Genetic insights into the epidemiology of cataracts, prevention and alternative treatment

A Thesis Submitted for the Degree of Doctor of Philosophy

By

Munisa Hashimi

Department of Biosciences,
Brunel University London

2025

Abstract

Cataract, a leading cause of visual impairment and blindness, remains a significant global health challenge, particularly in the context of an aging population. As the global population continues to age, the burden of cataract on healthcare systems, especially in developing countries, is expected to increase. Cataract involves the clouding of the lens, and surgical extraction remains the sole treatment option. Understanding the genetic mechanisms underlying cataract, identifying preventive measures, and exploring alternative treatments are critical to reducing this burden.

This thesis utilised UK Biobank generated genome-wide association study (GWAS), and publicly available GWAS data to investigate the shared genetic mechanisms between cataract subtypes and cataract-associated risk factors. It also assessed alcohol consumption, vitamin D levels and deficiency, and lanosterol as potential modifiable risk factors or alternative treatment options.

Genetic correlations were identified between overall cataract and type 2 diabetes (T2D), asthma and diabetic cataract, senile and diabetic cataract, and asthma and overall cataract. Co-localisation analysis highlighted genes of interest, including *WWP2* and *CDKN2B-AS1* between overall cataract and T2D, and *HLA-DQB1* between asthma and overall cataract. Mendelian randomisation analyses found no evidence of a causal relationship between vitamin D levels, vitamin D deficiency, or alcohol consumption and cataract. Similarly, lanosterol was not supported as a viable alternative treatment option.

In summary, while this study identified genetic links between cataracts and associated risk factors; it did not provide supporting evidence for vitamin D, alcohol, and lanosterol as effective preventive measures or alternative treatment options.

Declaration

I, Munisa Hashimi, hereby declare that this thesis, titled - Genetic Insights into the Epidemiology of Cataracts, Prevention and Alternative Treatment, is an original work.

All work presented in this thesis, unless otherwise referenced or acknowledged, is my own.

Acknowledgements

I would first like to thank Dr Fotios Drenos and Dr Alexander C Day for their incredible supervision throughout my PhD.

Dr Alexander C Day provided direction, guidance, and clinical relevance to my work, ensuring that each result was carefully considered. His attention to detail helped me maintain the highest standards in my research, and his unwavering support enabled me to overcome many challenges along the way.

I first met Dr Fotios Drenos during my undergraduate dissertation, and our collaboration truly transformed my professional journey. It was during this time that I was introduced to the world of data research, an experience that profoundly influenced my decision to pursue a PhD. Dr Drenos showed me what it means to be a successful scientist, while consistently embodying kindness, humility, and genuine care for his students. He has shaped my understanding not only of what it means to be a professional, but also of the kind of person others want to work with.

Beyond the invaluable scientific guidance he has provided over the years, Dr Drenos also offered steadfast support in my journey into public involvement and engagement. He has continuously encouraged me through both successes and setbacks, and for that, I will be forever grateful.

I would further like to thank Dr Hasnat A Amin, who I had the opportunity to work with since my dissertation project. Dr Amin provided support, mentorship and guidance throughout my academic journey.

I would like to thank all my friends and family, but especially the following individuals.

First, I would like to thank my parents, Mohammed Abdul Awal and Momata Begum, for their unwavering support over many years. Higher education is a privilege, one that is often taken for granted. My parents worked incredibly hard to provide their children with the best opportunities in life, something I will never take lightly. Their sacrifices have enabled me to pursue my education to the highest levels.

I would also like to thank my niece, Amelia Eva Rahman, for bringing so much joy into my life. She continues to inspire me to be the best scientist I can be. Through her aspirations for me as her aunt, I strive to live up to her expectations and lead by example.

And of course, my sister, Lily Begum, who was the big sister I never knew I needed, thank you for helping to create a path for me to reach my achievements.

I would also like to thank Annamaria Orsi for her support over so many years, for always reminding me what is truly important in life, celebrating every success, and cheering me on through it all.

Lastly, and most importantly, I would like to thank Luca Makeham. Without you, this thesis and all my successes would not have been possible. Your constant support has carried me through every step of this journey more than anyone else. I share this achievement with you.

Publication section

Published manuscripts

The following publications were produced from work conducted as part of this thesis:

Hashimi, M. *et al.* (2024) 'Using genetics to investigate the association between lanosterol and cataract', *Frontiers in Genetics*, 15. Available at: https://doi.org/10.3389/fgene.2024.1231521.

Hashimi, M. *et al*. (2025) 'Exploring the causal relationship between vitamin D levels and deficiency with the risk of cataract: A Mendelian Randomisation study'*, Karger Ophthalmic Research.* Available at: https://doi.org/10.1159/000545332

In Preparation

Hashimi, M. *et al*. 'Investigating the Relationship Between Alcohol Consumption and Cataract Risk: Findings from Observational and Genetic Analyses'*, pending*

Hashimi, M. *et al*. 'Genetic correlation and co-localisation analysis between cataract subtypes and risk factors*, pending*

Table of Contents

1 Introduction

Cataract is one of the leading causes of blindness and is estimated to account for 1 in 3 cases of blindness (Khairallah *et al.*, 2015; Flaxman *et al.*, 2017). In the UK, 25% of the population are predicted to develop cataracts before the age of 75 (Frampton *et al.*, 2014). With surgery the only effective treatment for cataract, internationally it is estimated 26 million cataract surgeries are performed each year, placing immense stress on health care systems (Chen *et al.*, 2021). The prevalence of cataract surgery is also increasing due to an ageing population amongst other environmental factors (Purola *et al.*, 2022). The cost of operations can be estimated by private procedure fees ranging from £2500-£5000 per eye (Watford, 2020). According to the NHS Payment Scheme, the unit price for intermediate to complex cataract procedures can range from £1130 to £2653 (NHS England, 2023). A combination of these statistics highlights the importance of establishing prevention methods for cataracts.

1.1 Structure of the eye

The eye is a small but complicated organ, with its structure in Figure 1.1. Vision occurs when light passes through the pupil to the retina. Light is refracted by the cornea and lens and the iris regulates the amount of light entering the eye. The lens focusses the light onto the retina (Willoughby *et al.*, 2010). The rods and cones are photoreceptors located in the retina. The function of rod cells is to respond to low light settings, while cone cells are responsible for functioning under higher light intensity and producing high visual acuity and colour vision. When light is directed onto the retina, the internal outer segment structure of the rods and cones undergo a process of phototransduction to produce electrical signals that reach the brain via the optic nerve (Willoughby *et al.*, 2010; Molday and Moritz, 2015). Information is processed through segments of the visual cortex to produce an image (Huff, Mahabadi and Tadi, 2024).

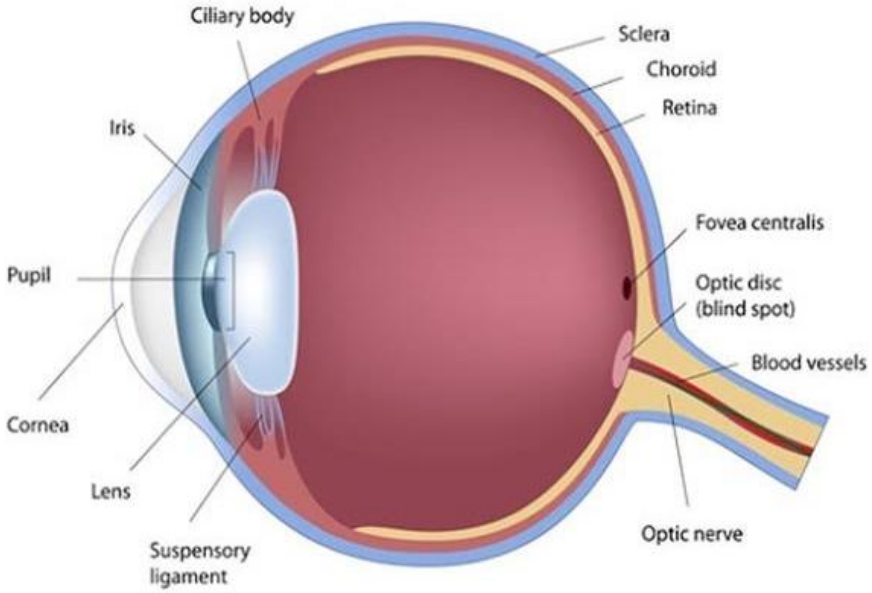


Figure 1.1: A labelled diagram of the eye. The diagram shows several features of the eye, such as the ciliary body, retina, and lens (Infinite Eyecare, 2020).

1.2 Pathology and aetiology

Cataract affects the lens of the eye. As shown in Figure 1.1, the lens is located behind the iris. During the development of cataract, the transparent lens becomes progressively cloudy, causing a loss of clarity in vision. Cataract development can be observed through lens opacifications patterns which are clinically described as nuclear, cortical and posterior subcapsular and are thought to represent different pathogenesis. Cataract symptoms range depending on subtype or stage of development, as noted in Section 1.3. In the early stages of cataract development, symptoms can include changes in the refractive state of the eye, loss of contrast sensitivity, progressing to visual impairment and even blindness when very advanced (Nizami and Gulani, 2021).

Whilst the most common type of cataract is age-related, risk factors, such as diabetes, trauma, steroid intake, sunlight exposure or rarer causes such as malnutrition, can induce early-onset cataracts (Praveen *et al.*, 2010; Liu *et al.*, 2017).

1.3 Different types of cataracts and detailed pathogenesis

1.3.1 Pathogenesis

The transparency of the lens is crucial to perform its function. The lens is composed of different layers, which include the nucleus, cortex, and capsule. The nucleus and cortex are lined with fibre cells. Fibre cells originate from epithelial cells located in the anterior part of the lens. These epithelial cells differentiate and elongate into fibre cells, when there is a

high expression of soluble crystallin proteins synthesised (Moreau and King, 2012). The α-crystallin and β-crystallin are formed from the amino acid residue polypeptides encoded by the αA and αB genes. The functions of the proteins include the maintenance of the refractive index and transparency of the lens (Horwitz *et al.*, 1999).

The lens has various functions to avoid light scattering structures forming on the tissue, such as the crystallin proteins. Crystallin proteins can be found in abundance on the lens of the eye and fill the gaps between the fibre cells. The same crystallins are present in the lens from birth and throughout an individual's life without being replaced or repaired, as the mature fibre cells lacks the cellular machinery for protein turnover, thus the same crystallin proteins are responsible for maintaining the function of the lens (Wistow, 2012). For example, the fibre cells are packed and compressed to avoid intercellular substances. Organelle degradation processes remove organelles, such as mitochondria and nuclei, during cell maturation. The crystallin proteins formed are structured and organised to help with lens transparency and avoid crystallisation (Moreau and King, 2012).

However, protein build-up occurs when crystallin proteins are damaged, leading them to misfold and aggregate into insoluble clumps. Mature fibre cells are unable to remove damaged proteins from the lens. This protein damage occurs over time and leads to cataract development. The primary function of the crystallin proteins is to maintain solubility of other lens proteins (Makley *et al.*, 2015).

1.3.2 Classification based on lens opacification location

Cataract can be classified by their aetiology and by the location of opacification on the lens. Morphological classifications of cataract can include nuclear sclerotic, cortical, and posterior subcapsular cataract (Liu *et al.*, 2017). Nuclear sclerotic cataract is the most common type and first occurs in the nucleus of the lens. In the eye, nuclear cataract can be seen as the hardening and yellowing of the lens nucleus, which progressively spreads to the remaining areas. Although initial symptoms can include the improvement of near-sighted vision (myopic refractive shift), if left untreated in later stages, can result in worsening vision (Sabhapandit, 2019; Albert and Gamm, 2024). A mechanism of nuclear cataract involves the deposition of urochrome in-between fibres (Nizami and Gulani, 2021).

In contrast, cortical cataracts are more common for younger patients exposed to conditions such as diabetes. Cortical cataract occurs on the outer edge of the lens, moving towards the nucleus (Albert and Gamm, 2024), creating a bicycle wheel spoke-like appearance. The mechanism of cortical cataracts is associated with cortical hydration occurring between lens fibres (Nizami and Gulani, 2021).

Posterior subcapsular cataract occurs towards the back of the lens and is characterised by a small opaque area that blocks light, causing glare both in sunlight and at night. Both diabetes and higher blood pressure have been previously observed as risk factors for posterior subcapsular cataract (Richter *et al.*, 2012). The cataract progresses much faster than other forms, and more commonly presents symptoms such as difficulty focusing on objects, due to its effect on directing light onto the retina (Fong, 2008). The key mechanism causing subcapsular cataract is defined through the fibrous metaplasia of the lens epithelium (Nizami and Gulani, 2021).

1.3.3 Classification based on cause

Age-related cataracts

Being the most common form of cataract, age-related cataracts can begin from the age of 40 but develops slowly, typically leading to a noticeable disturbance in vision from 60 onwards (Truscott and Friedrich, 2019). Environmental factors such as UV light, corticosteroid use and alcohol consumption have previously been associated with an increased risk of developing age-related cataracts (Ang and Afshari, 2021; Cicinelli *et al.*, 2023).

Dysfunctional lens syndrome (DLS) describes the natural aging of the crystalline lens and can be used to characterise a spectrum of age-related conditions, including cataract, affecting the lens through three broad stages (Fernández *et al.*, 2018; Waring IV, 2020). The first stage is defined as presbyopia and typically occurs between the ages of 42 to 50 and is characterised by loss of accommodation, leading to difficulty focusing on near objects (Fernández *et al.*, 2018). Presbyopia is widely accepted as the increased stiffness of the crystalline lens, leading to a loss of focusing power and near vision (Medeiros, 2016; Singh and Tripathy, 2023). The condition is the most common cause of visual impairment in older adults, with a 2008 population-based survey from the Brazilian Amazon discovering that presbyopia accounted for 71.8% of total cases of visual impairment, with cataract only accounting for 16.5% (Holden *et al.*, 2008; Singh and Tripathy, 2023). The second stage of DLS typically occurs for individuals aged 50 years or older and involves the progressive decline in accommodation, accompanied by an increase in ocular scatter and optical aberrations, resulting in further deterioration of visual quality (Medeiros, 2016; Fernández *et al.*, 2018; Motlagh and Geetha, 2022). The final stage of DLS is the full development of cataract through significant opacity and aberrations, severely impacting visual quality and leading to potential blindness (Medeiros, 2016; Waring IV, 2020).

Age-related cataracts can involve complex genetic effects, where multiple genetic variants interact with environmental factors to influence its progression (Shiels and Hejtmancik, 2015).

Metabolic cataracts

Every two minutes in the UK, an individual is diagnosed with diabetes, most commonly type 2 diabetes (T2D) (Diabetes UK, 2017). Patients with diabetes have a greater chance of developing cataracts. Causes of metabolic cataracts include the association with the aldose reductase pathway leading to osmatic stress on the lens. The aldose reductase enzyme in diabetic patients converts glucose into sorbitol and then to fructose. In the lens, the production of sorbitol exceeds the amount of sorbitol converting to fructose. The increase in sorbitol on the lens interferes with the osmotic gradient, leading to damage to the lens fibres, resulting in lens opacities (Pollreisz and Schmidt-Erfurth, 2010).

Congenital cataracts

Although cataracts commonly occur in the age-related form, some cases occur in infancy. Congenital cataracts can be described clinically as the development of lens opacity due to stressors applied to lens proteins *in utero* and is identified following birth. Different types of congenital cataracts include anterior polar, posterior polar, nuclear, and cerulean cataracts. Additional issues, such as amblyopia and nystagmus, can also occur due to congenital cataracts (Taylor, 1998; Bell *et al.*, 2020; Kim *et al.*, 2023). Amblyopia, also referred to as “lazy eye”, is a typically unilateral disorder characterised by a developmental disadvantage from one eye over another. In comparison, nystagmus is the involuntary and rapid oscillatory movement of the eye. Both secondary conditions are associated with poorer visual acuity (Sekhon, Rocha Cabrero and Deibel, 2023; Blair *et al.*, 2024).

The aetiology of congenital cataracts varies and is complex. The mechanism of development is associated with disturbances that occur during the development of the lens (Nizami and Gulani, 2021). A study conducted in India highlighted causes such as trauma, secondary disease, both ocular or systemic, and congenital rubella infection, which are also preventable. Furthermore, for the cases categorised as idiopathic, 67% of mothers had a history of family illness such as pulmonary tuberculosis and arthritis, and 22% were taking non-specific medication during pregnancy. Genes associated with congenital cataracts include crystallin proteins, gap junction channel protein, membrane protein, cytoskeletal protein, transcription factor, ferritin light and fibroblast growth factor genes (Johar *et al.*, 2004; Yi *et al.*, 2011).

1.3.4 Treatment and prevention

Currently the only effective treatment for cataract is surgery, typically involving making a small incision at the edge of the cornea, opening of the anterior capsule so the lens can be removed and replaced with a clear artificial lens (termed intraocular lens). Although different compounds, such as lanosterol and N-acetylcarnosine, have shown promise in animal or pre-clinical studies as potential cataract treatments, there is currently insufficient evidence from large-scale clinical trials to confirm their safety and efficacy in reversing established cataracts (Zhao *et al.*, 2015, 2021; Dubois and Bastawrous, 2017). Therefore, cataract surgery remains the only proven and effective treatment option. The success rate of cataract surgery is greater than 95%, with success defined as no instances of complications. For reported unsuccessful cases, this is commonly due to postoperative issues such as infection and retinal detachment (LESH, 2017).

In the case of treating congenital cataract, surgery is undertaken between 6 weeks and 3 months of age, whereby an intraocular lens is typically implanted to replace the crystalline lens. Following surgery, the child may still be required to wear spectacles to correct any residual refractive error to ensure as close to normal visual development as possible (Drack, 2005; Vijayalakshmi and Njambi, 2016; Self *et al.*, 2020).

The Royal College of Ophthalmologists' National Ophthalmology Database (RCOphth NOD) investigated surgical outcomes of 127,685 patients undergoing cataract surgery between 2006 and 2010. The study measured intraoperative/postoperative complications and preoperative/postoperative visual acuities across these individuals. The study highlighted the risk of complications during cataract surgery, identifying intraoperative complications across 4.2% of patients, with posterior capsular ruptures (PCR) being the most common, at 1.9% (Day *et al.*, 2015). More recently, the NOD Audit annual report revealed that the PCR rate has fallen to 0.79% in 2022 (Donachie *et al.*, 2024). Furthermore, data published by the European Society of Cataract & Refractive Surgeons (ESCRS) found that in 2022 across 271,387 observed cataract surgeries, 0.76% displayed postoperative complications (Behndig *et al.*, 2023) Among patients included in the study by Day *et al.*, 0.03% of 139 537 cases required additional surgery for retinal detachment, and 0.03% of 145,868 cases developed endophthalmitis within three months following cataract surgery. The risk of postoperative complications was significantly higher in patients who experienced a PCR, with the risk of retinal detachment and endophthalmitis being 42 and 8 times higher, respectively (Day *et al.*, 2015).

Advances in technology and methodology have significantly reduced the number of cases experiencing complications. According to the 2024 NOD Audit annual report, since 2010

there has been a 58% reduction in PCR incidences to 0.79%, during cataract surgery (Donachie *et al.*, 2024). Similar trends are also observed within the Swedish National Cataract Register (NCR), where from 1992 to 2021 complications associated with endophthalmitis and PCR have decreased from 0.10% to less than 0.02% and 2.8% to 0.6%, respectively (Bro *et al.*, 2023). For a developed country such as the UK, cataract surgery is extremely common, being the most common surgery conducted by the NHS. In 2022, approximately 608,000 cataract surgeries were performed by the NHS in England (Donachie *et al.*, 2024). Although the surgery is cost-free for patients under the NHS, the necessary criteria to be eligible for it can lead to patients waiting until their vision has severely deteriorated and normal daily activities, such as reading, become increasingly difficult. Furthermore, under the NHS, patients are typically only offered monofocal lenses resulting in individuals requiring corrective eyewear post-operatively for reading (Watford, 2020). However, while some NHS providers are beginning to provide more premium options, advanced intraocular lenses (IOLs), such as extended depth of focus (EDOF) and multifocal IOLs, are available in private healthcare settings, and can provide greater spectacle independence for both distance and near vision (CHEC, 2024; Kabbani *et al.*, 2024).

The success rate of cataract surgery in developing countries is significantly lower. This can be attributed to both resource availability and surgical techniques utilised. While the use of phacoemulsification, where ultrasound is used to emulsify and remove the lens, with IOL implantation is considered the gold standard to cataract surgery, it comes at a significant cost with expensive technology and human resource required. This makes the technique available only to countries with a developed health infrastructure and significant economic resources. In contrast, alternative techniques such as Small Incision Cataract Surgery (SICS) and Extra Capsular Cataract Extraction (ECCE) are generally accepted in developing countries where resources and infrastructure are minimal. While these practices are accepted they are less effective than those used across the developed world (Gogate, 2010; Malhotra *et al.*, 2014). Among all eye diseases, the blindness rate caused by cataracts in poor and remote regions is estimated to be greater than 50%, compared to 5% in developed countries. Across some African regions, access to cataract services is estimated to be a tenth of what is available for high-income countries (Chen *et al.*, 2021).

A recent study conducted at an eye hospital in Pakistan, covering the period from 2010 to 2020, evaluated the outcomes of 38,616 cataract surgeries. The findings showed that 4.26% of patients had severe visual impairment (< 6/60), while 13.86% experienced moderate visual impairment (< 6/18 to 6/60). Similarly, earlier studies have reported comparable findings, such as an earlier study conducted in Pakistan observed cataract

surgery in 1,317 subjects. The results of the study showed a third of patients having a visual acuity of less than 6/60 (Bourne *et al.*, 2006). A similar study conducted in Bangladesh observed cataract surgery in 12,782 subjects. The study concluded that the use of eye camps (non-governmental organisations performing primarily intracapsular cataract surgery) to perform cataract surgery correlated with the number of cases that resulted in visual acuity of less than 6/60 (Bourne *et al.*, 2003). Both studies highlighted the need for drastic improvement in the quality and quantity of surgeries in regulated and safer settings (Bourne *et al.*, 2003, 2006). This was supported further by a 2019 systematic review of cataract surgeries in low and middle income countries which highlighted an inadequate number of operations performed, quoting both supply issues due to poor infrastructure and weak health systems as well as demand issues surrounding fear of surgery and insufficient incomes (Mailu *et al.*, 2020).

More recent studies have shown improvements in surgery; however, complications can still occur (Chan, Mahroo and Spalton, 2010; Naeem *et al.*, 2012). For example, an analysis of cataract surgery outcomes within India was conducted in the Aravind Eye Hospital in Tamil Nadu. The data used in the study was collected between January 2012 and December 2018 and consisted of 1.86 million cataract surgeries (Ravindran *et al.*, 2021). SICS is commonly preferred in settings with high demands and limited access to surgical instruments, such as developing countries, as the cost and time of performing a SICS is significantly less than the phacoemulsification technique, a more modern cataract surgery (Bhargava *et al.*, 2015). The results of the study highlighted an increased use of the phacoemulsification surgical technique as opposed to SICS. However, the phacoemulsification was only offered to patients who were paying for the surgery, and its increased use may be attributable to an increase in paying capacity and insurance coverage. Overall, the study found improvements for visual acuity outcomes and a decreased rate of intraoperative complications (Ravindran *et al.*, 2021). Although surgery in hospitals have fewer complications, this may not be an available option, or could be too expensive, for some patients.

General prevention of age-related cataract can include wearing sunglasses/hats, to block sunlight, when exposed to the sun to reduce UV exposure, quitting smoking and eating healthily. These common preventive measures and other environmental and lifestyle factors, such as air pollution, hypertension, and alcohol consumption, have also been found to be associated to reduced cataract risk (Yu *et al.*, 2014; S. Y. Chua *et al.*, 2021; S. Y. L. Chua *et al.*, 2021). Additional lifestyle adjustments, such as updating glasses/contacts prescription, using bright lights or magnifying glasses to help conduct tasks such as reading, can also aid in cataract management (NEI, 2024). However, the effectiveness of prevention

and management techniques is currently unclear, and more work is required to validate which risk factors can be targeted as potential preventive measures against cataract risk.

1.4 Risk factors

1.4.1 Non-modifiable risk factors

Numerous modifiable risk factors have previously been suggested as associated with cataract development. While modifiable risk factors are important for establishing interventions, non-modifiable risk factors can help raise awareness for affected groups. A combination of modifiable and non-modifiable risk factors would help identify targeted interventions (Ho *et al.*, 2020). Most notably, studies have found that an increase in age significantly raises the risk of cataract incidence (Nirmalan *et al.*, 2004; Chen *et al.*, 2020). A cross-sectional population-based study of 5,150 participants across southern India found 79.4% of individuals aged 70 years and above to have age-related cataracts, compared to 15.7% amongst of those aged between 40-49. Although a small number of individuals had early-onset cataracts, the difference in cataract diagnoses between groups may be due to risk factors not accounted for in the study (Nirmalan *et al.*, 2004).

