construction in the *NLMR sample*, using the PLINK 1.9 ('--score' command) (Figure 6.2). Non-linear MR analysis was carried out as described above.

6.2.5 Simulations

A simulation study was carried out in order to assess the power of non-linear MR to detect a causal effect of refractive error on educational attainment in the *NLMR sample*. A (simulated) specific period of education was subtracted from the existing educational level of participants with hyperopia exceeding a specific threshold level, followed by a test of whether the resulting non-linear relationship between refractive error and educational level could be detected by non-linear Mendelian randomisation. Power was calculated as the proportion of simulations in which the non-linear relationship was detected. To implement the simulation, first, the observed *EduYears* values were permuted (i.e. randomly re-assigned amongst participants without replacement) in order to break any existing link between refractive error and education level. Next, a specific period of education (*'ΔEduYears'*; varying from 0.05 to 1.00 year of education, in steps of 0.05) was subtracted from the individual's permuted *EduYears* value for a proportion of participants (*'ΔParticipants'*; 0.5 to 1.0 in steps of 0.1) with refractive error equal to or higher than a threshold level of hyperopia (*'τHyperopia'*; +1.00 D, +2.00 D, or +3.00 D). The resulting level of education, or *'EduYearNew'* can be presented as follows: E*duYearsNew* = permuted *EduYears* – *ΔEduYears*. The steps described above were repeated 100 times per condition (100 replicates × 20 *ΔEduYears* levels × 6 *ΔParticipants* levels × 3 *τHyperopia* levels = 36,000 permutations in total). Statistical power was calculated as the proportion of permutations in which two conditions were met: (1) the null hypothesis of no heterogeneity was rejected ($p < 0.05$ for Cochran's Q test) and (2) the MR analysis detected a causal effect ($p < 0.05$ for the test of a causal effect in that decile) for at least one of the hyperopic deciles, i.e. deciles 6, 7, 8, 9 or 10.

6.3 Results

In the *NLMR sample*, 52.5% of participants were female and 36.6% had a University or college degree (Table 6.2). Mean educational attainment was 18.3 years, with a standard deviation (SD) of 2.5 years. Mean age of participants in UK Biobank who met the study inclusion criteria was 57.8 (SD 7.8) years and the mean refractive error was -0.27 (SD 2.67) Dioptres (D).

Characteristic		All n = 502,543	NLMR n = 60,988	Excluded n = 441,555	P-value
Female gender	N (%)	264,579 (54.2%)	31,990 (52.5%)	232,589 (54.5%)	4.40E-20
University or college degree	N (%)	163,534 (33.0%)	22,342 (36.6%)	141,192 (32.4%)	1.30E-94
Refractive error(D)*	Mean (SD)	-0.33 (2.71)	-0.27 (2.67)	-0.38 (2.73)	4.10E-14
Age (years)	Mean (SD)	57.12 (8.11)	57.77 (7.83)	57.03 (8.14)	<1.0E-99
Townsend Deprivation Index	Mean (SD)	-1.29 (3.10)	-1.42 (2.81)	-1.28 (3.13)	3.70E-31
EduYears (years)†	Mean (SD)	18.07 (2.54)	18.27 (2.50)	18.04 (2.55)	<1.0E-99

Table 6.2 Demographics of the full UK Biobank sample, NLMR sample and the sample excluded from the main analyses

*Refractive error data was available for 129,739 UK Biobank participants;

† *EduYears* data was available for 488,295 UK Biobank participants

Maximum and minimum values of *residual refractive error* for the *residual refractive error* deciles are presented in Table 6.3.

Decile	EduYears (Years)	Refractive error (D)		Residual refractive error (D)	
		Minimum	Maximum	Minimum	Maximum
1	19.26	-22.60	-1.82	-21.89	-3.29
2	18.93	-5.81	0.50	-3.29	-1.53
3	18.54	-3.76	1.42	-1.53	-0.61
4	18.38	-2.86	1.81	-0.61	-0.07
5	18.28	-2.37	2.43	-0.07	0.36
6	18.18	-2.09	2.47	0.36	0.77
7	18.04	-1.55	3.30	0.77	1.20
8	17.90	-1.08	3.73	1.20	1.73
9	17.72	-0.66	5.21	1.73	2.58
10	17.50	0.18	13.96	2.58	14.76

Table 6.3 Mean EduYears and minimum and maximum refractive error values for each decile after stratifying by residual refractive error.

In the observational analysis (OLS regression) in the full *NLMR sample*, a negative association between refractive error and educational attainment was found; a more positive refractive error was associated with a reduction in *EduYears*. Each additional Dioptre of refractive error was associated with 9.5 weeks (β = -0.18 years, 95% CI -0.19 to -0.17, p < 2E-16) less time in full-time education. The estimated effect was similar after adjusting for the potential confounders age, gender, and socioeconomic position (β = -0.15 years/D, 95% CI -0.16 to -0.15, p < 2.2E-16).

The PRS for refractive error explained 6.7% of the variance in refractive error in the *NLMR sample*. With linear Mendelian randomisation, which assumed the refractive error vs. years of education relationship was linear, the estimated causal effect of refractive error on educational attainment was β = -0.02 years/D (95% CI -0.04 to 0.01, p = 0.15). There was strong statistical support to suggest the linear MR estimate was different from the OLS regression effect size (Wu-Hausman test statistics = 94.4, p < 2E-16), suggesting the influence of unmeasured confounders on the OLS estimate. The linear MR analysis had 80% power to detect a causal effect of refractive error on educational attainment of at least -0.04 years/D (2 weeks reduction in *EduYears* per Dioptre) in the same *NLMR sample* of 60,988 UK Biobank individuals (*Figure 6.2*).

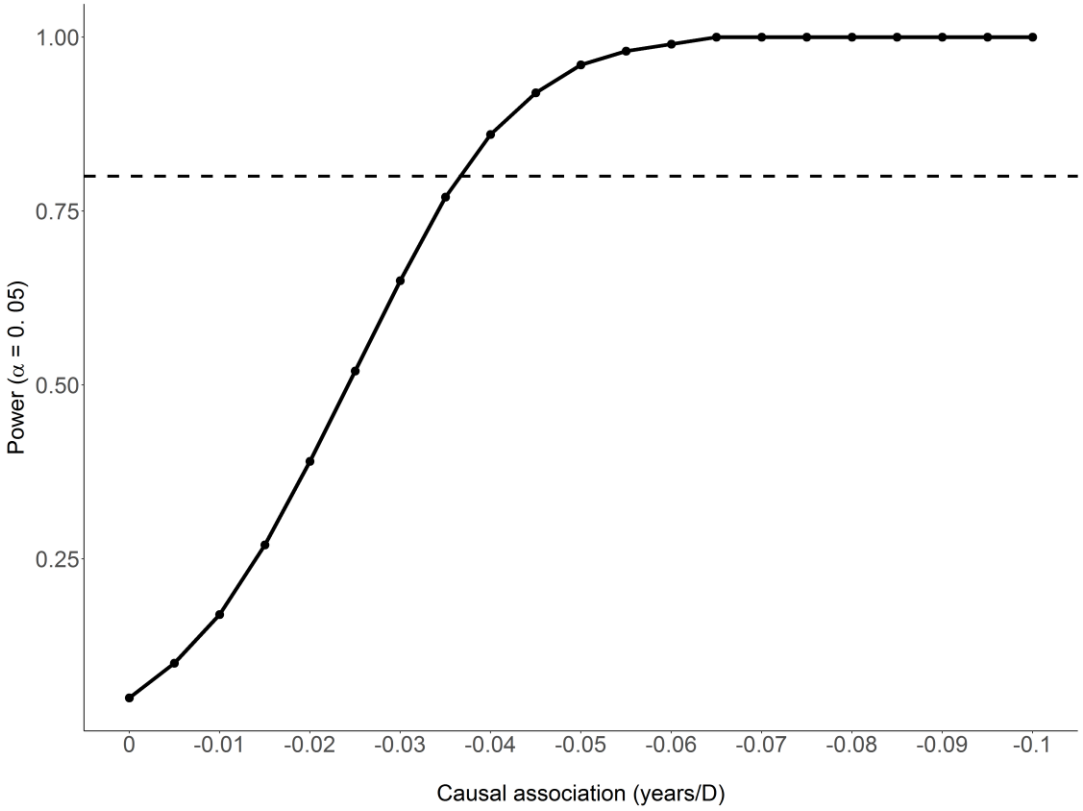

Figure 6.2 Power calculation of linear Mendelian randomisation for sample of 60,988 participants. The dotted line indicates 80% power.

After regressing out the effect of the genetic predisposition to refractive error and dividing the *NLMR sample* into deciles of residual refractive error, myopic participants were restricted to deciles 1-4 while hyperopic participants were restricted to deciles 5-10. Non-linear Mendelian randomisation using the PRS for refractive error as an instrumental variable estimated a causal effect ranging from -0.15 years/D to +0.09 years/D in each decile (*Figure 6.3*, panel A). There was evidence supporting a non-zero causal effect only for deciles 3 and 4 (p = 2.3E-04 and p = 8.7E-03, respectively), which corresponded to participants with refractive errors within the range -0.07 D to -1.53 D. There was no evidence to suggest that having hyperopia (deciles 6 to 10) had a causal effect on educational attainment. There was evidence of heterogeneity across strata (I^2 = 62.8%; Cochran's Q = 23.3, p = 5.5E-03).

The non-linear MR estimated effect of refractive error on educational attainment within refractive error deciles are presented in *Table 6.4*-*Table 6.5*.

Quantile	Effect	SE	p-value	95% CI
1	0.016	0.031	0.611	-0.045 to 0.077
2	-0.046	0.038	0.233	-0.121 to 0.029
3	-0.151	0.040	0.0002	-0.229 to -0.072
4	-0.114	0.047	0.016	-0.208 to -0.021
5	-0.066	0.052	0.210	-0.168 to 0.037
6	0.020	0.052	0.706	-0.083 to 0.123
7	0.000	0.053	0.999	-0.103 to 0.103
8	0.074	0.051	0.143	-0.025 to 0.174
9	0.011	0.046	0.803	-0.079 to 0.102
10	-0.020	0.034	0.549	-0.087 to 0.046

Table 6.4 Estimated NLMR causal effect of refractive error on educational attainment within each decile, 2SLS approach. The effect is expressed in units EduYears per Dioptre.

Quantile	Effect	SE	p-value	95% CI
1	0.019	0.029	0.517	-0.038 to 0.076
2	-0.052	0.037	0.158	-0.123 to 0.020
3	-0.147	0.039	0.0002	-0.223 to -0.070
4	-0.107	0.053	0.042	-0.210 to -0.004
5	-0.079	0.059	0.177	-0.195 to 0.036
6	0.026	0.059	0.663	-0.090 to 0.142
7	0.017	0.065	0.794	-0.110 to 0.143
8	0.076	0.054	0.159	-0.030 to 0.182
9	0.019	0.047	0.685	-0.073 to 0.111
10	-0.026	0.030	0.387	-0.086 to 0.033

Table 6.5 Estimated NLMR causal effect of refractive error on educational attainment within each decile, IVW approach. The effect is expressed in units EduYears per Dioptre.

It has been shown previously that most genetic variants associated with refractive error have effect sizes that vary markedly between individuals, indicating possible influence of gene-gene or gene-environment interactions (394). In that study (394), the effect of genetic variants was higher in individuals with more myopic or more hyperopic refractions in comparison with those with emmetropic refractions. As the non-linear MR analysis using the PRS for refractive error as an instrumental variable did not account for this non-uniform effect size across deciles, a non-linear IVW MR with 131 discrete genetic variants as instrumental variables (rather than the PRS) was carried out. Hence, the difference in the SNP vs. refractive error association across deciles was taken into account.

The pattern of results (*Figure 6.3*, panel B) was very similar to that of the original non-linear MR analysis (*Figure 6.3*, panel A). There was evidence of heterogeneity across strata (I^2 = 61.9%: Cochran's Q test statistics = 22.8, p = 6.7E-03) and the only deciles with evidence of a causal relationship between refractive error and educational attainment were deciles 3 and 4 (p = 1.2E-04 and p = 2.8E-02, respectively), corresponding to individuals with low levels of myopia.

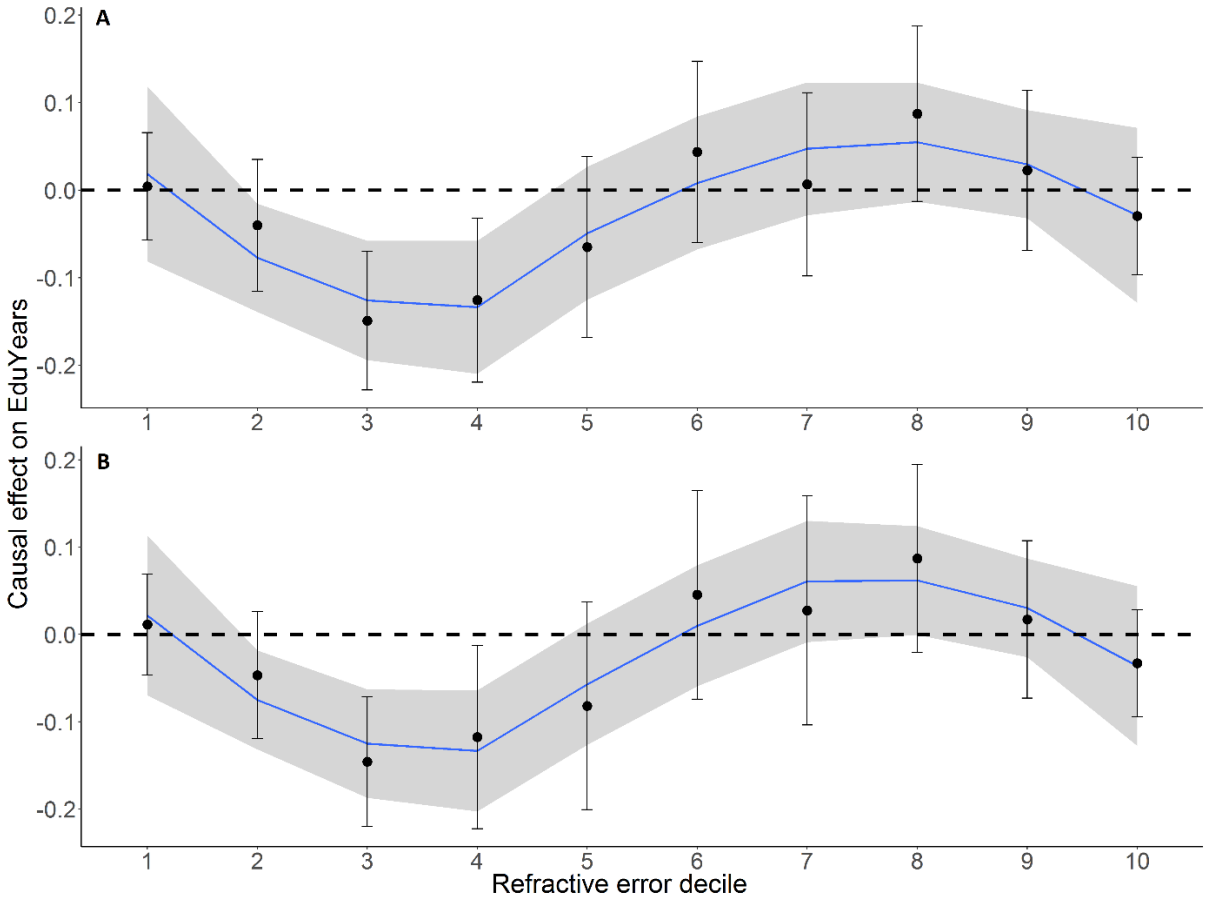

Figure 6.3 Non-linear Mendelian randomisation analysis in the NLMR sample stratified by residual refractive error. (A) 2SLS using the PRS for refractive error as an instrumental variable and (B) IVW meta-analysis of 131 discrete SNP used as instrumental variables. Points represent the estimated causal effect of refractive error on educational attainment (EduYears per Dioptre) within each decile; error bars represent 95% confidence intervals; the dashed line depicts no causal effect; the blue line represents the smoothed conditional mean; the grey shaded region represents 95% confidence interval of the smoothed conditional mean.

A 2 SLS non-linear Mendelian randomisation using a PRS with SNP weights derived from a GWAS for AOSW-inferred refractive error was performed as a

sensitivity analysis to avoid over-fitting which may arise when using 1-sample MR, i.e. the same discovery and validation sample. The estimated effect for each stratum is presented in *Table 6.6*. The pattern of results was similar to that of the main analysis, indicating minimal bias due to the use of a non-independent sample for deriving the SNP weights for the PRS (*Figure 6.4*).

Quantile	Effect	SE	p-value	95% CI
1	0.044	0.042	0.298	-0.038 to 0.126
2	-0.024	0.050	0.638	-0.122 to 0.074
3	-0.181	0.050	0.0003	-0.280 to -0.082
4	-0.188	0.060	0.002	-0.304 to -0.071
5	-0.074	0.064	0.251	-0.199 to 0.051
6	-0.120	0.066	0.069	-0.249 to 0.009
7	-0.072	0.064	0.262	-0.198 to 0.054
8	0.013	0.062	0.833	-0.110 to 0.137
9	0.030	0.060	0.621	-0.087 to 0.146
10	-0.019	0.044	0.661	-0.106 to 0.067

Table 6.6 Estimated effect of refractive error on educational attainment within each decile, 2SLS approach using a PRS with SNP weights from an independent GWAS for AOSW-inferred refractive error.