Myopia, a refractive error within the eye that results in short-sightedness, is a recently suggested risk factor for age-related cataracts (Chakraborty, Read and Vincent, 2020; Hugosson and Ekström, 2020). The association between myopia and age-related cataracts was investigated in a systematic review and meta-analysis on 12 population-based studies across 38,007 participants aged between 30-97. The study's results suggested an association between myopia and prevalent nuclear, possibly resulting from myopic shift, and posterior subcapsular cataracts; however, no association was discovered with cortical cataract. Although the study was conclusive for prevalent cataract subgroups, the study's limitations, including the difficulty in grading posterior subcapsular cataracts, lack of cohort studies, and potential biases present in the original population-based studies, may affect the results from the meta-analysis (Pan *et al.*, 2013). Overall, a potential bias in myopia investigations with cataract incidence may be due to patients with myopia being more likely to go to the opticians, thus increasing the likelihood of cataract detection and diagnosis. Other studies have also found an association between myopia and the risk of age-related cataracts, although more research is required to better understand the potential causal association (Younan *et al.*, 2002; Hugosson and Ekström, 2020).

Alongside myopia and age, other non-modifiable risk factors have been associated with age-related cataracts such as biological sex (Chen *et al.*, 2020).

Female sex is currently considered a non-modifiable risk factor due to the potential influence of oestrogen in cataract development. It is suggested that women are more susceptible to cataract formation post-menopause due to a decline in oestrogen production. Oestrogen possesses several properties, including anti-oxidative effects, which are known to prevent lens opacifications (Zetterberg and Celojevic, 2015). This is supported by further studies that also imply the role of the female sex in cataract risk (Zetterberg and Celojevic, 2015; Lou *et al.*, 2018; Chen *et al.*, 2020). In addition, hormone replacement therapy is considered to act as an intervention to decrease cataract risk (Aina *et al.*, 2006; Lai *et al.*, 2013). However, as previously discussed, aging is a known risk factor for cataract; therefore, as women typically have a longer life expectancy than men, this could explain observed associations between biological sex and cataract risk (Zarulli *et al.*, 2018).

Studies have also investigated the influence of ethnicity on cataract risk. It has been observed that specific ethnicities may be more susceptible to cataract formation after adjusting for known cataract risk factors (Storey *et al.*, 2013; Chua *et al.*, 2015; Awidi *et al.*, 2024; Patnaik *et al.*, 2024). However, studies are limited due to their small sample size and lack of accounting for all known environmental risk factors, for example, participants' diets (Storey *et al.*, 2013). However, correlations between ethnic groups and cataract risk may result from genetic variation between groups. Unless all non-genetic environmental exposures are accounted for, results would remain a correlation, vulnerable to the effect of external confounders.

Inherited forms of cataract (commonly categorised as congenital, infantile, or juvenile) occur between birth and up to the age of 40. Cataracts can be inherited via several modes of Mendelian inheritance, such as autosomal dominant or autosomal recessive patterns. Currently, 42 genomic loci have been associated with inherited primary cataract, with primary indicating no involvement of a secondary disease. 12 of these loci have no identified gene, known as "orphan" loci. The remaining 30 known genes can be categorised into one of the following: cytoplasmic crystalline, membrane proteins, cytoskeletal proteins, and DNA/RNA-bounding proteins. Some genes have been found to have underlying associations with both inherited and age-related cataracts. These genes include *EPHA2, GJA3, GJA8, MIP, HSF4, LIM2* and *CRYAA.* Notably, *EPHA2* was found in multiple ethnic populations for cortical and posterior subcapsular cataracts. Identifying genes and loci associated with inherited cataracts would improve our ability to provide personal diagnosis and enhanced genetic counselling for individuals and effected families. The diagnosis of inherited cataract can help families become better informed about their medical future. Going forward, a greater understanding of these genes could allow for molecular genetic

links to be established to provide knowledge on genes that could influence the predisposition of age-related cataract (Shiels and Hejtmancik, 2015).

1.4.2 Modifiable risk factors

Diabetes

Diabetes is caused by the body's inability to use or produce insulin, resulting in a hyperglycaemic state. There are different forms of the disease, which include gestational, type 1 and type 2 diabetes. Most diabetes cases are related to type 2 and are believed to result from risk factors such as obesity and lack of physical exercise (American Diabetes Association, 2010). Diabetes is considered the direct cause of 1.6 million deaths in 2021 and is a significant risk factor for several conditions such as strokes and diabetic retinopathy resulting in blindness (World Health Organisation, 2024). A consistent association between cataracts and diabetes has previously been suggested (Li, Wan and Zhao, 2014). Landmark studies such as the Blue Mountain Eye Study, the Wisconsin Epidemiological Study and the Beaver Dam Eye Study have explored the relationship between cataracts and diabetes. All three studies concluded some form of association between diabetes and an increased risk of cataracts. However, the studies differ in their results, suggesting only specific types of cataracts are impacted by diabetes (Klein *et al.*, 1995; Klein, Klein and Moss, 1995; Rowe *et al.*, 2000).

The Wisconsin Epidemiological Study, including 2,366 participants, observed an increased risk of cataract surgery amongst diabetic patients. Participants were split into two groups based on individual's age at diabetes diagnosis. A diagnosis before the age of 30 was classified within the younger-onset group, but excluded anyone under the age of 18. The older-onset group were those diagnosed 30 years and older. The baseline assessment included ocular examination to ensure participants had not previously undergone cataract surgery. The participants were re-examined after 4- and 10-years. After the 10-year examination, the study concluded a 27% increase in incidences of cataract surgery for individuals aged 45 years and above and a 44% increase in incidences for those aged 75 years or older. The study explored several confounding factors that may have interfered with results, such as diabetic medication and blood pressure. However, it was not possible to adjust the study for antecedent factors (Klein, Klein and Moss, 1995).

The Beaver Dam Eye Study discovered a higher risk of cortical and posterior subcapsular cataracts amongst diabetic patients. The Blue Mountain Eye study group also found a statistically significant association between posterior subcapsular cataracts and diabetes (Klein *et al.*, 1995; Rowe *et al.*, 2000).

A more recent study conducted in 2018 involving 56,510 diabetic patients from the UK-based Clinical Practice Research Datalink explored the rate of incident cataract and its link to diabetes. The comparison between diabetic patients and a control group highlighted a two-fold increased incident rate of cataract in diabetic participants. The results of the study also suggested that a previous diagnosis of macular oedema, longer duration of having diabetes and poor diabetic control may also lead to an increased risk of cataract (Becker *et al.*, 2018). Other cross-sectional studies have also reported the increased risk of cataract amongst diabetic patients compared to non-diabetic patients. For example, whilst diabetic individuals are at significant risk of developing cataracts earlier than the general population, Memon *et al.* found this risk to increase further as patients become older (Memon *et al.*, 2016). This is further supported by a cross-sectional study conducted across the Longitudinal Ageing Study in India (LASI) which observed diabetic adults, aged above 60, being 1.5 times more likely to develop cataracts after controlling for socio-economic and demographic factors (Khan and Shaw, 2023). Overall, there is an evident increased risk of cataract amongst younger and older sample groups with diabetes. However, current literature may benefit from further studies limiting the effects of confounding factors to establish a direct causal association between diabetes and cataracts. Although it is suggested that better metabolic control and good diabetic management can decrease the risk of cataract, limited studies have been conducted on the association of poor diabetic control with cataract risk (Pollreisz and Schmidt-Erfurth, 2010; Yuan, Wolk and Larsson, 2022).

Obesity

Obesity is defined as a body mass index (BMI) greater than 30 and the accumulation of excessive body fat. Obesity can lead to health issues such as coronary heart disease and type 2 diabetes. The increasing rate of obesity places enormous stress on health care systems (Scarborough *et al.*, 2011; Agha and Agha, 2017; Tiwari and Balasundaram, 2023). The association between obesity and cataract has been investigated in several studies, with varying results. Some studies have suggested an association between obesity and cataracts (Reddy, Giridharan and Reddy, 2012; Pan and Lin, 2014; Ye *et al.*, 2014; Lee *et al.*, 2015; Niazi *et al.*, 2023), whilst others suggest a lack of association (Park *et al.*, 2013; Mohammadi *et al.*, 2017). Pan and Lin conducted a meta-analysis, consisting of 163,013 individuals between the age of 40 to 84, investigating the relationship between obesity and age-related cataract. The study concluded that obesity was associated with a 12%, 34%, and 52% increase in risk of developing nuclear, cortical, and posterior subcapsular cataracts, respectively (Pan and Lin, 2014). A similar meta-analysis conducted to determine the relationship between BMI and age-related cataract found an association between

increasing BMI and the development of posterior subcapsular cataract. Whereas no associations with nuclear or cortical cataracts was identified (Ye *et al.*, 2014). These findings were supported by an additional meta-analysis across 16 studies with a total sample size of 1,607,125 individuals, that found an increasing association between BMI age-related and posterior subcapsular cataract. However, whilst the study also found no association with nuclear cataract, a positive association was observed with cortical cataract (Niazi *et al.*, 2023). A study on a sample group of 3258 individuals, from data derived from the fourth Korea National Health and Nutrition Examination Survey (2009) also found that obesity may lead to a lower risk of cataract formation. Although adjustments were made for confounding factors, such as age and smoking status, the results highlighted disproportionate vitamin levels between the normal-weight and overweight groups, which may have interfered with the observations between obesity and cataract formation (Park *et al.*, 2013).

A Mendelian randomisation (MR) study, a genetic analysis method utilising genetic variants as proxies for risk factors to establish causal relationships, has also been conducted to evaluate a potential causal relationship between obesity and age-related cataract (Tan *et al.*, 2019). The FTO single nucleotide polymorphism (SNP) was used as the marker of genetic predisposition for obesity, and cataract data was collected from surgical information. The FTO SNP is a strong marker for obesity, as shown by several studies associating it to the disease (Berulava and Horsthemke, 2010; Huang, Chen and Wang, 2023). Tan *et al.* found an association was only seen between posterior subcapsular cataract and the FTO SNP when adjusting for protein intake, suggesting a lower protein intake interaction with the FTO SNP may increase the risk of cataract development (Tan *et al.*, 2019).

Corticosteroids

Corticosteroids have a wide range of medical uses, ranging from daily use by asthma sufferers to the treatment of inflammatory bowel disease. The adrenal cortex produces steroid hormones in the form of glucocorticoid and mineralocorticoids. Glucocorticoids have an anti-inflammatory and immunosuppressive effect, while mineralocorticoids affect the renal tube of the kidneys, which allows the regulation of electrolytes and water balance. Corticosteroid medication can be synthetically designed to mirror naturally occurring steroid hormones and has become one of the most prescribed forms of medicine in the US (Raissy *et al.*, 2010; Waljee *et al.*, 2017). The chronic use of corticosteroids, especially glucocorticoids, can lead to a range of side effects, including fractures, diabetes, and cataract (Van Staa *et al.*, 2000; Liu and Manche, 2011; Hwang and Weiss, 2014).

Corticosteroid use has previously been suggested as a risk factor for cataract development. (Ericson-Neilsen and Kaye, 2014). A meta-analysis of randomised control trials (RCTs) and observational studies for rheumatoid arthritis patients investigated the risk of glucocorticoids on cataract development. Although an association was not observed in the RCTs, an association was evident within observational studies, which could imply a lack of cataract examination in the RCTs or the risk of bias in observational studies due to confounding factors (Black *et al.*, 2016). A similar study investigated patients undergoing inhaled corticosteroid therapy for asthma and associated risk factors. The systematic review highlighted a statistically significant 5% increase in cataract development due to the therapy. However, the paper noted the possibility of bias in the studies analysed due to the lack of adjustment for confounders (Patel *et al.*, 2020). A further meta-analysis study investigated the side effects of intranasal corticosteroids use during allergic rhinitis treatment. The study did not find evidence to suggest an association between the treatment and posterior subcapsular cataract (Valenzuela *et al.*, 2019).

Overall, further research in this area is required as studies included in the systematic reviews and meta-analysis possessed various limitations, such as the risk of bias from confounders and data collection.

Vitamin C

Vitamin C, also referred to as L-ascorbate or L-ascorbic acid, has been suggested to be protective within eye health, including reducing the risk of cataract development. Vitamin C is found in high concentrations on the lens and surrounding ocular humors. Vitamin C is believed to protect the eye from oxidative damage caused by exposure to ultraviolet light and aids in the anabolism of other antioxidants, such as vitamin E and lutein (Fan *et al.*, 2020). Antioxidants, such as vitamin C, have previously been shown to improve light-induced oxidative stress through the neutralisation of reactive oxygen species and restoring stability to lens cell membranes (Goyal *et al.*, 2009; Kisic *et al.*, 2012).

Several studies have observed an association between vitamin C intake and the development of cataract (Valero *et al.*, 2002; Rautiainen *et al.*, 2010; Ravindran *et al.*, 2011; Jiang *et al.*, 2019). A case-control study in a Mediterranean population using 347 cases of nuclear, cortical, and posterior subcapsular cataract investigated the role of dietary vitamin C with cataract. Patients diagnosed with nuclear, cortical, or posterior subcapsular cataract were selected with an age range of 55-74. The selected patients underwent interviews and blood sample collections to better understand potential confounding lifestyle factors and their vitamin C intake. The results suggested a greater intake of vitamin C in an individual's diet displayed an inverse association with cataract risk (Valero *et al.*, 2002). Similar results

have been found in other studies, where a low consumption of vitamin C has been associated with an increased risk of cataract development (Ravindran *et al.*, 2011; Wei *et al.*, 2016). Further studies using the UK Biobank cohort have suggested an association between vitamin C intake and cataract risk. A study of 72,160 individuals, cataract-free at baseline, assessed their fruit and vegetable intake using a web-based 24-hour dietary questionnaire between 2009 and 2012. The findings indicated that higher fruit and vegetable consumption was associated with a reduced risk of cataract, potentially due to their high antioxidant content, including vitamin C. However, the study did not examine the effects of vitamin C in isolation (Fan *et al.*, 2023).

However, further research across a meta-analysis of RCTs and cohort studies found that an association was only observed amongst cohort studies. These cohort studies suggested a high consumption of vitamin C was associated with a decreased risk of cataract development. These results, as opposed to those seen in RCTs, suggested potential limitations in both forms of investigations. For example, in RCTs there may be a lack of focus on a cataract outcome and potential confounders may be unaccounted for within cohort studies (Jiang *et al.*, 2019). For example, a cohort study using 24,593 female participants between the ages of 49-83 found that a greater consumption of vitamin C caused an elevated risk of cataract development; however, results were limited due to the interference of confounding factors, such as corticosteroid and hormone replacement therapy use among participants (Rautiainen *et al.*, 2010). Overall, the findings amongst observational studies remain inconsistent, while meta-analyses identify weaker evidence for an association with cataract and vitamin C when using RCTs.

Smoking

Smoking tobacco is a known risk factor for numerous conditions such as respiratory and cardiovascular diseases. Hundreds of harmful toxic chemicals are found in cigarettes and produced as by-products in tobacco smoke. Some substances include carbon monoxide and nicotine, amongst other known carcinogens (West, 2017). Oxidative stress plays a significant role in the pathogenesis of age-related cataract, particularly nuclear and cortical cataract. Chemical toxins from smoking increase oxidative stress on the lens, leading to cataract development (Beebe, Holekamp and Shui, 2010). A meta-analysis was conducted in 2012 by Ye *et al.* to observe the relationship between smoking and age-related cataracts. Following an extensive study selection process, the meta-analysis included 8 case-control studies and 13 cohort studies. The results were statistically significant and suggested an elevated risk of developing age-related cataract for past and present smokers. However, a stronger association was identified for current smokers when compared to past smokers. Further analysis was conducted into the subtypes of cataract, finding both nuclear and

posterior subcapsular cataract to be associated with smoking, but no association with cortical cataract (Ye *et al.*, 2012). A population-based cohort study conducted by Han *et al.* further observed the association between smoking and age-related cataract in diabetic patients, but only for nuclear and subcapsular cataract. The results suggest that oxidative stress may affect the various subtypes of cataract differently (Han *et al.*, 2020). Overall, a clear association exists between smoking (both past and present) and age-related cataract. Whilst the meta-analysis provided strong evidence, literature may benefit from a causal understanding of the relationship and effect on pathogenesis for the different subtypes with smoking.

Air Pollution

Air pollution, defined as the presence of harmful substances in the air, such as particulates, nitrogen dioxide (NO_2) and nitrogen oxides (NO_X), has also been observed to be linked with increased instances of cataract surgery. An observational study on UK Biobank participants discovered a 5% increased risk of cataract surgery after exposure to $PM_{2.5}$ (particulate matter 2.5 micrometres and smaller), NO_2, and NO_X, with the likelihood of cataract surgery becoming progressively higher with greater air pollution exposure. However, in line with other risk factors discussed, the study noted that more research is required to identify whether the association is causal (S. Y. L. Chua *et al.*, 2021).

Hypertension

Hypertension is defined by systolic blood pressure values of 130mmHg or more and diastolic blood pressure values of greater than 80mmHg (Iqbal and Jamal, 2023). Hypertension affects 1.28 billion individuals globally, with an estimated 46% of adults unaware they suffer from the condition. Hypertension is a significant risk factor for several diseases and an important cause of premature death across the globe (World Health Organisation, 2023). A meta-analysis conducted by Yu *et al.* between 1990 and 2014 investigated the role of hypertension in cataract risk. The analysis involved 9 cohorts, 5 case-control and 11 cross-sectional studies. Across the 14 cross-sectional and case-control studies, hypertension significantly increased the risk of developing any cataract type. A cataract subgroup analysis, including 7 studies, found the association present between hypertension and posterior subcapsular and cortical cataracts, but not nuclear. An additional analysis on the 9 cohort studies also found a significant association between cataract risk and hypertension, when considering all cataract types. However, when investigating cataract subgroups across 6 of 9 cohort studies this association was only present for posterior subcapsular cataract, and not cortical or nuclear cataract (Yu *et al.*, 2014). Although, it should be noted that the role of steroid medication was not accounted for during the original observational studies. Steroids can lead to increased hypertension

and are an established risk factor for cataracts, thus a potential confounder (Whitworth *et al.*, 1989; Ericson-Neilsen and Kaye, 2014). Although RCT studies are typically considered to produce the highest level of evidence, they do not rule out limitations from confounders (Petticrew and Roberts, 2003; Spieth *et al.*, 2016; Bhide, Shah and Acharya, 2018). In addition to the meta-analyses, more recent cross-sectional studies have been completed. A study conducted by Mylona *et al.* investigated the prominence of hypertension, along with diabetes and dyslipidaemia, as a risk factor for cataract. The study, including 454 females and 380 males, found the presence of hypertension amongst all three cataract types, both alone and alongside diabetes and dyslipidaemia. Notably, diabetes alone or in combination with dyslipidaemia was only present for posterior subcapsular and nuclear cataract. Although limited as a cross-sectional study, the results highlight the importance of understanding other potential conditions which may be confounders (Mylona *et al.*, 2019). To use hypertension as a modifiable risk factor, it would first need to be established as an independent causal factor through further research.

Ultraviolet radiation exposure - sunlight

Ultraviolet radiation (UVR) is classified as a carcinogen as it has both tumour promotor and initiator properties. UVR is considered a modifiable environmental risk factor for some conditions, such as skin cancer. Although UVR is a risk factor for several diseases, it is crucial in vitamin D synthesis. UVR exposure can occur from sunlight or recreational uses, such as sunbeds (D'Orazio *et al.*, 2013). UVR causes oxidative damage to proteins in the lens cells via oxidative stress. Commonly oxidative stress is the by-product of the process of cells using oxygen to produce energy; however, in the case of UVR, glycation takes place, thus increasing oxidative damage to the lens proteins and leading to cataract development (Linetsky *et al.*, 2014). Studies have established an association between the risk of cataracts and UV exposure (Delcourt *et al.*, 2014; Miyashita *et al.*, 2019; Vashist *et al.*, 2020), although a systematic review or RCT has not yet been achieved between the two. Whilst population-based studies conducted so far provide evidence of an association, a case-control study in a Mediterranean population conducted by Pastor-Valero *et al.* found no association between cataract and UVR exposure in adult life (Pastor-Valero *et al.*, 2007). Studies are limited due to the requirement of self-reported data, which introduces possible inaccuracies in data collection. For example, a questionnaire asking about sunlight exposure may not consider an individual's clothing or presence of eyewear such as sunglasses, which would inherently impact their level of UVR exposure (Rees, 2004; Roberts, 2011). Further research must be conducted to validate the results of observational studies, such as comparing cataract outcomes across countries of varying latitude and common occupations, of which contribute to changes in UVR intensity across populations

(Pastor-Valero *et al.*, 2007; Chang *et al.*, 2009). However, an RCT may not be feasible to see the long-term effects of UVR on cataract due to the time it takes for the cataract to develop.

Traumatic cataracts

A subcategory of cataracts caused by ocular trauma is called traumatic cataract. Examples of ocular trauma include penetrating or blunt injuries to the eye. Cataract development can occur immediately after the instance of trauma due to the lens capsule being ruptured. Even if the lens capsule fails to rupture, direct damage to the lens fibres can still stimulate cataract formation (Diego Zamora-de la Cruz *et al.*, 2016). A prospective study on 48 cases from North India concluded traumatic cataract to be most prevalent in young males. However, the study was limited due to a small sample size and insufficient follow-up period (Sharma *et al.*, 2016). Although surgery is an established treatment for traumatic cataracts, secondary issues such as glaucoma development and surgical complications can lead to permanent visual impairment. A study conducted by Du *et al.* in Shanghai, China, aimed to better understand traumatic cataract cases in children. The investigation included 321 children with a mean age of 6.3 years old. The study concluded that children between 2-4 were most likely to suffer from traumatic cataracts. In these cases, the leading causes of trauma involved metal objects, toys, and wooden sticks. The study's results highlighted the importance of awareness and prevention of traumatic cataracts, notably in the case of children (Du *et al.*, 2018).

Vitamin D

Vitamin D is crucial for regulating calcium, magnesium, and phosphate to allow healthy bone development and maintenance. After conversion, vitamin D is found in the body as 25hydroxyvitamin D (25(OH)D) (Sizar *et al.*, 2023). A healthy level for vitamin D is considered to be 25(OH)D > 50 nmol/L (NICE, 2022). Globally, vitamin D deficiency affects an estimated 1 billion individuals (Palacios and Gonzalez, 2014). Vitamin D deficiency can result from decreased dietary intake and lack of sun exposure, chronic liver disease, medication inducing hepatic p450 enzyme (a catabolic enzyme) and end-organ resistance. Through primarily sun exposure and dietary intake, vitamin D is obtained in the form of ergocalciferol (D2) and cholecalciferol (D3). They are biologically inactive and require enzymatic conversion to reach their active form. Being converted by the enzyme hepatic enzyme 25–hydroxylase in the liver they produce 25-hydroxy-vitamin D2 (25-OH-D2) and 25hydroxy-vitamin D3 (25-OH-D3). The reaction products are converted to 1,25 dihydroxyvitamin D in the kidneys by the enzyme 1-alpha-hydroxylase (Gil, Plaza-Diaz and Mesa, 2018; Chang and Lee, 2019; Sizar *et al.*, 2023).

The anti-inflammatory and antioxidant properties of vitamin D have previously been suggested to decrease oxidative stress and chronic inflammation. Both outcomes are known to aid the pathogenesis of cataract (Öktem and Aslan, 2021).

Vitamin D deficiency has been linked with conditions such as osteomalacia, diabetes, cardiovascular and autoimmune diseases (Holick, 2007; Mailhot and White, 2020; Costenbader, 2022; Zhou, Selvanayagam and Hyppönen, 2022). An association between vitamin D and cataract risk has previously been observed (Abdellah *et al.*, 2019; Öktem and Aslan, 2021). Studies have also suggested a preventive role in increased vitamin D levels through supplementation to reduce the risk of age-related cataract (Jee and Kim, 2015). A case-control study conducted by Öktem and Aslan investigated the risk of early-onset cataracts with vitamin D deficiency. The investigation discovered a statistically significant risk of cataract across the sample of cases when compared to a healthy control group. Although, the study was subject to limitations. For example, there may be reverse causation whereby individuals in the case group stay inside and have limited sunlight exposure due to their cataract, which in-turn causes vitamin D deficiency. This is one of many possible confounders for which the study does not adjust for (Öktem and Aslan, 2021). A similar case-control study by Abdellah *et al.* used a sample of participants aged 50 and above further suggested an association between vitamin D deficiency and the risk of age-related cataracts. However, the study is also subject to limitations and does not provide strong evidence of a modifiable causal role (Abdellah *et al.*, 2019). The literature would benefit from further evidence provided by more reliable analysis, such as a meta-analysis or RCT, to draw a more conclusive role of vitamin D with cataract.

Alcohol consumption

Alcohol consumption has previously been associated with cataract and a wider range of ocular conditions, such as age-related macular degeneration and other chronic systemic conditions, including cancer and cardiovascular diseases (Karimi, Arabi and Shahraki, 2021; X. Zhang *et al.*, 2021). Alcohol consumption is a potential modifiable protective factor that can potentially aid in preventing diseases if a causal association can be established (Karimi, Arabi and Shahraki, 2021). Findings vary across current studies on the association between alcohol consumption and cataracts (Lindblad *et al.*, 2007; Xu, You and Jonas, 2009; Kanthan *et al.*, 2010; Wang and Zhang, 2014; Gong *et al.*, 2015; S. Y. Chua *et al.*, 2021; Fukai *et al.*, 2022; Kanclerz, Hecht and Tuuminen, 2023). A meta-analysis conducted on 5 case-control and 5 cohort studies discovered that heavy alcohol consumption correlated with an increased risk of age-related cataracts. The study also found a correlation implying a protective role of moderate alcohol consumption; however, this was not statistically significant (Gong *et al.*, 2015).

A study conducted using the Blue Mountains Eye cohort, including 3,654 participants aged 49 years or older, observed a significant increase in cataract surgery amongst those who consumed high levels of alcohol. Similar to the Gong *et al.* meta-analysis, the study also identified an association between moderate alcohol consumption and a decreased risk of cataract surgery (Kanthan *et al.*, 2010; Gong *et al.*, 2015). Lindblad *et al.* suggested a 7% increased risk of cataract surgery with an intake of 13g of alcohol per day (Lindblad *et al.*, 2007). In contrast, other studies have found a lack of evidence for an association between alcohol consumption and cataracts (Xu, You and Jonas, 2009; Wang and Zhang, 2014; Kanclerz, Hecht and Tuuminen, 2023). A recently published observational study, using data from the UK Biobank and European Prospective Investigation of Cancer (EPIC)-Norfolk studies, investigated the association between alcohol type and consumption with incident rates of cataract surgery. The EPIC-Norfolk dataset consists of 25,639 UK residents, whilst the UK Biobank is a population-based prospective study that included 502,504 UK residents. The study concluded that individuals with low to moderate alcohol consumption, particularly wine consumption, showed a lower risk of cataract surgery. However, the study highlighted the requirement of further research due to the influence of confounding factors, such as members of higher social classes consuming more alcohol and having access to better medical care, which may have interfered in the study's results (S. Y. Chua *et al.*, 2021).

Dependent on the availability of data associated with relevant risk factors can explore their relationship with cataract risk further. Therefore, further detail on literature and analysis of the relationship between vitamin D and alcohol consumption, respectively, with cataract can be found in Sections 4 and 5**.**

1.5 The genetics of cataract

As previously mentioned, cataract is a complex disease influenced by both environmental and genetic risk factors. Research has demonstrated that genetic studies can significantly enhance our understanding of cataract by uncovering hereditary patterns and identifying associated genetic loci. The following section summarises the current knowledge on cataract genetics, from family-based studies to advancements in sequencing technologies.

Family based studies

Twin studies have provided significant insights to the heritability of cataract. Comparisons between monozygotic (identical) and dizygotic (fraternal) twins allow researchers to understand what proportion genetic versus environment factors contribute to the variance in cataract risk. One study focusing on age-related cortical cataract determined 53-58% of the variability of cortical cataract risk was attributed to genetic effects, whereas 26-37% and 11-

16% was explained by environmental factors and age, respectively; therefore, highlighting the importance of genetic effects in cataract development (Hammond *et al.*, 2001).

Historically, family-based studies have played a crucial role in identifying genetic mutations associated with inherited cataract, particularly in cases of congenital cataracts. For example, one study involving slit-lamp examination and genetic sequencing across a three-generation family with congenital cataract cases identified a mutation in the *CRYAA* gene (Su *et al.*, 2012). Similar studies conducted amongst varying ethnic populations have also expanded knowledge in this area. This includes a European cohort of 25 families where researchers identified 20 distinct genetic variants associated with inherited cataract (Rechsteiner *et al.*, 2021).

Studies like these have uncovered numerous genes implicated in cataract development and can be broadly categorised into two groups. The first group includes mutations in crystallin genes, such as *CRYAA* and *CRYAB*, which play critical roles in maintaining lens transparency, and the second involves gap junctional proteins, such as *GJA3*. Additionally, mutations in genes outside of these groups, such as the *HSF4* gene, which encodes a heat shock transcription factor, have also been linked to cataract development, further showing the genetic complexity of hereditary cataract cases (Hejtmancik, 2008).

While family-based studies primarily focus on rarer forms of cataract, such as congenital cataract, they offer valuable insights into the hereditary patterns underlying cataract development.

Mouse models

Mouse models have been widely used to investigate the biological and genetic factors underlying human diseases, due to the genetic similarity between mice and humans, including specific gene contributions to cataract development. Their use has been enhanced through further advancements, such as transgenic, knock-out, and gene knock-in techniques, allowing them to provide unique genetic insights that could be applied to human diseases (Perlman, 2016). These models have also been instrumental in improving our understanding of ophthalmological conditions, including cataract.

Mouse models have highlighted many of the genetic findings previously described, particularly in genes, such as those encoding crystallin and gap junction proteins. While these models are primarily used to study congenital cataract, efforts have also been made to develop mouse models for age-related cataract (Graw, 2019). These models are increasingly relied upon to evaluate potential treatments aimed at preventing cataract onset. However, developing mouse models for age-related cataract is challenging due to the nature of this condition being predominantly caused by aging. The short lifespan of mice limits their ability to replicate age-

related changes seen in humans without associated difficulties and expenses to keep them alive beyond the age cataract typically develops (Rowan *et al.*, 2021).

Despite these challenges, certain mouse models have been developed to study age-related cataract, providing valuable insights into the functional mechanisms and genetic factors involved. For instance, a knock-out model targeting the *CRYBB2* gene was used to study cortical cataract, a subtype of age-related cataract. The model revealed that the cataract resulted from the aging process of the mice and their reduced ability to handle oxidative stress (Zhang *et al.*, 2008). Similarly, a knock-out model for the *GHR* gene demonstrated decreased development of age-related cataract, offering insights into how growth hormone pathways may influence lens aging (Wolf *et al.*, 2005).

While these models have significantly advanced our understanding of cataract development and the roles of specific genes, several limitations remain. Differences in biology and gene-environment interactions between mice and humans make it challenging to generalise findings. Additionally, accurately reproducing the multi-subtype nature (cortical, nuclear and posterior subcapsular) of human age-related cataract in mice remains a significant limitation. Therefore, additional techniques must be utilised to understand the genetic links between various risk factors and cataract.

Genome-wide association studies

More recently, genome-wide association studies (GWAS) have been instrumental in identifying specific genetic loci associated with cataract risk. Details of the methodology used to conduct GWAS are provided in Section 2 Materials and methods.

To date, two large-scale GWAS have significantly advanced our understanding of the genetics of cataract.

Choquet *et al.* conducted a meta-analysis using data from the Genetic Epidemiology Research in Adult Health and Aging (GERA) cohort and the UK Biobank (UKB), producing multi-ethnic GWAS results for cataract. Cataract cases were defined through a combination of diagnostic and self-reported data from the UKB and diagnostic information, including cases of cataract surgery, from Kaiser Permanente Northern California (KPNC). This analysis identified 47 genetic loci, 37 of which were novel loci discovered through the multi-ethnic meta-analysis. Additionally, several loci associated with potential drug targets were identified, including *RARB*, *KLF10*, *DNMBP*, *HMGA2*, *MVK*, *BMP4*, *CPAMD8*, and *JAG1*. These findings highlighted important pathways involved in lens development, oxidative stress responses, and other biological processes relevant to cataract (Choquet *et al.*, 2021).

Building on these findings, a subsequent study by Diaz-Torres *et al.* expanded the multi-ethnic GWAS by incorporating additional cohorts. This meta-analysis combined data from GERA and UKB with new data from the Mass General Brigham Biobank (MGBB) and FinnGen, significantly increasing the sample size to 121,725 cases and 821,856 controls. Cataract cases in MGBB and FinnGen were identified using ICD codes, ensuring consistent diagnostic criteria. The expanded GWAS identified 101 independent loci, including 57 novel loci, furthering the understanding of the genetic basis of cataracts. The larger sample size allowed for the identification of more robust associations and reinforced the role of previously implicated pathways (Diaz-Torres *et al.*, 2024).

Copy Number Variants

While most genetic investigations of cataract have focused on SNPs more recent work has also examined the role of copy number variants (CNVs). CNVs are structural variations in the genome where the number of copies of a given DNA segment differs between individuals. These segments can range in size from a few base pairs to thousands of base pairs and typically arise through duplication or deletion events. While some CNVs have no phenotypic consequences, others can disrupt coding or regulatory regions and play a key role in disease development (Pös *et al.*, 2021).

CNVs have been implicated in cataract aetiology, particularly in congenital and early-onset cases. A study of 347 patients, aged 18 months to 35 years with early-onset bilateral cataract, identified specific CNVs in genomic regions containing collagen genes, which are critical for maintaining lens transparency and refractive properties (Fox *et al.*, 2024). CNVs have also been explored in age-related cataract, with evidence implicating variants in *HSF4* and *WRN*, both genes involved in DNA repair pathways, in disease pathogenesis (Jiang *et al.*, 2013). Overall, current knowledge of CNVs in cataract remains focused on congenital subtypes with limited exploration of age-related cataract.

1.6 Limitations in current literature

Amongst the current literature, both modifiable and non-modifiable risk factors have been investigated via observational studies; however, both limitations and gaps in prevailing knowledge remain.

1.6.1 Confounding factors

Although several observational studies have been able to find an association between cataract and various proposed risk factors, their findings are limited by the presence of confounding factors. Confounding factors create difficulties in distinguishing between correlation and causation (Skelly, Dettori and Brodt, 2012).

Figure 1.2: Confounding factors example diagram, alongside the Alcohol Consumption and Cataracts case example.

As shown in Figure 1.2, we may observe a relationship between A and B, but the association from C affecting both. For example, the observed relationship between heavy alcohol consumption and the cataract risk may result from confounding factors, such as smoking, where smoking is a known risk factor for cataracts and increases alcohol consumption (Benjamin, Burns and Proctor, 2013; Gong *et al.*, 2015). Therefore, a causal relationship between heavy alcohol consumption and cataract cannot be verified if smoking is not accounted for. Without evidence of causality, risk factors like alcohol consumption cannot be assessed as a modifiable risk factor due of the presence of potential confounding (S. Y. Chua *et al.*, 2021).

Another example lies within a 2014 study conducted by Wise *et al.* (2014) where an increased risk of cataract was observed across patients who used statins (Wise *et al.*, 2014). However, it was discovered that study participants who developed cataract were more likely to have diabetes, cardiovascular disease, and chronic obstructive pulmonary disease, all known risk factors for cataract, in the year before the study took place. In addition, the study failed to consider the time taken to develop cataracts. It was likely that participants were exposed to these confounding factors for a significant period before they began statin therapy. Therefore, rather than statins causing an increased risk of cataract, the increased exposure of risk factors for cataract may have led to the participants use of statins (Spence, 2015). This is further supported through the establishment of a causal link between statins and type 2 diabetes (Swerdlow *et al.*, 2015).

Reverse causality is a further limitation of observational studies, where an exposure causes an outcome, but the outcome may also cause the exposure. Reverse causality may lead to difficulty in establishing a meaningful causal relationship between two factors. A potential example related to cataract could come in establishing a causal relationship between vitamin D deficiency and cataract risk. Individuals with cataracts may spend less time outdoors because of impaired vision or increased light sensitivity, which in turn reduces

sunlight exposure and lowers vitamin D levels. This could create the appearance that vitamin D deficiency increases cataract risk, when in fact the causal direction is the reverse (Öktem and Aslan, 2021).

Observational studies can be limited further due to the risk of bias. Selection bias occurs when participants are not selected at random and do not represent the intended population, causing results to be systematically skewed. Observational bias occurs because of how data and information are collected during the study, such as an interviewer emphasising different questions, thus influencing observations (Delgado-Rodríguez and Llorca, 2004; Cook, 2010; Sterrantino, 2024). Notably, healthy-volunteer bias has been well documented across studies, such as the UK Biobank, whereby participants have healthier lifestyles, better education and overall health than the target population. Such bias has been found to distort phenotype-outcome associations in observational studies as well as the associations between variants and outcomes in genetic studies (Schoeler *et al.*, 2023).

Due to confounding factors, reverse causality and bias risk, alternative methods of investigation are required to draw informative conclusions.

1.7 Research questions, aims and objectives

Overall, this investigation establishes the following research questions, aims, and objectives, developed in response to identified gaps in the existing literature.

1.7.1 Research questions:

1. Do different forms of cataracts share underlying genetic mechanisms?
2. Can modifiable risk factors be identified to delay or prevent the development of cataracts?
3. Surgery forms the primary treatment for cataracts. Can alternative intervention occur to delay the development or treatment of cataracts?

1.7.2 Research aims:

1. Identify and understand the common genetic mechanisms underlying congenital, senile, and diabetic-related cataracts.
2. Test modifiable lifestyle factors to aid in the prevention of cataract development.
3. Investigate the use of oxysterol eye drops as an alternative treatment option.

1.7.3 Research objectives:

1. Produce genome-wide association results for cataracts in the UK Biobank. Specifically for congenital, senile, and diabetic-related cataracts.
2. Complete genetic correlation and co-localisation testing to identify the shared genetic components between different traits and explore if the same change in the DNA causes the observed overlap.
3. Using the generated cataract GWAS results, applying Mendelian randomisation to all available phenotypic traits in the UK Biobank (outcome) to identify the causally associated modifiable risk factors.
4. Test selected possible pharmacological interventions for the treatment, or delay, of cataracts.

2. Materials and methods

2.1 Introduction

This section describes the materials used across the analyses presented in the results of Sections 3, 4, 5 and 6. The primary data source for these analyses was the UK Biobank cohort and this section details the participant characteristics, the genotyping, and the rigorous quality control (QC) steps applied to the data prior to analysis.

The investigations detailed within this thesis focuses on two primary analytical approaches: genome-wide association studies (GWAS) and Mendelian randomisation (MR). This section provides an overview of the use of REGENIE for conducting GWAS, along with the theoretical background for both GWAS and MR.

2.2 The UK Biobank cohort description

The UK Biobank (UKB) is a large-scale, population-based prospective study. Participants were recruited from 2006 to 2010. Its purpose is to investigate risk factors, including genetic predispositions and environmental exposures, for a wide range of diseases, to identify new ways to prevent and treat various conditions (UK Biobank, 2007; Allen *et al.*, 2012).

The UKB includes over 500,000 participants from across the UK. Data were anonymised and collected at 22 locations in England, Scotland, and Wales, minimising geographic bias, as shown in Figure 2.1. Participants were aged 40 to 69 at recruitment, with a mean age of 56.5 and males represented 45.8% of the sample population. The age range of 40-69 supports the inclusion of participants with established health conditions, those with emerging symptoms, and individuals in early stages of disease, allowing for comparison of disease onset and progression. Participants contributed detailed environmental, lifestyle, and medical history information, and biological samples such as blood, urine, and saliva samples, as well as a range of physical measures such as blood pressure and hand grip. Figure 2.2, shows a summary of the range of data that was collected (Allen *et al.*, 2012).

To enhance the available phenotypic information, comprehensive longitudinal follow-up examinations are undertaken, with additional examples of more regular follow-up in smaller sample subsets (e.g., physical activity data) (Sudlow *et al.*, 2015). The data are continually updated via follow-up, allowing for baseline comparisons as well as the exploration of emerging data types. Follow-up within the UKB cohort can occur in two different forms, via direct follow-up with the participants through additional data collection such as updated questionnaires or through data linkage to other records such as primary care or hospital

admission data. Due to its large sample size, comprehensive baseline data and longitudinal follow up, the UKB is suitable to study a range of age-related and complex diseases such as cataract (Allen *et al.*, 2012).

Figure 2.1: Map of UKB baseline assessment centres (UK Biobank, 2021).

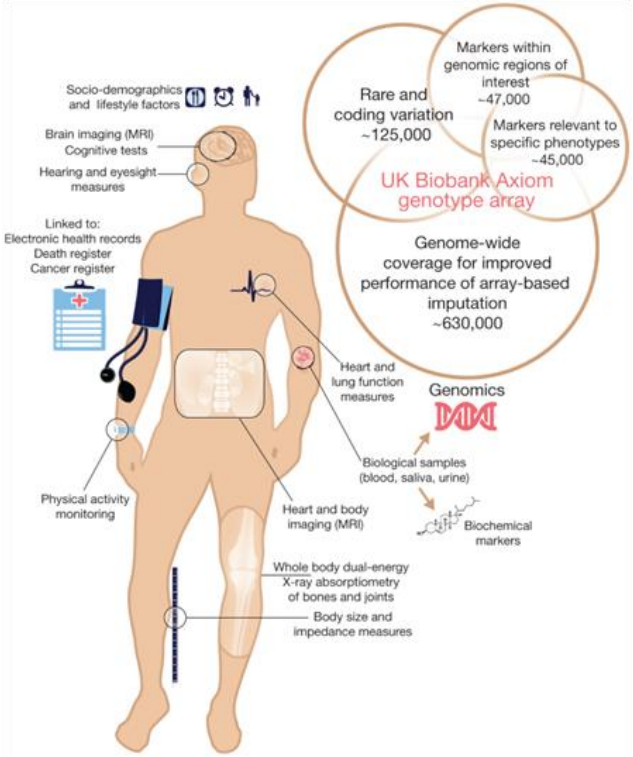


Figure 2.2: Diagram of data collected by the UKB within their baseline assessment (Bycroft et al.*, 2018).*

2.3 Genotyping and imputation of the UKB

The UKB has genotype information for 488,377 participants. UKB participants provided blood samples collected at assessment centres, and genotyping was performed by the Affymetrix Research Services Laboratory in 106 sequential batches of approximately 4,700 samples. Two similar genotyping arrays were used to assay the biological samples of the 488,377 participants (Bycroft *et al.*, 2018). The first 49,950 participants, selected based on smoking behaviour and lung functions from the UK Biobank Lung Exome Variant Evaluation (UK BiLEVE) study, were genotyped using the Applied Biosystems UK BiLEVE Axiom Array with 807,411 markers (Wain *et al.*, 2015). The remaining 438,427 participants were genotyped using the Applied Biosystems UK Biobank Axiom Array with 825,927 markers. Both arrays shared 95% of their content across markers. The selected markers on both arrays were designed to detect genetic variants, including single nucleotide polymorphisms (SNPs) and insertions/deletions (indels), and to explore potential associations with specific diseases. The arrays included coding variants spanning a range of minor allele frequencies (MAF), incorporating rare variants (< 1% MAF) and markers suited for imputation in a European population, covering both common (> 5%) and low (1-5%) frequency MAF ranges (Bycroft *et al.*, 2018).

Genotyped data can be further imputed to infer variants that were not directly obtained from the assayed sample population. Imputation relies on a reference panel constructed from whole-genome sequencing data that provides a comprehensive map of genetic variants including SNPs, insertions/deletions (indels), copy number variants, and human leukocyte antigen (*HLA*) alleles. Imputation can also help correct genotyping errors, improving data accuracy and coverage. The UKB genotyped data was imputed using the following reference panels: Haplotype Reference Consortium (HRC), UK10K haplotype panel and 1000 Genomes phase 3 reference panel, each providing comprehensive information common and rare variants (Huang *et al.*, 2015; McCarthy *et al.*, 2016). By using these reference panels, the imputation process can predict genotype data that were not directly obtained through genotyping. This "in silico" genotype data enhances the overall power of analyses using the imputed UKB data.

2.4 UKB ethical approval

The UKB has ethical approval from the North West Multi-centre Research Ethics Committee (MREC) for Research Tissue Bank (RTB) approval (REC reference: 11/NW/0382), allowing researchers to conduct their studies under UKB's access policies and ethical guidelines. The UKB also obtained informed consent from all participants.

Ethical approval was further approved by the Brunel Research Ethics committee (Review reference - 30904-LR-Jun/2021- 32998-1) on 09/07/2021.

The UKB approved the data used for this thesis under application 72850.

2.5 Quality Control Procedures of UKB data

The samples underwent QC by the UKB during the DNA extraction and genotype calling stages, and relevant samples were removed prior to data access. Additional QC steps outlined by Bycroft *et al.* were applied in this study prior to performing analyses using the UKB data. Although these represent the typical QC procedures for UKB data, any modifications or omissions of these steps during the analyses will be clearly outlined within the relevant section. The sample and genetic QC metrics applied were based on the widely accepted QC measured established by Bycroft *et al.,* which provides rigorous standards to ensure data reliability and consistency when using the UKB cohort (Bycroft *et al.*, 2018).

2.5.1 Sample QC

To ensure the reliability and consistency within the sample data, common QC metrics were applied to filter for missingness, heterozygosity, and ancestry. The following sample QC (as shown in Chart 1) ensures that only accurate and reliable data were included in the analysis.

To adhere to ethical standards and ensure the UKB participant right to request data withdrawal is upheld, individuals who requested data withdrawal from UKB (as of April 2023) were removed from analyses.

Sample QC measures included mismatched sex from baseline characteristics, which had self-reported sex (Data-Field 31) differing from genetic sex (Data-Field 22001). Mismatched sex may arise from rare genetic variations or transgender individuals. However, such mismatches can also indicate errors in sample handling or labelling, which may compromise the reliability of the analysis. To ensure data quality and consistency, these samples were removed. Further samples were removed from the analysis for the presence of sex chromosome aneuploidy (Data-Field 22019) as individuals with such anomalies may exhibit poor genotyping quality. Individuals with low-quality genotyped data, were identified and removed via outliers in heterozygosity and missing rates (Data-Field 22027), as heterozygosity outliers indicate potential contamination, while high missing rates suggest genotyping errors or poor-quality samples.

Kinship coefficients for all pairs of samples in the UKB were estimated using the genetic data. 30.3% of participants were inferred to be related to at least another UK biobank participant with a third-degree relationsip or closer (Bycroft *et al.*, 2018). To filter for related individuals

in the UKB dataset, kinship information provided by the UKB was used (Data-Field 22021). Participants who were excluded from the kinship inference process or had ten or more third-degree relatives were removed.

In addition, the UKB also provides a relatives file to identify and filter related individuals within the cohort, which allows researchers to exclude one individual from each related pair and was used as an additional relatedness filter. Relatedness is filtered when using UKB data as phenotypic similarity between related individuals in analyses may skew results through shared genetic variants being falsely associated with traits that are also similar due to close relatedness (Coleman *et al.*, 2016). In some instances, the relatedness filter is not applied, for example when the model accounts for relatedness. Where relatedness is not filtered this is made clear in the relevant section.

Samples were filtered to include only European individuals. Individuals' ethnicity was established through self-reported data that they identified as "White British" and had appropriate genetic ancestry to match (Data-Field 22006).

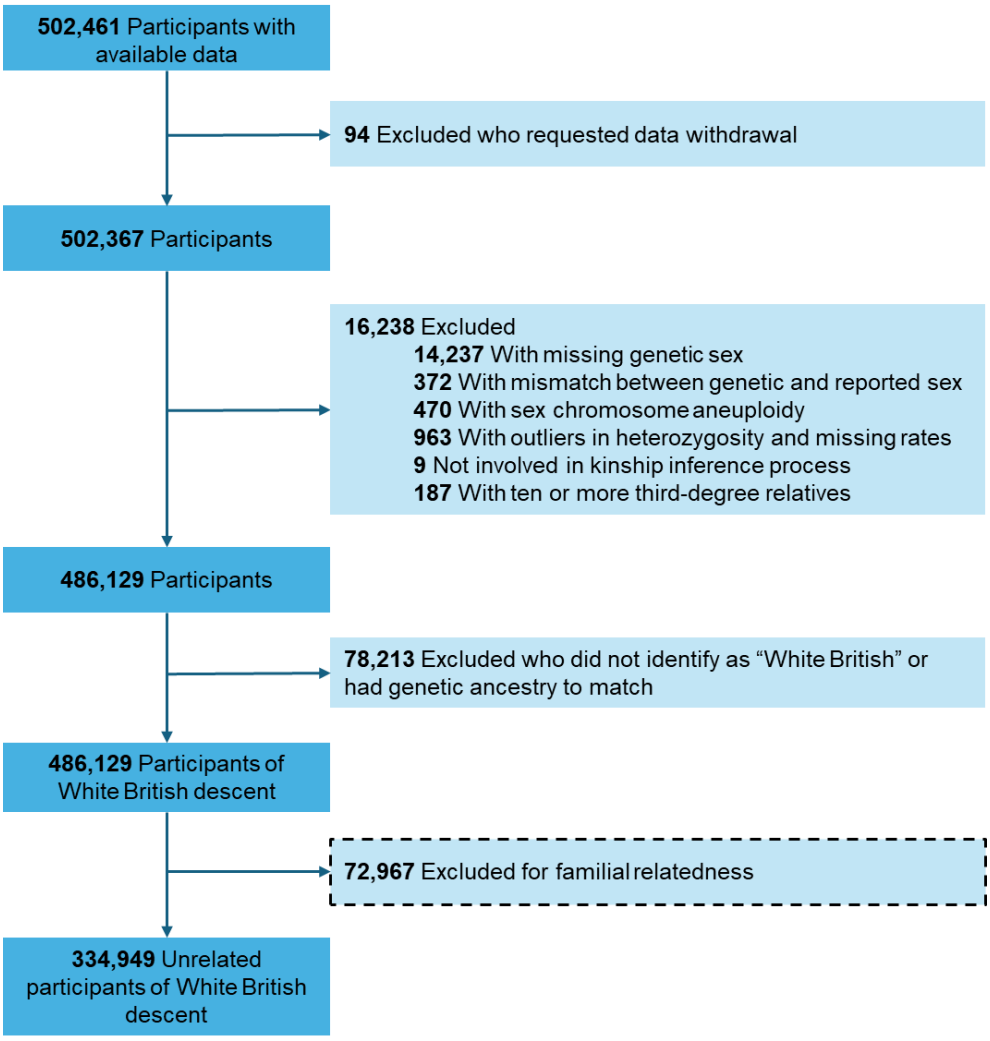


Chart 1: Flowchart showing sample quality control analysed from the UK Biobank as described in Section 2.5.1. Filter for relatedness, as shown in the dotted border, only applied to relevant analyses.