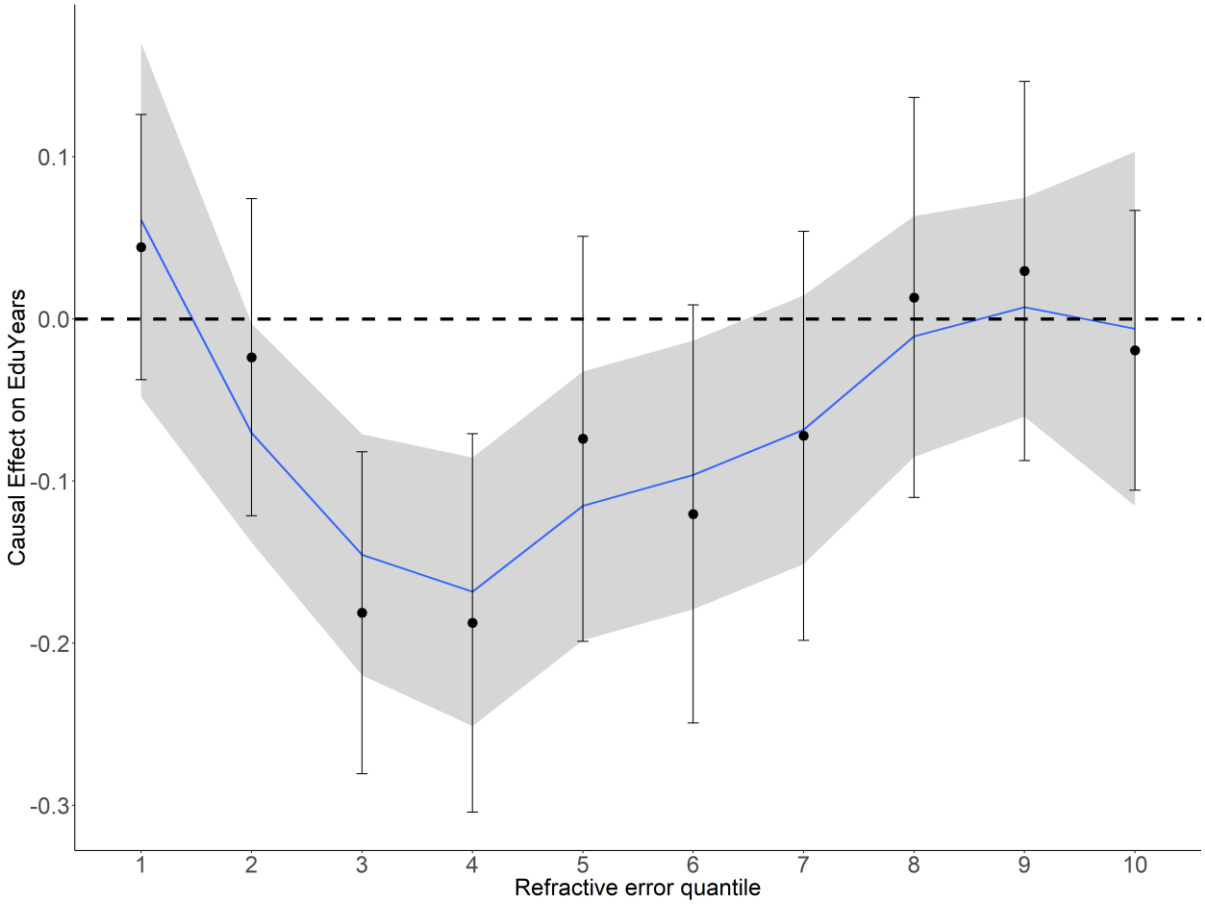

Figure 6.4 Non-linear Mendelian randomisation analysis in the NLMR sample stratified by residual refractive error. 2SLS using the PRS for refractive error (using SNP weights derived from the GWAS for AOSW-inferred avMSE) as an instrumental variable. Points represent the estimated causal effect of refractive error on educational attainment (EduYears per Dioptre) within each decile; error bars represent 95% confidence intervals; the dashed line depicts no causal effect; the blue line represents the smoothed conditional mean; the grey shaded region represents 95% confidence interval of the smoothed conditional mean.

A comprehensive set of simulations was performed to assess the statistical power of the non-linear MR analyses to detect a causal effect of hyperopia on educational attainment. The results, shown in Figure 6.4, demonstrated that power increased when one of the following conditions was met:

1) the simulated causal effect size (*ΔEduYears*) increased,

2) when a larger proportion of participants (*ΔParticipants*) was affected by the causal effect.

3) and when the threshold level of hyperopia (*τHyperopia*) at which a causal effect arose decreased.

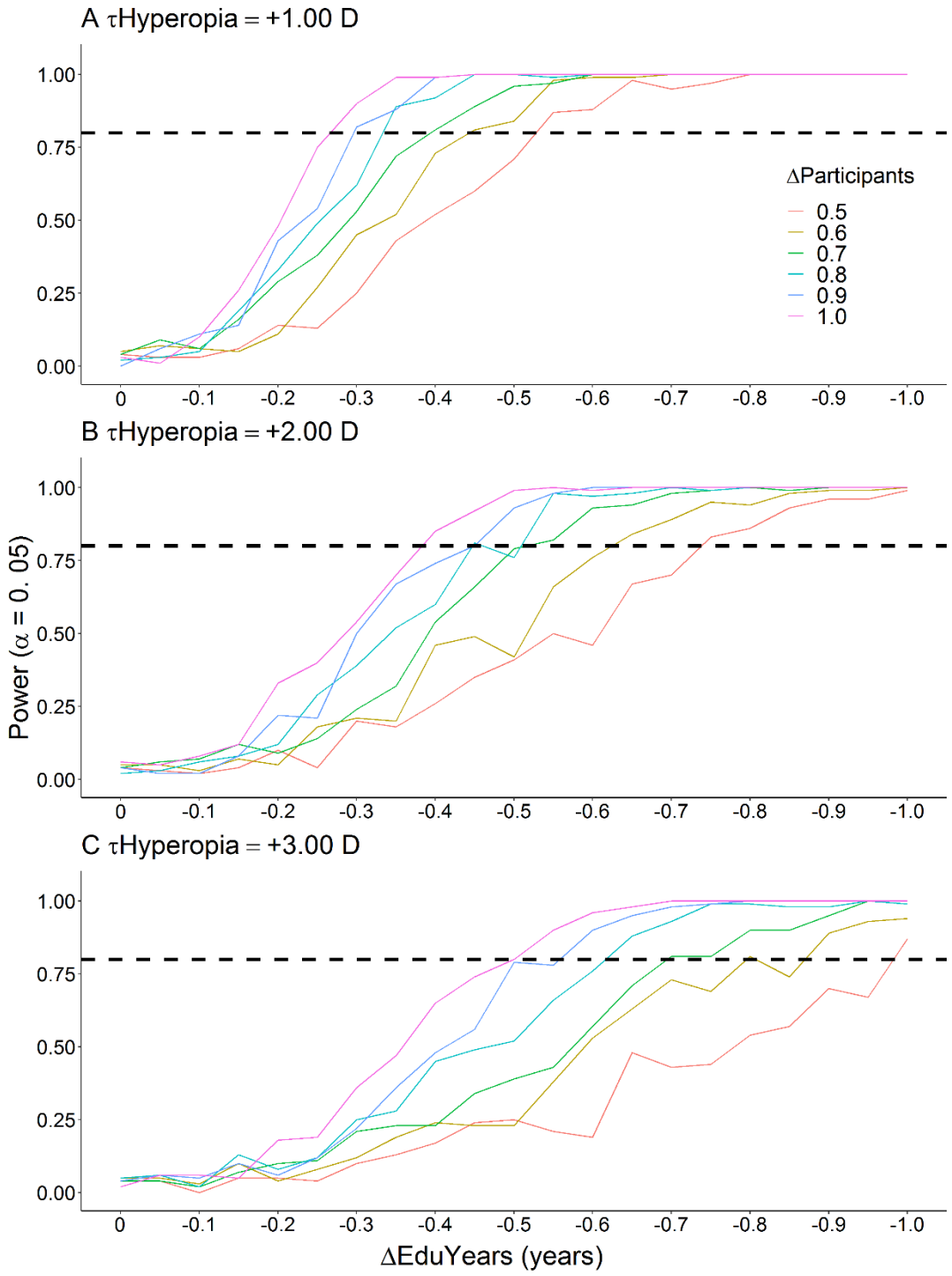

*Figure 6.5 Non-linear MR power simulations. Power to detect both non-linearity of the causal relationship between refractive error and educational attainment and a causal effect in a hyperopic decile. Power is plotted as a function of causal effect size (ΔEduYears). Coloured lines represent the proportion of participants (ΔParticipants) with refractive error above a threshold level of hyperopia (τHyperopia; +1.00 D, +2.00 D and +3.00 D in panels **A-C**, respectively).*

When *τHyperopia* was +1.00 D and *ΔParticipants* was 1.0, the study had 80% power to detect an effect of *ΔEduYears* ≤ -0.3 years, i.e. when all individuals with hyperopia of at least +1.00 D received 4 months or less education. When *τHyperopia* was +3.00 D and *ΔParticipants* was 1.0, the study had 80% power to detect an effect of *ΔEduYears* ≤ -0.55 years.

6.4 Discussion

This analysis supported the hypothesis that the relationship between refractive error and *EduYears* is non-linear, but did *not* support the hypothesis that hyperopia is a causal risk factor for educational attainment. Specifically, the results of this study suggested a small causal effect of short-sightedness on education of approximately 8 weeks more education per Dioptre of less myopia (i.e. -0.15 years/D, 95% CI -0.23 to -0.07), but no causal effect of far-sightedness on education.

The non-linear Mendelian randomisation undertaken here is, to date, the first study to assess the causality in the association between refractive error and education attainment without making the assumption that this relationship is linear.

The estimated causal effect had 95% confidence intervals that did *not* overlap zero only in deciles with a moderate and low level of myopia, not in deciles of participants with hyperopia. Thus, the current analysis did not support the main hypothesis of hyperopia being a causal risk factor for lower educational attainment.

The current study was performed as a one-sample MR, where the estimated causal effect of the outcome on the exposure can be described as a ratio of the two estimates (Wald estimate; Section 2.4.2) (272). The numerator in this case is the effect of the PRS on the outcome (βy) and the denominator is the effect of the PRS on the exposure (βx) (*Figure 6.6*).

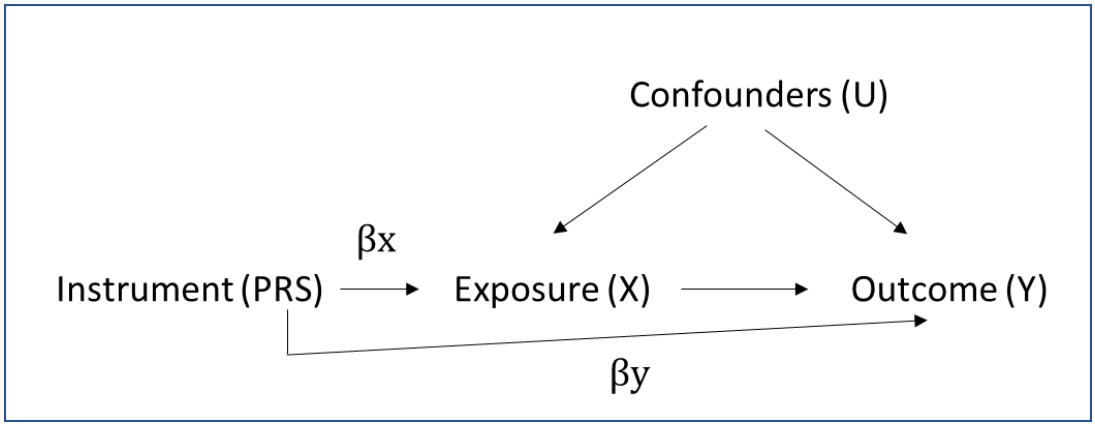

Figure 6.6 Causal Diagram. U is the set of unmeasured confounders of the exposure-outcome association.

Data on the instrumental variable (PRS), exposure (*avMSE*) and the outcome (*EduYears*) were available in UK Biobank, making it possible to use the 1-sample MR design.

Strengths of this study were the large sample size of 60,988 UK Biobank participants and that the refractive error was measured using an objective, standardised method (autorefractometry). Given the sample size, the MR analysis assuming a *linear* relationship between refractive error and years in education had approximately 80% power to detect even a modest causal effect; namely, a 2-week decrease in full-time education per Dioptre of hyperopia. However, power was much lower for the *non-linear* MR analyses. For instance, if only individuals with at least +3.00 D of hyperopia were affected, then the study had approximately 80% power to detect a causal effect of between a half and a full year less education.

To reduce the likelihood of bias due to population stratification, the study was limited to individuals of white British genetic ancestry. In addition, the first five ancestry principal components were included in the main analysis, thereby providing further protection from population stratification-induced bias. The genetic variants selected for constructing the PRS for refractive error were strongly associated with the phenotype of interest in a GWAS (238), satisfying the first MR assumption. However, as their biological roles are not clear, strong assumptions had to be made that these variants satisfied the second and the third MR assumptions. Indeed, the second and the third MR assumptions are untestable in practice (340) .

One of the limitations of the study was that the effect size quantifying the association with refractive error was calculated within deciles using the same data

as the main analysis, which can lead to ‘over-fitting’ and biased estimates. However, the sensitivity analysis in which SNP-exposure weights were derived using a GWAS for AOSW-inferred refractive error provided reassurance that over-fitting had minimal influence on the results. Another limitation was the fact that UK Biobank participants were reported to have higher values of educational attainment than the general population of the UK (234). This could be a potential source of a bias when estimating the effect in observational, linear and non-linear MR analyses. A further limitation of the current study was the fact that refractive error was measured in adulthood while investigating the causal effect on education during childhood. Refractive error development across the lifespan in general depends on two mechanisms (395): (1) emmetropization (Section 1.2.2) and, (2) a trend towards myopia development (396).The first mechanism typically occurs in childhood and is completed during the school years (397). The second one describes the myopic shift in refraction in adolescence and in adulthood. In elderly groups, refraction remains stable, despite changes in the crystalline lens (398). Thus, a person who develops myopia in childhood will typically remain myopic in adulthood; and a hyperopic refraction in adulthood typically reflects far-sightedness in childhood; various models were proposed to predict the adult refractive error given the refraction in childhood (399, 400). An elegant formula, proposed by Flitcroft et al. (396) (*Figure 6.7*) allows the researcher to calculate a refraction at a given age and shows that different factors contribute to myopia and hyperopia development. It can be assumed that the main driver in refraction development after emmetropization stops is ‘myopic drift’. Hence, if the refractive error in adulthood is known, then at least in theory, the refractive error in childhood would have been of the same sign or even more positive (hyperopic).

From the genetic point of view, a previous study showed that a PRS constructed using the CREAM Consortium GWAS summary statistics from adult individuals was able to explain a portion of the refractive error variance in a paediatric cohort (ALSPAC) (401). This suggests there is a genetic correlation between refraction in childhood and in adulthood.

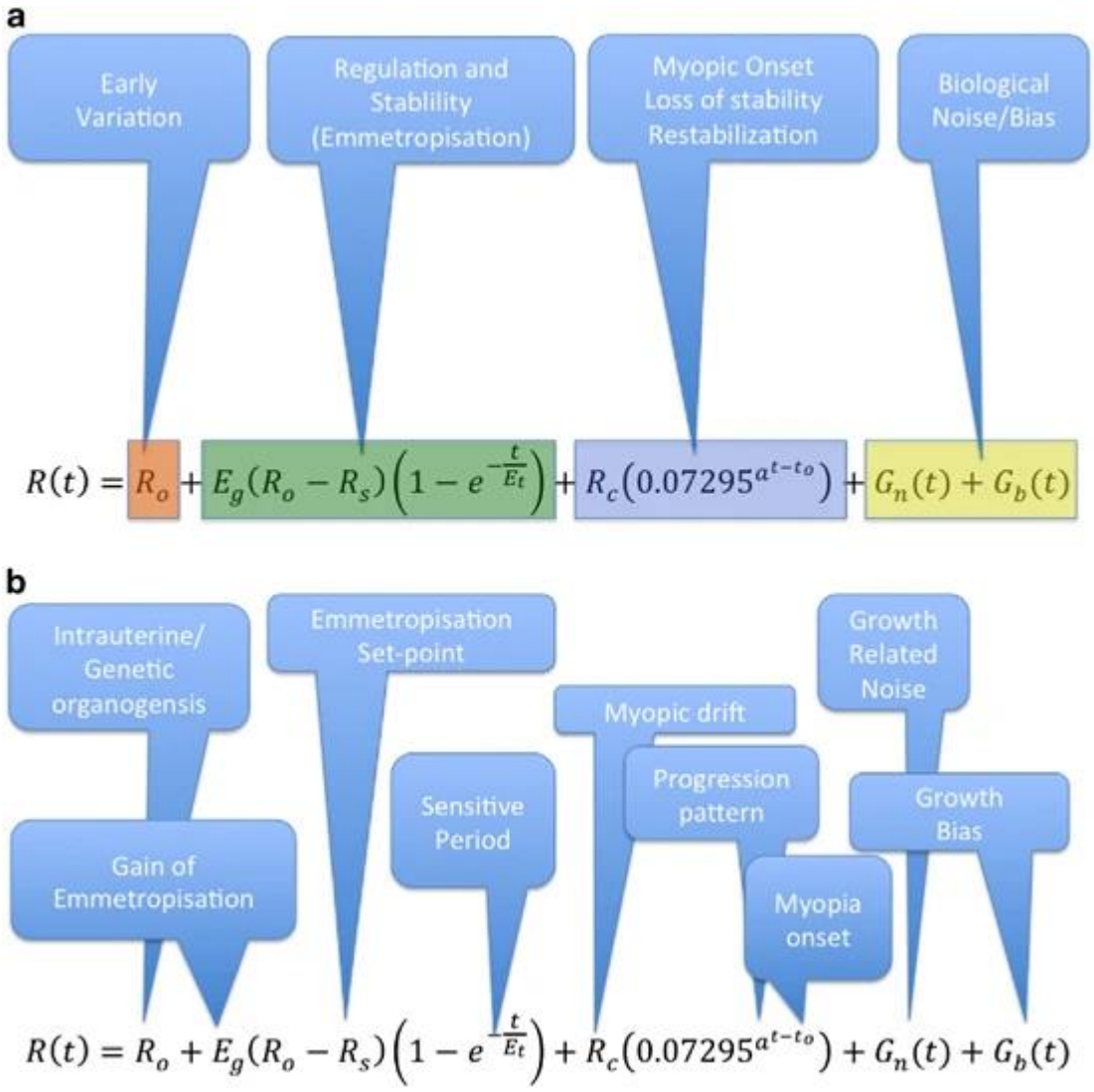

Figure 6.7 Annotation of an equation describing the biological mechanisms associated with each component of the model (a) and annotation of an equation describing the biological relevance of each component parameter (b). Taken from: Flitcroft, D. Emmetropisation and the aetiology of refractive errors. Eye ***28,*** *169–179 (2014).* https://doi.org/10.1038/eye.2013.276

In conclusion, this study suggested that the causal effect of refractive error on schooling is non-linear. However it revealed no evidence of a major causal role of hyperopia in reducing educational attainment, although statistical power was limited to detecting effects of half a year or more of schooling per Dioptre of hyperopia.

Chapter 7. General Discussion and Future Work

A diverse series of hypotheses have been tested in this thesis in order to better understand the aetiology of paediatric eye disorders. In this chapter, I review my findings and discuss the implications of the work and how they interrelate.

7.1 Genetic Predisposition to Strabismus

A novel genetic locus conferring susceptibility to strabismus was identified, supporting the hypothesis that commonly occurring polymorphisms increase the risk of non-syndromic strabismus. Over the past 2-3 decades, research into the genetics of strabismus has lagged behind that of ocular traits such as refractive error and astigmatism, possibly because strabismus is comparatively rare and is a challenging trait to diagnose and quantify. Prior efforts focusing on linkage analysis, coupled with the occurrence of strabismus as a feature of numerous disease syndromes, may have led ophthalmic researchers to regard strabismus as being primarily monogenic. My results suggest a shift in perspective may be needed, since the locus I discovered on 17q25.3 supports an oligogenic or polygenic contribution to strabismus. The discovery of a locus with such a large effect (in an analysis of a comparatively small number of strabismus cases) implies that additional strabismus susceptibility loci may exist. Genetic discoveries, such as the identification of the locus on chromosome 17 that confers a more than 50% increased risk, offer researchers a completely new avenue of investigation into the aetiology of the disorder.

Nevertheless, despite its importance, the identification of the 17q25.3 genetic locus does not immediately shed light on the functional mechanism underlying the increased risk of strabismus. This is because the strabismus-associated causal variant at the locus could not be pin-pointed exactly. Fine-mapping and functional prediction highlighted the *TSPAN10* gene as the most likely candidate. *TSPAN10* is a member of a tetraspanin gene family, which encodes a series of cell-surface proteins containing 4 trans-membrane domains, forming 2 (1 short and 1 long) extracellular loops. Tetraspanins directly and indirectly interact with a broad spectrum of transmembrane molecules: other tetraspanins, integrins, cytokine receptors, immunoglobulins and proteolytic enzymes, and are instrumental in forming tetraspanin-enriched micro-domains with diverse functions (402-404). Members of the tetraspanin family function as molecular scaffolds and are involved in such processes as growth of the organism, immune response and reproduction

(405-408). A total of 33 tetraspanins have been identified in mammals, 37 in Drosophila melanogaster (403), approximately 50 members of tetraspanin family in zebrafish (409) with 22 of them being expressed in embryos and suggested to have an effect on zebrafish development. Some of the tetraspanins show cell and organ specificity (410). Tetraspanin-10, also named *Oculospanin,* was discovered in cDNA libraries from retinal pigment epithelium and choroid tissues (411). It belongs to the TSpanC8 subfamily (tetraspanins with 8 cysteine residues in their large extracellular domain). TSpanC8 tetraspanins are evolutionary conserved and regulate the trafficking and function of metalloproteinase ADAM10 (412), which plays an important role in Notch signalling and is essential for the cell differentiation processes through the cleavage of Notch by γ-secretase (413). Polymorphic variant rs6420484 causes a C177Y substitution in *TSPAN10* (Section 4.3.4); however, the cysteine at position 177 is not located in the large extracellular loop, suggesting that the SNP would *not* directly impair γ-secretase cleavage activity.

TSPAN22 (also known as *PRPH2* or *RDS*) and *TSPAN23* (also known as *ROM1*) are expressed in the photoreceptors of the retina and in the rim of the rod outer segment disks; mutations in these photoreceptor-specific genes were reported to cause retinitis pigmentosa (414). In an animal study, *TSPAN12* (*EVR5*) mutant mice exhibited a phenotype similar to the familial exudative vitreoretinopathy (FEVR) mice (415); investigation in humans supported the hypothesis that mutations in *TSPAN12* are causal for FEVR development.

Further studies of *TSPAN10* and more broadly, the 17q25.3 locus, should provide additional information regarding the biological and physiological mechanisms of strabismus development.

7.2 Causality in the Relationship Between Environmental Risk Factors and Myopia

Two studies in this thesis investigated the causality of the relationship between refractive error and: (1) education, and (2) birth weight.

Educational attainment is among the environmental factors showing the strongest effects on refractive error development (Section 1.2.4.1.1). For instance, there is a 4-fold higher likelihood of myopia development in individuals with a university degree vs. those with only primary level education (49). The robust association

between educational attainment and refractive error has been established in numerous observational studies (43, 416, 417), with 2 MR studies testing this relationship for *causality* (45, 46). Both of these MR studies, despite being less prone to bias than standard observational approaches, have limitations. The SNP used as instrumental variables in the MR setting were strongly associated with the exposure (education), but knowledge of their functional effects was extremely limited. Hence, the possibility of SNP having an effect on the outcome via 2 or more independent pathways (horizontal pleiotropy; Section 2.4.4) is highly plausible; this can lead to a biased estimate of the true causal effect. Sensitivity analysis such as the MR-Egger test (Section 2.4.4) allow the analyst to test for unbalanced horizontal pleiotropy; however, there are several issues (e.g. influence of outlying genetic variants) leading to a biased MR-Egger estimate. Thus, “triangulation”, which in epidemiological research typically refers to the use of different sources of information or different scientific approaches in order to increase the validity of the estimate (418, 419) is arguably the best way to overcome the limitations mentioned above. The regression discontinuity study design (Section 2.4.5) indeed does have a different source of bias compared to MR and it is appropriate to triangulate them to answer the same causal question. The RD design has been used to assess the effect of education on adult health outcomes and mortality in the UK (371, 420) including a recent study in UK Biobank participants (372); with, in general, a positive effect of additional education on health being found. The RD study in this thesis has reported a *negative* (adverse) effect of the ROSLA 1972 educational reform on refractive error; this finding is in agreement with observational and MR studies. However, the interpretation of the results of an RD study is not straightforward; the estimated local average treatment effect (302)(Section 2.4.5) can be interpreted as the effect on those affected by the reform (those who remained at school due to the reform who otherwise would have left school). This low external validity makes it difficult to generalise the results to the general population. Despite providing a *causal* effect estimate, neither the MR or RD analyses allow one to disentangle the exact biological mechanism(s) underlying the estimated effect. I speculate that time outdoors and near work activities mediate the effect of educational attainment on myopia. Higher rates of incidence and progression of myopia in children from economically developed South-East Asian countries coincide with the reported decrease in time spent outdoors (47, 58, 421-423); with time outdoors being reported as a protective factor for myopia development (424). One plausible pathway linking time outdoors with negative refractive error is light intensity, which is typically higher outdoors (425). Increased amounts of near work may also mediate

the effect of education on myopia via excessive accommodation; however animal models (426, 427) have not confirmed the hypothesis of a direct effect of accommodation on myopia, and the results of human studies are conflicting (50, 53, 56).

The RD study performed in the current thesis reported the effect of the additional year of schooling due to ROSLA 1972 on those affected by the reform; this finding adds evidence of there being a *causal* effect of education on myopia.

Refractive errors may have a strong influence on the learning process. The increased use of electronic devices and increasing duration of class activities lead to excessive pressure on the accommodative-vergence system in school children, causing numerous adverse effects (Section 1.4.3). Children with hyperopic refraction errors, especially uncorrected hyperopes are more vulnerable to these effects, which could be expected to have an adverse impact on learning. Such an adverse effect of hyperopia on educational performance has been reported in observational epidemiological studies (225) (214); with the only one attempt to establish the causality in the relationship between refractive error and educational attainment (to test the hypothesis that each additional Dioptre in refraction is *causally* associated with change in time spent in education) (46). The latter MR study was performed under the assumption of *linearity* in the relationship between refractive error and educational attainment and found no causal effect. However, my non-linear MR study (Chapter 6) explored the possibility that the association between the exposure and the outcome was non-linear. Whilst a linear MR analysis estimates a causal effect which is the averaged effect in the population, it still can be used in the context of a non-linear exposure-outcome relationship when the condition of monotonicity is met (392). In other words, when the direction in the change of outcome is the same across the exposure distribution (more or less years spent in education for all values of refractive error). However, if the assumption of monotonicity is not met, the estimated effect will be biased. The non-linear MR in the current thesis assessed the shape of the exposure-outcome relationship and tested this relationship for linearity using Cochran's Q statistics. My analysis supported the hypothesis of a *non-linear* relationship between refractive error and years spent in education; whereas the null hypothesis of *no causal effect* of hyperopia on educational attainment could not be fully rejected. In particular, I identified a small causal effect of *myopia* on education of -0.15 years/D (95% CI -0.23 to -0.07). This result implies that children who have a low degree of myopia

are more likely to continue in schooling than would otherwise be the case. I speculate that the mechanism underlying this relationship may involve psychological effects, e.g. children wearing glasses may be relatively keener to study and relatively less keen to participate in sports or other non-scholarly activities.

The non-linear MR analysis was performed with stratification of the sample based on residual refractive error ('genetic risk-free'; Section 6.2.3) with subsequent estimation of the association between the exposure and the outcome in each stratum. Individual-level data are required to assess this genetic risk-free relationship (i.e. a one-sample MR analysis; Section 2.4.4). Given the limited statistical power for the current non-linear MR to detect a causal effect, it may prove fruitful to perform such an analysis in a larger sample. However, sharing of individual level data across different studies is challenging for data privacy reasons, and meta-analysis across sites would be difficult because of inconsistency in strata boundaries and also because of potential differences in exposure-outcome relationships in different countries.

Despite a steady increase in the number of epidemiological studies using MR analysis to assess the causality of exposure-outcome relationships, non-linear MR has not been widely adopted. As regards research in the vision sciences there have been few MR studies of any kind (428). The MR study in the current thesis (Chapter 3) reported a small positive causal effect of birth weight on refractive error; however, it had limited statistical power, meaning that the estimation was imprecise. The self-reported nature of the birth weight variable could have influenced the estimate; however, if the exposure is non-differentially measured with error it typically does not bias the estimate in the IV settings (429). The MR-Egger sensitivity analyses showed no evidence for unbalanced pleiotropy, although it is not possible to completely rule it out (as the biological mechanisms of the chosen genetic variants influencing birth weight are not clear). A more biologically-based approach (430) to conducting an MR study with multiple instruments (e.g. the use of genetic variants from genes with known biological effects on the exposure), or use of pathway-specific PRS (when PRS are calculated using SNP associated with a specific molecular pathway) (431) could gain more information on the variants selected as instruments.

7.3 Future Work

The findings from the studies conducted in this thesis pave the way for further research in the field of causal epidemiology in vision sciences.

Animal experiments may shed light on the role of *TSPAN10* in strabismus and amblyopia. For example, it would be feasible to generate a mouse mutant model with a disease-carrying mutation with subsequent assessment of the visual functions and the morphology of the eye. However, despite of the widespread use of mouse models, the limited binocularity of mice makes them a poor animal model for strabismus. Non-human primates are considered to be a better model to investigate human diseases as they are more closely related to humans (432). Due to similarities in genetics, developmental biology, anatomy and developed binocularity, primates can be used to investigate the oculomotor and/or neural visual disruption associated with strabismus. To determine the mechanism downstream of the rs75078292 genetic variant I would suggest making 3 lines of gene knock-out monkeys: a *TSPAN10*-KO line, a *NPLOC4*-KO line and a *PDE6G*-KO line. I would carry out an assessment of tropia, VA, refractive error, RNFL thickness in each KO line (by measuring cover test, optokinetic response, cycloplegic autorefraction, and OCT).

Possible mechanisms of strabismus development in the animal study proposed above would include but would not be limited to: 1) the role of tetraspanins in synapse function, 2) the role of tetraspanins in axon guidance.

For the first mechanism, the regulation of synaptic protein function by tetraspanin has been reported recently (433-435). Tetraspanins, acting through their scaffold protein properties, are able to interact with various molecules including integrins; the latter proteins are known to participate in synaptic function (436). One of the members of tetraspanin family, TSPAN7, has been reported to regulate glutamatergic and dopaminergic synaptic function via associations with the immunoglobulin-like superfamily (IGSF) member IGSF3 (437), PICK1 protein (433), and β1-integrin (435). Glutamatergic and dopaminergic signalling systems play an important role in the functioning of the central nervous system (438) and in vision (439-441). As I mentioned in Section 7.1, TSpanC8 tetraspanins are known to regulate the trafficking and function of ADAM10 (412). Cleavage of some of the proteins (e.g. N-cadherin) involved in synapse formation and/or function is regulated

by ADAM10 (442). I speculate that the excitatory/inhibitory synaptic disbalance could be the reason for binocular dysfunction.

Next, there is a growing body of evidence that mutations in genes regulating axon guidance in the ocular-motor system result in the development of eye movement disorders (443) and may be associated with strabismus (444). The role of tetraspanin family members in axon guidance and axonal growth has been hypothesized (445) based on the expression of CD9 and the *Drosophila* protein *late bloomer* (*lbl*) in motoneurons (446, 447) and the ability of tetraspanins to interact with integrins; they, in turn are promotors of axonal growth.

That would help to decipher molecular pathways of strabismus and, potentially, amblyopia development and would open new perspectives in the treatment of these eye conditions. Also the finding of the proposed study may demonstrate gene-environment interaction in strabismus development.

Results of the GWAS conducted in this thesis confirmed a hypothesis that commonly occurring genetic variants are associated with the increased risk of strabismus. Given the limitations of this study (e.g. unbalanced case-control ratio, self-reported phenotype in an adult cohort) it might prove fruitful to carry out a case-control association study in a paediatric cohort. It would require a recruiting a sample of 2,000 children: 1,000 assigned to the case cohort; and 1,000 assigned to the control cohort. Assignment would based on the results of careful phenotyping, including strabismus assessment performed by an orthoptist. For obtaining the genetic data, I would ideally use the whole genome sequencing (WGS) as it allows to expand the horizons of the research and to test not only common variants, but also rare and low-frequency variants and variants not in LD with SNP in the genotyping array. However, the hypothesis could be tested with genotyping technologies provided by two genotyping companies: Affymetrix and Illumina; which are typically used in genetics studies.

References

1. Dunham I, Kundaje A, Aldred SF, Collins PJ, Davis CA, Doyle F, et al. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012;489(7414):57-74.

2. Wang DG, Fan J-B, Siao C-J, Berno A, Young P, Sapolsky R, et al. Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. Science. 1998;280(5366):1077.

3. Mills RE, Luttig CT, Larkins CE, Beauchamp A, Tsui C, Pittard WS, et al. An initial map of insertion and deletion (INDEL) variation in the human genome. Genome Research. 2006;16(9):1182-90.

4. Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, et al. Global variation in copy number in the human genome. Nature. 2006;444(7118):444-54.

5. Plomin R, Haworth CMA, Davis OSP. Common disorders are quantitative traits. Nature Reviews Genetics. 2009;10(12):872-8.

6. Lander ES, Schork NJ. Genetic dissection of complex traits. Science. 1994;265(5181):2037.

7. Kemper KE, Goddard ME. Understanding and predicting complex traits: knowledge from cattle. Human Molecular Genetics. 2012;21(R1):R45-R51.

8. Falconer DS. Introduction to quantitative genetics. Harlow, England: Prentice Hall; 1996.

9. Visscher PM, Hill WG, Wray NR. Heritability in the genomics era — concepts and misconceptions. Nature Reviews Genetics. 2008;9(4):255-66.

10. Bulmer MG. The mathematical theory of quantitative genetics: Clarendon Press.; 1980.

11. Hill WG, Goddard ME, Visscher PM. Data and theory point to mainly additive genetic variance for complex traits. PLoS Genetics. 2008;4(2):e1000008-e.

12. Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, et al. Common SNPs explain a large proportion of the heritability for human height. Nature Genetics. 2010;42(7):565-9.