2.5.2 Variant QC

QC metrics for genetic variants were also applied to enhance the reliability of the genetic data for UKB participants and reduce bias associated with using poor quality genetic data.

Relevant genetic QC procedures are outlined in each specific analysis section. However, an overview of the genetic filters applied is as follows: Imputed variants with an INFO score < 0.8, or an effect allele frequency (EAF) < 0.01 were also removed.

2.6 Genome-wide association study using REGENIE

Genetic variants associated with the risk of cataract form the foundation for the sequence of analyses conducted in this study. To identify these variants, GWAS analyses were performed. This section describes the theoretical background, analytical approach, and tools used to conduct GWAS in the UKB cohort.

2.6.1 Genetic models

To convert genotyped data into numerical form, different genetic models can be assumed: additive, dominant and recessive. These models define how genetic variants are coded and interpreted in analyses. The additive model is the most commonly used in GWAS because of its simplicity and biological plausibility. It assumes that the effect of carrying an additional copy of the effect allele is constant. For example, the SNP rs1042725 in the *HMGA2* gene, associated with height (Yang *et al.*, 2009), can be coded as 0 for homozygotes with the common allele (TT), 1 for heterozygotes (CT), and 2 for homozygotes with the effect allele (CC). If carrying one copy of the effect allele (C) increases height by 0.5 cm, the additive model predicts that heterozygotes (CT) will be 0.5 cm taller and homozygotes (CC) will be 1 cm taller, on average.

Alternative models, such as dominant or recessive, may better capture the genetic architecture of specific traits in some cases. For instance, under a dominant model, the phenotype is influenced by the presence of at least one effect allele (e.g., coding TT = 0, CT = 1, CC = 1), while the recessive model assumes an effect only for homozygotes of the effect allele (e.g., coding TT = 0, CT = 0, CC = 1).

Where applicable, the choice of genetic model used in the analysis is justified, such as in Section 2.6.5, which explains the use of the additive model in genetic analyses with REGENIE.

2.6.2 GWAS Theoretical Background

GWAS tests the association between genetic variants and specific diseases or traits, establishing links between genotypes and phenotypes, contributing to a better understanding of genetic mechanisms, potential treatments, and prevention strategies. These associations are assessed through regression models that account for the allele frequencies of genetic variants in populations that are ancestrally similar (such as Europeans) but differ phenotypically. To date, a wide range of traits have been explored using GWA analysis. As reviewed in Tam *et al.,* the first GWA analysis was in age-related macular degeneration in 2005 but this has been further expanded into other traits such as cardiometabolic traits, anorexia, cancer and type 2 diabetes (Tam *et al.*, 2019).

To conduct a GWAS, an initial sample of participants is required, including both genetic data (e.g. DNA samples) and non-genetic data (e.g. lifestyle factors and phenotypic characteristics like sex and age). Larger sample sizes are generally preferred, as they reduce the risk of false positives and increase statistical power. Depending on the trait, one approach is to form case and control groups within the sample, based on the presence or absence of the disease being studied. However, alternative methods, such as analysing the trait on a continuous scale, are

often utilised when appropriate data are available for example for height or body-max index (Uffelmann *et al.*, 2021).

GWAS typically uses linear or logistic regression models to test the association of each genetic variant with the trait. Linear regression is used for continuous traits, such as BMI or height, while logistic regression is used for binary traits, such as having a cataract operation versus not having a cataract operation. For example, for a binary trait using a logistic regression, each SNP is assigned a numerical value (based on the genetic model) and analysed to assess the strength of its association with cataract, depending on the combination of alleles present. This method provides both the *p*-value, indicating statistical significance, and the effect size alongside standard errors of the association for each SNP.

2.6.3 Multiple testing

To assess whether identified SNPs are statistically significant in a GWAS analysis, we refer to the *p*-value. The *p*-value represents the frequency of the observed relationship between a SNP and a trait under the null hypothesis, which assumes no true effect exists between the SNP and the trait. Setting a *p*-value threshold of 0.05 establishes a 5% significance level, representing the maximum acceptable probability of rejecting the null hypothesis when it is true. A result with a *p*-value below this threshold allows for the null hypothesis to be rejected, indicating that the observed result is unlikely to have occurred by chance.

In GWAS, a stricter *p*-value threshold of $p < 5x10^{-8}$ is widely accepted to account for multiple testing. This threshold, derived using the Bonferroni correction (Armstrong, 2014), accounts for the roughly 1,000,000 independent tests completed typically in a GWAS. Without this correction, the likelihood of false positives would increase due to the large number of statistical tests conducted, potentially leading to false positives (Altshuler, Donnelly, and The International HapMap Consortium, 2005; Pe'er *et al.*, 2008).

2.6.4 Independent SNPs and assessing linkage disequilibrium

In GWAS analysis, a large number of SNPs may show an association with the trait. However, it is important to assess which of these SNPs represent truly independent signals. Some SNPs may be in linkage disequilibrium (LD) with the SNP that has the true association. LD refers to the non-random association of alleles at two different loci, where they are inherited together more frequently than expected by chance, indicating a correlation between the SNPs (Slatkin, 2008).

LD is commonly quantified by r^2, which measures the correlation between alleles at two loci. An r^2 of 0 indicates no correlation, meaning the alleles are independent of each other. An r^2 of 1 indicates that the alleles are very correlated, where one SNP completely proxies the other.

The r^2 value can be interpreted as a representation of the degree of LD between SNPs; the closer the r^2 is to 1, the more likely it is that the SNPs are in strong LD.

To identify independent SNPs, SNPs are grouped into clumps based on the defined LD threshold (e.g., $r^2 < 0.1$) and base pair distance criteria (kilobases). A lead SNP is selected as the most significantly associated variant within a defined base pair range. SNPs in strong LD with the lead SNP (at a defined threshold such as $r^2 \geq 0.1$), are assigned to the same clumping group, and only the lead SNP is kept representing the independent signal. Various software tools can be used, such as PLINK v2.0 (https://www.cog-genomics.org/plink/2.0/), which performs LD clumping. In each analysis, the method used for clumping has been specified.

2.6.5 REGENIE

Genetic associations for relevant traits such as cataract development were obtained through a GWA analysis. The GWA analysis was conducted using REGENIE v3.2.8. REGENIE is a C++ programme used to conduct whole genome regression modelling of large genome-wide association studies, using a mixed-model-based approach (Mbatchou *et al.*, 2021). REGENIE's mixed-model approach addresses data complexities like relatedness and population stratification by combining fixed effects from known confounders with random effects that capture unobserved influences, such as genetic relatedness (Zhou *et al.*, 2018). Prior to performing a GWAS using REGENIE, QC steps were taken for the genotype input file using PLINK v2.0. QC removed SNPs with MAF < 0.01, minor allele count (MAC) < 100 (MAC refers to the total number of minor alleles observed across all individuals for a given SNP, MAC < 100 excludes variants with a low allele count), genotype missingness > 0.1, Hardy-Weinberg equilibrium *p*-value > 10^{-15}, and samples with > 0.1 missingness.

Additional QC steps, as outlined in Sections 2.5.1 and 2.5.2, were applied, excluding relatedness filters, on account of REGENIE's mixed-model approach. While REGENIE is able to account for population structure, we have maintained limiting our sample to a single ancestry group (European) to reduce the risk of Type 1 errors.

REGENIE provides an advanced method for conducting GWAS analyses accounts for population structure, relatedness, and case-control imbalance, while being more computationally efficient than other available methods. REGENIE is run over two steps:

1. Stacked block ridge regression: SNPs are grouped into blocks, and ridge regression is used to estimate genetic effects within each block. These are then combined to create a single genetic prediction, which is then decomposed into 23 chromosome predictions. To avoid bias from testing variants on the same chromosome used for prediction, a Leave-One-Chromosome-Out (LOCO) approach is applied.

2. Association testing: LOCO predictions from Step 1 are the used as covariates during association testing. Each SNP is tested for association with the trait using either linear regression (for continuous traits) or Firth logistic regression (for binary traits), the latter helping to reduce bias in unbalanced case-control settings.

In the REGENIE analysis the additive genetic model is used as default (Mbatchou *et al.*, 2021).

Further information regarding the REGNIE QC and steps can be found at https://rgcgithub.github.io/regenie.

2.7 Mendelian randomisation description

This section provides a brief overview of another commonly used method in this study: Mendelian randomisation.

2.7.1 Background

As outlined in the Introduction, numerous observational studies, including case-control designs, have investigated the effects of various exposures on cataracts. However, these studies are inherently limited by confounding factors and the potential for reverse causation, which prevents the interpretation of causal relationships. While RCTs are considered the gold standard for establishing causality, designed to reduce the risk of bias and interference of confounding factors, they often face significant challenges, including financial constraints and ethical concerns and face challenges with generalisability to diverse populations, thus are not always feasible. Causality is critical for advancing public health, as it identifies opportunities for targeted interventions to reduce disease risk and improve health outcomes (Kendall, 2003; Glass *et al.*, 2013; Hariton and Locascio, 2018; Monti *et al.*, 2018; Zabor, Kaizer and Hobbs, 2020).

MR makes use of the naturally occurring form of randomisation of genetic variants to explore the causal relationship between an exposure and an outcome (Davies, Holmes and Smith, 2018). This process broadly parallels RCTs. In an RCT, as shown in Figure 2.3, participants are randomly allocated into intervention or control group, ensuring that any differences in the outcome of the trial can be the direct result of the intervention rather than external factors. MR mimics this random allocation by using genetic variants, SNPs, as instrumental variables (Sobczyk *et al.*, 2023).

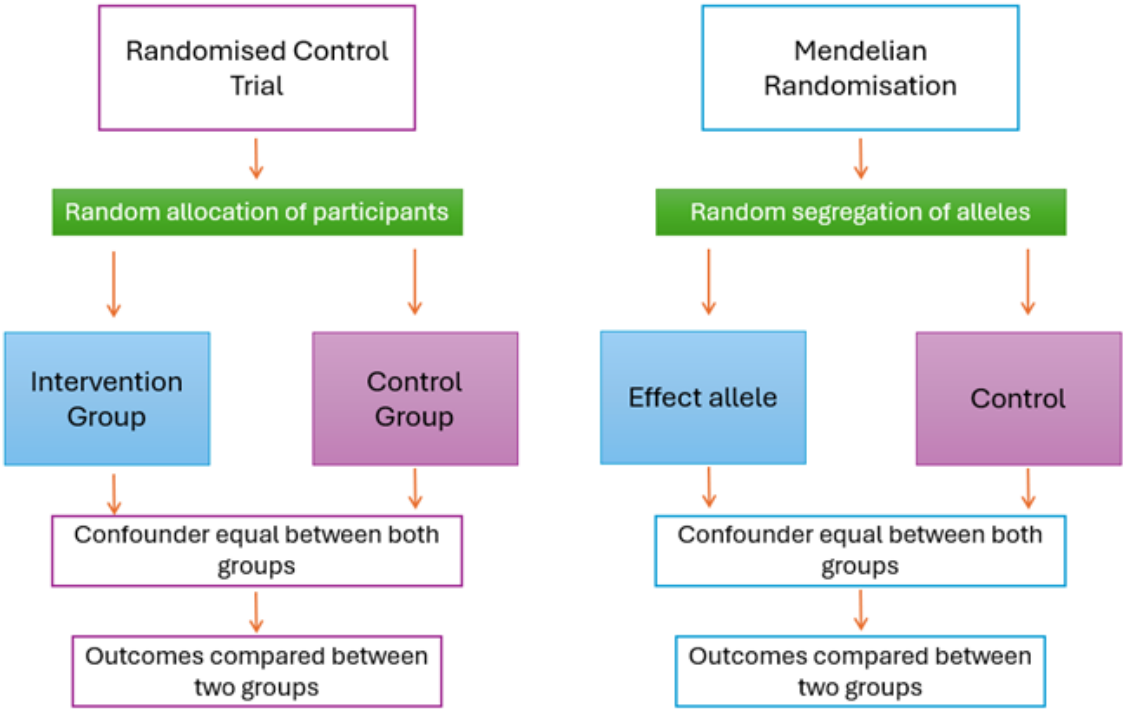


Figure 2.3: Illustration of differences in methodology between RCT and MR. adapted from concepts illustrated in (Howell et al., 2018)

The use of SNPs is possible as instrumental variables as it is based on Mendel's laws of inheritance, particularly the laws of segregation and independent assortment. These principles state that alleles are inherited randomly and independently (from variants not in the same gene locus or linkage disequilibrium block) during meiosis, ensuring that genetic variation is distributed across individuals without bias (Wolf, Ferguson-Smith, and Lorenz, 2022). For example, this naturally occurring randomisation, means that genetic variants associated with an exposure such as BMI, are independent of confounding factors (socioeconomic status or lifestyle choices) allowing for the relationship between exposure and outcome to be observed without interference.

MR serves as a powerful alternative to RCTs for testing causal relationships, especially within investigations were conducting an RCT may be unethical, impractical, or financially challenging. Notably, risk factors currently associated with cataracts would not be possible to be controlled in an RCT setting due to the duration of time it takes cataract to be fully developed and diagnosed.

The validity of in the MR estimates depends on the SNPs meeting three key assumptions, which determine whether the genetic variants can serve as valid instruments (Davies, Holmes and Smith, 2018):

1. Relevance Assumption: The SNPs must be associated with the exposure.
2. Independence Assumption: The SNPs must be independent of confounding factors.
3. Exclusion Restriction Assumption: The SNPs must influence the outcome only through the exposure.

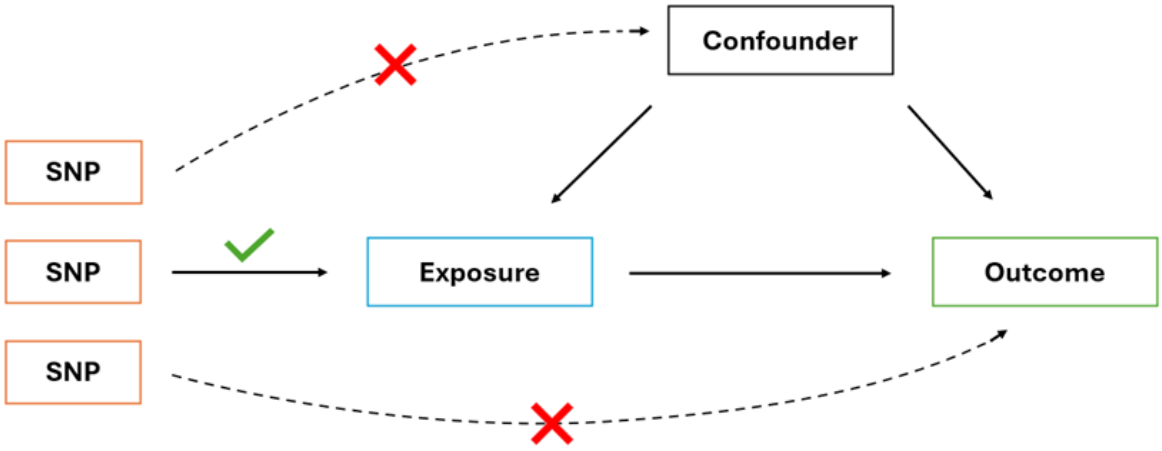


Figure 2.4: Assumptions underpinning the use of MR analyses.

Figure 2.4 illustrates the assumptions required for genetic variants to serve as valid instrumental variables in a Mendelian randomisation analysis.

2.7.2 Evaluating instrumental variable assumptions

Violations of instrumental variable assumptions can introduce bias into MR results, making it crucial to understand how these assumptions can be violated and how violations can be tested.

The relevance assumption is violated when weakly associated genetic instruments are used, leading to weak instrument bias. This bias can result in either confounded observational estimates or bias towards the null, depending on the type of MR study. To uphold this assumption, MR analyses typically use SNPs that meet the genome-wide significance threshold ($p < 5\times10^{-8}$) from GWA analyses. SNPs at this threshold are strongly associated with the exposure of interest, ensuring their suitability as instruments. Furthermore, these SNPs are often used in groups rather than individually to increase the power and reliability of the analysis.

The Independence assumption requires that genetic instruments are not associated with confounding factors that affect both the exposure and the outcome. SNPs can be checked for associations with known confounders or other variables to identify potential violations of this assumption.

The exclusion restriction assumption ensures that SNPs influence the outcome only through the exposure. Violations of this assumption often arise from horizontal pleiotropy, when a genetic instrument is associated to the outcome via a pathway independent of the exposure. Sensitivity analyses, such as MR-Egger regression and the Weighted Median, can help detect and account for such violations. It is possible to examine the intercept of the MR-Egger regression. A non-zero intercept indicates the presence of pleiotropy. While the Weighted Median estimator provides a robust causal estimate even if up to 50% of instruments are invalid (Zheng *et al.*, 2017).

2.7.3 Conducting and interpreting the results of an MR

Different statistical approaches can be taken when conducting MR depending on the study design. For example, a one-sample MR, the genetic instruments for the exposure and the outcome are measured within the same population, yielding the causal estimate of the risk factor on the outcome. In contrast, a two-sample MR utilises summary-level data from two independent samples, one for the exposure and one for the outcome helping to improve statistical power of the analysis (Monti *et al.*, 2018)

Outlined below are the different statistical approaches used in MR, which can accommodate various study designs, ranging from single-SNP analyses to those incorporating multiple SNPs. Each of the methods below, assume that the SNPs are valid instruments under the instrumental variable assumptions (Relevance, Independence and Exclusion-restriction assumption).

Ratio of Coefficients Method (Wald Ratio): This method, also known as the Wald ratio, estimates the causal effect of the exposure (X) on the outcome (Y) using a single genetic variant.

The formula for calculating the Wald ratio is:

$$Wald\ Ratio = \frac{Association\ between\ SNP\ and\ Outcome\ (Y)}{Association\ between\ SNP\ and\ Exposure\ (X)}$$

While the Wald ratio is suitable for single-SNP analyses, most MR studies require a method that can incorporate multiple SNPs (Boehm and Zhou, 2022).

Two-Stage Least Squares

In the two-stage approach, used for a one-sample MR. Two regressions are performed, typically using either linear or logistic regression, depending on outcome variable type (binary or continuous). In the first stage, the exposure variable is regressed on the instrumental variable (the SNP), producing fitted values for the exposure. In the second stage, these fitted

values are used as the independent variable, regressed against the outcome variable, to provide a causal effect estimate (Boehm and Zhou, 2022).

Inverse Variance Weighted (IVW) method

The IVW method extends the Wald ratio approach to include multiple SNPs, typically applied for a two-sample MR method. This approach combines the Wald estimates from each SNP by weighting them according to the inverse of their variance, creating a combined estimate of the causal effect, where each SNP is treated as an independent instrumental variable and the IVW causal effect combines each variants ratio estimate in a meta-analysis (Burgess, Butterworth and Thompson, 2013). While IVW is the most widely accepted method for summarising multiple SNPs, it can be biased if pleiotropy is present, thus tests such as the MR Egger and Weighted Median are included as sensitivities (Burgess, Butterworth and Thompson, 2013). MR-Egger can be used as a sensitivity analysis to assess directional pleiotropy and test for a causal effect while also estimating its size. MR-Egger uses the Instrument Strength Independent of Direct Effect (InSIDE) assumption, which allows for a weaker assumption when a significant number of instrumental variables are invalid. The MR Egger intercept indicates the presence of a pleiotropy effect. Where the intercept value is 0, there is no pleiotropic effect. Alternatively, a non-zero intercept would reveal the results of the MR to be biased due to the presence of directional pleiotropy (Burgess and Thompson, 2017).

When instrumental variable assumptions are likely violated, alternative methods can provide more reliable estimates. For instance, the weighted median method can produce a consistent causal estimate even if up to 50% of the instruments are invalid, as long as the majority of the weight comes from valid instruments (Bowden *et al.*, 2016). The simple mode estimator groups SNPs based on similarities in their individual causal effect estimates and derives the causal effect from the largest cluster of SNPs. The weighted mode estimator extends this by giving more weight to SNPs within each cluster according to the inverse variance of their effect on the outcome, giving greater influence to more precisely estimated SNPs (Hwang *et al.*, 2019).

3 Using genetics to investigate the association between lanosterol and cataract

3.1 Introduction

As stated in Section 1.3.4, the only effective treatment for cataract is surgery, whereby the cataract is removed and replaced with an artificial intraocular lens. While this is typically a safe and successful procedure (Davis, 2016), complications can still occur and can be serious (Chan, Mahroo and Spalton, 2010; Naeem *et al.*, 2012), with significant disparities in surgical success rates between developed and developing countries (Gogate, 2010; Malhotra *et al.*, 2014). Whilst there are no non-surgical treatments for cataracts, non-invasive options would significantly lessen the burden on public services with the volume of cataract procedures growing at an annual rate of 3.1% (Chen *et al.*, 2021). Medication to treat cataract would lessen the need for surgery and avoid its associated complications, thus providing a safer alternative treatment that individuals can easily access in poor and remote regions.

Recent studies (including studies in whole humans, animal models and isolated tissues) have suggested that oxysterols, such as lanosterol, are able to restore the transparency of lenses affected by cataracts (Makley *et al.*, 2015; Zhao *et al.*, 2015; Molnar *et al.*, 2019; Wang *et al.*, 2022). In addition, studies have reported that lanosterol was effective in redissolving aggregates of bound proteins and restoring lens clarity in human lenses (Qi *et al.*, 2016; Chen *et al.*, 2018).

Zhao *et al.* found that lanosterol reduced cataract severity and increased lens clarity in animals, specifically *in vitro* in rabbit cataractous lenses and *in vivo* in dogs (Zhao *et al.*, 2015). Defects in lanosterol synthase (*LSS*), which synthesises lanosterol, have also been found to be associated with congenital cataracts in humans (Zhao *et al.*, 2021). Following the Zhao *et al.* publication, lanosterol eyedrops have been marketed with claims to reverse the effects of cataracts, more commonly for the use in animals (Balashova *et al.*, 2018). However, further research on lanosterols has questioned their effectiveness as a cataract treatment. Whilst Balashova *et al.* found laboratory-based evidence for stabilising rapidly progressive cataracts in a human patient using lanosterol eyedrops, Shanmugam *et al.* and Daszynski *et al.* found no evidence that lanosterol reverses lens opacification or affects lens protein solubilisation in cataractous human and animal lenses (Shanmugam *et al.*, 2015; Balashova *et al.*, 2018; Daszynski *et al.*, 2019).

The production of drug-based treatments has historically been limited by the weak predictive efficacy found in preclinical experiments using cell, tissue, and animal models. Genomic data

used for analysis is becoming an increasingly important part of drug development and benefit the process by facilitating target validation and being increasingly relevant to human biology rather than studying animal models of diseases (Spreafico *et al.*, 2020). SNPs associated with a gene affecting a protein of interest can be used as proxies to investigate a drug's potential impact on the respective protein (Finan *et al.*, 2017). The use of genetic evidence in selecting and assessing the efficacy of drug targets can significantly increase the likelihood of a drug reaching phase III of trials and entering the market (Nelson *et al.*, 2015). Due to the recent success of genetic evidence, the use of genetic association results in drug development has become increasingly popular (Liou, 2014). However, genetic evidence on the effect of lanosterol on cataract has not yet been established.

In this investigation, using data from the UKB we apply different genetic analysis approaches to investigate the relationship between lanosterol and cataracts. In brief, we tested whether genetic variants in the lanosterol synthase gene region have a statistically significant association with cataract risk. We then extended our search to include genomic regions previously associated with lanosterol production and tested their association with cataract risk. Finally, we generated a genetic risk score using independent genetic variants previously associated with lanosterol, to test if their combined effect can provide evidence for lowering the risk of cataracts.

3.2 Methods

3.2.1 Sample and Variant QC

QC steps as set out in Sections 2.5.1 and 2.5.2 were applied.

3.2.2 Cataract definition

Individual level data was used to define cataract cases in the UKB by identifying participants' hospital inpatient records relating to diagnostic codes (Data-Field 41270 – ICD-10: H25 Senile cataract, H26 Other cataract, H28 Cataract and other disorders of lens in diseases classified elsewhere and Q12.0 Congenital cataract; and Data-Field 41271 – ICD-9: 7433 Congenital cataract and lens anomalies) and operation codes (Data-Field 41272 – OPCS4: C75.1 Insertion of prosthetic replacement for lens NEC and C71.2 Phacoemulsification of lens). Any individuals with self-reported cataracts (Data-Field 6148) or self-reported cataract surgery (Data-Field 5324 and 20004), but no supporting diagnostic or operation codes, were removed to reduce the risk of misclassification in the cases and controls.

3.2.3 Data for phytosterol traits

Genetic association estimates for the production of lanosterol from other phytosterols were obtained from Scholz *et al.* In this study, a genome-wide meta-analysis of the metabolism of phytosterols was performed using 9,758 individuals from six studies: KORA, LIFE-Adult, LIFE-Heart, LURIC and Sorbs. The conversion of phytosterols, which included brassicasterol, campesterol, sitosterol and stigmasterol, to lanosterol was described using the ratio of both total and free concentrations of each phytosterol-to-lanosterol ratio in reaction equilibria. In total, 8 phytosterol-to-lanosterol ratio traits were included (Scholz *et al.*, 2022).

3.2.4 Data for replication

Two cohorts were used to replicate any identified statistically significant associations:

1. GWAS summary statistics from FinnGen (R9) for cataract senile (59,522 cases and 312,864 controls) and cataract other (17,699 cases and 312,864 controls) (Kurki *et al.*, 2023).
2. Multi-ethnic meta-analysed cataract GWAS summary statistics from Choquet *et al.* using UK Biobank and Genetic Epidemiology Research in Adult Health and Aging (GERA) cohorts. This included 585,243 individuals (67,844 cases and 517,399 controls) (Choquet *et al.*, 2021)

3.2.5 Statistical analysis

Analysis was performed using the statistical software R v4.0.5 (R Core Team, 2021), unless otherwise stated. PLINK v2.0 (https://www.cog-genomics.org/plink/2.0/) was used to identify independent SNPs and generate the genetic risk scores across the 8 phytosterol-to-lanosterol ratio traits.

3.2.6 Genome-wide association study

Genetic associations for cataract development were obtained through a GWAS using the previously described cataract phenotype. The cataract GWAS was conducted using REGENIE using QC and methods previously outlined in Section 2.6.5.

Following QC, 45,449 cases were identified, alongside 353,371 controls. A description of cases and controls can be found in Table 3.1. The covariates included were sex, age, agexsex, age squared and the first 10 principal components of the genetic data.

To assess potential inflation in test statistics and distinguish between polygenicity and confounding, we applied LD Score Regression (LDSC) using the provided scripts (munge_sumstats.py and ldsc.py) available at https://github.com/bulik/ldsc. The LDSC

intercept reflects inflation attributable to confounding (e.g. population stratification, relatedness, or technical artefacts), supported by an associated ratio that represents the proportion of inflation not explained by polygenicity (B. K. Bulik-Sullivan *et al.*, 2015).

Relevant Manhattan and QQ plots with corresponding genomic inflation factor are found in Supplementary Figure 3.2 and 3.3 of the Appendix.

3.2.7 Approach 1: Identifying SNPs in the region of the *LSS* gene

The drug target, lanosterol, is synthesised by lanosterol synthase (*LSS*) (Zhao *et al.*, 2021). A list of SNPs for *LSS* was obtained through the National Library of Medicine gene database using assembly GRCh37.p13 (PubChem, 2025). The *LSS* gene is located on chromosome 21 within the base pair region of 47608360 and 47648688. The base pair region was expanded by 5Kb to identify SNPs between 47603360 and 47653688. The cataract GWAS results were filtered according to the expanded *LSS* coordinates. Statistical significance was accepted at a multiple testing adjusted *p*-value < (0.05 / the number of independent SNPs present in the selected sample). Independent SNPs were identified using PLINK v2.0 to produce a pruned list of variants in approximate linkage equilibrium, using a r^2 threshold of < 0.1, in the *LSS* gene region.

3.2.8 Approach 2: Comparison of phytosterol-to-lanosterol ratios GWAS with cataract GWAS results

The published GWAS results for the phytosterol-to-lanosterol ratio traits were compared to the generated cataract GWAS results. The phytosterol-to-lanosterol ratio summary statistics were filtered to the GWAS accepted *p*-value < $5x10^{-8}$ and independent SNPs for the look-up analysis were identified for each individual phytosterol-to-lanosterol ratio summary statistic at an r^2 threshold of < 0.1. Independent SNPs were identified using the LD pruning function in PLINK v2.0.

The cataract summary statistics were filtered for a *p*-value < (0.05 / the number of independent SNPs present across all phytosterol-to-lanosterol traits). Independent SNPs for the multiple testing correction were identified by combining variants across all 8 phytosterol-to-lanosterol traits and LD pruning at an r^2 threshold of < 0.1 using PLINK v2.0.

3.2.9 Approach 3: Genetic risk score analysis

An unweighted genetic risk score (GRS), calculated by the sum of the risk alleles representing decreasing lanosterol, was used to represent a summary of the genetic predisposition for lower levels of lanosterol (Igo, Kinzy and Bailey, 2019; Seral-Cortes *et al.*, 2021). To calculate the GRS for the lanosterol related traits observed in Scholz *et al.* (Scholz *et al.*, 2022), we

identified independent SNPs by combing all SNPs across the 8 phytosterol-to-lanosterol traits and then pruning the full list of variants using an r^2 threshold of 0.1. 9 independent SNPs were identified across all 8 traits. All effect sizes were set to 1 to perform an unweighted GRS, representing a decreased amount of lanosterol within the ratio. The GRS was generated for all individuals in the UKB using PLINK v2.0 (Chang *et al.*, 2015). The association between the GRS and cataract outcome was estimated using a logistic regression analysis adjusting for sex, age, agexsex, age squared and the first 10 principal components**.** Post QC, as described in Section 2.5.1, an additional filter for relatedness was applied, resulting in 36,952 cases and 290,623 controls present in the regression analysis.

3.3 Results

3.3.1 Case-control description

Table 3.1: UK Biobank case-control baseline characteristics by cataract status. If applicable, standard deviations are presented in round brackets.

	Females			**Males**		
	All	**Cases**	**Controls**	**All**	**Cases**	**Controls**
N	215,429	26,301	189,128	183,591	19,148	164,443
Mean Age (years)	56.61 (7.93)	62.62 (5.49)	55.78 (7.85)	57.06 (8.10)	62.63 (5.69)	56.41 (8.09)
Mean BMI (kg/m^2)	27.05 (5.14)	27.68 (5.21)	26.96 (5.12)	27.85 (4.23)	28.23 (4.36)	27.81 (4.21)
Ever Smoked (%)	40.61	44.78	40.03	51.02	60.08	49.97
Have Diabetes (%)	3.38	7.19	2.85	6.45	13.19	5.67
Employed/Self-Employed (%)	54.28	29.13	57.78	59.98	35.96	62.77

After QC, 399,020 individuals remained (215,429 Females and 183,591 Males) in the sample. The sample included 45,449 cases and 353,571 controls. Table 3.1 summarises the sample's baseline characteristics, including sex, age, BMI and lifetime smoker, diabetes, and employment status at the initial point of assessment.

3.3.2 UKB GWAS

Replication of independent SNPs in UKB cataract GWAS in FinnGen

Using the UKB cataract GWAS summary statistics, we assessed replication of the independent genome-wide significant loci in the available FinnGen cataract GWAS. A total of 32 SNPs were available for replication. The comparison between effect estimates is shown in Supplementary Figure 3.1, which includes a scatter plot of UKB versus FinnGen betas and the corresponding correlation coefficient (r = 0.938).

Genomic inflation assessment

As shown in Supplementary Figure 3.3, the genomic inflation factor for the UKB cataract GWAS is λ = 1.157, suggesting modest inflation. LDSC estimated an intercept of 1.027, which is low, and a ratio of 0.136. These results indicate that only 13.6% of the observed inflation is consistent with confounding, while the majority is attributable to polygenicity.

3.3.3 Identifying genetic variants in the *LSS* gene region

We identified 203 SNPs available in our summary statistics results and present within the region of the *LSS* gene. A locus specific Manhattan plot for the 203 SNPs can be seen in Figure 3.1 and full results can be found in Supplementary Table 3.1 of the Appendix. Overall, 13 independent SNPs were identified from the sample of 203 SNPs. SNPs with *p*-value < (0.05 / 13) were considered statistically significant for affecting cataract outcomes. One SNP, rs191009864, met the significance threshold.

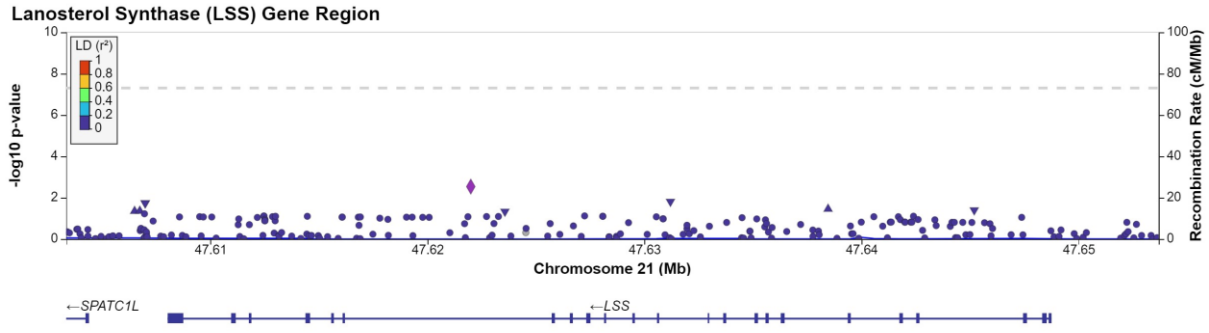


Figure 3.1: Manhattan plot produced using LocusZoom (Pruim et al., 2010) displaying the results of Supplementary Table 3.1 of the Appendix. The dotted line is set at a significance threshold of $5x10^{-8}$. Left y-axis displays -log10(p-value), and x-axis displays the LSS base pair region in chromosome 21 (GRCh37).

SNP rs191009864 was not present in the replication cohorts. Proxy SNPs, rs137865789, rs147393304, rs141632640 were identified using LD Link (Machiela and Chanock, 2015) in a European population, ± 500kb of the rs191009864, with an r^2 = 1.0, and were present in the FinnGen summary statistics, none of the proxy SNPs tested were able to replicate our findings. The proxies identified were not present in the Choquet *et al.* multi-ethnic GWAS. Approach 1 was repeated in the Choquet *et al.* multi-ethnic cataract GWAS and no statistically significant SNPs were identified within the *LSS* gene region. These lookups suggest that the observed association of SNP rs191009864 may represent a false positive.

Table 3.2: Description of rs191009864 and proxies (r^2 = 1.0) located in FinnGen summary statistics.

SNP	Summary Statistics	Effect Allele	Other Allele	EAF	BETA	SE	P
rs191009864	REGENIE cataract	A	G	0.013	-0.104	0.035	0.003
rs137865789*	FINNGEN cataract senile	A	G	0.004	0.030	0.063	0.630
	FINNGEN cataract other	A	G	0.004	-0.109	0.096	0.258
rs147393304	FINNGEN cataract senile	A	G	0.002	0.044	0.095	0.643
	FINNGEN cataract other	A	G	0.002	-0.138	0.143	0.332
rs141632640*	FINNGEN cataract senile	A	G	0.004	0.021	0.061	0.737
	FINNGEN cataract other	A	G	0.004	-0.109	0.093	0.245

**Alleles adjusted to reflect complementary strand.*

3.3.4 Investigating genetic variants across phytosterol-to-lanosterol ratios and cataract GWAS results

The GWAS summary statistics provided by Scholz *et al.* featured 8 phytosterol-to-lanosterol ratios in the cholesterol synthesis pathway. Using the generally accepted *p*-value threshold of < $5x10^{-8}$ we identified independent SNPs associated with each of the 8 phytosterol-to-lanosterol ratios. We followed-up these SNPs in our cataract GWAS in the UKB cohort. In total 23 SNPs were identified across both sets of results. The full results of the look-up analysis can be found in Supplementary Table 3.2 of the Appendix.

Within the phytosterol-to-lanosterol ratio summary statistics, we can infer that SNPs with a beta < 0 imply an increased presence of lanosterol within the reaction equilibria. For the cataract GWAS, SNPs with a beta < 0 indicate a reduced risk of cataracts. The presence of SNPs with a cataract GWAS beta < 0 and phytosterol-to-lanosterol beta < 0 would indicate increasing lanosterol is protective against cataract risk. Furthermore, SNPs with a cataract GWAS beta > 0 and phytosterol-to-lanosterol beta > 0 indicates decreased lanosterol is associated to cataract development.

All statistically significant SNPs across all 8 phytosterol-to-lanosterol traits can be tagged by just 9 independent SNPs. Therefore, statistical significance for the cataract associations, adjusted for multiple testing, was set at a *p*-value threshold of < (0.05 / 9). No statistically significant SNPs were identified to overlap between both cataract and phytosterol-to-lanosterol GWAS results.

3.3.5 Genetic risk score analysis

We used the identified 9 independent SNPs (rs10205879, rs10208987, rs11057839, rs145288624, rs3846662, rs612169, rs6735229, rs67734975 and rs7599981) to create our unweighted genetic risk score.

The logistic regression for the unweighted genetic risk score on cataract outcomes yielded no statistically significant relationship between individuals' GRS and cataract risk (OR = 1.002, ln(OR) SE= 0.003, P(>|z|) = 0.568).

3.4 Discussion

Current studies surrounding the effect of lanosterol as a treatment for cataract formation are divided. While some studies have supported the effectiveness of lanosterol in treating cataracts (Zhao *et al.*, 2015; Balashova *et al.*, 2018), others have suggested lanosterol is ineffective in the breakdown of cataracts on the lens (Shanmugam *et al.*, 2015; Daszynski *et al.*, 2019). This study aimed to investigate the genetic evidence for an association between lanosterol and cataract to help assess lanosterol as a possible drug treatment option for cataract. Using previously published genetic results for phytosterol-to-lanosterol ratios and generated UKB genetic data for cataract risk, we tested for genetic evidence around the *LSS* gene region, compared overlapping genetic variants in GWAS results related to the presence of lanosterol and conducted a genetic risk score analysis. We found the *LSS* SNP rs191009864 to be statistically significant at a multiple testing adjusted *p*-value < (0.05 / 13) with risk of cataract in our UKB data. However, we were unable to replicate this, or the identified proxies, in either FinnGen or Choquet *et al.* summary statistics. Therefore, we cannot claim that robust statistically significant evidence was found to support an association between lanosterol and cataract risk. We did not identify any other SNPs associated with lanosterol metabolism as associated with cataract risk. A score based on these genetic variants also had no evidence of association with cataract risk. Overall, the results of this investigation do not support the use of lanosterol as a treatment for cataract.

Our classification for cataract cases for the cataract GWAS were derived differently than other published results. Various cataract definitions have been used across different studies, for example, Choquet *et al.* produced a multi-ethnic cataract GWAS using self-reported cataract operation and ICD-10 diagnostic codes (Choquet *et al.*, 2021), whilst an observational study using UKB cataract cases conducted by Chua *et al.* used operative codes to define incident cataract (S. Y. Chua *et al.*, 2021). Our study utilised a combination of cataract definitions, previously used in the UKB, to maximise available cataract cases. The definition used for the cataract GWAS included operation codes C75.1 Insertion of prosthetic replacement for lens

NEC and C71.2 Phacoemulsification of lens from Data-Field 41272 and the diagnostic codes H25, H26, H28 and Q12.0 from Data-Field 41270. This definition does not account for the difference in cataract subtypes, for example, cortical, nuclear, or posterior subcapsular. Therefore, our genetic analyses would not account for lanosterol potentially being more effective in treating a specific subtype, a limitation also identified by Chua *et al.* (S. Y. Chua *et al.*, 2021).

Furthermore, our cataract GWAS was conducted over a European cohort. However, associations between lanosterol and cataract risk have been identified among other ethnic groups. For example, Zou *et al.* discovered evidence that *LSS*-rs2968 A allele is associated with nuclear age-related cataract risk within a Chinese population. However, *LSS*-rs2968 was not found to be statistically significant at a Bonferroni corrected *p*-value within an overall age-related cataract definition (Zou *et al.*, 2020). In our study, *LSS*-rs2968 also did not reach statistical significance ($p = 0.085$) within our cataract GWAS. Additionally, Zou *et al.* reported a protective effect of *LSS*-rs2968 A allele, while our results suggest the opposite, indicating a causative effect with cataract; however, not statistically significant. These results suggest the need to investigate the effects of lanosterol on specific cataract subtypes, such as nuclear cataract. Further analysis across additional ethnic groups would also be useful in understanding lanosterol's potential role as a treatment for cataract.

When generating the UKB cataract GWAS, REGENIE facilitated a mixed-model-based approach, allowing for the inclusion of related individuals. This provided 45,449 cases alongside 353,371 controls. Another benefit of using REGENIE is that it accounts for the case-control imbalance that was present within the cataract phenotype to reduce the risk of Type 1 errors and inflated estimates, whilst also improving statistical power (Mbatchou *et al.*, 2021). To further assess the robustness of the GWAS results, we calculated the genomic inflation factor as well as the LDSC score regression intercept and ratio. The observed genomic inflation factor ($\lambda = 1.157$) indicates modest test-statistic inflation, which can arise both from polygenicity and confounding biases such as population stratification, relatedness, or technical artefacts. Since the genomic inflation is sensitive to sample size, using LDSC to distinguish between these causes is critical in large biobank-based GWAS. In our analysis, the LDSC intercept (1.027) was close to 1, suggesting minimal residual confounding. A ratio of 0.136, supports this interpretation that the majority of the inflation reflects polygenicity of cataract rather than confounding bias. Thus, the genomic inflation observed is most likely attributable to the large sample size in the presence of polygenicity, rather than confounding.

The effect of lanosterol could also differ depending on the severity and maturity of the cataract. Given that the cataract definition includes operation codes, we can infer the presence of

mature cataract cases in the analysis. Considering the severity of cataracts, splitting age-related and early onset cases, and assessing both separately could change the drug efficacy observed. However, surgery can occur due to external factors other than visual impairment and may not be indicative of a cataract endpoint. Furthermore, our genetic analysis assessed the effect of lanosterol on cataract risk through genetic predisposition and should be unaffected by our cataract endpoint.

Using genetics to understand the clinical application of a potential treatment can be difficult but has been successful in recent literature. For example, an investigation on lowering cholesterol levels successfully utilised genetic proxies to mimic the effect of enzyme inhibitors and statins (Ference *et al.*, 2019). A challenge faced in this study was that we considered the exposure to lanosterol over a lifetime at low concentrations, as opposed to the far higher concentrations that would be used in pharmacological interventions. Therefore, we may have observed a lack of association due to a low effect size of lanosterol. However, genetics can still be utilised to ascertain lanosterol's role in reducing cataract risk using different approaches.

For example, Xu *et al.*, in a review of pharmacotherapy of cataracts, concluded that the use of lanosterol derivatives in steroid eye drops could be more efficient in reversing protein aggregations than lanosterol. As lanosterol is naturally occurring and is a component of the synthesis of cholesterol, it is unable to maintain a high concentration on the lens. Furthermore, lanosterol's low solubility limits its clinical application. However, Xu *et al.* found that lanosterol derivatives were able to effectively break down protein aggregations while avoiding the limitations surrounding lanosterol itself (Xu *et al.*, 2020). This suggests that lanosterol derivatives could form a viable cataract treatment, rather than lanosterol itself. Therefore, a genetic analysis of lanosterol derivatives is required to validate its use in cataract prevention.

A further limitation was observing the GWAS results of phytosterol-to-lanosterol ratios rather than using a GWAS on lanosterol levels, due to a lack of data availability. Using the results of the GRS analysis as an example, the independent SNPs used to generate risk scores only indicated the presence of lanosterol within the biosynthesis of cholesterol. A lanosterol levels GWAS is required to analyse the effect of lanosterol itself. Our analysis was conducted on blood measurements of lanosterol and their extension onto the lens. The lens capsule is unique in its selective permeability of macromolecules, and proteins, therefore substances found in the blood may not necessarily reflect what occurs locally on the lens (Danysh and Duncan, 2009). Reyes *et al.* examined *LSS* expression levels directly from the lens and discovered its overexpression on cataractous lenses, suggesting potential limitations in assessing blood measurements of lanosterol against cataract risk (Reyes *et al.*, 2023).

Expression quantitative trait loci (eQTL) were considered to assess the causal effect of lanosterol on cataract risk. However, current eQTL data is representative of the gene expression of lanosterol in the blood as opposed to the lens. An expansion of eQTL data available would allow for additional genetic analysis.

The results from different genetic analyses found no genetic evidence to support lanosterol's potential role as a treatment for cataract. Further genetic understanding of the direction of effect of lanosterol levels and its derivatives on cataracts would be beneficial in establishing its role as a non-surgical treatment. Additional analysis across different ethnicities and cataract subtypes may be needed to better understand the effect of lanosterol on cataract risk.

4 Exploring the causal relationship between vitamin D levels and deficiency with the risk of cataract: A Mendelian randomisation study

4.1 Introduction

As noted in Section 1.3.4, surgical extraction of the lens is the only current treatment for cataracts. Therefore, establishing preventive measures may ease the burden on healthcare services globally (Shu *et al.*, 2023). Exploring possible alternatives to surgery, via modifiable risk factors, including vitamin D levels and deficiency, could provide cost-effective solutions with established global availability, such as vitamin D supplementation (Chugh and Dabas, 2021).

Vitamin D is a fat-soluble vitamin, primarily obtained through sunlight exposure via synthesis in the skin which accounts for 90% of total vitamin D replenishment. Vitamin D can also be sourced through the consumption of animal- and plant-based foods such as oily fish and egg yolks. Medical properties of vitamin D are well documented within current literature, notably for calcium homeostasis and bone metabolism (Chang and Lee, 2019).

Vitamin D serum levels of 25(OH)D are used as a marker of vitamin D status. The National Institute for Health and Care Excellence (NICE) states that the UK recommended thresholds for deficiency, insufficiency and sufficiency are 25(OH)D < 25 nmol/L, 25-50 nmol/L, and > 50 nmol/L, respectively (NICE, 2022). Recent studies have also highlighted the prevalent nature of vitamin D deficiency and its current burden on healthcare systems worldwide (Cui *et al.*, 2023). It has been estimated that 40.4% of Europeans suffer from vitamin D deficiency and insufficiency, 13.0% with 25(OH)D concentrations below 30nmol/L (Cashman *et al.*, 2016). Globally, an approximate of 1 billion people suffer from vitamin D deficiency (Nair and Maseeh, 2012).

While vitamin D levels have been previously associated with several diseases, studies have also highlighted the association between vitamin D levels and ocular conditions, such as cataract. Chan *et al.* suggests that higher vitamin D levels reduces the risk of cataract development (Chan *et al.*, 2022). Other observational studies have suggested that lower vitamin D levels are associated with increased cataract risk (Brown and Akaichi, 2015; Abdellah *et al.*, 2019). As mentioned in Section 1.4.2, a case-control study investigating vitamin D levels and cataract amongst individuals with a mean age of 48.1 ± 8.5, observed lower vitamin D levels in cases compared to controls (Öktem and Aslan, 2021).

Although observational studies conducted thus far have indicated an association, due to the potential presence of unmeasured confounding factors, it is important to note that these relationships are based on correlation rather than causation. For example, in the context of vitamin D and cataract, type 2 diabetes is a known confounding factor associated with both variables (Kiziltoprak *et al.*, 2019; Khudayar *et al.*, 2022). Unadjusted confounding factors can lead to misinterpreted associations within observational studies (Davies, Holmes and Smith, 2018).

Furthermore, while sun exposure is the primary source of obtaining vitamin D, UV exposure on the lens is suggested to increase the risk of developing cataracts (Roberts, 2011; Delcourt *et al.*, 2014; Vashist *et al.*, 2020). This suggests that, although some observational studies have indicated that vitamin D and its antioxidative properties may help reduce the risk of cataracts, prolonged UV exposure can damage lens proteins through glycation processes, potentially increasing the risk of cataract development (Linetsky *et al.*, 2014). Therefore, further analysis is required to investigate a possible causal relationship.

As discussed in Section 2.7, evidence for causality can be established through using MR. There are different approaches to conducting MRs, including one- and two-sample MR analyses. In one-sample MR, both the exposure and outcome data are obtained from the same population sample and individual level data is used, while in two-sample MR, the genetic variants for the exposure and outcome are derived from summary level data of different samples. Both approaches are susceptible to similar bias and are dependent on the availability and quality of data (Lawlor, 2016).

In this investigation, we aim to use available genetic data from European and multi-ethnic cohorts and explore the potential causal association between vitamin D levels and deficiency with cataract. Leveraging data from the UKB, we conduct observational analyses to explore the relationship between vitamin D and incident cataract. However, to address the limitations of observational studies, we further conduct MR analyses using UKB data and publicly available and generated GWAS, to investigate the possible causal relationship between vitamin D and cataract.

4.2 Methods

4.2.1 Data for observational and one-sample MR analysis

Vitamin D phenotype: UKB

Vitamin D levels (nmol/L) were extracted from the UKB cohort using Data-Field 30890. Vitamin D measurements were available for 448,311 UKB participants, obtained from biological

samples from their initial assessment. Information regarding vitamin D supplementation was obtained from Data-Field 6155 and encoded as cases where individuals declared that they take vitamin D or multivitamin supplements when asked “Do you regularly take any of the following? (You can select more than one answer)”.

Incident cataract phenotype: UKB

In the UKB, incident cataract cases were defined using both diagnostic and operation codes. This includes ICD-10 classifications of senile cataract (H25), other cataract (H26), and cataract and other disorders of lens in diseases classified elsewhere (H28) (Data-Field 41270). Operation classifications were determined using OPCS-4 for surgical cataract cases of insertion of prosthetic replacement for lens NEC (C75.1) and phacoemulsification of lens (C71.2) (Data-Field 41272).