13. Yang J, Zeng J, Goddard ME, Wray NR, Visscher PM. Concepts, estimation and interpretation of SNP-based heritability. Nature Genetics. 2017;49(9):1304-10.

14. Grosvenor T. A review and a suggested classification system for myopia on the basis of age-related prevalence and age of onset. American Journal of Optometry and Physiological Optics. 1987;64(7):545-54.

15. Fredrick DR. Myopia. BMJ (Clinical research ed). 2002;324(7347):1195-9.

16. Flitcroft DI, He M, Jonas JB, Jong M, Naidoo K, Ohno-Matsui K, et al. IMI–Defining and classifying myopia: a proposed set of standards for clinical and epidemiologic studies. Investigative Ophthalmology & Visual Science. 2019;60(3):M20-M30.

17. Goss D, Eskridge J. Myopia. In: Amos J, editor. Diagnosis and management in vision care. Boston: Butterworths; 1987. p. 121-71.

18. Pesudovs K, Elliott DB. Refractive error changes in cortical, nuclear, and posterior subcapsular cataracts. British Journal of Ophthalmology. 2003;87(8):964.

19. Furushima M, Imaizumi M, Nakatsuka K. Changes in Refraction Caused by Induction of Acute Hyperglycemia in Healthy Volunteers. Japanese Journal of Ophthalmology. 1999;43(5):398-403.

20. Troilo D, Wallman J. The regulation of eye growth and refractive state: An experimental study of emmetropization. Vision Research. 1991;31(7–8):1237-50.

21. Gwiazda J, Thorn F, Bauer J, Held R. Emmetropization and the progression of manifest refraction in children followed from infancy to puberty. Clinical Vision Sciences. 1993;8(4):337-44.

22. Morgan IG. The biological basis of myopic refractive error. Clinical and Experimental Optometry. 2003;86(5):276-88.

23. Mutti DO, Mitchell GL, Jones LA, Friedman NE, Frane SL, Lin WK, et al. Axial growth and changes in lenticular and corneal power during emmetropization in infants. Investigative Ophthalmology & Visual Science. 2005;46(9):3074-80.

24. Pennie FC, Wood ICJ, Olsen C, White S, Charman WN. A longitudinal study of the biometric and refractive changes in full-term infants during the first year of life. Vision Research. 2001;41(21):2799-810.

25. Meng W, Butterworth J, Malecaze F, Calvas P. Axial length of myopia: A review of current research. Ophthalmologica. 2011;225(3):127-34.

26. Smith EL, III, Hung LF, Arumugam B. Visual regulation of refractive development: insights from animal studies. Eye. 2014;28(2):180-8.

27. Bourne RRA, Stevens GA, White RA, Smith JL, Flaxman SR, Price H, et al. Causes of vision loss worldwide, 1990–2010: a systematic analysis. The Lancet Global Health. 2013;1(6):e339-e49.

28. Holden BA, Fricke TR, Wilson DA, Jong M, Naidoo KS, Sankaridurg P, et al. Global prevalence of myopia and high myopia and temporal trends from 2000 through 2050. Ophthalmology. 2016;123(5):1036-42.

29. Smith TST, Frick KD, Holden BA, Fricke TR, Naidoo KS. Potential lost productivity resulting from the global burden of uncorrected refractive error. Bulletin of the World Health Organization. 2009;87(6):431-7.

30. Foster PJ, Jiang Y. Epidemiology of myopia. Eye. 2014;28(2):202-8.

31. Rosenfield M, Gilmartin B, Goldschmidt E. Myopia and nearwork. Oxford: Butterworth-Heinemann; 1998.

32. Wu LJ, You QS, Duan JL, Luo YX, Liu LJ, Li X, et al. Prevalence and associated factors of myopia in high-school students in Beijing. PLOS ONE. 2015;10(3):e0120764.

33. Sawada A, Tomidokoro A, Araie M, Iwase A, Yamamoto T. Refractive errors in an elderly Japanese population: The Tajimi study. Ophthalmology. 2008;115(2):363-70.e3.

34. Vitale S, Sperduto RD, Ferris FL, Iii. Increased prevalence of myopia in the united states between 1971-1972 and 1999-2004. Archives of Ophthalmology. 2009;127(12):1632-9.

35. Logan NS, Shah P, Rudnicka AR, Gilmartin B, Owen CG. Childhood ethnic differences in ametropia and ocular biometry: the Aston Eye Study. Ophthalmic and Physiological Optics. 2011;31(5):550-8.

36. Donoghue L, McClelland JF, Logan NS, Rudnicka AR, Owen CG, Saunders KJ. Refractive error and visual impairment in school children in Northern Ireland. British Journal of Ophthalmology. 2010;94(9):1155.

37. Williams KM, Bertelsen G, Cumberland P, Wolfram C, Verhoeven VJM, Anastasopoulos E, et al. Increasing prevalence of myopia in Europe and the impact of education. Ophthalmology. 2015;122(7):1489-97.

38. Saw S-M, Shankar A, Tan S-B, Taylor H, Tan DT, Stone RA, et al. A cohort study of incident myopia in Singaporean children. Investigative Ophthalmology & Visual Science. 2006;47(5):1839-44.

39. Jung S-K, Lee JH, Kakizaki H, Jee D. Prevalence of myopia and its association with body stature and educational level in 19-year-old male conscripts in Seoul, South Korea. Investigative Ophthalmology & Visual Science. 2012;53(9):5579-83.

40. Harman NB. The Education of high myopes. SAGE Publications; 1913.

41. Holm E. Investigations of myopia in Danish secondary schools. Acta Ophthalmologica. 1925;3(1-2):121-30.

42. H C. The hygiene of the eye in schools. (Transl. and ed W.P. Turnbull). London: Simkin, Marshall & Co.; 1886 (Originally published in German, 1983).

43. Eong KGA, Tay TH, Lim MK. Education and myopia in 110,236 young Singaporean males. Singapore Medical Journal. 1993;34(6):489-92.

44. Cumberland PM, Bao Y, Hysi PG, Foster PJ, Hammond CJ, Rahi JS, et al. Frequency and distribution of refractive error in adult life: Methodology and findings of the UK Biobank study. PLOS ONE. 2015;10(10):e0139780.

45. Cuellar-Partida G, Lu Y, Kho PF, Hewitt AW, Wichmann HE, Yazar S, et al. Assessing the genetic predisposition of education on myopia: a Mendelian randomization study. Genetic Epidemiology. 2016;40(1):66-72.

46. Mountjoy E, Davies NM, Plotnikov D, Smith GD, Rodriguez S, Williams CE, et al. Education and myopia: assessing the direction of causality by Mendelian randomisation. BMJ. 2018;361.

47. French AN, Ashby RS, Morgan IG, Rose KA. Time outdoors and the prevention of myopia. Experimental Eye Research. 2013;114:58-68.

48. Ip JM, Saw S-M, Rose KA, Morgan IG, Kifley A, Wang JJ, et al. Role of near work in myopia: Findings in a sample of Australian school children. Investigative Ophthalmology & Visual Science. 2008;49(7):2903-10.

49. Morgan I, Rose K. How genetic is school myopia? Progress in Retinal and Eye Research. 2005;24(1):1-38.

50. Mutti DO, Mitchell GL, Moeschberger ML, Jones LA, Zadnik K. Parental myopia, near work, school achievement, and children's refractive error. Investigative Ophthalmology & Visual Science. 2002;43(12):3633-40.

51. Hepsen IF, Evereklioglu C, Bayramlar H. The effect of reading and near-work on the development of myopia in emmetropic boys: a prospective, controlled, three-year follow-up study. Vision Research. 2001;41(19):2511-20.

52. Duke-Elder S, Abrams D. Ophthalmic optics and refraction: Mosby; 1970.

53. Guo L, Yang J, Mai J, Du X, Guo Y, Li P, et al. Prevalence and associated factors of myopia among primary and middle school-aged students: a school-based study in Guangzhou. Eye (London, England). 2016;30(6):796-804.

54. Woodman EC, Read SA, Collins MJ, Hegarty KJ, Priddle SB, Smith JM, et al. Axial elongation following prolonged near work in myopes and emmetropes. British Journal of Ophthalmology. 2011;95(5):652.

55. Hsu C-C, Huang N, Lin P-Y, Fang S-Y, Tsai D-C, Chen S-Y, et al. Risk factors for myopia progression in second-grade primary school children in Taipei: a population-based cohort study. British Journal of Ophthalmology. 2017;101(12):1611.

56. Low W, Dirani M, Gazzard G, Chan Y-H, Zhou H-J, Selvaraj P, et al. Family history, near work, outdoor activity, and myopia in Singapore Chinese preschool children. British Journal of Ophthalmology. 2010;94(8):1012.

57. Huang H-M, Chang DS-T, Wu P-C. The Association between near work activities and myopia in children-A systematic review and meta-analysis. PLOS ONE. 2015;10(10):e0140419-e.

58. Rose KA, Morgan IG, Ip J, Kifley A, Huynh S, Smith W, et al. Outdoor Activity Reduces the Prevalence of Myopia in Children. Ophthalmology. 2008;115(8):1279-85.

59. He M, Xiang F, Zeng Y, et al. Effect of time spent outdoors at school on the development of myopia among children in china: A randomized clinical trial. JAMA. 2015;314(11):1142-8.

60. Jin J-X, Hua W-J, Jiang X, Wu X-Y, Yang J-W, Gao G-P, et al. Effect of outdoor activity on myopia onset and progression in school-aged children in northeast China: the Sujiatun Eye Care Study. BMC Ophthalmology. 2015;15:73-.

61. Guggenheim JA, Northstone K, McMahon G, Ness AR, Deere K, Mattocks C, et al. Time outdoors and physical activity as predictors of incident myopia in childhood: a prospective cohort study. Investigative Ophthalmology & Visual Science. 2012;53(6):2856-65.

62. Guggenheim JA, Williams C, Northstone K, Howe LD, Tilling K, St Pourcain B, et al. Does vitamin D mediate the protective effects of time outdoors on myopia? Findings from a prospective birth cohort. Investigative Ophthalmology & Visual Science. 2014;55(12):8550-8.

63. Wu P-C, Chen C-T, Lin K-K, Sun C-C, Kuo C-N, Huang H-M, et al. Myopia prevention and outdoor light intensity in a school-based cluster randomized trial. Ophthalmology. 2018;125(8):1239-50.

64. Read SA, Collins MJ, Vincent SJ. Light exposure and eye growth in childhood. Investigative Ophthalmology & Visual Science. 2015;56(11):6779-87.

65. Cohen Y, Belkin M, Yehezkel O, Solomon AS, Polat U. Dependency between light intensity and refractive development under light–dark cycles. Experimental Eye Research. 2011;92(1):40-6.

66. Karouta C, Ashby RS. Correlation between light levels and the development of deprivation myopia. Investigative Ophthalmology & Visual Science. 2015;56(1):299-309.

67. Wang M, Schaeffel F, Jiang B, Feldkaemper M. Effects of light of different spectral composition on refractive development and retinal dopamine in chicks. Investigative Ophthalmology & Visual Science. 2018;59(11):4413-24.

68. Qian Y-F, Dai J-H, Liu R, Chen M-J, Zhou X-T, Chu R-Y. Effects of the chromatic defocus caused by interchange of two monochromatic lights on refraction and ocular dimension in guinea pigs. PLOS ONE. 2013;8(5):e63229.

69. Ward AH, Norton TT, Huisingh CE, Gawne TJ. The hyperopic effect of narrow-band long-wavelength light in tree shrews increases non-linearly with duration. Vision Research. 2018;146-147:9-17.

70. Cordain L, Eaton SB, Brand Miller J, Lindeberg S, Jensen C. An evolutionary analysis of the aetiology and pathogenesis of juvenile-onset myopia. Acta Ophthalmologica Scandinavica. 2002;80(2):125-35.

71. Jacobsen N, Jensen H, Lund-Andersen H, Goldschmidt E. Is poor glycaemic control in diabetic patients a risk factor of myopia? Acta Ophthalmologica. 2008;86(5):510-4.

72. Fledelius HC, Miyamoto K. Diabetic myopia - is it lens-induced? Acta Ophthalmologica. 1987;65(4):469-73.

73. Løgstrup N, Sjølie AK, Kyvik KO, Green A. Long-term influence of insulin dependent diabetes mellitus on refraction and its components: a population based twin study. The British journal of ophthalmology. 1997;81(5):343-9.

74. Mäntyjärvi M. Myopia and diabetes: A review. Acta Ophthalmologica. 1988;66(S185):82-5.

75. Edwards MH, Leung SSF, Lee WTK. Do variations in normal nutrition play a role in the development of myopia? Optometry and Vision Science. 1996;73(10):638-43.

76. Teikari JM, Kaprio J, Koskenvuo MK, Vannas A, Rao DC. Heritability estimate for refractive errors—a population-based sample of adult twins. Genetic Epidemiology. 1988;5(3):171-81.

77. Sanfilippo PG, Hewitt AW, Hammond CJ, Mackey DA. The heritability of ocular traits. Survey of Ophthalmology. 2010;55(6):561-83.

78. Lopes MC, Andrew T, Carbonaro F, Spector TD, Hammond CJ. Estimating heritability and shared environmental effects for refractive error in twin and family studies. Investigative Ophthalmology & Visual Science. 2009;50(1):126-31.

79. Dirani M, Chamberlain M, Shekar SN, Islam AFM, Garoufalis P, Chen CY, et al. Heritability of refractive error and ocular biometrics: The Genes in Myopia (GEM) twin study. Investigative Ophthalmology & Visual Science. 2006;47(11):4756-61.

80. Hysi PG, Choquet H, Khawaja AP, Wojciechowski R, Tedja MS, Yin J, et al. Meta-analysis of 542,934 subjects of European ancestry identifies new genes and mechanisms predisposing to refractive error and myopia. Nature Genetics. 2020;52(4):401-7.

81. Guggenheim JA, St Pourcain B, McMahon G, Timpson NJ, Evans DM, Williams C. Assumption-free estimation of the genetic contribution to refractive error across childhood. Molecular Vision. 2015;21:621-32.

82. Shah RL, Guggenheim JA, Eye UB, Consortium V. Genome-wide association studies for corneal and refractive astigmatism in UK Biobank demonstrate a shared role for myopia susceptibility loci. Human Genetics. 2018;137(11-12):881-96.

83. Schwartz M, Haim M, Skarsholm D. X-linked myopia: Bornholm eye disease: linkage to DNA markers on the distal part of Xq. Clinical Genetics. 1990;38(4):281-6.

84. Guo X, Xiao X, Li S, Wang P, Jia X, Zhang Q. Nonsyndromic high myopia in a Chinese family mapped to MYP1: linkage confirmation and phenotypic characterization. Archives of Ophthalmology. 2010;128(11):1473-9.

85. Ratnamala U, Lyle R, Rawal R, Singh R, Vishnupriya S, Himabindu P, et al. Refinement of the X-linked nonsyndromic high-grade myopia locus MYP1 on Xq28 and exclusion of 13 known positional candidate genes by direct sequencing. Investigative Ophthalmology & Visual Science. 2011;52(9):6814-9.

86. Orosz O, Rajta I, Vajas A, Takács L, Csutak A, Fodor M, et al. Myopia and late-onset progressive cone dystrophy associate to LVAVA/MVAVA exon 3 interchange haplotypes of Opsin genes on Chromosome X. Investigative Ophthalmology & Visual Science. 2017;58(3):1834-42.

87. Young TL, Ronan SM, Drahozal LA, Wildenberg SC, Alvear AB, Oetting WS, et al. Evidence that a locus for familial high myopia maps to chromosome 18p. The American Journal of Human Genetics. 1998;63(1):109-19.

88. Young TL, Ronan SM, Alvear AB, Wildenberg SC, Oetting WS, Atwood LD, et al. A second locus for familial high myopia maps to chromosome 12q. The American Journal of Human Genetics. 1998;63(5):1419-24.

89. Wang B, Liu Y, Chen S, Wu Y, Lin S, Duan Y, et al. A novel potentially causative variant of NDUFAF7 revealed by mutation screening in a Chinese family with pathologic myopia. Investigative Ophthalmology & Visual Science. 2017;58(10):4182-92.

90. Okui S, Meguro A, Takeuchi M, Yamane T, Okada E, Iijima Y, et al. Analysis of the association between the LUM rs3759223 variant and high myopia in a Japanese population. Clinical Ophthalmology (Auckland, NZ). 2016;10:2157.

91. Stambolian D, Ibay G, Reider L, Dana D, Moy C, Schlifka M, et al. Genomewide linkage scan for myopia susceptibility loci among Ashkenazi Jewish families shows evidence of linkage on chromosome 22q12. American Journal of Human Genetics. 2004;75(3):448-59.

92. Klein AP, Duggal P, Lee KE, Klein R, Bailey-Wilson JE, Klein BEK. Confirmation of linkage to ocular refraction on Chromosome 22q and identification of a novel linkage region on 1q. Archives of Ophthalmology. 2007;125(1):80-5.

93. Hammond CJ, Andrew T, Tat Mak Y, Spector TD. A susceptibility locus for myopia in the normal population is linked to the PAX6 gene region on chromosome 11: A genomewide scan of dizygotic twins. The American Journal of Human Genetics. 2004;75(2):294-304.

94. Wojciechowski R, Moy C, Ciner E, Ibay G, Reider L, Bailey-Wilson JE, et al. Genomewide scan in Ashkenazi Jewish families demonstrates evidence of linkage of ocular refraction to a QTL on chromosome 1p36. Human Genetics. 2006;119(4):389-99.

95. Li Y-J, Guggenheim JA, Bulusu A, Metlapally R, Abbott D, Malecaze F, et al. An international collaborative family-based whole-genome linkage scan for high-grade myopia. Investigative Ophthalmology & Visual Science. 2009;50(7):3116-27.

96. Naiglin L, Gazagne C, Dallongeville F, Thalamas C, Idder A, Rascol O, et al. A genome wide scan for familial high myopia suggests a novel locus on chromosome 7q36. Journal of Medical Genetics. 2002;39(2):118.

97. Paget S, Julia S, Vitezica ZG, Soler V, Malecaze F, Calvas P. Linkage analysis of high myopia susceptibility locus in 26 families. Molecular Vision. 2008;14:2566-74.

98. Ciner E, Wojciechowski R, Ibay G, Bailey-Wilson JE, Stambolian D. Genomewide scan of ocular refraction in African-American families shows significant linkage to chromosome 7p15. Genetic Epidemiology. 2008;32(5):454-63.

99. Xiao X, Li S, Jia X, Guo X, Zhang Q. X-linked heterozygous mutations in ARR3 cause female-limited early onset high myopia. Molecular Vision. 2016;22:1257-66.

100. Ouyang J, Sun W, Xiao X, Li S, Jia X, Zhou L, et al. CPSF1 mutations are associated with early-onset high myopia and involved in retinal ganglion cell axon projection. Human Molecular Genetics. 2019;28(12):1959-70.

101. Young TL, Atwood LD, Ronan SM, Dewan AT, Alvear AB, Peterson J, et al. Further refinement of the MYP2 locus for autosomal dominant high myopia by linkage disequilibrium analysis. Ophthalmic Genetics. 2001;22(2):69-75.

102. Lin HJ, Wan L, Tsai Y, Chen WC, Tsai SW, Tsai FJ. The association between lumican gene polymorphisms and high myopia. Eye. 2010;24(6):1093-101.

103. Paluru P, Ronan SM, Heon E, Devoto M, Wildenberg SC, Scavello G, et al. New Locus for Autosomal Dominant High Myopia Maps to the Long Arm of Chromosome 17. Investigative Ophthalmology & Visual Science. 2003;44(5):1830-6.

104. Zhang Q, Guo X, Xiao X, Jia X, Li S, Hejtmancik J. A new locus for autosomal dominant high myopia maps to 4q22-q27 between D4S1578 and D4S1612. Molecular Vision. 2005;11:554-60.

105. Paluru PC, Nallasamy S, Devoto M, Rappaport EF, Young TL. Identification of a Novel Locus on 2q for Autosomal Dominant High-Grade Myopia. Investigative Ophthalmology & Visual Science. 2005;46(7):2300-7.

106. Zhang Q, Guo X, Xiao X, Jia X, Li S, Hejtmancik JF. Novel locus for X linked recessive high myopia maps to Xq23–q25 but outside MYP1. Journal of Medical Genetics. 2006;43(5):e20.

107. Zhang Q, Li S, Xiao X, Jia X, Guo X. Confirmation of a genetic locus for X-linked recessive high myopia outside MYP1. Journal of Human Genetics. 2007;52(5):469-72.

108. Nallasamy S, Paluru PC, Devoto M, Wasserman NF, Zhou J, Young TL. Genetic linkage study of high-grade myopia in a Hutterite population from South Dakota. Molecular Vision. 2007;13:229-36.

109. Lam CY, Tam POS, Fan DSP, Fan BJ, Wang DY, Lee CWS, et al. A Genome-wide Scan Maps a Novel High Myopia Locus to 5p15. Investigative Ophthalmology & Visual Science. 2008;49(9):3768-78.

110. Yang Z, Xiao X, Li S, Zhang Q. Clinical and linkage study on a consanguineous Chinese family with autosomal recessive high myopia. Molecular Vision. 2009;15:312-8.

111. Ma J-H, Shen S-H, Zhang G-W, Zhao D-S, Xu C, Pan C-M, et al. Identification of a locus for autosomal dominant high myopia on chromosome 5p13.3-p15.1 in a Chinese family. Molecular Vision. 2010;16:2043-54.

112. Shi Y, Qu J, Zhang D, Zhao P, Zhang Q, Tam POS, et al. Genetic variants at 13q12.12 are associated with high myopia in the Han Chinese population. American Journal of Human Genetics. 2011;88(6):805-13.

113. Shi Y, Li Y, Zhang D, Zhang H, Li Y, Lu F, et al. Exome sequencing identifies ZNF644 mutations in high myopia. PLoS Genetics. 2011;7(6):e1002084-e.

114. Tran-Viet K-N, St Germain E, Soler V, Powell C, Lim S-H, Klemm T, et al. Study of a US cohort supports the role of ZNF644 and high-grade myopia susceptibility. Molecular Vision. 2012;18:937-44.

115. Zhao F, Wu J, Xue A, Su Y, Wang X, Lu X, et al. Exome sequencing reveals CCDC111 mutation associated with high myopia. Human Genetics. 2013;132(8):913-21.

116. Aldahmesh MA, Khan AO, Alkuraya H, Adly N, Anazi S, Al-Saleh AA, et al. Mutations in LRPAP1 are associated with severe myopia in humans. The American Journal of Human Genetics. 2013;93(2):313-20.

117. Jiang D, Li J, Xiao X, Li S, Jia X, Sun W, et al. Detection of mutations in LRPAP1, CTSH, LEPREL1, ZNF644, SLC39A5, and SCO2 in 298 families with early-onset high myopia by exome sequencing. Investigative Ophthalmology & Visual Science. 2015;56(1):339-45.

118. Guo H, Jin X, Zhu T, Wang T, Tong P, Tian L, et al. SLC39A5 mutations interfering with the BMP/TGF-β pathway in non-syndromic high myopia. Journal of Medical Genetics. 2014;51(8):518.

119. Guo H, Tong P, Liu Y, Xia L, Wang T, Tian Q, et al. Mutations of P4HA2 encoding prolyl 4-hydroxylase 2 are associated with nonsyndromic high myopia. Genetics in Medicine. 2015;17(4):300-6.

120. Nakanishi H, Yamada R, Gotoh N, Hayashi H, Yamashiro K, Shimada N, et al. A genome-wide association analysis identified a novel susceptible locus for pathological myopia at 11q24.1. PLOS Genetics. 2009;5(9):e1000660.

121. Nishizaki R, Ota M, Inoko H, Meguro A, Shiota T, Okada E, et al. New susceptibility locus for high myopia is linked to the uromodulin-like 1 (UMODL1) gene region on chromosome 21q22.3. Eye. 2009;23(1):222-9.

122. Li Y-J, Goh L, Khor C-C, Fan Q, Yu M, Han S, et al. Genome-wide association studies reveal genetic variants in CTNND2 for high myopia in Singapore Chinese. Ophthalmology. 2011;118(2):368-75.

123. Hysi PG, Young TL, Mackey DA, Andrew T, Fernández-Medarde A, Solouki AM, et al. A genome-wide association study for myopia and refractive error identifies a susceptibility locus at 15q25. Nature genetics. 2010;42(10):902-5.

124. Solouki AM, Verhoeven VJ, Van Duijn CM, Verkerk AJ, Ikram MK, Hysi PG, et al. A genome-wide association study identifies a susceptibility locus for refractive errors and myopia at 15q14. Nature Genetics. 2010;42(10):897-901.

125. Verhoeven VJM, Hysi PG, Wojciechowski R, Fan Q, Guggenheim JA, Hohn R, et al. Genome-wide meta-analyses of multiancestry cohorts identify multiple new susceptibility loci for refractive error and myopia. Nature Genetics. 2013;45(3):314-8.

126. Wojciechowski R, Hysi PG. Focusing in on the complex genetics of myopia. PLOS Genetics. 2013;9(4):e1003442.

127. Kiefer AK, Tung JY, Do CB, Hinds DA, Mountain JL, Francke U, et al. Genome-Wide Analysis Points to Roles for Extracellular Matrix Remodeling, the Visual Cycle, and Neuronal Development in Myopia. PLOS Genetics. 2013;9(2):e1003299.

128. Pickrell JK, Berisa T, Liu JZ, Segurel L, Tung JY, Hinds DA. Detection and interpretation of shared genetic influences on 42 human traits. Nat Genet. 2016;48(7):709-17.

129. Fan Q, Barathi VA, Cheng C-Y, Zhou X, Meguro A, Nakata I, et al. Genetic variants on chromosome 1q41 influence ocular axial length and high myopia. PLOS Genetics. 2012;8(6).

130. Cheng C-Y, Schache M, Ikram MK, Young TL, Guggenheim JA, Vitart V, et al. Nine loci for ocular axial length identified through genome-wide association studies, including shared loci with refractive error. American Journal of Human Genetics. 2013;93(2):264-77.

131. Gao XR, Huang H, Kim H. Genome-wide association analyses identify 139 loci associated with macular thickness in the UK Biobank cohort. Human Molecular Genetics. 2018;28(7):1162-72.

132. Hosoda Y, Yoshikawa M, Miyake M, Tabara Y, Shimada N, Zhao W, et al. CCDC102B confers risk of low vision and blindness in high myopia. Nature Communications. 2018;9(1):1782.

133. Schaeffel F, Glasser A, Howland HC. Accommodation, refractive error and eye growth in chickens. Vision research. 1988;28(5):639-57.

134. Wildsoet C, Wallman J. Choroidal and scleral mechanisms of compensation for spectacle lenses in chicks. Vision research. 1995;35(9):1175-94.

135. Barathi V, Boopathi V, Yap EP, Beuerman RW. Two models of experimental myopia in the mouse. Vision research. 2008;48(7):904-16.

136. Moderiano D, Do M, Hobbs S, Lam V, Sarin S, Alonso-Caneiro D, et al. Influence of the time of day on axial length and choroidal thickness changes to hyperopic and myopic defocus in human eyes. Experimental eye research. 2019;182:125-36.

137. Stone RA, Lin T, Laties AM, Iuvone PM. Retinal dopamine and form-deprivation myopia. Proceedings of the National Academy of Sciences of the United States of America. 1989;86(2):704-6.

138. McCarthy CS, Megaw P, Devadas M, Morgan IG. Dopaminergic agents affect the ability of brief periods of normal vision to prevent form-deprivation myopia. Experimental Eye Research. 2007;84(1):100-7.

139. Schmid KL, Wildsoet CF. Inhibitory Effects of Apomorphine and Atropine and Their Combination on Myopia in Chicks. Optometry and Vision Science. 2004;81(2):137-47.

140. Brainard GC, Morgan WW. Light-induced stimulation of retinal dopamine: a dose-response relationship. Brain Research. 1987;424(1):199-203.

141. Ashby RS, Schaeffel F. The effect of bright light on lens compensation in chicks. Investigative ophthalmology & visual science. 2010;51(10):5247-53.

142. Smith EL, Hung L-F, Arumugam B, Huang J. Negative lens–induced myopia in infant monkeys: effects of high ambient lighting. Investigative ophthalmology & visual science. 2013;54(4):2959-69.

143. Feldkaemper M, Diether S, Kleine G, Schaeffel F. Interactions of spatial and luminance information in the retina of chickens during myopia development. Experimental eye research. 1999;68(1):105-15.

144. Gwiazda J, Thorn F, Bauer J, Held R. Myopic children show insufficient accommodative response to blur. Investigative Ophthalmology & Visual Science. 1993;34(3):690-4.

145. Guo SS, Sivak JG, Callender MG, Diehl-Jones B. Retinal dopamine and lens-induced refractive errors in chicks. Current Eye Research. 1995;14(5):385-9.

146. Hung L-F, Crawford ML, Smith EL. Spectacle lenses alter eye growth and the refractive status of young monkeys. Nature medicine. 1995;1(8):761-5.

147. Chakraborty R, Read SA, Collins MJ. Hyperopic defocus and diurnal changes in human choroid and axial length. Optometry and Vision Science. 2013;90(11):1187-98.

148. Wildsoet CF. Neural pathways subserving negative lens-induced emmetropization in chicks – Insights from selective lesions of the optic nerve and ciliary nerve. Current Eye Research. 2003;27(6):371-85.

149. Li SY, Li S-M, Zhou YH, Liu LR, Li H, Kang MT, et al. Effect of undercorrection on myopia progression in 12-year-old children. Graefe's Archive for Clinical and Experimental Ophthalmology. 2015;253(8):1363-8.

150. Chung K, Mohidin N, O'Leary DJ. Undercorrection of myopia enhances rather than inhibits myopia progression. Vision Research. 2002;42(22):2555-9.

151. Leung JTM, Brown B. Progression of myopia in Hong Kong Chinese schoolchildren is slowed by wearing progressive lenses. Optometry and Vision Science. 1999;76(6):346-54.

152. Cheng D, Woo GC, Drobe B, Schmid KL. Effect of bifocal and prismatic bifocal spectacles on myopia progression in children: Three-year results of a randomized clinical trial. JAMA Ophthalmology. 2014;132(3):258-64.

153. Wolffsohn JS, Flitcroft DI, Gifford KL, Jong M, Jones L, Klaver CCW, et al. IMI – Myopia control reports overview and introduction. Investigative Ophthalmology & Visual Science. 2019;60(3):M1-M19.

154. Cheng X, Xu J, Chehab K, Exford J, Brennan N. Soft contact lenses with positive spherical aberration for myopia control. Optometry and Vision Science. 2016;93(4):353-66.

155. Aller TA, Liu M, Wildsoet CF. Myopia control with bifocal contact lenses: A randomized clinical trial. Optometry and Vision Science. 2016;93(4):344-52.

156. Si J-K, Tang K, Bi H-S, Guo D-D, Guo J-G, Wang X-R. Orthokeratology for myopia control: a meta-analysis. Optometry and Vision Science. 2015;92(3):252-7.

157. Sun Y, Xu F, Zhang T, Liu M, Wang D, Chen Y, et al. Orthokeratology to control myopia progression: a meta-analysis. PloS one. 2015;10(4).

158. Dyer JA. Role of cycloplegics in progressive myopia. Ophthalmology. 1979;86(5):692-4.

159. Huang J, Wen D, Wang Q, McAlinden C, Flitcroft I, Chen H, et al. Efficacy comparison of 16 interventions for myopia control in children: a network meta-analysis. Ophthalmology. 2016;123(4):697-708.

160. Chia A, Lu Q, Tan D. Five-year clinical trial on atropine for the treatment of myopia 2. Ophthalmology. 2015;123.

161. Hung L-F, Arumugam B, Ostrin L, Patel N, Trier K, Jong M, et al. The Adenosine receptor antagonist, 7-Methylxanthine, alters emmetropizing responses in infant macaques. Investigative Ophthalmology & Visual Science. 2018;59(1):472-86.

162. Trier K, Munk Ribel-Madsen S, Cui D, Brøgger Christensen S. Systemic 7-methylxanthine in retarding axial eye growth and myopia progression: a 36-month pilot study. Journal of Ocular Biology, Diseases, and Informatics. 2008;1(2-4):85-93.

163. Robaei D, Rose KA, Kifley A, Cosstick M, Ip JM, Mitchell P. Factors associated with childhood strabismus: findings from a population-based study. Ophthalmology. 2006;113(7):1146-53.

164. Lueder GT. Orbital causes of incomitant strabismus. Middle East African journal of ophthalmology. 2015;22(3):286-91.

165. Khan AO. A Modern Approach to Incomitant Strabismus. Middle East African journal of ophthalmology. 2015;22(3):263-4.

166. Cotter SA, Varma R, Tarczy-Hornoch K, McKean-Cowdin R, Lin J, Wen G, et al. Risk factors associated with childhood strabismus: the multi-ethnic pediatric eye disease and Baltimore pediatric eye disease studies. Ophthalmology. 2011;118(11):2251-61.

167. Pathai S, Cumberland PM, Rahi JS. Prevalence of and early-life influences on childhood strabismus: Findings from the millennium cohort study. Archives of Pediatrics & Adolescent Medicine. 2010;164(3):250-7.

168. Atkinson J, Braddick O, Bobier B, Anker S, Ehrlich D, King J, et al. Two infant vision screening programmes: Prediction and prevention of strabismus and amblyopia from photo- and videorefractive screening. Eye. 1996;10:189.

169. Bremer DL, Palmer EA, Fellows RR, Baker JD, Hardy RJ, Tung B, et al. Strabismus in premature infants in the first year of life. Archives of Ophthalmology. 1998;116(3):329-33.

170. O'Connor AR, Stephenson TJ, Johnson A, Tobin MJ, Ratib S, Fielder AR. Change of refractive state and eye size in children of birth weight less than 1701 g. The British Journal of Ophthalmology. 2006;90(4):456-60.

171. Pennefather PM, Clarke MP, Strong NP, Cottrell DG, Dutton J, Tin W. Risk factors for strabismus in children born before 32 weeks' gestation. British Journal of Ophthalmology. 1999;83(5):514.

172. VanderVeen DK, Bremer DL, Fellows RR, Hardy RJ, Neely DE, Palmer EA, et al. Prevalence and course of strabismus through age 6 years in participants of the Early Treatment for Retinopathy of Prematurity randomized trial. Journal of AAPOS : the official publication of the American Association for Pediatric Ophthalmology and Strabismus. 2011;15(6):536-40.

173. Chew E, Remaley NA, Tamboli A, Zhao J, Podgor MJ, Klebanoff M. Risk factors for esotropia and exotropia. Archives of Ophthalmology. 1994;112(10):1349-55.