Phenotypic data outlining the corresponding date of each relevant diagnostic and operation code were recorded (Data-Fields 41280, 41281 and 41282) and compared against the date participants first attended the UKB assessment centre (Data-Field 53) to identify incident cataract cases. Prevalent cases, occurring prior to the initial UKB assessment centre test date, were removed from the analysis.

QC steps as outlined in Sections 2.5.1 and 2.5.2 were applied, excluding the additional relatedness filters. In this instance, to maximise the number of cases available, relatedness was filtered in the phenotype by removing one individual from each related pair was removed at random from our cases and all related pairs were removed from the controls. After filtering, 31,231 cases and 239,870 controls remained for incident cataract.

UKB individual level data was also used in the analysis to generate genetic risk scores and genome-wide association results.

4.2.2 Data for two-sample MR analysis

4.2.2.1 Exposure data

Vitamin D deficiency: UKB

Genetic instruments representing vitamin D deficiency were obtained from the GWA analysis conducted by Amin and Drenos in the UKB cohort (Amin and Drenos, 2021). Vitamin D deficiency cases were identified by vitamin D levels < 25nmol/L and controls were defined by vitamin D levels ≥ 50nmol/L. Post QC, 35,079 cases and 140,898 controls remained, with 17 independent genetic variants identified (Amin and Drenos, 2021).

Vitamin D levels: SUNLIGHT Consortium

Genetic instruments for vitamin D levels were obtained from the publicly available GWA analysis from the SUNLIGHT (Study of Underlying Genetic Determinants of Vitamin D and Highly Related Traits) consortium (Wang *et al.*, 2010; Jiang *et al.*, 2018). The study by Jiang *et al.* comprised a meta-analysis across 31 GWA studies and 79,366 individuals. Further details on the studies, including QC and analysis, are provided by Jiang *et al.* (Jiang *et al.*, 2018).

Vitamin D levels: Manousaki et al.

We further utilised genetic instruments for vitamin D levels using publicly available GWA analysis by Manousaki *et al.* using 401,460 European individuals from the UKB cohort. The study measured baseline vitamin D levels, adjusting for vitamin D supplementation across 24,874 individuals. This was meta-analysed, using previous 25(OH)D GWAS published by Manousaki *et al.* (Manousaki *et al.*, 2017). The analysis provided 138 conditionally independent SNPs for serum 25(OH)D. Additional analysis and details on QC are documented by Manousaki *et al.* elsewhere (Manousaki *et al.*, 2020).

4.2.2.2 Outcome data

UKB cataract GWAS

Genetic associations for all cataract outcomes, including prevalent cases, were obtained from the previous REGENIE GWAS as discussed in Section 3.2.6. Corresponding Manhattan and QQ plots can be found in Supplementary Figure 3.2 and 3.3.

UKB and GERA multi-ethnic cataract GWAS

Choquet *et al.* conducted a meta-analysis using a multi-ethnic cataract GWAS including 585,243 individuals (67,844 cases and 517,399 controls) from UKB and Genetic Epidemiology Research in Adult Health and Aging (GERA) cohorts. Cases were determined using ICD-10 diagnostic codes and self-reported cataract operations. Protocols on GERA's participant genotyping and previous QC include the exclusion of variants according to poor clustering, ethnic-specific low batch representation and effect allele frequency, extreme heterozygosity (EUR only) and call rates < 90%. Further details of the analysis and QC is provided by Choquet *et al.* (Choquet *et al.*, 2021).

4.2.3 Statistical analysis

Unless specified otherwise, analysis was conducted using the R statistical software v4.0.5 (R Core Team, 2021) with MR conducted through the "TwoSampleMR" package (Hemani *et al.*, 2018) and results visualised using "ggplot2" (Villanueva and Chen, 2019) and "forestploter" packages (Dayimu, 2024). PLINK v2.0 (https://www.cog-genomics.org/plink/2.0/) was used to

generate genetic risk scores for vitamin D levels in the UKB (Chang *et al.*, 2015). Clumping was performed at a linkage disequilibrium threshold of $r^2 = 0.001$ using the "TwoSampleMR" package.

4.2.4 Observational analysis

Three observational analyses were performed using our incident cataract definition, with all analyses performed using a logistic regression. The first analysis regressed incident cataract on baseline vitamin D levels. 23,974 individuals were removed due to missing data. Therefore, the analysis included 247,127 individuals, including 28,305 cases for incident cataract (218,822 controls).

The second observational analysis tested incident cataract on vitamin D status (sufficient and deficient). Vitamin D deficiency was defined as individuals with vitamin D levels ≤ 25 nmol/L, with sufficient levels ≥ 50 nmol/L. Individuals with insufficient vitamin D levels or missing data were removed from this analysis (125,850 individuals). The logistic regression used 145,251 individuals, including 29,233 cases of vitamin D deficiency (116,018 controls) and 16,934 cases of incident cataracts (128,317 controls).

Both observational analyses were controlled for sex, age, agexsex and age squared.

The third observational analysis investigated the association between the vitamin D supplementation and incident cataract. UKB participants who reported taking vitamin D or multivitamins were considered to supplement vitamin D in their diet. Individuals who preferred not to respond to the question were excluded from the analysis.

The analysis was performed twice: first controlling for sex, age, agexsex, age squared, and baseline vitamin D levels, while the second analysis excluded baseline vitamin D levels from the covariates. Both analyses utilised 28,141 incident cataract cases and 218,063 controls.

4.3 Mendelian randomisation

This investigation uses both one-sample and two-sample MR approaches to ensure the robustness of findings, with details on each approach previously discussed in Section 2.7. The use of both approaches strengthens the reliability of our results by addressing their respective limitations, as discussed later in the manuscript (Davies, Holmes and Smith, 2018).

4.3.1 One-sample MR

A weighted genetic risk score (GRS) was generated, calculated by the sum of risk alleles representing a genetic predisposition to higher vitamin D levels based on their respective effect sizes in the Jiang *et al.* GWAS. Summary statistics were filtered for *p*-value ≤ 5×10^{-8} and

independent SNPs were identified using the "TwoSampleMR" R package using the "clump_data" function ($r^2 = 0.001$). All SNP effect sizes were made positive for vitamin D levels, reassigning the effect allele where appropriate. Therefore, individuals with larger GRS values had an increased genetic predisposition to higher vitamin D levels. The one-sample MR (1SMR) was conducted in two steps:

1. First, a linear regression on vitamin D levels from the UKB on the generated GRS for vitamin D levels.
2. Second, a logistic regression was run on the fitted values from the previous step against incident cataract cases in the UKB, controlling for sex, age, agexsex, age squared and the first 10 genetic principal components (PCs).

A GRS for vitamin D deficiency was created using the 17 independent SNPs identified by Amin and Drenos, representing a genetic predisposition to vitamin D deficiency (Amin and Drenos, 2021). SNP effect sizes were made positive in relation to the exposure, and effect alleles were changed as appropriate when generating the GRS. The above steps for the 1SMR analysis were repeated, modifying step 1 to a logistic regression for vitamin D deficiency (cases were defined as levels ≤ 25 nmol/L and controls defined as levels ≥ 50 nmol/L) on the generated GRS for vitamin D deficiency.

As a sensitivity analysis, an additional GRS for vitamin D levels was created using the 138 conditionally independent SNPs identified in the Manousaki *et al.* GWAS (Manousaki *et al.*, 2020) and present in the individual-level data from the UKB dataset. The steps for the 1SMR analysis were then repeated accordingly.

4.3.2 Two-sample MR

A two-sample MR (2SMR) analysis was conducted to evaluate the causal relationship between vitamin D levels and vitamin D deficiency (exposures) with cataract (outcome). Exposure data consisted of vitamin D levels retrieved from Jiang *et al.* GWA-meta analysis of the SUNLIGHT consortium (Jiang *et al.*, 2018) and vitamin D deficiency instrumental variables from the Amin and Drenos GWA analysis of the UKB cohort (Amin and Drenos, 2021). Two sources of outcome data for cataract were used, one generated in the UKB and the other obtained from a multi-ethnic cataract GWAS (Choquet *et al.*, 2021; Hashimi *et al.*, 2024).

SNPs associated with vitamin D levels were filtered for a *p*-value ≤ 5×10^{-8}. SNPs were then clumped to identify independent SNPs ($r^2 = 0.001$). Genetic variants for vitamin D deficiency from Amin and Drenos were filtered for statistical significance and independence, so no further thresholds were applied (Amin and Drenos, 2021). Exposure and outcome data were

harmonised to ensure that SNP effects on the exposure and outcome correspond to the same allele. The 2SMR was performed with the “TwoSampleMR” R package (Hemani *et al.*, 2018).

As detailed in Section 2.7.3, a total of 5 MR methods can be used to interpret the results of a 2SMR. In this analysis we interpreted from all 5 methods: MR Egger, weighted median, IVW, simple mode and weighted mode. Each method applies distinct assumptions and criteria to generate the causal effect of the exposure on an outcome (Slob and Burgess, 2020). The results for this study are primarily interpreted by the IVW method. The MR Egger intercept was used to assess the presence of horizontal pleiotropy. Horizontal pleiotropy occurs when variants of the MR can affect the outcome through an alternative pathway, outside of the exposure (Burgess and Thompson, 2017). If horizontal pleiotropy is detected, results can be interpreted using other MR methods, which account for the presence of pleiotropy within the genetic variants used (Davies, Holmes and Smith, 2018).

Using the 138 independent SNPs identified by Manousaki *et al.* as instruments for vitamin D levels, we performed two additional 2SMR as sensitivity analyses. To remain consistent to the previous 2SMR analysis, these analyses utilised outcome data from the UKB cataract and a multi-ethnic cataract GWAS. Causal estimates were derived using IVW and three additional MR methods, while MR-Egger was conducted to assess pleiotropy.

4.3.3 Gene-based analysis

Recent studies have identified four primary genes *DHCR7*, *CYP2R1*, *GC*, and *CYP24A1* as key regulators of vitamin D biological processes. Using the National Library of Medicine Genome Browser (https://www.ncbi.nlm.nih.gov/datasets/genome/), we retrieved the genomic coordinates for each gene based on the GRCh37.p13 assembly, with an additional 5 kb added.

We then extracted SNPs within these coordinates from the vitamin D GWAS conducted by the SUNLIGHT consortium (Jiang *et al.*, 2018). To ensure alignment with GRCh37.p13 in our datasets, we cross-referenced the SNPs with base pair positions in the UKB cataract GWAS. After verifying the coordinates, the identified SNPs from all four genes were combined and clumped to retain independent variants. A 2SMR analysis was then performed using the methods outlined earlier.

A non-linear Mendelian randomisation (NLMR) analysis was performed as a supplementary investigation to assess potential non-linear effects of vitamin D on cataract risk. Due to the known biases in current NLMR methods, these results are provided in Supplementary Analysis 4.1.

4.4 Results

4.4.1 Case-control description

After QC, 271,101 individuals remained (144,843 Females and 126,258 Males) in our sample. The sample included 31,231 cases of incident cataract and 239,870 controls. Table 4.1 summarises the sample's baseline characteristics, including sex, age, BMI, and vitamin D supplementation, ever-smoker, diabetes, and employment status at recruitment.

Table 4.1: UKB case-control baseline characteristics by cataract status.

	Females			**Males**		
	All	**Cases**	**Controls**	**All**	**Cases**	**Controls**
N	144,843	18,181	126,662	126,258	13,050	113,208
Mean Age (Years)	56.55 (7.89)	62.52 (5.47)	55.69 (7.81)	56.99 (8.10)	62.58 (5.66)	56.34 (8.09)
Mean BMI (kg/m2)	27.00 (5.15)	27.61 (5.20)	26.91 (5.13)	27.80 (4.21)	28.23 (4.34)	27.75 (4.19)
Take Vitamin D supplements (%)	27.54	27.76	27.51	19.66	19.50	19.68
Ever Smoked (%)	40.35	44.42	39.76	50.61	60.00	49.53
Have Diabetes (%)	3.33	6.92	2.81	6.31	12.89	5.55

4.4.2 Observational analysis

Vitamin D levels with incident cataract

The observational analysis found a statistically significant relationship between vitamin D levels and incident cataract, indicating increased vitamin D is associated with lower incident cataract risk (OR = 0.998, ln(OR) SE = 3.23×10^{-4}, $p = 6.72 \times 10^{-14}$). Therefore, per unit increase in vitamin D levels, an individual's risk of cataract decreases by approximately 0.2%.

Deficient vs. sufficient levels with incident cataract

A statistically significant relationship between vitamin D deficiency and incident cataract was observed, indicating vitamin D deficiency was associated with increasing incident cataract risk (OR = 1.237, ln(OR) SE = 0.022, $p = 9.05 \times 10^{-23}$).

Supplement use with incident cataract

The observational analysis found insufficient evidence of an association between the consumption of supplements containing vitamin D and the risk of incident cataract (OR = 0.971, ln(OR) SE = 0.016, $p = 0.057$). Similar results were observed when accounting for vitamin D levels as a covariate (OR = 0.993, ln(OR) SE = 0.016, $p = 0.636$).

4.4.3 One-sample MR

8 independent SNPs were used to generate a weighted GRS representing elevated levels of vitamin D in individuals within the UKB cohort. The GRS produced explained 2.6% of the variation of vitamin D levels.

The results of the two-step regression analysis showed no statistically significant relationship between vitamin D levels GRS and incident cataract (OR = 1.001, ln(OR) SE = 0.002, p = 0.541).

A weighted GRS representing the genetic prediction of vitamin D deficiency was generated using 17 independent SNPs in the UKB cohort. 2.1% of the variation (using the McFadden r^2 approximation) of vitamin D deficiency was explained by the generated GRS.

No statistically significant evidence was found for the relationship between vitamin D deficiency GRS, representing an increasing predisposition to vitamin D deficiency, and incident cataract risk (OR = 1.095, ln(OR) SE = 0.145, p = 0.534).

The additional sensitivity analysis, using the GRS for vitamin D based on the 138 conditionally independent SNPs from Manousaki *et al.,* did not identify a statistically significant association with incident cataract (OR = 1.000, ln(OR) SE = 0.002, p = 0.85).

4.4.4 Two-sample MR

Post harmonisation and clumping with UKB cataract results, 7 SNPs were present for vitamin D levels and 17 SNPs for vitamin D deficiency. For the 2SMR analysis with multi-ethnic cataract, post harmonisation, 7 SNPs were present for vitamin D levels and 15 SNPs for vitamin D deficiency. Palindromic SNPs with intermediate allele frequencies were removed, 1 from vitamin D levels, for both analyses, and 2 from vitamin D deficiency when using the multi-ethnic cataract outcome.

Vitamin D levels SUNLIGHT consortium and cataract

Using genetic instruments from the SUNLIGHT consortium GWAS, no evidence was found to suggest a causal relationship between vitamin D levels and UKB cataract (IVW: OR = 1.122, 95% CI: 0.968-1.301, p = 0.125). Additionally, no evidence was found to suggest a causal relationship between vitamin D levels and multi-ethnic cataract (IVW: OR = 1.097, 95% CI: 0.963- 1.251, p = 0.165).

Vitamin D deficiency UKB and cataract

Furthermore, no evidence was found to suggest that vitamin D deficiency causally effects cataracts across the UKB GWAS (IVW: OR= 0.987, 95% CI: 0.959-1.015, p = 0.344). No evidence of a causal effect was found between vitamin D deficiency (IVW: OR= 0.988, 95% CI: 0.964-1.014, p = 0.361) and multi-ethnic cataract. Four additional robust MR methods were also used in this analysis. None of the additional methods found evidence of a causal association between vitamin D levels and deficiency with UKB cataract or between vitamin D levels and deficiency with Choquet *et al.* multi-ethnic cataract.

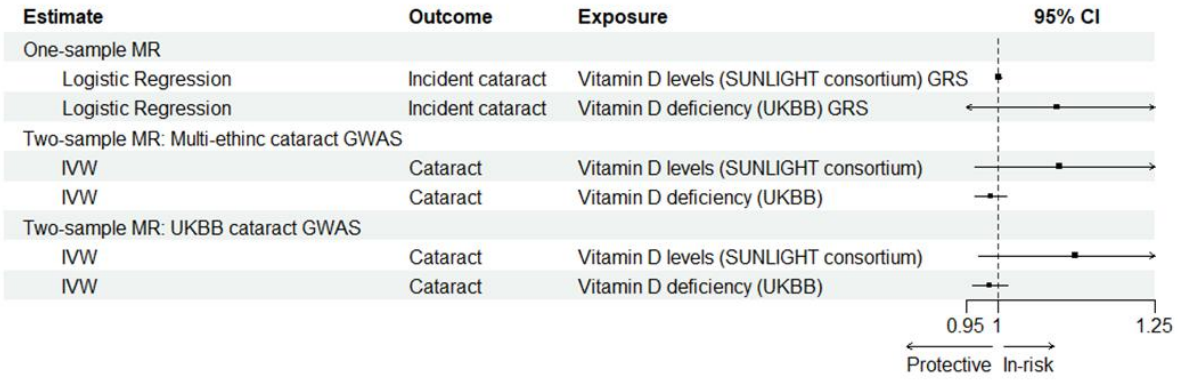


Figure 4.1: Forest plot depicting logistic regression and IVW results for one-sample and two-sample MR analyses. The analyses examine the relationship between genetic risk scores (GRS) for vitamin D levels and deficiency, derived from the SUNLIGHT consortium GWAS and UK Biobank, and the risk of incident cataract. Results are presented separately for vitamin D levels (SUNLIGHT consortium) and deficiency (UKB) across both two-sample MR analyses using UKB cataract and a multi-ethnic cataract GWAS. The black swathe represents the 95% confidence intervals.

Figure 4.1 displays the IVW results for vitamin D levels (SUNLIGHT consortium) and deficiency (UKB) for both UKB and multi-ethnic cataract GWAS.

Pleiotropy was also tested using the intercept of the MR Egger model for all analyses. No evidence of pleiotropy bias was found in any analysis of the MR Egger intercept at a p-value threshold of 0.05.

Full results of all MR methods and pleiotropy tests can be found in Supplementary Tables 4.1.1, 4.1.2, and 4.1.3 for UKB cataract and Supplementary Tables 4.2.1, 4.2.2, and 4.2.3 for Choquet *et al.* multi-ethnic cataract.

Using the genetic instruments for vitamin D levels identified by Manousaki *et al.* 90 SNPs were available after harmonisation for the UKB cataract GWAS, and 62 SNPs were available for the multi-ethnic cataract GWAS. The IVW analysis did not provide evidence of an association between vitamin D levels and cataract risk in either the UKB dataset (IVW: OR = 1.014, 95%

CI: 0.961–1.068, *p* = 0.617) or the multi-ethnic dataset (IVW: OR = 1.026, 95% CI: 0.966–1.090, *p* = 0.396). However, the weighted median analysis for vitamin D levels and UKB cataract suggested a potentially statistically significant association (OR = 1.076, 95% CI: 1.004–1.153, *p* = 0.038). Furthermore, the MR-Egger analysis indicated no evidence of bias due to pleiotropy in the results. Full results of all MR methods and pleiotropy tests can be found in Supplementary Tables 4.3.1, 4.3.2 and 4.3.3 for UKB cataract and Choquet *et al.* multi-ethnic cataract.

4.4.5 Gene-focused analysis results

After harmonisation and clumping, 4 SNPs were identified within the vitamin D-related genes from the SUNLIGHT GWAS and were present in both the UKB and multi-ethnic cataract GWAS. No evidence of a causal association was observed in the IVW analyses for either dataset (UKB: IVW OR = 1.164, 95% CI: 0.995–1.361*, p* = 0.058; Multi-ethnic: IVW OR = 1.114, 95% CI: 0.969–1.280*, p* = 0.129).

For full results, including MR-Egger intercept analyses, refer to Supplementary Tables 4.4.1, 4.4.2 and 4.4.3 for the UKB cataract and multi-ethnic cataract dataset.

Results from the non-linear Mendelian randomisation (NLMR) analysis are presented in Supplementary File 4.1.

4.5 Discussion

This study aimed to investigate the correlation and possible causal association between vitamin D levels and deficiency with incident cataract risk. By leveraging publicly available and generated GWA studies (Jiang *et al.*, 2018; Manousaki *et al.*, 2020; Amin and Drenos, 2021; Choquet *et al.*, 2021; Hashimi *et al.*, 2024), we conducted a comprehensive investigation into the association between vitamin D and cataract. Within this investigation we first explored an observational relationship between vitamin D and incident cataract. We then proceeded to conduct a 1SMR analysis within the UKB, further supplemented by a 2SMR analysis to identify any potential causal relationship between vitamin D levels or deficiency and cataract risk. While our observational analyses suggested an association between vitamin D levels and incident cataract, genetic analyses yielded no robust evidence of a causal relationship.

The observational analysis conducted in this investigation, although limited due to possible interference of confounding factors and reverse causality, aligns with current literature (Brown and Akaichi, 2015; Abdellah *et al.*, 2019; Öktem and Aslan, 2021; Chan *et al.*, 2022). We find a negative correlation between vitamin D levels and incident cataract risk. While deficient vitamin D levels were positively correlated with incident cataract risk when compared to

sufficient levels. When observing the effect of vitamin D supplementation on incident cataract risk, insufficient evidence was present to suggest an association between vitamin D levels and incident cataract is driven through the use of vitamin D supplements.

In contrast, our 1SMR suggested no evidence of a causal association between genetically predicted vitamin D levels (SUNLIGHT consortium or Manousaki *et al.* UKB GWAS), vitamin D deficiency (UKB) and incident cataract. Additionally, the 2SMR IVW analyses also showed no statistically significant association for vitamin D levels (SUNLIGHT consortium or Manousaki *et al.* UKB GWAS) or deficiency (UKB) with cataract risk. Whilst not statistically significant, we did observe vitamin D levels and deficiency impacting cataract risk in the opposite direction to what was suggested by our observational analysis. However, opposite effects when comparing estimates from observational and MR studies (as well as RCTs) have previously been investigated and are common, suggesting direction of effects are not comparable between analyses (Janiaud *et al.*, 2021). As previously stated, observational analyses are limited by confounding factors and the potential presence of reverse causality. Öktem and Aslan note that patients with cataracts may stay indoors longer and receive less exposure to sunlight, suggestive of a potential source of reverse causality where cataract development could induce vitamin D deficiency (Öktem and Aslan, 2021).

Some of our genetic analyses provided evidence suggesting a potential association between vitamin D levels and increased cataract risk. For example, our 2SMR weighted-median results found evidence of a statistically significant association between genetically predicted vitamin D levels, based on instruments identified by Manousaki *et al.* and cataract risk in the UKB GWAS (Weighted Median: OR = 1.076, 95% CI: 1.004–1.153, p = 0.038). Furthermore, our gene-focused analysis of four key vitamin D-related genes (*DHCR7*, *CYP2R1*, *GC*, and *CYP24A1*), showed weak evidence of an association between vitamin D levels and increased cataract risk in the UKB dataset (IVW: OR = 1.164, 95% CI: 0.995–1.361, p = 0.058) but no evidence of an association in the multi-ethnic dataset.

Similar findings were reported in a different 2SMR study investigating the relationship between the Manousaki *et al.* GWAS for vitamin D (predominantly European ancestry) with the Ishigaki et al. GWAS for cataracts (East Asian ancestry) (OR = 1.11, 95% CI: 1.00–1.22, p = 0.032) (Wang and Xin, 2024). This study also observed an increased risk of cataracts associated with higher vitamin D levels and suggested that this effect might be attributed to UV exposure. However, this mismatch in ancestry may introduce bias due to differences in linkage disequilibrium and allele frequencies between populations. Therefore, both 2SMR analyses are limited by bias introduced by the datasets used.

Rahman *et al.* investigated vitamin D supplementation from 2014 to 2020 using an RCT to test for a causal relationship between vitamin D and cataract risk. The RCT included individuals from the general Australian population ranging between 60-84 years old. Individuals were split into two groups, one given 60,000iU of vitamin D3 and the remaining given a placebo. Supplements were taken once a month over a period of 5 years. A total of 19,925 individuals remained eligible across the trial. The RCT found no evidence to suggest that vitamin D supplementation lowered the risk of cataract surgery (Rahman *et al.*, 2023). At present, no other RCTs have been conducted for vitamin D and cataract development. The results of our MR approach complement the findings of the Rahman *et al.* RCT and build on them further. As noted, RCTs have limitations, including restricted sample sizes, limited generalisability to different populations, and a constrained time frame for observation (Monti *et al.*, 2018). In contrast, our MR approach leverages a larger sample size and examines the lifelong genetic predisposition to higher vitamin D levels, offering insights into its potential impact on cataract risk. This approach extends our understanding of the causal relationship between vitamin D and cataract, by addressing limitations that can occur in RCT studies.

The results of this investigation and current RCTs suggest no effect of vitamin D supplementation on cataract risk; however, the identified association in the observational studies may be influenced by unaccounted confounding factors.