174. Pike M, Holmstrom G, De Vries L, Pennock J, Drew K, Sonksen P, et al. Patterns Of Visual Impairment Associated With Lesions Of The Preterm Infant Bran. Developmental Medicine & Child Neurology. 1994;36(10):849-62.

175. van Hof-Van Duin J, Evenhuis-Van Leunen A, Mohn G, Baerts W, Fetter WPF. Effects of very low birth weight (VLBW) on visual development during the first year after term. Early Human Development. 1989;20(3):255-66.

176. Torp-Pedersen T, Boyd HA, Poulsen G, Haargaard B, Wohlfahrt J, Holmes JM, et al. Perinatal risk factors for strabismus. International Journal of Epidemiology. 2010;39(5):1229-39.

177. Hakim RB, Tielsch JM. Maternal cigarette smoking during pregnancy: a risk factor for childhood strabismus. Archives of Ophthalmology. 1992;110(10):1459-62.

178. Ernst M, Moolchan ET, Robinson ML. Behavioral and neural consequences of prenatal exposure to nicotine. Journal of the American Academy of Child & Adolescent Psychiatry. 2001;40(6):630-41.

179. Mohney BG, Erie JC, Hodge DO, Jacobsen SJ. Congenital esotropia in Olmsted County, Minnesota. Ophthalmology. 1998;105(5):846-50.

180. Major A, Maples W, Toomey S, DeRosier W, Gahn D. Variables associated with the incidence of infantile esotropia. Optometry-Journal of the American Optometric Association. 2007;78(10):534-41.

181. Snir M, Nissenkorn I, Sherf I, Cohen S, Ben IS. Visual acuity, strabismus, and amblyopia in premature babies with and without retinopathy of prematurity. Annals of Ophthalmology. 1988;20(7):256-8.

182. Holmström G, Rydberg A, Larsson E. Prevalence and development of strabismus in 10-year-old premature children: a population-based study. Journal of Pediatric Ophthalmology and Strabismus. 2006;43(6):346-52.

183. Donders FC. An essay on the nature and the consequences of anomalies of refraction: P. Blakiston's Son & Company; 1899.

184. Walsh LA, LaRoche GR, Tremblay F. The use of binocular visual acuity in the assessment of intermittent exotropia. Journal of American Association for Pediatric Ophthalmology and Strabismus. 2000;4(3):154-7.

185. Tang SM, Chan RYT, Bin Lin S, Rong SS, Lau HHW, Lau WWY, et al. Refractive Errors and Concomitant Strabismus: A Systematic Review and Meta-analysis. Scientific Reports. 2016;6(1):35177.

186. Graeber CP, Hunter DG, Engle EC. The Genetic Basis of Incomitant Strabismus: Consolidation of the Current Knowledge of the Genetic Foundations of Disease. Semin Ophthalmol. 2013;28(5-6):427-37.

187. Kruger JM, Mansouri B, Cestari DM. An update on the genetics of comitant strabismus. Seminars in Ophthalmology. 2013;28(5-6):438-41.

188. Maconachie GE, Gottlob I, McLean RJ. Risk factors and genetics in common comitant strabismus: A systematic review of the literature. JAMA Ophthalmology. 2013;131(9):1179-86.

189. Podgor MJ, Remaley NA, Chew E. Associations Between Siblings for Esotropia and Exotropia. Archives of Ophthalmology. 1996;114(6):739-44.

190. Reynolds JD, Wackerhagen M. Strabismus in Monozygotic and Dizygotic Twins. American Orthoptic Journal. 1986;36(1):113-9.

191. Sanfilippo PG, Hammond CJ, Staffieri SE, Kearns LS, Melissa Liew SH, Barbour JM, et al. Heritability of strabismus: Genetic influence is specific to eso-deviation and independent of refractive error. Twin Research and Human Genetics. 2012;15(5):624-30.

192. Ye XC, Pegado V, Patel MS, Wasserman WW. Strabismus genetics across a spectrum of eye misalignment disorders. Clinical Genetics. 2014;86(2):103-11.

193. Georges A, Cambisano N, Ahariz N, Karim L, Georges M. A Genome Scan Conducted in a Multigenerational Pedigree with Convergent Strabismus Supports a Complex Genetic Determinism. PLOS ONE. 2013;8(12):e83574.

194. Parikh V, Shugart YY, Doheny KF, Zhang J, Li L, Williams J, et al. A strabismus susceptibility locus on chromosome 7p. Proceedings of the National Academy of Sciences of the United States of America. 2003;100(21):12283-8.

195. Rice A, Nsengimana Jrm, Simmons IG, Toomes C, Hoole J, Willoughby CE, et al. Replication of the recessive STBMS1 locus but with dominant inheritance. Investigative Ophthalmology & Visual Science. 2009;50(7):3210-7.

196. Shaaban S, MacKinnon S, Andrews C, Staffieri SE, Maconachie GDE, Chan W-M, et al. Genome-wide association study identifies a susceptibility locus for comitant esotropia and suggests a parent-of-origin effect. Investigative Ophthalmology & Visual Science. 2018;59(10):4054-64.

197. Chia A, Lin X, Dirani M, Gazzard G, Ramamurthy D, Quah B-L, et al. Risk factors for strabismus and amblyopia in young Singapore Chinese children. Ophthalmic Epidemiology. 2013;20(3):138-47.

198. Awan M, Proudlock FA, Gottlob I. A randomized controlled trial of unilateral strabismic and mixed amblyopia using occlusion dose monitors to record compliance. Investigative Ophthalmology & Visual Science. 2005;46(4):1435-9.

199. Erdem E, Çınar GY, Somer D, Demir N, Burcu A, Örnek F. Eye patching as a treatment for amblyopia in children aged 10–16 years. Japanese Journal of Ophthalmology. 2011;55(4):389-95.

200. Calcutt C, Murray A. Untreated essential infantile esotropia: factors affecting the development of amblyopia. Eye. 1998;12(2):167-72.

201. Leat SJ. To prescribe or not to prescribe? Guidelines for spectacle prescribing in infants and children. Clinical and Experimental Optometry. 2011;94(6):514-27.

202. Ingram RM, Arnold PE, Dally S, Lucas J. Results of a randomised trial of treating abnormal hypermetropia from the age of 6 months. The British Journal of Ophthalmology. 1990;74(3):158-9.

203. Newsham D. The effect of recent amblyopia research on current practice in the UK. The British Journal of Ophthalmology. 2010;94(10):1352-7.

204. Loudon SE, Simonsz HJ. The History of the Treatment of Amblyopia. Strabismus. 2005;13(2):93-106.

205. Parks MM, Friendly DS. Treatment of Eccentric Fixation in Children Under Four Years of Age *. American Journal of Ophthalmology. 1966;61(3):395-9.

206. Scully J. Early intensive occlusion in strabismus with non-central fixation. British Medical Journal. 1961;2(5267):1610.

207. Oliver M, Neumann R, Chaimovitch Y, Gotesman N, Shimshoni M. Compliance and Results of Treatment for Amblyopia in Children More Than 8 Years Old. American Journal of Ophthalmology. 1986;102(3):340-5.

208. Repka MX, Beck RW, Holmes JM, Birch EE, Chandler DL, Cotter SA, et al. A randomized trial of patching regimens for treatment of moderate amblyopia in children. Archives of Ophthalmology (Chicago, Ill: 1960). 2003;121(5):603-11.

209. Holmes JM, Kraker RT, Beck RW, Birch EE, Cotter SA, Everett DF, et al. A randomized trial of prescribed patching regimens for treatment of severe amblyopia in children. Ophthalmology. 2003;110(11):2075-87.

210. Williams C, Northstone K, Harrad RA, Sparrow JM, Harvey I, Team AS. Amblyopia treatment outcomes after screening before or at age 3 years: follow up from randomised trial. BMJ (Clinical research ed). 2002;324(7353):1549-.

211. Ip JM, Robaei D, Kifley A, Wang JJ, Rose KA, Mitchell P. Prevalence of hyperopia and associations with eye findings in 6-and 12-year-olds. Ophthalmology. 2008;115(4):678-85. e1.

212. Augsburger AR. Hyperopia. In: Amos JF, editor. Diagnosis and Management in Vision Care. Boston: Butterworths; 1987. p. 101-19.

213. Dobson V, Sebris S. Longitudinal study of acuity and stereopsis in infants with or at-risk for esotropia. Investigative Ophthalmology & Visual Science. 1989;30(6):1146-58.

214. Williams WR, Latif AHA, Hannington L, Watkins DR. Hyperopia and educational attainment in a primary school cohort. Archives of Disease in Childhood. 2005;90(2):150-3.

215. Benjamin WJ, Borish IM. Borish's clinical refraction. 2006.

216. Tarczy-Hornoch K. The epidemiology of early childhood hyperopia. Optometry and Vision Science. 2007;84(2):115-23.

217. O'Donoghue L, Mcclelland JF, Logan NS, Rudnicka AR, Owen CG, Saunders KJ. Refractive error and visual impairment in school children in Northern Ireland. British Journal of Ophthalmology. 2010;94(9):1155-9.

218. Ingram RM, Walker C, Wilson JM, Arnold PE, Dally S. Prediction of amblyopia and squint by means of refraction at age 1 year. The British Journal of Ophthalmology. 1986;70(1):12-5.

219. Banks MS. Infant refraction and accommodation. International Ophthalmology Clinics. 1980;20(1):205-32.

220. Castagno VD, Fassa AG, Carret MLV, Vilela MAP, Meucci RD. Hyperopia: a meta-analysis of prevalence and a review of associated factors among school-aged children. BMC Ophthalmology. 2014;14(1):163.

221. Narayanasamy S, Vincent SJ, Sampson GP, Wood JM. Visual demands in modern Australian primary school classrooms. Clinical and Experimental Optometry. 2016;99(3):233-40.

222. Stewart-Brown S, Haslum MN, Butler N. Educational attainment of 10-year-old children with treated and untreated visual defects. Developmental Medicine & Child Neurology. 1985;27(4):504-13.

223. Collier JD, Rosenfield M. Accommodation and convergence during sustained computer work. Optometry-Journal of the American Optometric Association. 2011;82(7):434-40.

224. Narayanasamy S, Vincent SJ, Sampson GP, Wood JM. Impact of simulated hyperopia on academic-related performance in children. Optometry and Vision Science. 2015;92(2):227-36.

225. Kulp MT, Ciner E, Maguire M, Moore B, Pentimonti J, Pistilli M, et al. Uncorrected hyperopia and preschool early literacy: Results of the Vision in Preschoolers-Hyperopia in Preschoolers (VIP-HIP) Study. Ophthalmology. 2016;123(4):681-9.

226. Garzia RP, Nicholson SB, Gaines CS, Murphy MA, Kramer A, Potts J. Effects of nearpoint visual stress on psycholinguistic processing in reading. Journal of the American Optometric Association. 1989.

227. Walton HN, Schubert DG, Clark D, Burke W. Effects of induced hyperopia. American Journal of Optometry and Physiological Optics. 1978;55(7):451-5.

228. Morjaria P, Murali K, Evans J, Gilbert C. Spectacle wearing in children randomised to ready-made or custom spectacles, and potential cost savings to programmes: study protocol for a randomised controlled trial. Trials. 2016;17:36-.

229. Horwood AM, Riddell PM. Hypo-accommodation responses in hypermetropic infants and children. British Journal of Ophthalmology. 2011;95(2):231-7.

230. Caloroso EE, Rouse MW, Cotter SA. Clinical management of strabismus. 1993.

231. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. PLoS Medicine. 2015;12(3):e1001779.

232. Conroy M, Sellors J, Effingham M, Littlejohns TJ, Boultwood C, Gillions L, et al. The advantages of UK Biobank's open-access strategy for health research. Journal of Internal Medicine. 2019;286(4):389-97.

233. Delgado-Rodriguez M, Llorca J. Bias. Journal of Epidemiology & Community Health. 2004;58(8):635-41.

234. Fry A, Littlejohns TJ, Sudlow C, Doherty N, Adamska L, Sprosen T, et al. Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants With Those of the General Population. American journal of epidemiology. 2017;186(9):1026-34.

235. Allen N, Sudlow C, Downey P, Peakman T, Danesh J, Elliott P, et al. UK Biobank: Current status and what it means for epidemiology. Health Policy and Technology.1(3):123-6.

236. Coughlin SS. Recall bias in epidemiologic studies. Journal of clinical epidemiology. 1990;43(1):87-91.

237. Cumberland PM, Chianca A, Rahi JS. Accuracy and utility of self-report of refractive error. JAMA Ophthalmology. 2016;134(7):794-801.

238. Tedja MS, Wojciechowski R, Hysi PG, Eriksson N, Furlotte NA, Verhoeven VJM, et al. Genome-wide association meta-analysis highlights light-induced signaling as a driver for refractive error. Nature Genetics. 2018;50(6):834-48.

239. Qin X-J, Margrain TH, To CH, Bromham N, Guggenheim JA. Anisometropia is independently associated with both spherical and cylindrical ametropia. Investigative Ophthalmology & Visual Science. 2005;46(11):4024-31.

240. Howie B, Marchini J, Stephens M. Genotype imputation with thousands of genomes. G3 (Bethesda). 2011;1(6):457-70.

241. Walter K, Min JL, Huang J, Crooks L, Memari Y, McCarthy S, et al. The UK10K project identifies rare variants in health and disease. Nature. 2015;526(7571):82-90.

242. Wain LV, Shrine N, Miller S, Jackson VE, Ntalla I, Artigas MS, et al. Novel insights into the genetics of smoking behaviour, lung function, and chronic obstructive pulmonary disease (UK BiLEVE): a genetic association study in UK Biobank. The Lancet Respiratory Medicine.3(10):769-81.

243. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank resource with deep phenotyping and genomic data. Nature. 2018;562(7726):203-9.

244. McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood AR, Teumer A, et al. A reference panel of 64,976 haplotypes for genotype imputation. Nature Genetics. 2016;48(10):1279-83.

245. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank resource with deep phenotyping and genomic data. Nature. 2018;562(7726):203-9.

246. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience. 2015;4:7.

247. Boyd A, Golding J, Macleod J, Lawlor DA, Fraser A, Henderson J, et al. Cohort Profile: The 'Children of the 90s'--the index offspring of the Avon Longitudinal Study of Parents and Children. International Journal of Epidemiology. 2013;42(1):111-27.

248. Fraser A, Macdonald-Wallis C, Tilling K, Boyd A, Golding J, Davey Smith G, et al. Cohort Profile: The Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort. International Journal of Epidemiology. 2012;42(1):97-110.

249. Williams C, Northstone K, Howard M, Harvey I, Harrad RA, Sparrow JM. Prevalence and risk factors for common vision problems in children: data from the ALSPAC study. British Journal of Ophthalmology. 2008;92(7):959.

250. Zhao J, Mao J, Luo R, Li F, Pokharel GP, ELLWEIN LB. Accuracy of noncycloplegic autorefraction in school-age children in China. Optometry and vision science. 2004;81(1):49-55.

251. Fotedar R, Rochtchina E, Morgan I, Wang JJ, Mitchell P, Rose KA. Necessity of cycloplegia for assessing refractive error in 12-year-old children: a population-based study. American journal of ophthalmology. 2007;144(2):307-9.

252. Williams C, Miller L, Northstone K, Sparrow JM. The use of non-cycloplegic autorefraction data in general studies of children's development. British Journal of Ophthalmology. 2008;92(5):723.

253. Horikoshi M, Yaghootkar H, Mook-Kanamori DO, Sovio U, Taal HR, Hennig BJ, et al. New loci associated with birth weight identify genetic links between intrauterine growth and adult height and metabolism. Nature Genetics. 2013;45(1):76-82.

254. Lee JJ, Wedow R, Okbay A, Kong E, Maghzian O, Zacher M, et al. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. Nature Genetics. 2018;50(8):1112-21.

255. Parascandola M, Weed DL. Causation in epidemiology. Journal of Epidemiology and Community Health. 2001;55(12):905.

256. Hill AB. The environment and disease: association or causation? : Sage Publications; 1965.

257. Klungel OH, Martens EP, Psaty BM, Grobbee DE, Sullivan SD, Stricker BHC, et al. Methods to assess intended effects of drug treatment in observational studies are reviewed. Journal of clinical epidemiology. 2004;57(12):1223-31.

258. Evans DM, Smith GD. Mendelian Randomization: New Applications in the Coming Age of Hypothesis-Free Causality. Annual Review of Genomics and Human Genetics. 2015;16(1):327-50.

259. Lawlor DA, Harbord RM, Sterne JAC, Timpson N, Davey Smith G. Mendelian randomization: Using genes as instruments for making causal inferences in epidemiology. Statistics in Medicine. 2008;27(8):1133-63.

260. Bender R. Introduction to the Use of Regression Models in Epidemiology. In: Verma M, editor. Cancer Epidemiology. Totowa, NJ: Humana Press; 2009. p. 179-95.

261. Schneider A, Hommel G, Blettner M. Linear regression analysis: part 14 of a series on evaluation of scientific publications. Deutsches Arzteblatt international. 2010;107(44):776-82.

262. Kleinbaum DG, Kupper LL, Chambless LE. Logistic regression analysis of epidemiologic data: theory and practice. Communications in Statistics - Theory and Methods. 1982;11(5):485-547.

263. Bush WS, Moore JH. Chapter 11: Genome-wide association studies. PLOS Computational Biology. 2012;8(12):e1002822-e.

264. Anderson CA, Pettersson FH, Clarke GM, Cardon LR, Morris AP, Zondervan KT. Data quality control in genetic case-control association studies. Nature protocols. 2010;5(9):1564-73.