Vitamin D as an exposure variable has previously been studied using non-linear MR (NLMR) analyses. An additional analysis NLMR analysis was conducted to investigate a potential non-linear relationship between vitamin D and cataract risk (see Supplementary Analysis 4.1). However, these analyses are severely vulnerable to bias when the key assumption that genetic effects are homogenous across the population is violated (Sofianopoulou *et al.*, 2024). However, it has been reported, that this bias can be reduced by implementing varying stratification approaches in NLMR analyses and triangulating results to develop reliable inferences (Burgess, 2023). We used three proposed stratification methods which showed inconsistent results, with the latest improved method not supporting a statistically significant non-linear relationship. Furthermore, our negative control analysis identified bias in the other methods (Hamilton *et al.*, 2024). Given the reported biases and methodological challenges associated with NLMR, we chose not to include these analyses in the main results and instead focused our investigation on the more robust linear MR analyses. Therefore, we cannot claim any non-linear relationship between vitamin D levels and cataract risk based on non-linear methods. Future improvements in NLMR methodologies, particularly those addressing biases inherent to current approaches, may provide further insights into the complex non-linear effects of vitamin D levels on cataract risk.

Additional limitations were also present during this investigation. Pleiotropy is a reoccurring limitation of MR based studies. While our analysis of the MR Egger intercept suggests no horizontal pleiotropy present during our 2SMR analysis, we cannot fully eliminate the possible presence of pleiotropy. The cataract definitions used during our analyses may have further limited the study. It has been observed that different risk factors have varying effects on the development of differing cataract subtypes (nuclear, cortical and poster subscapular) (Vashist *et al.*, 2020; Rahman *et al.*, 2023). The definitions used in this study utilised diagnostic and operation codes; therefore, did not distinguish between cataract subtypes. As an opportunity for further research, investigating the effects of vitamin D on different cataract etiological subtypes (age-related, traumatic) and anatomical subtypes (nuclear, cortical, posterior subcapsular) may provide different evidence for vitamin D supplementation, as observed by Abdellah *et al.* in a case-control study amongst adults aged 50 years and over (Abdellah *et al.*, 2019).

Furthermore, while we utilised both the 1SMR approach, which uses individual levels data from the same sample for both the exposure and outcome, and the 2SMR approach using summary level data, each method has its respective strengths and limitations. The 2SMR approach can enhance statistical power and reduce biases such as weak instrument bias often observed in 1SMR; however, it relies on the critical assumption that the exposure and outcome datasets are independent and do not overlap. In this study, some degree of sample overlap was present within the 2SMR, which may introduce bias (Davies, Holmes and Smith, 2018; Burgess *et al.*, 2023). Due to the lack of data availability, optimal outcome data was not available in the 2SMR analyses. Through using my cataract and Amin and Drenos vitamin D deficiency GWAS results we introduced overlapping UKB samples which may bias the MR towards observational estimates; however, in this case it has not affected our overall conclusion of no evidence of association. In addition, using Choquet *et al.* multi-ethnic cataract GWAS, our 2SMR analysis is limited by differing ancestries across exposure and outcome data, potentially weakening the possibility of observing an association, but not likely to affect the direction of effect (Burgess *et al.*, 2023). To mitigate this risk, the analysis using vitamin D levels from the SUNLIGHT consortium and cataract outcomes from my UKB GWAS was considered the most robust and should be prioritised for interpretation.

4.6 Conclusion

In conclusion, whilst there is a correlative association between vitamin D and cataract, we found no robust evidence of a causal relationship between vitamin D and incident cataract risk. The results of our investigation do not support any clinical use of vitamin D supplementation for preventing the development of cataract.

5 Investigating the Relationship Between Alcohol Consumption and Cataract Risk: Findings from Observational and Genetic Analyses

5.1 Introduction

As referenced in Section 1.4.2, alcohol consumption has been suggested as a modifiable risk factor associated with various diseases such as several types of cancer, cardiovascular disease, diabetes, and liver disease (GBD 2016 Alcohol Collaborators, 2018; Ingold, Amin and Drenos, 2019; Karimi, Arabi and Shahraki, 2021; X. Zhang *et al.*, 2021). Higher levels of alcohol consumption have typically been more prevalent in older populations compared to younger populations (Veerbeek *et al.*, 2019). However, while the UK government recommends an alcohol limit of 14 units per week, exceeding this limit has become increasingly common among young adults (Bhatti *et al.*, 2020).

There have been inconsistent findings across studies reporting the association between alcohol intake and cataract risk. Heavy alcohol consumption has previously been linked to increased production of reactive oxygen species through the metabolism of ethanol in the liver by CYP2E1 enzyme, which may contribute to cataract formation (Gong *et al.*, 2015). Further research has also suggested that lifetime alcohol consumption is linked to higher risk of cataract surgery (Lindblad *et al.*, 2007; Kanthan *et al.*, 2010; Gong *et al.*, 2015; S. Y. Chua *et al.*, 2021; Fukai *et al.*, 2022). Other studies suggest mixed results. Kanthan *et al.* found no association between alcohol intake and long-term cataract risk, but discovered an increased likelihood of cataract surgery amongst heavy consumers of alcohol (Kanthan *et al.*, 2010). However, other studies have suggested a lack of evidence to suggest a link with alcohol consumption, depicting no protective or causal effects of moderate and high consumption, respectively (Wang and Zhang, 2014).

Studies have also suggested specific effects of general wine consumption on cataract risk, finding an association between wine consumption and reduced risk of cataract (Ritter *et al.*, 1993; S. Y. Chua *et al.*, 2021). The antioxidants found within wine, such as polyphenols and resveratrol, have been previously hypothesised to reduce the oxidative process leading to cataract formation on the lens (Prickett *et al.*, 2004; Arranz *et al.*, 2012; Abu-Amero, Kondkar and Chalam, 2016).

Observational associations between alcohol consumption and cataract development remain inconsistent and limited, due to limitations surrounding unmeasured confounding factors that

can suggest an association between the exposure and outcome when the association is due to an unaccounted variable, and risk of reverse causation, where the outcome influences the exposure, rather than exposure influencing the outcome. For example, in the case of alcohol consumption and cataract risk, sociodemographic factors may play a significant role. This link to sociodemographic level could influence factors such as access to healthcare, potentially driving the observed association between alcohol consumption and cataract risk (S. Y. Chua *et al.*, 2021). Other studies suggest that higher socioeconomic levels are associated with more frequent drinking, while lower socioeconomic levels are linked to heavier episodic drinking (Beard *et al.*, 2019). A further example of a confounding factor in investigations examining the relationship between alcohol consumption and cataract risk is smoking. Previous evidence highlights a strong association between smoking and alcohol consumption, as well as between smoking and an increased risk of cataracts (Ye *et al.*, 2012; Meader *et al.*, 2016; Saunders *et al.*, 2022).

To avoid the limitations of observational studies, as noted in Section 2.7, MR analyses can help distinguish causation from correlation while mitigating the limitations of both observational studies and traditional RCTs (Burgess *et al.*, 2020). A well-known genetic proxy for alcohol consumption is the variant rs1229984 in the alcohol dehydrogenase 1B gene (*ADH1B*), which encodes the ADH1B enzyme, a primary pathway for alcohol metabolism. Individuals carrying the rare variant of rs1229984 often experience a flush response to alcohol, leading them to consume less alcohol and possess lower blood ethanol levels (Holmes *et al.*, 2014).

Using individual-level data from the UKB and publicly available GWAS results for cataracts and alcohol consumption; we aim to understand the causal relationship between alcohol consumption and cataract. This investigation will first explore the association between alcohol consumption and cataract risk through observational analyses, contributing to and building upon the existing literature. These analyses will be followed by genetic investigations, utilising the rs1229984 variant as a proxy for alcohol consumption. Furthermore, GWAS data will be leveraged to conduct two-sample and multivariable MR analyses to assess causality.

5.2 Methods

5.2.1 UKB Exposure data (weekly alcohol intake)

Alcohol intake data was obtained from the UKB through touchscreen questionnaires for lifestyle and environmental factors. An example Assessment Centre Environment (ACE) touchscreen question included "In an average WEEK, how many glasses of RED wine would you drink? (There are six glasses in an average bottle)". Therefore, data was gathered for the consumption of specific beverage types, such as average weekly red wine intake (Data-Field

1568), average weekly beer plus cider intake (Data-Field 1588), average weekly champagne plus white wine intake (Data-Field 1578), average weekly fortified wine intake (Data-Field 1608) and average weekly spirits intake (Data-Field 1598). The commonly used conversion scale provided by the NHS (https://www.lanarkshirelinks.org.uk/wp-content/uploads/2015/12/HWL-ALCOHOL-KNOW-YOUR-LIMITS-SHEET.pdf) was used to calculate the units of alcohol consumed per individual based on their weekly average intake of each alcohol type. Participants who answered, "Do not know", "Prefer not to answer" or had data missing for any of the above data fields, were removed from the analysis to ensure accuracy across alcohol intake data. Exclusions were also made for those who reduced alcohol intake due to “Illness or ill health” or “Doctor's advice”, to ensure accurate exposure data and reduce the likelihood of confounding bias. Total units of weekly alcohol intake were calculated as the sum of weekly intake for each beverage. As the previously mentioned data fields did not explicitly identify all non-drinkers, therefore alcohol intake frequency (Data Field 1558) was used. In Data-Field 1558 participants were asked in an ACE touchscreen question "About how often do you drink alcohol?", those who answered “Never” were assigned with a weekly alcohol intake of 0. To control for extreme outliers, individuals were removed from the phenotype with weekly alcohol intake beyond a threshold of six times the interquartile range above the mean.

Alcohol intake frequency was used to form an ordinal phenotype, ranking individuals based on increasing their intake frequency, assigning “Never” as 0, “Special occasions only” as 1, “One to three times a month” as 2, “Once or twice a week” as 3, “Three or four times a week” as 4, and “Daily or almost daily” as 5.

5.2.2 Outcome data (incident cataract)

Cataract cases were obtained from UKB data using a combination of diagnosis and operational codes. Diagnostic codes included the ICD-10 classifications for senile cataract (H25), other cataract (H26), and cataract and other disorders of the lens in diseases classified elsewhere (H28) (Data-Field 41270). Operational classifications were determined using OPCS-4 codes for surgical cataract cases, specifically for the insertion of prosthetic replacement for lens NEC (C75.1) and phacoemulsification of the lens (C71.2) (Data-Field 41272).

Phenotypic data detailing the date each relevant diagnostic and operational code was recorded (Data-Fields 41280, 41281, and 41282) were compared against the date participants first attended the UKB assessment centre (Data-Field 53) to identify incident cataract cases. Any prevalent cases of diagnosis or operation, which occurred prior to the initial baseline UKB assessment centre visit, were excluded from the analysis.

5.2.3 Genome-wide association study – exposure and outcome

Alcohol consumption GWAS

Additional genetic variants associated with alcohol were obtained from a European GWAS meta-analysis on the number of alcoholic drinks consumed per week. Saunders *et al.* identified genetic variants associated with alcohol intake using data from 60 cohorts, comprising a total of 2,965,643 individuals. Further information regarding the GWA analysis can be found elsewhere (Saunders *et al.*, 2022).

UKB - European cataract GWAS

European UKB cataract GWAS results were obtained from the previous the previous REGENIE GWAS as discussed in Section 3.2.6. In total, after QC filters outlines in Sections 2.5.1, 2.5.2 and 2.6.5, 45,449 cases and 353,371 controls were identified.

UKB and GERA - Multi-ethnic cataract GWAS

An additional cataract GWAS conducted by Choquet *et al.*, was used to supplement the analyses and ensure robustness of our results when using the UKB European cataract GWAS. This is a multi-ethnic GWAS conducted over the GERA and UKB cohorts, which included a total of five ethnic cohorts. The phenotype included diagnostic codes and self-reported cataract operations. A combination of both cohorts provided 67,844 cases and 517,399 controls. Additional information is available (Choquet *et al.*, 2021).

5.2.4 Statistical analysis

Analysis was conducted using the R statistical software v4.0.5 (R Core Team, 2021), the “MendelianRandomization” (Yavorska and Burgess, 2017) and the “TwoSampleMR” packages (Hemani *et al.*, 2018). The graphical representation of the results was completed using ”ggplot2” (Villanueva and Chen, 2019).

Individual-level genotype data for rs1229984 was extracted using PLINK v2.0 (https://www.cog-genomics.org/plink/2.0/) within 487,409 individuals. The “--extract” command, alongside “--export A”, was used to generate a .raw file with additive coding of genotype values for rs1229984. Due to previously established evidence of dominance for rs1229984, this decision is supported by previous research showing that carriers of the minor T allele of rs1229984 consume significantly less alcohol on average compared to non-carriers, with a strong association observed between the variant and alcohol intake in a dominant model (Ingold, Amin and Drenos, 2019).

5.2.5 Characteristics of UKB incident cataract

To assess the relationship between baseline characteristics and incident cataract, we tested the presence of correlation with several previously suggested risk factors using t-tests for continuous variables and chi-squared tests for categorical variables. Any baseline characteristics significantly associated with cataract risk were identified using a *p*-value threshold of 0.05.

5.2.6 Observational analysis

Observational analyses were conducted to examine the relationship between incident cataract and alcohol consumption variables, including total weekly alcohol intake, weekly red wine intake and the following strata: individuals with total weekly alcohol >14 units and between 1-14 units, based on recommended thresholds. Alcohol consumption was separated based on recommended levels allowing for the exploration of a potential non-linear relationship.

These analyses were adjusted for relevant covariates, sex, age, agexsex, age squared, smoking status and BMI using the logistic regression model.

Additional analysis was performed to assess the effect of alcohol intake frequency on cataract risk, incorporating total weekly alcohol intake as a covariate. This allowed us to isolate the specific impact of drinking frequency on cataract risk by controlling for the total alcohol consumed.

5.2.7 Genetic analysis

To understand the variability between our variant and alcohol intake, we regressed weekly alcohol intake on the rs1229984 variant using the common variant as dominant. We also tested the correlation of the variant with previously suggested cataract risk factors as listed in Table 5.1. Each potential risk factor was regressed against the rs1229984, adjusting for age, sex, the 10 genetic principal components (PCs), and other potential confounders (e.g. BMI and diabetes, smoking and employment status). Any variables that were also found to be significantly associated with the rs1229984 were included as additional covariates in the MR analysis.

An instrumental variable regression analysis was conducted using incident cataract and rs1229984 (the instrument variable) with the inclusion of our covariates sex, age, agexsex, age squared, smoking status, BMI and 10 genetic principal components, using the logistic regression model. We used the incident cataract phenotype for consistency with the observational analyses. We expanded our analysis testing other alcohol consumption variables, including alcohol frequency, and the following strata based on recommended

thresholds: high alcohol consumption (> 14 units) and moderate alcohol consumption (1-14 units). Each analysis was conducted using the two-step instrumental variable regression approach, where the alcohol variable was regressed on the dominant rs1229984 genotype to calculate fitted values. The incident cataract phenotype was then regressed on these fitted values. Both steps of the regression included the same covariates as mentioned above. Statistical significance was accepted at a *p*-value threshold < 0.05.

In addition, 2SMR analyses were conducted using GWA results of alcohol consumption and the two previously described cataract GWA results (Choquet *et al.*, 2021; Saunders *et al.*, 2022; Hashimi *et al.*, 2024). Statistical significance was accepted at a *p*-value < 0.05. Five MR methods were used for this analysis, including MR Egger, weighted median, IVW, simple mode and weighted mode, with additional tests for the presence of pleiotropy and heterogeneity also conducted (Hemani *et al.*, 2018; Slob and Burgess, 2020; Yuan *et al.*, 2020). While the primary interpretation of the results is based on the IVW method, the additional MR analyses were conducted as sensitivity tests to assess the robustness of the findings. These tests help detect pleiotropy such as MR Egger and evaluate the results in the presence of invalid or weak genetic instruments (Slob and Burgess, 2020). To further investigate the impact of other risks potentially involved in the causal relationship between alcohol consumption and cataract risk, such as BMI, we conducted a multivariable Mendelian randomisation (MVMR) analysis. This analysis was performed using the MendelianRandomization package, incorporating the same alcohol consumption data used in the 2SMR (Saunders *et al.*, 2022) and publicly available BMI GWAS data from the GIANT consortium, which includes 322,154 individuals of European descent across multiple cohorts (Locke *et al.*, 2015) with the UKB cataract GWAS (Hashimi *et al.*, 2024). To ensure consistency between datasets, we harmonised the exposure and outcome data by ensuring alignment across effect alleles and then performed clumping to obtain independent SNPs.

5.3 Results

5.3.1 Case control description

Table 5.1: UK Biobank incident cataract case-control baseline characteristics, and statistical significance against incident cataract cases. Key variables include age, sex, BMI, smoking status, alcohol consumption status, diabetes status, and employment status. Standard deviations are provided in round brackets, where appropriate.

	Females			**Males**		
	All	**Cases**	**Controls**	**All**	**Cases**	**Controls**
N	144,843	18,181	126,662	126,258	13,050	113,208
Mean Age (Years)	56.55 (7.89)	62.52 (5.47)	55.69 (7.81)	56.99 (8.10)	62.58 (5.66)	56.34 (8.09)
Mean BMI (kg/m2)	27.00 (5.15)	27.61 (5.20)	26.91 (5.13)	27.80 (4.21)	28.23 (4.34)	27.75 (4.19)
Take Vitamin D supplements (%)	27.54	27.76	27.51	19.66	19.50	19.68
Ever Smoked (%)	40.35	44.42	39.76	50.61	60.00	49.53
Have Diabetes (%)	3.33	6.92	2.81	6.31	12.89	5.55
Employed/Self-Employed (%)	54.64	30.01	58.17	60.44	36.90	63.15

Post QC and removal of prevalent cataract cases, 144,843 Females and 126,258 Males remained in the sample, providing a total of 271,101 individuals (31,231 cases of incident cataract and 239,870 controls). Table 5.1 summarises the incident cataract sample baseline characteristics, including sex, age, BMI, and ever-consumed alcohol, ever-smoker, diabetes, and employment status at recruitment. All baseline characteristics were statistically significant for incident cataract ($p < 0.05$).

As shown in Table 5.1, ever-consumed alcohol status at baseline was observed to be 92.2% and 95.2% for Females and Males, respectively.

5.3.2 Observational analysis

Total weekly alcohol intake

The analysis found no statistically significant relationship between incident cataract and total weekly alcohol intake (OR = 1.001, ln(OR) SE = 4.075×10^{-4}, $p = 0.125$).

Total weekly alcohol intake above the recommended limit (> 14 units)

When investigating incident cataract against total alcohol intake amongst individuals above the recommended limit of 14 units per week, we identify a statistically significant relationship (OR = 1.003, ln(OR) SE = 5.334×10^{-4}, $p = 8.79 \times 10^{-10}$). This suggests that every unit of alcohol

consumed above 14 units a week is correlated with a rise in an individual's incident cataract risk by 0.3%.

Total weekly alcohol intake within the recommended limit (1-14 units)

However, no statistically significant relationship was observed between incident cataract risk and total weekly alcohol intake within the recommended limit of 1-14 units (OR = 0.998, ln(OR) SE = 3.739×10^{-3}, $p = 0.560$).

Weekly red wine intake

The analysis indicated a statistically significant relationship between weekly red wine intake and incident cataract, suggesting that a per unit increase in red wine consumption per week is correlated with a fall of 0.5% in an individual's odds of developing incident cataract (OR = 0.995, ln(OR) SE = 7.609×10^{-4}, $p = 2.08 \times 10^{-12}$).

Alcohol intake frequency

Alongside all other covariates and controlling for total weekly alcohol, the analysis indicates a statistically significant relationship between incident cataract risk and alcohol intake frequency (OR = 0.937, ln(OR) SE = 6.320×10^{-3}, $p = 4.41 \times 10^{-25}$). This suggests that increasing the frequency of alcohol intake, adjusting for units consumed, thus representing reduced heavy episodic drinking, decreases an individual's risk of developing incident cataract.

5.3.3 *ADH1B* variant rs1229984 analyses (dominant model)

As stated in Section 2.7.1, genetic variants used as instrumental variables must satisfy MR assumptions, including the Independence assumption which states that associations between SNPs and outcomes must be independent of confounding factors. Given the statistical significance of baseline characteristics shown in Table 5.1, BMI, diabetes, smoker and employment status were regressed against rs1229984 to identify potential confounders to be considered as additional covariates.

Table 5.2: Results for regression analysis against significant baseline characteristics and rs1229984.

rs1229984	**OR**	***p* -value**
Mean BMI (kg/m2)	0.72	2.9E-13
Have Diabetes (%)	0.93	0.23
Ever Smoked (%)	1.04	0.07
Employed/Self-Employed (%)	1.01	0.61

As shown in Table 5.2, controlling for sex, age, 10 genetic principal components and remaining variables in this analysis, we only observe a statistically significant association between

rs1229984 and BMI ($p < 0.05$). In contrast, we find no relationship between rs1229984 and diabetes, smoking or employment status at $p = 0.05$. As we already control for BMI and smoking status no additional covariates were added to our analysis.

rs1229984 was shown to be strongly associated with alcohol consumption when regressing total weekly alcohol intake on rs1229984. Using a dominant model, rs1229984 was strongly associated with alcohol intake (Beta = -4.247, F-statistic = 2508, $p = 6.897 \times 10^{-133}$,). The presence of rs1229984 explains an additional 0.2% of the variability of alcohol consumption, after considering sex, age, agexsex, age squared, smoking status, BMI and the first 10 principal components, suggesting the variant's appropriateness as an instrument.

However, at a *p*-value threshold of 0.05, no statistically significant evidence was found for an association between rs1229984 and incident cataract risk (OR = 1.008, ln(OR) SE = 0.009, $p = 0.380$). While the odds ratio suggests a slight increasing trend, the evidence does not support a statistically significant association between alcohol intake and cataract risk.

Furthermore, no statistically significant causal association was observed between incident cataract risk and rs1229984 and alcohol consumption for individuals who consumed above 14 units a week (OR = 1.020, ln(OR) SE = 0.016, $p = 0.193$), within 1-14 units a week (OR = 1.066, ln(OR) SE = 0.178, $p = 0.718$) or when analysing alcohol intake frequency (OR = 1.207, ln(OR) SE = 0.104, $p = 0.069$).

Table 5.3: All results from each observational and genetic analysis, detailing the cases and controls present for each analysis.

Analysis	Cases	Controls	OR	ln(OR) SE	*p*-value
Observational					
Incident cataract ~ total weekly alcohol intake	22,611	180,427	1.001	4.08E-04	0.125
Incident cataract ~ total weekly alcohol intake above the recommended limit (>14 units)	11,969	103,760	1.003	0.001	**8.79E-10**
Incident cataract ~ total weekly alcohol intake within the recommended limit (1-14 units)	7,905	61,749	0.998	0.004	0.560
Incident cataract ~ weekly red wine intake	22,812	181,603	0.995	0.001	**2.08E-12**
Incident cataract ~ alcohol intake frequency	22,550	179,991	0.937	0.006	**4.41E-25**
Genetic					
Incident cataract ~ rs1229984 dominant model	22,611	180,427	1.008	0.009	0.380
Incident cataract ~ rs1229984 dominant model for those above the recommended limit (>14 units)	11,969	103,760	1.020	0.016	0.193
Incident cataract ~ rs1229984 dominant model for those within the recommended limit (1-14 units)	7,905	61,749	1.066	0.178	0.718
Incident cataract ~ rs1229984 dominant model for alcohol intake frequency	30,929	238,250	1.207	0.104	0.069

All cases and controls for each observational and genetic analysis are detailed in Table 5.3.

5.3.4 Two-sample MR analysis

After harmonising the exposure and outcome data, 10 independent SNPs were present for alcohol consumption and UKB cataract GWAS. Similarly, 10 SNPs were present after harmonising when using the multi-ethnic cataract GWAS as the outcome.

As shown in Figure 5.1, no evidence was found to suggest a causal relationship between alcohol consumption and cataracts in the UKB (IVW: OR = 1.403, 95% CI: 0.976-2.016, p = 0.068) or in the multi-ethnic cohort (IVW: OR = 1.110, 95% CI: 0.787-1.565, p = 0.552). Additional MR methods were also applied to examine the relationship between alcohol consumption and cataract in both the UKB and the multi-ethnic cohort to assess the robustness of the IVW results, but no evidence of a causal effect was found (See Supplementary Table 5.1 and 5.2 of the Appendix).