265. Ma C, Blackwell T, Boehnke M, Scott LJ, the Go TDi. Recommended joint and meta-analysis strategies for case-control association testing of single low-count variants. Genetic Epidemiology. 2013;37(6):539-50.

266. Firth D. Bias reduction of maximum likelihood estimates. Biometrika. 1993;80(1):27-38.

267. Wang X. Firth logistic regression for rare variant association tests. Frontiers in Genetics. 2014;5:187-.

268. Loh P-R, Tucker G, Bulik-Sullivan BK, Vilhjálmsson BJ, Finucane HK, Salem RM, et al. Efficient Bayesian mixed-model analysis increases association power in large cohorts. Nature Genetics. 2015;47:284.

269. Loh P-R, Kichaev G, Gazal S, Schoech AP, Price AL. Mixed-model association for biobank-scale datasets. Nature Genetics. 2018;50(7):906-8.

270. Hernán MR, Robins JM. Causal inference. Boca Raton, U.S.: Taylor & Francis; 2019.

271. Lousdal ML. An introduction to instrumental variable assumptions, validation and estimation. Emerg Themes Epidemiol. 2018;15:1-.

272. Wald A. The fitting of straight lines if both variables are subject to error. Annals of Mathematical Statistics. 1940;11(3):284-300.

273. Johnson T. Efficient calculation for multi-SNP genetic risk scores. Technical report, The Comprehensive R Archive Network Retrieved from http://cranr-projectorg/web/packages/gtx/vignettes/ashg2012pdf. 2003.

274. Evangelou E, Ioannidis JPA. Meta-analysis methods for genome-wide association studies and beyond. Nature Reviews Genetics. 2013;14(6):379-89.

275. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics. 2010;26(17):2190-1.

276. Katan M. Apolipoprotein E isoforms, serum cholesterol, and cancer. The Lancet. 1986;327(8479):507-8.

277. Gray R, Wheatley K. How to avoid bias when comparing bone marrow transplantation with chemotherapy. Bone Marrow Transplantation. 1991;7 Suppl 3:9-12.

278. Davey Smith G, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease?*. International Journal of Epidemiology. 2003;32(1):1-22.

279. Hernán MA. A definition of causal effect for epidemiological research. Journal of Epidemiology & Community Health. 2004;58(4):265-71.

280. Pingault J-B, O'Reilly PF, Schoeler T, Ploubidis GB, Rijsdijk F, Dudbridge F. Using genetic data to strengthen causal inference in observational research. Nature Reviews Genetics. 2018;19(9):566-80.

281. Imbens GW, Rubin DB. Causal inference in statistics, social, and biomedical sciences: Cambridge University Press; 2015.

282. Wang T, Xu L. Circulating Vitamin E Levels and Risk of Coronary Artery Disease and Myocardial Infarction: A Mendelian Randomization Study. Nutrients. 2019;11(9):2153.

283. Frikke-Schmidt R, Nordestgaard BG, Stene MCA, Sethi AA, Remaley AT, Schnohr P, et al. Association of Loss-of-Function Mutations in the ABCA1 Gene With High-Density Lipoprotein Cholesterol Levels and Risk of Ischemic Heart Disease. JAMA. 2008;299(21):2524-32.

284. Garg A, Sharma A, Krishnamoorthy P, Garg J, Virmani D, Sharma T, et al. Role of niacin in current clinical practice: a systematic review. The American journal of medicine. 2017;130(2):173-87.

285. Ference BA. How to use Mendelian randomization to anticipate the results of randomized trials. European Heart Journal. 2017;39(5):360-2.

286. Mach F, Baigent C, Catapano AL, Koskinas KC, Casula M, Badimon L, et al. 2019 ESC/EAS Guidelines for the management of dyslipidaemias: lipid modification to reduce cardiovascular risk: The Task Force for the management of dyslipidaemias of the European Society of Cardiology (ESC) and European Atherosclerosis Society (EAS). European Heart Journal. 2019;41(1):111-88.

287. Burgess S, Thompson SG. Use of allele scores as instrumental variables for Mendelian randomization. International Journal of Epidemiology. 2013;42(4):1134-44.

288. Pierce BL, Burgess S. Efficient design for Mendelian randomization studies: subsample and 2-sample instrumental variable estimators. American Journal of Epidemiology. 2013;178(7):1177-84.

289. Verbanck M, Chen C-Y, Neale B, Do R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. Nature Genetics. 2018;50(5):693-8.

290. Zhu Z, Zheng Z, Zhang F, Wu Y, Trzaskowski M, Maier R, et al. Causal associations between risk factors and common diseases inferred from GWAS summary data. Nature Communications. 2018;9(1):224.

291. Corbin LJ, Richmond RC, Wade KH, Burgess S, Bowden J, Smith GD, et al. BMI as a Modifiable Risk Factor for Type 2 Diabetes: Refining and Understanding Causal Estimates Using Mendelian Randomization. Diabetes. 2016;65(10):3002-7.

292. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. International Journal of Epidemiology. 2015;44(2):512-25.

293. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. Genetic Epidemiology. 2016;40(4):304-14.

294. Hartwig FP, Davey Smith G, Bowden J. Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. International Journal of Epidemiology. 2017;46(6):1985-98.

295. Bor J, Moscoe E, Mutevedzi P, Newell M-L, Bärnighausen T. Regression discontinuity designs in epidemiology: causal inference without randomized trials. Epidemiology (Cambridge, Mass). 2014;25(5):729-37.

296. Lee DS, Card D. Regression discontinuity inference with specification error. Journal of Econometrics. 2008;142(2):655-74.

297. Chaplin DD, Cook TD, Zurovac J, Coopersmith JS, Finucane MM, Vollmer LN, et al. The internal and external validity of the Regression Discontinuity design: A meta-analysis of 15 within-study comparisons. Journal of Policy Analysis and Management. 2018;37(2):403-29.

298. Wing C, Bello-Gomez RA. Regression discontinuity and beyond: Options for studying external validity in an internally valid design. American Journal of Evaluation. 2018;39(1):91-108.

299. Trochim WM. Research design for program evaluation: The regression-discontinuity approach: SAGE Publications, Inc; 1984.

300. Imbens G, Lemieux T. Regression discontinuity designs: A guide to practice. Journal of Econometrics. 2008;142(2):615-35.

301. Hahn J, Todd P, Van der Klaauw W. Identification and estimation of treatment effects with a Regression-Discontinuity design. Econometrica. 2001;69(1):201-9.

302. Imbens GW, Angrist JD. Identification and estimation of local average treatment effects. Econometrica. 1994;62(2):467-75.

303. Ludwig J, Miller DL. Does head start improve children's life chances? Evidence from a Regression Discontinuity design. National Bureau of Economic Research Working Paper Series. 2005;No. 11702.

304. Imbens G, Kalyanaraman K. Optimal bandwidth choice for the Regression Discontinuity estimator. The Review of Economic Studies. 2012;79(3):933-59.

305. Dudbridge F, Gusnanto A. Estimation of significance thresholds for genomewide association scans. Genetic Epidemiology. 2008;32(3):227-34.

306. Gao X, Starmer J, Martin ER. A multiple testing correction method for genetic association studies using correlated single nucleotide polymorphisms. Genetic Epidemiology. 2008;32(4):361-9.

307. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. American Journal of Human Genetics. 2007;81(3):559-75.

308. Gibson G. Hints of hidden heritability in GWAS. Nature Genetics. 2010;42(7):558-60.

309. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, et al. LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics (Oxford, England). 2010;26(18):2336-7.

310. Yang J, Weedon MN, Purcell S, Lettre G, Estrada K, Willer CJ, et al. Genomic inflation factors under polygenic inheritance. European Journal of Human Genetics. 2011;19(7):807-12.

311. Hellwege JN, Keaton JM, Giri A, Gao X, Velez Edwards DR, Edwards TL. Population stratification in genetic association studies. Current protocols in human genetics. 2017;95:1.22.1-1..3.

312. Smith JG, Newton-Cheh C. Genome-wide association study in humans. Methods in Molecular Biology. 2009;573:231-58.

313. Devlin B, Roeder K. Genomic control for association studies. Biometrics. 1999;55(4):997-1004.

314. Lande R, Thompson R. Efficiency of marker-assisted selection in the improvement of quantitative traits. Genetics. 1990;124(3):743-56.

315. International Schizophrenia C, Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC, et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature. 2009;460(7256):748-52.

316. Dudbridge F. Polygenic Epidemiology. Genetic Epidemiology. 2016;40(4):268-72.

317. Ghorbani Mojarrad N, Plotnikov D, Williams C, Guggenheim JA, Eye ftUB, Consortium V. Association between polygenic risk score and risk of myopia. JAMA Ophthalmology. 2020;138(1):7-13.

318. Vilhjálmsson Bjarni J, Yang J, Finucane Hilary K, Gusev A, Lindström S, Ripke S, et al. Modeling linkage disequilibrium increases accuracy of polygenic risk scores. The American Journal of Human Genetics. 2015;97(4):576-92.

319. Benner C, Spencer CCA, Havulinna AS, Salomaa V, Ripatti S, Pirinen M. FINEMAP: efficient variable selection using summary data from genome-wide association studies. Bioinformatics. 2016;32(10):1493-501.

320. Yates A, Akanni W, Amode MR, Barrell D, Billis K, Carvalho-Silva D, et al. Ensembl 2016. Nucleic Acids Res. 2016;44(D1):D710-D6.

321. Cunningham F, Achuthan P, Akanni W, Allen J, Amode M R, Armean IM, et al. Ensembl 2019. Nucleic Acids Research. 2019;47(D1):D745-D51.

322. Rentzsch P, Witten D, Cooper GM, Shendure J, Kircher M. CADD: predicting the deleteriousness of variants throughout the human genome. Nucleic Acids Research. 2018;47(D1):D886-D94.

323. Vaser R, Adusumalli S, Leng SN, Sikic M, Ng PC. SIFT missense predictions for genomes. Nature protocols. 2016;11(1):1.

324. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. Nature methods. 2010;7(4):248-9.

325. Staley JR, Blackshaw J, Kamat MA, Ellis S, Surendran P, Sun BB, et al. PhenoScanner: a database of human genotype-phenotype associations. Bioinformatics (Oxford, England). 2016;32(20):3207-9.

326. Landrum MJ, Kattman BL. ClinVar at five years: Delivering on the promise. Human Mutation. 2018;39(11):1623-30.

327. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature. 2020;581(7809):434-43.

328. Saunders KJ, McCulloch DL, Shepherd AJ, Wilkinson AG. Emmetropisation following preterm birth. The British Journal of Ophthalmology. 2002;86(9):1035-40.

329. Gallo JE, Holmström G, Kugelberg U, Hedquist B, Lennerstrand G. Regressed retinopathy of prematurity and its sequelae in children aged 5-10 years. The British Journal of Ophthalmology. 1991;75(9):527-31.

330. Quinn GE, Dobson V, Kivlin J, Kaufman LM, Repka MX, Reynolds JD, et al. Prevalence of myopia between 3 months and 5 12 years in preterm infants with and without retinopathy of prematurity. Ophthalmology. 1998;105(7):1292-300.

331. Chen T-C, Tsai T-H, Shih Y-F, Yeh P-T, Yang C-H, Hu F-C, et al. Long-term evaluation of refractive status and optical components in eyes of children born prematurely. Investigative Ophthalmology & Visual Science. 2010;51(12):6140-8.

332. Goisis A, Özcan B, Myrskylä M. Decline in the negative association between low birth weight and cognitive ability. Proceedings of the National Academy of Sciences. 2017;114(1):84-8.

333. Norris T, Seaton SE, Manktelow BN, Baker PN, Kurinczuk JJ, Field D, et al. Updated birth weight centiles for England and Wales. Archives of Disease in Childhood - Fetal and Neonatal Edition. 2018;103(6):F577.

334. O'Donoghue L, Kapetanankis VV, McClelland JF, Logan NS, Owen CG, Saunders KJ, et al. Risk factors for childhood myopia: Findings from the NICER study. Investigative Ophthalmology & Visual Science. 2015;56(3):1524-30.

335. Chua SYL, Ikram MK, Tan CS, Lee YS, Ni Y, Shirong C, et al. Relative contribution of risk factors for early-onset myopia in young Asian children. Investigative Ophthalmology & Visual Science. 2015;56(13):8101-7.

336. Rahi JS, Cumberland PM, Peckham CS. Myopia over the lifecourse: Prevalence and early life influences in the 1958 British Birth cohort. Ophthalmology. 2011;118(5):797-804.

337. Fieß A, Schuster AK, Pfeiffer N, Nickels S. Association of birth weight with corneal power in early adolescence: Results from the National Health and Nutrition Examination Survey (NHANES) 1999–2008. PLOS ONE. 2017;12(10):e0186723.

338. Pierce BL, Ahsan H, Vanderweele TJ. Power and instrument strength requirements for Mendelian randomization studies using multiple genetic variants. International journal of epidemiology. 2011;40(3):740-52.

339. Palmer TM, Lawlor DA, Harbord RM, Sheehan NA, Tobias JH, Timpson NJ, et al. Using multiple genetic variants as instrumental variables for modifiable risk factors. Statistical Methods in Medical Research. 2012;21(3):223-42.

340. Didelez V, Sheehan N. Mendelian randomization as an instrumental variable approach to causal inference. Statistical Methods in Medical Research. 2007;16(4):309-30.

341. Hemani G, Bowden J, Davey Smith G. Evaluating the potential role of pleiotropy in Mendelian randomization studies. Human Molecular Genetics. 2018;27(R2):R195-R208.

342. Fall CHD, Sachdev HS, Osmond C, Restrepo-Mendez MC, Victora C, Martorell R, et al. Association between maternal age at childbirth and child and adult outcomes in the offspring: a prospective study in five low-income and middle-income countries (COHORTS collaboration). The Lancet Global Health. 2015;3(7):e341-e422.

343. Lin Z, Mao GY, Vasudevan B, Jin ZB, Ciuffreda KJ, Jhanji V, et al. The association between maternal reproductive age and progression of refractive error in urban students in Beijing. PLOS ONE. 2015;10(9):e0139383.

344. Hinkle SN, Albert PS, Mendola P, Sjaarda LA, Yeung E, Boghossian NS, et al. The association between parity and birthweight in a longitudinal consecutive pregnancy cohort. Paediatric and Perinatal Epidemiology. 2014;28(2):106-15.

345. Guggenheim JA, McMahon G, Northstone K, Mandel Y, Kaiserman I, Stone RA, et al. Birth order and myopia. Ophthalmic Epidemiology. 2013;20(6):10.3109/09286586.2013.848457.

346. Ruiz M, Goldblatt P, Morrison J, Kukla L, Švancara J, Riitta-Järvelin M, et al. Mother's education and the risk of preterm and small for gestational age birth: a DRIVERS meta-analysis of 12 European cohorts. Journal of Epidemiology and Community Health. 2015;69(9):826.

347. Silvestrin S, da Silva CH, Hirakata VN, Goldani AAS, Silveira PP, Goldani MZ. Maternal education level and low birth weight: a meta-analysis. Jornal de Pediatria. 2013;89(4):339-45.

348. Williams KM, Hysi PG, Yonova-Doing E, Mahroo OA, Snieder H, Hammond CJ. Phenotypic and genotypic correlation between myopia and intelligence. Scientific Reports. 2017;7:45977.

349. Ehrenberg HM, Mercer BM, Catalano PM. The influence of obesity and diabetes on the prevalence of macrosomia. American Journal of Obstetrics & Gynecology. 2004;191(3):964-8.

350. Li C, Zeng L, Wang D, Dang S, Chen T, Watson V, et al. Effect of maternal pre-pregnancy BMI and weekly gestational weight gain on the development of infants. Nutrition Journal. 2019;18(1):6.

351. Dubois L, Girard M. Early determinants of overweight at 4.5 years in a population-based longitudinal study. International Journal of Obesity. 2006;30(4):610-7.

352. Barker DJ, Bull AR, Osmond C, Simmonds SJ. Fetal and placental size and risk of hypertension in adult life. BMJ (Clinical research ed). 1990;301(6746):259-62.

353. Gluckman PD, Hanson MA, Cooper C, Thornburg KL. Effect of in utero and early-life conditions on adult health and disease. The New England journal of medicine. 2008;359(1):61-73.

354. Gaillard R. Maternal obesity during pregnancy and cardiovascular development and disease in the offspring. European journal of epidemiology. 2015;30(11):1141-52.

355. Davey Smith G, Ebrahim S. What can mendelian randomisation tell us about modifiable behavioural and environmental exposures? BMJ. 2005;330(7499):1076-9.

356. Cotter SA, Varma R, Tarczy-Hornoch K, McKean-Cowdin R, Lin J, Wen G, et al. Risk factors associated with childhood strabismus: the multi-ethnic pediatric eye disease and Baltimore pediatric eye disease studies. Ophthalmology. 2011;118(11):2251-61.

357. The GTEx Consortium. The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. Science. 2015;348(6235):648-60.

358. Bagos PG. Genetic model selection in genome-wide association studies: robust methods and the use of meta-analysis. Statistical Applications in Genetics and Molecular Biology. 2013;12(3):285-308.

359. Qi L, Ma J, Qi Q, Hartiala J, Allayee H, Campos H. A genetic risk score and risk of myocardial infarction in hispanics. Circulation. 2011;123(4):374-80.

360. The 1000 Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes. Nature. 2012;491(7422):56-65.

361. Chiang C, Scott AJ, Davis JR, Tsang EK, Li X, Kim Y, et al. The impact of structural variation on human gene expression. Nature Genetics. 2017;49(5):692-9.

362. Plotnikov D, Shah RL, Rodrigues JN, Cumberland PM, Rahi JS, Hysi PG, et al. A commonly occurring genetic variant within the NPLOC4-TSPAN10-PDE6G gene cluster is associated with the risk of strabismus. Human Genetics. 2019;138(7):723-37.

363. Fritsche LG, Igl W, Bailey JNC, Grassmann F, Sengupta S, Bragg-Gresham JL, et al. A large genome-wide association study of age-related macular degeneration highlights contributions of rare and common variants. Nature Genetics. 2015;48(2):134-43.

364. Currant H, Hysi P, Fitzgerald TW, Gharahkhani P, Bonnemaijer PWM, Atan D, et al. Genetic variation affects morphological retinal phenotypes extracted from UK Biobank Optical Coherence Tomography images. medRxiv. 2020:2020.07.20.20157180.

365. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature. 2007;447(7145):661-78.

366. Zhou W, Nielsen JB, Fritsche LG, Dey R, Gabrielsen ME, Wolford BN, et al. Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. Nature Genetics. 2018;50(9):1335-41.

367. Sul JH, Martin LS, Eskin E. Population structure in genetic studies: Confounding factors and mixed models. PLOS Genetics. 2018;14(12):e1007309.

368. Woodin T, McCulloch G, Cowan S. Secondary Education and the Raising of the School-Leaving Age: Coming of Age?: Palgrave Macmillan US; 2013.

369. Del Bono E, Galindo-Rueda F. The long term impacts of compulsory schooling: evidence from a natural experiment in school leaving dates. Institute for Social and Economic Research; 2006.

370. Calonico S, Cattaneo MD, Titiunik R. Robust nonparametric confidence intervals for Regression-Discontinuity designs. Econometrica. 2014;82(6):2295-326.

371. Clark D, Royer H. The Effect of Education on Adult Mortality and Health: Evidence from Britain. Am Econ Rev. 2013;103(6):2087-120.

372. Davies NM, Dickson M, Davey Smith G, van den Berg GJ, Windmeijer F. The causal effects of education on health outcomes in the UK Biobank. Nature Human Behaviour. 2018;2(2):117-25.

373. Wilson T. Compulsory education and teenage motherhood. Stirling Economics Discussion Paper. 2017.

374. Lee DS, Lemieux T. Regression Discontinuity designs in economics. National Bureau of Economic Research Working Paper Series. 2009;No. 14723.

375. International HapMap Consortium, Altshuler DM, Gibbs RA, Peltonen L, Altshuler DM, Gibbs RA, et al. Integrating common and rare genetic variation in diverse human populations. Nature. 2010;467(7311):52-8.

376. Hernán MA, Robins JM. Causal inference, Part I. Boca Raton: Chapman & Hall/CRC, forthcoming. 2019:25-40.

377. Wooldridge JM. Asymptotic properties of weighted M-estimators for standard stratified samples. Econometric Theory. 2001;17(2):451-70.

378. Solon G, Haider SJ, Wooldridge JM. What are we weighting for? Journal of Human Resources. 2015;50(2):301-16.

379. McCrary J. Manipulation of the running variable in the regression discontinuity design: A density test. Journal of Econometrics. 2008;142(2):698-714.

380. Sanderson E, Windmeijer F. A weak instrument F-test in linear IV models with multiple endogenous variables. Journal of Econometrics. 2016;190(2):212-21.

381. Machin S, Marie O, Vujić S. The crime reducing effect of education*. The Economic Journal. 2011;121(552):463-84.

382. Moscoe E, Bor J, Bärnighausen T. Regression discontinuity designs are underutilized in medicine, epidemiology, and public health: a review of current and best practice. Journal of Clinical Epidemiology. 2015;68(2):132-43.

383. Russell RJ, Startup MJ. Month of birth and academic achievement. Personality and Individual Differences. 1986;7(6):839-46.

384. McMahon G, Zayats T, Chen Y-P, Prashar A, Williams C, Guggenheim JA. Season of birth, daylight hours at birth, and high myopia. Ophthalmology. 2009;116(3):468-73.

385. Rosner J. The relationship between moderate hyperopia and academic achievement: how much plus is enough? Journal of the American Optometric Association. 1997;68(10):648-50.

386. Shankar S, Evans MA, Bobier WR. Hyperopia and emergent literacy of young children: Pilot study. Optometry and Vision Science. 2007;84(11):1031-8.

387. Sun Y-Q, Burgess S, Staley JR, Wood AM, Bell S, Kaptoge SK, et al. Body mass index and all cause mortality in HUNT and UK Biobank studies: linear and non-linear mendelian randomisation analyses. BMJ. 2019;364:l1042.

388. Aune D, Sen A, Prasad M, Norat T, Janszky I, Tonstad S, et al. BMI and all cause mortality: systematic review and non-linear dose-response meta-analysis of 230 cohort studies with 3.74 million deaths among 30.3 million participants. BMJ. 2016;353:i2156.

389. Wood A, Guggenheim JA. Refractive error has minimal influence on the risk of age-related macular degeneration: A Mendelian randomization study. American Journal of Ophthalmology. 2019;206:87-93.

390. Plotnikov D, Guggenheim J. Is a large eye size a risk factor for myopia? A Mendelian randomization study. bioRxiv. 2017.

391. Brion M-JA, Shakhbazov K, Visscher PM. Calculating statistical power in Mendelian randomization studies. International Journal of Epidemiology. 2013;42(5):1497-501.

392. Burgess S, Davies NM, Thompson SG, Consortium EPICI. Instrumental variable analysis with a nonlinear exposure-outcome relationship. Epidemiology (Cambridge, Mass). 2014;25(6):877-85.

393. Wray NR, Yang J, Hayes BJ, Price AL, Goddard ME, Visscher PM. Pitfalls of predicting complex traits from SNPs. Nature Reviews Genetics. 2013;14(7):507-15.

394. Pozarickij A, Williams C, Hysi PG, Guggenheim JA, Aslam T, Barman SA, et al. Quantile regression analysis reveals widespread evidence for gene-environment or gene-gene interactions in myopia development. Communications Biology. 2019;2(1):167.

395. Thorn F, Gwiazda J, Held R. Myopia progression is specified by a double exponential growth function. Optometry and Vision Science. 2005;82(4):E286.

396. Flitcroft DI. Emmetropisation and the aetiology of refractive errors. Eye. 2014;28(2):169-79.

397. Grosvenor T, Goss DA. Role of the Cornea in Emmetropia and Myopia. Optometry and Vision Science. 1998;75(2).

398. Hemenger RP, Garner LF, Ooi CS. Change with age of the refractive index gradient of the human ocular lens. Investigative Ophthalmology & Visual Science. 1995;36(3):703-7.

399. Medina A. A model for emmetropization: Predicting the progression of ametropia. Ophthalmologica. 1987;194(2-3):133-9.

400. Hung GK, Ciuffreda KJ. Model of human refractive error development. Current Eye Research. 1999;19(1):41-52.

401. Ghorbani Mojarrad N, Williams C, Guggenheim JA. A genetic risk score and number of myopic parents independently predict myopia. Ophthalmic and Physiological Optics. 2018;38(5):492-502.

402. Seipold L, Saftig P. The emerging role of tetraspanins in the proteolytic processing of the amyloid precursor protein. Frontiers in Molecular Neuroscience. 2016;9:149.

403. Charrin S, Jouannet S, Boucheix C, Rubinstein E. Tetraspanins at a glance. Journal of Cell Science. 2014;127(17):3641.

404. Berditchevski F, Odintsova E. Tetraspanins as regulators of protein trafficking. Traffic. 2007;8(2):89-96.

405. Termini CM, Gillette JM. Tetraspanins Function as Regulators of Cellular Signaling. Frontiers in Cell and Developmental Biology. 2017;5(34).

406. van Spriel AB. Tetraspanins in the humoral immune response. Portland Press Ltd.; 2011.

407. Kaji K, Oda S, Shikano T, Ohnuki T, Uematsu Y, Sakagami J, et al. The gamete fusion process is defective in eggs of Cd9-deficient mice. Nature Genetics. 2000;24(3):279-82.

408. Le Naour F, Rubinstein E, Jasmin C, Prenant M, Boucheix C. Severely reduced female fertility in CD9-deficient mice. Science. 2000;287(5451):319-21.

409. Marsay K, Roehl H, Monk P, Partridge L, Carney T, editors. Tetraspanins in zebrafish development. Mechanisms of Development; 2017: Elsevier.

410. Maecker HT, Todd SC, Levy S. The tetraspanin superfamily: molecular facilitators. The FASEB Journal. 1997;11(6):428-42.

411. Wistow G, Berstein SL, Wyatt MK, Farriss RN, Behal A, Touchman J, et al. Expressed sequence tag analysis of human RPE/choroid for the NEIBank Project: Over 6000 non-redundant transcripts, novel genes and splice variants. Molecular Vision. 2002;8(21):205-20.

412. Eschenbrenner E, Jouannet S, Clay D, Chaker J, Boucheix C, Brou C, et al. TspanC8 tetraspanins differentially regulate ADAM10 endocytosis and half-life. Life Science Alliance. 2020;3(1):e201900444.

413. Dornier E, Coumailleau F, Ottavi JF, Moretti J, Boucheix C, Mauduit P, et al. TspanC8 tetraspanins regulate ADAM10/Kuzbanian trafficking and promote Notch activation in flies and mammals. Journal of Cell Biology. 2012;199(3):481-96.

414. Kajiwara K, Berson EL, Dryja TP. Digenic retinitis-pigmentosa due to mutations at the unlinked peripherin/RDS and ROM1 loci. Science. 1994;264(5165):1604-8.

415. Junge HJ, Yang S, Burton JB, Paes K, Shu X, French DM, et al. TSPAN12 regulates retinal vascular development by promoting Norrin-but not Wnt-induced FZD4/β-catenin signaling. Cell. 2009;139(2):299-311.

416. Rosner M, Belkin M. Intelligence, education, and myopia in males. Archives of Ophthalmology. 1987;105(11):1508-11.

417. Sperduto RD, Seigel D, Roberts J, Rowland M. Prevalence of Myopia in the United States. JAMA Ophthalmology. 1983;101(3):405-7.

418. Lawlor DA, Tilling K, Davey Smith G. Triangulation in aetiological epidemiology. International Journal of Epidemiology. 2017;45(6):1866-86.

419. Patton MQ. Enhancing the quality and credibility of qualitative analysis. Health services research. 1999;34(5 Pt 2):1189.

420. Silles M. The causal effect of education on health: Evidence from the United Kingdom. Economics of Education Review. 2009;28(1):122-8.

421. He M, Xiang F, Zeng Y, Mai J, Chen Q, Zhang J, et al. Effect of time spent outdoors at school on the development of myopia among children in China: A randomized clinical trial. JAMA. 2015;314(11):1142-8.

422. Jones LA, Sinnott LT, Mutti DO, Mitchell GL, Moeschberger ML, Zadnik K. Parental History of Myopia, Sports and Outdoor Activities, and Future Myopia. Investigative Ophthalmology & Visual Science. 2007;48(8):3524-32.

423. Rose KA, Morgan IG, Smith W, Burlutsky G, Mitchell P, Saw S-M. Myopia, lifestyle, and schooling in students of Chinese ethnicity in Singapore and Sydney. Archives of ophthalmology. 2008;126(4):527-30.

424. Wu P-C, Tsai C-L, Wu H-L, Yang Y-H, Kuo H-K. Outdoor activity during class recess reduces myopia onset and progression in school children. Ophthalmology. 2013;120(5):1080-5.

425. Wu P-C, Chen C-T, Lin K-K, Sun C-C, Kuo C-N, Huang H-M, et al. Myopia Prevention and Outdoor Light Intensity in a School-Based Cluster Randomized Trial. Ophthalmology. 2018;125(8):1239-50.

426. McBrien NA, Moghaddam HO, New R, Williams LR. Experimental myopia in a diurnal mammal (Sciurus carolinensis) with no accommodative ability. The Journal of Physiology. 1993;469(1):427-41.

427. Schmid KL, Wildsoet CF. Effects on the compensatory responses to positive and negative lenses of intermittent lens wear and ciliary nerve section in chicks. Vision Research. 1996;36(7):1023-36.

428. Plotnikov D, Guggenheim JA. Mendelian randomisation and the goal of inferring causation from observational studies in the vision sciences. Ophthalmic and Physiological Optics. 2019;39(1):11-25.

429. Burgess S, Small DS, Thompson SG. A review of instrumental variable estimators for Mendelian randomization. Statistical Methods in Medical Research. 2017;26(5):2333-55.

430. Burgess S, Davey Smith G, Davies N, Dudbridge F, Gill D, Glymour M, et al. Guidelines for performing Mendelian randomization investigations [version 2; peer review: 1 approved, 1 approved with reservations]. Wellcome Open Research. 2020;4(186).

431. Darst BF, Koscik RL, Racine AM, Oh JM, Krause RA, Carlsson CM, et al. Pathway-specific polygenic risk scores as predictors of Amyloid-β deposition and cognitive function in a sample at increased risk for Alzheimer's disease. Journal of Alzheimer's disease. 2017;55(2):473-84.

432. Chen Y, Niu Y, Ji W. Genome editing in nonhuman primates: approach to generating human disease models. Journal of Internal Medicine. 2016;280(3):246-51.

433. Murru L, Moretto E, Martano G, Passafaro M. Tetraspanins shape the synapse. Molecular and Cellular Neuroscience. 2018;91:76-81.

434. Salas IH, Callaerts-Vegh Z, Arranz AM, Guix FX, D'Hooge R, Esteban JA, et al. Tetraspanin 6: A novel regulator of hippocampal synaptic transmission and long term plasticity. PLOS ONE. 2017;12(2):e0171968.

435. Bassani S, Cingolani LA, Valnegri P, Folci A, Zapata J, Gianfelice A, et al. The X-linked intellectual disability protein TSPAN7 regulates excitatory synapse development and AMPAR trafficking. Neuron. 2012;73(6):1143-58.

436. Park YK, Goda Y. Integrins in synapse regulation. Nature Reviews Neuroscience. 2016;17(12):745-56.

437. Usardi A, Iyer K, Sigoillot SM, Dusonchet A, Selimi F. The immunoglobulin-like superfamily member IGSF3 is a developmentally regulated protein that controls neuronal morphogenesis. Developmental neurobiology. 2017;77(1):75-92.

438. Noble EP. D2 dopamine receptor gene in psychiatric and neurologic disorders and its phenotypes. American Journal of Medical Genetics Part B: Neuropsychiatric Genetics. 2003;116(1):103-25.

439. Richter FG, Fendl S, Haag J, Drews MS, Borst A. Glutamate Signaling in the Fly Visual System. iScience. 2018;7:85-95.

440. Nebbioso M, Plateroti AM, Pucci B, Pescosolido N. Role of the Dopaminergic System in the Development of Myopia in Children and Adolescents. Journal of Child Neurology. 2014;29(12):1739-46.

441. Tian N, Xu H-p, Wang P. Dopamine D2 receptors preferentially regulate the development of light responses of the inner retina. The European journal of neuroscience. 2015;41(1):17-30.

442. Jouannet S, Saint-Pol J, Fernandez L, Nguyen V, Charrin S, Boucheix C, et al. TspanC8 tetraspanins differentially regulate the cleavage of ADAM10 substrates, Notch activation and ADAM10 membrane compartmentalization. Cellular and Molecular Life Sciences. 2016;73(9):1895-915.

443. Ferrario JE, Baskaran P, Clark C, Hendry A, Lerner O, Hintze M, et al. Axon guidance in the developing ocular motor system and Duane retraction syndrome depends on Semaphorin signaling via alpha2-chimaerin. Proceedings of the National Academy of Sciences. 2012;109(36):14669.

444. Chilton JK, Guthrie S. Axons get ahead: Insights into axon guidance and congenital cranial dysinnervation disorders. Developmental Neurobiology. 2017;77(7):861-75.

445. Perron JC, Bixby JL. Tetraspanins expressed in the embryonic chick nervous system. FEBS Letters. 1999;461(1):86-90.

446. Tole S, Patterson PH. Distribution of CD9 in the developing and mature rat nervous system. Developmental Dynamics. 1993;197(2):94-106.

447. Kopczynski CC, Davis GW, Goodman CS. A neural tetraspanin, encoded by *late bloomer*, that facilitates synapse formation. Science. 1996;271(5257):1867.

Appendix.

Appendix papers:

Plotnikov D & Guggenheim JA. Mendelian randomisation and the goal of inferring causation from observational studies in the vision sciences. *Ophthalmic and Physiological Optics* 2019; 39: 11– 25. https://doi.org/10.1111/opo.12596 (Chapter 2)

Plotnikov D, Williams C, Guggenheim JA. Association between birth weight and refractive error in adulthood: a Mendelian randomisation study. *British Journal of Ophthalmology*; Published Online First: 16 May 2019. http://dx.doi.org/10.1136/bjophthalmol-2018-313640 (Chapter 3)

Plotnikov D, Shah RL, Rodrigues JN, Cumberland PM, Rahi JS, Hysi PG, Atan D, Williams C, Guggenheim JA; UK Biobank Eye and Vision Consortium. A commonly occurring genetic variant within the NPLOC4-TSPAN10-PDE6G gene cluster is associated with the risk of strabismus. *Human Genetics.* 2019 Jul; 138(7):723-737. https://doi.org/10.1007/s00439-019-02022-8 (Chapter 4)