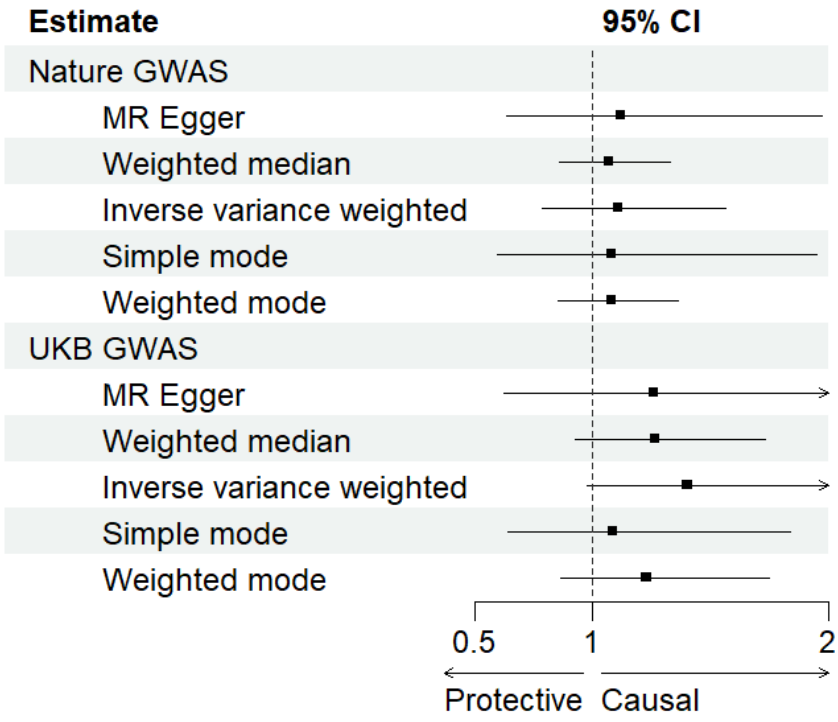


Figure 5.1: Two-sample MR results forest plot for alcohol consumption against incident cataract.

Pleiotropy was tested using the intercept of the MR Egger model for all analyses. No evidence of pleiotropy bias was found in any analysis of the MR Egger intercept at a p-value threshold of 0.05.

Full results of all MR methods and pleiotropy tests can be found in Supplementary Tables 5.1 and 5.2 of the Appendix for UKB cataract and multi-ethnic cataract cohorts, respectively.

In addition, we tested for heterogeneity across both analyses using the Cochran Q test to assess the heterogeneity between the genetic variants from the alcohol and cataract GWAS. Evidence for heterogeneity was found across both UKB (Q-statistic = 22.9, p = 0.006) and multi-ethnic (Q-statistic = 33.2, $p = 1.209 \times 10^{-4}$) MR results.

Across both cohorts, the SNPs rs72794102 and rs1260326, associated with the genes *RP11-89K21.1* and *GCKR*, respectively, depict a protective association with incident cataract risk. In contrast, the SNPs rs34839, rs210593, and rs1583973, associated with the genes *CLN3*, *AUTS2*, and *ADH1C* (nearest upstream gene), suggest a causal relationship with increased cataract risk in the MR analyses.

To address the heterogeneity identified, we performed an additional sensitivity analysis by excluding all five SNPs showing heterogenous properties in the analysis (rs72794102, rs1260326, rs34839, rs210593 and rs1583973). The results remained consistent with the main analysis with cataracts in the UKB (IVW: OR = 1.194, 95% CI: 0.91-1.568, p = 0.200) and multi-ethnic cohort (IVW: OR = 1.041, 95% CI: 0.8.4-1.277, p = 0.700), providing further confidence in the conclusion that there is no evidence for a causal relationship between alcohol consumption and cataract risk.

5.3.5 MVMR results

We further assessed the potential causal relationship between alcohol consumption and cataract risk by conducting an MVMR analysis, including BMI as an additional exposure variable. This approach aimed to clarify the direct effect of alcohol consumption on cataract risk while accounting for the indirect effects of BMI, as highlighted in our previous genetic analysis (see Table 5.2).

The MVMR analysis included 65 independent SNPs that were identified within the alcohol, BMI, and cataract GWAS datasets. These SNPs were associated with either alcohol consumption or BMI and were harmonised with the cataract GWAS to ensure effect allele consistency across datasets. Overall, the MVMR found no statistically significant evidence of an association between alcohol consumption and cataract when accounting for BMI (OR = 1.064, ln(OR) SE = 0.238, p = 0.795).

For our previously generated UKB cataract GWAS, Manhattan and QQ plots are found in Supplementary Figure 3.2 and 3.3, respectively.

5.4 Discussion

Our study aimed to investigate the potential causal relationship between alcohol consumption and incident cataract by leveraging genetic data. We used publicly available GWA results for alcohol consumption and cataract (Choquet *et al.*, 2021; Saunders *et al.*, 2022), as well as previously generated UKB cataract associations (Hashimi *et al.*, 2024). We first conducted observational analyses to compare findings with current literature and explore observed associations between incident cataract and various alcohol consumption measures. In this analysis we explored weekly total alcohol, red wine consumption and alcohol intake frequency. We also included variables for different alcohol ranges (1-14 units and > 14 units). Using genetic data, we investigated causality through the rs1229984 variant. We examined potential non-linear effects of alcohol consumption, as suggested in previous literature, as well as alcohol intake frequency (Wang and Zhang, 2014; S. Y. Chua *et al.*, 2021). We supplemented our findings with a 2SMR analysis to explore the causal relationship between alcohol consumption and cataract risk, further expanded by an MVMR analysis accounting for the role of BMI.

All baseline characteristics were found to be statistically significant ($p < 0.05$) when tested against incident cataract, consistent with literature that have also tested UKB baseline characteristics against cataract endpoints (S. Y. Chua *et al.*, 2021). Our observational results broadly were also consistent with other previously conducted analyses; for example, we also observed a statistically significant association between weekly red wine intake and decreased risk of incident cataracts (Gong *et al.*, 2015; S. Y. Chua *et al.*, 2021). Furthermore, we observed results for a statistically significant association between high alcohol consumption and alcohol frequency with incident cataract risk. Other studies have observed similar patterns, with high alcohol consumption linked to an increased risk of cataract surgery and reductions in alcohol intake slowing cataract development (Lindblad *et al.*, 2007; Gong *et al.*, 2015; Fukai *et al.*, 2022). For example, Gong *et al.*, reported that heavy alcohol consumption, defined as more than two standard drinks (20 g of alcohol) per day, was associated with an increased risk of age-related cataract (pooled relative risk, 1.26; 95% confidence interval, 1.06–1.50) highlighting the dose-dependent relationship between alcohol intake and cataract risk (Gong *et al.*, 2015).

However, similar to current literature our findings were mixed. Some observational analyses did not show a statistically significant association, specifically for moderate alcohol consumption and total alcohol intake which both present *p*-value > 0.05, consistent with Kanthan *et al.* (2010) and Wang and Zhang (2014), who respectively found no association between alcohol consumption or moderate alcohol consumption and cataract risk. These

results suggest that the effects of alcohol consumption on cataract risk may be dose-dependent and potentially non-linear, with associations only detectable at higher levels of alcohol intake. However, although our observational analysis includes large sample sizes, the analysis may still lack sufficient statistical power to detect a weak correlation.

However, our genetic analyses, using the rs1229984 variant and the two-sample MR, found no statistically significant evidence for a causal relationship between alcohol consumption and cataract risk. We used the rs1229984 variant, which is associated with lower alcohol consumption, to assess the causal relationship between alcohol intake and cataract risk. We investigated alcohol consumption within and above recommended limits, as well as alcohol frequency, to evaluate their potential causal effects on cataract risk. However, no statistically significant associations were observed between any level of alcohol consumption frequency and cataract. The instrumental variable regression between alcohol intake frequency and cataract risk, adjusting for covariates, yielded an OR of 1.207 (ln(OR) SE = 0.104, $p = 0.069$), indicating a positive association but did not reach statistical significance. While the results suggest a potential causal association between more frequent alcohol consumption and increased cataract risk, the lack of statistical significance ($p > 0.05$) implies this finding could reflect limited statistical power to detect the effect. Despite the strong association between rs1229984 and alcohol consumption, the variant did not show a significant causal effect on cataract risk, suggesting no evidence of a direct genetic influence of alcohol intake on cataract development.

Similarly, the 2SMR analysis notably produced a large OR of 1.4 but failed to reach statistical significance $p = 0.068$. However, after removing outliers to address heterogeneity, the association was lower and still did not reach statistical significance. This suggests that previously observed associations in observational studies may be due to unobserved confounding. Therefore, it is expected that different lifestyle behaviours or other socioeconomic factors associated with alcohol consumption are responsible for the increased risk of cataract with high alcohol consumption and similarly for the lower risk of cataract with red wine consumption, as seen in the observational analysis (OR = 0.995, $p = 2.08 \times 10^{-12}$). As observed in our analysis BMI has a relationship with both alcohol consumption and cataract, therefore, to assess the direct effect of alcohol consumption on cataract we performed the MVMR analysis. After controlling for BMI, the MVMR did not provide evidence of a statistically significant association between genetically predicted alcohol consumption and cataract risk (OR = 1.064, $p = 0.795$). These results suggest that any relationship between alcohol and cataract risk observed in prior analyses may be confounded by BMI. This aligns with the findings of the 2SMR analysis, where the observed effect could, in part, be explained by changes in BMI by alcohol consumption. These results highlight the potential confounding role

of BMI in this relationship. However, further analysis is required to clarify whether BMI or potential associated nutritional deficiencies act as a confounder or a mediator in the association between alcohol consumption and cataract risk (Falkowska *et al.*, 2023).

It has been previously reported that the relationship between alcohol and health outcomes resembles a “J-shaped curve” (Plunk *et al.*, 2014). This is a commonly observed effect of alcohol on cardiovascular disease and has also been previously reported for cataract outcomes, that low alcohol intake reduces the risk of cataract while consuming high levels of alcohol greatly increases associated risks (Wang and Zhang, 2014; Piano, 2017; Tsai, Gao and Wen, 2023). We observed a statistically significant observational association between consuming above 14 units of alcohol and increased cataract risk, but no evidence between moderate alcohol consumption (1 – 14 units) and cataract risk, partly confirming this “J-shaped” relationship.

Our study is not without limitations, healthy volunteer selection bias within the UKB sample may be an additional source of uncertainty, leading to more healthy volunteers within our sample than what is observed in the UK population (Fry *et al.*, 2017). Chu *et al.,* noted that observational results for alcohol consumption are susceptible to bias arising from methodological approaches and data limitations, stating the requirement of caution when interpreting such results (Chu *et al.*, 2020). In addition, it has been previously suggested that alcohol consumption may be underreported within sample collections. Any systemic underreporting of alcohol consumption with the UKB may lead to an overestimation in the association of moderate alcohol consumption and cataract risk, in traditional observational studies, this concern is mitigated when using genetic instruments in MR analyses (Vance, Caverly and Hayward, 2020). However, while no statistically significant association was found between moderate alcohol consumption and cataract risk in this study, underreporting and participation bias could still contribute to bias, as genetic correlations in UK Biobank data have been shown to be influenced by participation bias (Schoeler *et al.*, 2023). Thus, while the overall risk of bias is reduced, it cannot be entirely ruled out.

As observed across other studies and this paper’s observational results, red wine consumption displayed a negative association with incident cataract risk. While we did not observe statistically significant protective characteristics within our genetic analysis between the different alcohol variables, a genetic instrument specific to red wine consumption would allow us to perform an MR analysis to explore this relationship further (S. Y. Chua *et al.*, 2021). Conducting a genetic analysis for red wine would require additional specific genetic instruments, which are currently not available. This remains an important area for future research as it could help to differentiate the effects of different types of alcohol on cataract

risk. Future analyses, MVMR, using genetic instruments for specific alcohol types, could provide further insights by estimating the independent effects of each alcohol type. In addition, further investigation on the effect of frequency adjusting for alcohol consumption is required to understand the causal impact of regular alcohol intake, controlling for levels, on incident cataract risk.

In conclusion, while our observational analyses suggest a potential association between alcohol consumption and cataract risk, our genetic analyses using rs1229984 variant and publicly available GWAS results do not support a causal relationship. This highlights the limitations of observational studies, particularly with respect to confounding factors and reverse causation. Further analysis could investigate whether specific drinking patterns or types of alcohol, such as red wine, effect cataract risk, potentially providing deeper insights into the protective effects observed in these studies.

6 Exploring the genetic overlap between cataract subtypes and systemic risk factors: Genetic correlation and co-localisation analyses

6.1 Introduction

Cataract can be broken into subtypes based on the morphological development of the cataract on the lens (nuclear, cortical and posterior subcapsular) and further classified based on the aetiological subtypes (such as age-related, congenital, diabetic related, medication-induced, or trauma-induced) (Bixler, 2019; Shiels and Hejtmancik, 2019).

Age-related cataract is the most common form of cataract (Yonova-Doing *et al.*, 2016; Hashemi *et al.*, 2020). Age-related cataract is an aetiological subtype characterised by its onset in later life and can present with different morphological patterns, including nuclear, cortical, and posterior subcapsular cataracts. The development of age-related cataract is influenced by a combination of genetic and environmental risk factors, including smoking, obesity, and dietary patterns (Yonova-Doing *et al.*, 2020). While age-related cataract, particularly the nuclear type, is consistently reported as the most prevalent aetiological subtype of cataract, studies have highlighted challenges in accurately classifying this subtype compared to others. This has contributed to a limited understanding of its underlying genetic factors, highlighting the need for further work in this area (Yonova-Doing *et al.*, 2020).

However, a recent study has suggested an increasing prevalence in early-onset cataract cases (defined as occurring before the age of 60) (Sarkar *et al.*, 2023). Therefore, while age-related cataract is currently the most prevalent form of cataract, it is important to investigate prevention and non-surgical treatment options that capture multiple or all cataract subtypes.

Diabetic cataract is a major ocular complication that can occur in individuals with diabetes, who are reported to have a two- to five-fold increased risk of developing cataracts compared to those without diabetes (Chang *et al.*, 2016). Several mechanisms have been proposed for the pathogenesis of diabetic cataract. One key pathway is the polyol pathway, in which the enzyme aldose reductase catalyses the conversion of glucose into sorbitol. In individuals with diabetes, sorbitol accumulates more rapidly than in non-diabetic individuals, as its production exceeds its conversion to fructose. The accumulation of sorbitol inside the lens creates a hyperosmotic effect, causing an influx of fluid that leads to swelling of the lens fibres. This swelling disrupts their function and ultimately results in fibre degeneration. It has been proposed that patients with type 1 diabetes are thought to experience swelling of cortical lens

fibres due to this osmotic imbalance, increasing the risk of cataract development (Kiziltoprak *et al.*, 2019).

Childhood cataract, including congenital cataract, is much rarer. Congenital cataract is diagnosed within the first two months of life, while cases identified beyond this period are referred to as developmental cataracts (Katre *et al.*, 2022). The prevalence of congenital and childhood cataracts, continues to grow, with an estimate of 200,000 children worldwide blinded by congenital or childhood cataract (Sheeladevi *et al.*, 2016). In the UK, congenital cataracts affects approximately 2.5–3.5% of every 10,000 children, often occurring within the first year of life (Rahi, Dezateux, and British Congenital Cataract Interest Group, 2001). Congenital cataracts may result from infections during pregnancy, such as syphilis, rubella, and toxoplasmosis, but are more commonly caused by inherited genetic factors (Yi *et al.*, 2011).

While genetic inheritance is a common cause across different forms of cataracts, particularly congenital cataract, environmental risk factors also play a significant role in their prevalence and accelerated development of cataract. Established risk factors include: type 2 diabetes (T2D), high body mass index (BMI), and asthma, with numerous observational studies and genetic analyses suggesting a causal relationship between cataracts and both T2D and BMI, highlighting the metabolic and systemic impacts of these conditions on eye health (Li, Wan and Zhao, 2014; Harahap and Rania, 2019; Yuan, Wolk and Larsson, 2022; Savran and Ulrik, 2023). The association between asthma and cataract is supported by studies investigating the effects of steroid use, particularly the exposure to inhaled corticosteroids. Observational studies have shown that daily use of inhaled corticosteroids can increase the risk of developing cataract, underlining the need for careful clinical management and guidance of steroid use in asthma patients (Savran and Ulrik, 2023).

Genetic analysis has become a widely used approach for identifying the underlying mechanisms of complex diseases. For example, genetic data can be used to identify shared genetic variants between traits through genetic correlation analysis and to determine if this overlap is caused by the same genetic variant using co-localisation analysis. These methods have proven effective in uncovering shared biological pathways between conditions, such as the genetic overlap observed across different anxiety disorders (Friligkou *et al.*, 2024). Understanding the shared biological pathways between cataract subtypes, as well as between cataract and other risk factors, can provide valuable insights for therapeutic targets and alternative treatment options. Given that poor glucose control is a well-established causal factor in cataract formation, improving glycaemic regulation through existing anti-diabetic medications already lowers cataract risk. However, further investigation of the shared biological pathways could identify novel treatment. For example, if cataract and a risk factor

such as T2D share biological pathways, depending on the nature of the pathway such as common pathways and causal pathways, this could reveal overlapping drug targets, potentially offering alternative treatment for cataract. Furthermore, understanding the shared mechanisms across cataract aetiological subtypes is essential, as it will help the understanding of the basic underlying mechanisms of different cataract subtypes, broadening the identification of therapeutic targets which address multiple cataracts subtypes, improving patient outcomes and advancing preventive strategies.

While surgery is the only treatment for cataract, the increasing prevalence of the condition and demand for surgeries highlights the need for alternative approaches (Mailu *et al.*, 2020). Understanding the shared biological pathways between cataract subtypes and related conditions could reveal novel therapeutic options and provide insights into prevention, potentially easing the burden on healthcare systems (Berkowitz *et al.*, 2024). In this study, BMI, T2D and asthma have been chosen as risk factors to be investigated due to their apparent association to cataract and predefined treatments that could be identified as alternative treatments, or medications causal, to cataract if shared genetic mechanisms are discovered (Liang and Chao, 2023; Rothberg *et al.*, 2023).

In this investigation, we will explore the underlying and shared genetic mechanisms across cataract subtypes and cataract and associated risk factors to identify underlying genetic pathways through genetic correlation analysis. Additionally, we will use co-localisation analysis to determine whether these pathways are influenced by the same genetic variants, providing deeper insight into shared biological pathways.

6.2 Materials and methods

Information regarding the study population, genotyping and more detail on the REGENIE process can be found in Sections 2.2, 2.3 and 2.6, respectively.

6.2.1 Phenotypes in the UKB

Diabetic cataract

Diabetic cataract was defined using hospital inpatient diagnostic information (Data-Field 41270), operation codes (Data-Field 41272), and their respective diagnosis and operation dates (Data-Fields 41280 and 41282). Cases were identified as individuals diagnosed with cataracts (ICD-10: H25 Senile cataract, H26 Other cataract, H28 Cataract and other disorders of lens in diseases classified elsewhere, and OPCS4: C75.1 Insertion of prosthetic replacement for lens NEC and C71.2 Phacoemulsification of lens) at any point after an initial diagnosis of type 2 diabetes (ICD-10: E11 Non-insulin-dependent diabetes mellitus). Controls

consisted of individuals without cataracts. Following QC, 6,004 cases and 331,747 controls were identified.

Senile cataract

Senile cataract was defined using hospital inpatient diagnostic codes (Data-Field 41270 - ICD-10: H25 Senile cataract). To avoid overlapping samples with diabetic cataract, individuals who self-reported being diabetic (Data-Field 20002) or had a hospital record of a type 2 diabetes diagnosis (Data-Fields 41202 and 41204 - ICD-10: E11 Non-insulin-dependent diabetes mellitus) were excluded from both cases and controls. Additionally, any case of senile cataract also classified as congenital cataract (see below) were excluded. After these exclusions, 22,665 cases and 393,285 controls were identified.

Congenital cataract

Congenital cataract cases were also defined using hospital inpatient diagnostic codes (Data-Field 41271 – ICD-10: Q12.0 Congenital cataract, Data-Field 41270 – ICD-9: 7433 Congenital cataract and lens anomalies). Controls represented individuals without cataract. GWAS for congenital cataract was conducted using PLINK v2.0, consistent with previous studies examining rare phenotypes in UKB.

Genotyping quality control and imputation procedures in UKB have been described in Section 2.5. Following these standard QC steps, additional exclusions specific to the PLINK analysis were applied. Unlike REGENIE, PLINK does not include related individuals, which resulted in a lower number of controls for this analysis so participants were filtered for relatedness by removing one individual at random from related pairs using the genetic kinship data (Bycroft *et al.*, 2018).

After applying PLINK-based filtering, 140 congenital cataract cases and 290,739 controls were identified. Associations between genetic variants that passed QC and congenital cataract were assessed under an additive genetic model, adjusting for sex, age, agexsex and age squared.

Overall Cataract

A previously generated cataract definition, based on diagnostic and operation codes, was used to compare against risk factors. While the overall cataract definition may overlap with other classifications, such as age-related cataract, it includes diagnostic and operation codes regardless of subtype classification, ensuring that cases that might have otherwise been missed due to classification are captured. This definition is detailed in Section 3.2.2. For clarity this will be referred to as overall cataract.

6.2.2 Publicly available GWA results

Summary statistics from three peer-reviewed publicly available GWA studies were used within the genetic correlation and co-localisation analysis as risk factors against cataract, these studies covered BMI, T2D and asthma. SNPs from following GWA studies were included in the genetic correlation analysis to assess their shared genetic variants with cataract. These datasets were chosen because they align with the objectives of this study and allow for comparability within a European population within large sample sizes and robust QC.

Body mass index

The GWAS for BMI was conducted using 120,286 individuals of British ancestry from the UKB, using imputed genotype data from a combined 1000 Genomes/UK10K reference panel (Wood *et al.*, 2016). Variants were quality controlled for imputation quality (< 0.9), Hardy–Weinberg equilibrium ($p < 1\times10^{-6}$) and minor allele frequency (< 0.5%). Summary statistics are available online on the GWAS catalogue (https://www.ebi.ac.uk/gwas/studies/GCST006802).

Type 2 diabetes

Xue *et al.* (2018) GWAS for T2D was also used in this study. This GWAS included 659,316 individuals, predominately of a European ancestry (655,666), supplemented by a small Pakistani cohort (3,650) (Xue *et al.*, 2018). Despite being a mixed cohort, there is limited evidence of genetic heterogeneity between those of European and Pakistani ancestry for T2D (Morris *et al.*, 2012). The summary level results provided combined three GWAS data sets of European ancestry: DIAbetes Genetics Replication And Meta-analysis (DIAGRAM), Genetic Epidemiology Research on Adult Health and Aging (GERA), and UKB. Genotyped data was imputed using the HapMap2 and 1000 Genomes reference panels for DIAGRAM, and GERA and UKB, respectively. Data from GERA was quality controlled for SNPs and individuals with missing rate ≥ 0.02, Hardy-Weinberg equilibrium ($p < 1\times10^{-6}$) and minor allele count (≤ 1). UKB data was also controlled for imputation quality (< 0.3), Hardy-Weinberg equilibrium ($p < 1\times10^{-6}$), missing genotype rate (> 0.05) and minor allele count (< 5). Summary statistics are publicly available (https://www.ebi.ac.uk/gwas/studies/GCST006867).

Asthma

Summary statistics for Zhu *et al.* (2019) GWAS for Asthma were also used in the genetic correlation and co-localisation analysis. This GWAS included 394,283 individuals with European ancestry across the UKB (Zhu *et al.*, 2019). Imputation was performed using the Haplotype Reference Consortium (HRC) reference panel, with insertion–deletion mutations and variants with MAF < 1% removed as part of QC procedures. Summary statistics from this study are publicly available (https://www.ebi.ac.uk/gwas/studies/GCST008918).

6.2.3 Statistical analysis

Analysis was performed using the statistical software R v4.0.5 (R Core Team, 2021), unless otherwise stated. REGENIE v3.2.8 was used to produce GWA results for the UKB cataract phenotypes (senile, congenital and diabetic cataract) (Mbatchou *et al.*, 2021). PLINK v2.0 (https://www.cog-genomics.org/plink/2.0/) was used to clump GWAS results, identifying the lead SNPs across different loci and their associated groups of SNPs. Genetic correlations were ran using the LD Score Regression package (B. K. Bulik-Sullivan *et al.*, 2015) and co-localisation was run using the COLOC package v5.2.3 (Giambartolomei *et al.*, 2014).

6.2.4 Genetic correlation (LDSC) analysis

To estimate the genetic correlations the linkage disequilibrium score (LDSC) tool was used, scripts provided (munge_sumstat.py and ldsc.py can be downloaded at the following: https://github.com/bulik/ldsc?tab=readme-ov-file) (B. Bulik-Sullivan *et al.*, 2015). Genetic correlation scores were calculated for each variable against every other variable. The results of the genetic correlation analysis range from -1 to 1. A positive score, indicating a positive genetic correlation, meaning that genetic variants associated with one trait are more likely to also be associated with another trait in the same direction. A score between 0 and -1, a negative score, indicates a negative genetic correlation, suggesting that genetic variants associated with one trait are inversely associated with another trait. A score of 0 states no genetic correlation, indicating that there is no shared genetic influence between the two traits. The HapMap3 reference panel was used to provide LD scores for the analysis, focusing on common SNPs from individuals of European ancestry.

Pre-filters, implemented using the munge_sumstat.py script, were applied to both the generated and publicly available GWAS summary statistics. The filters applied to each GWAS dataset included the following criteria: imputation quality (INFO > 0.9), minor allele frequency (MAF > 0.01) based on the HapMap 3 panel, and adjustments for variation in sample size between SNPs. Additionally, indels and structural variants were excluded, as well as strand-ambiguous SNPs and SNPs with alleles that did not align with the 1000 Genomes Project. Further details on the pre-filters applied to the GWAS results can be found elsewhere (B. Bulik-Sullivan *et al.*, 2015).

When applying the ldsc.py script, SNPs are separated into blocks ensuring variants in close linkage disequilibrium are grouped. Across these blocks LD scores are calculated. Genetic covariance is estimated by regressing the product of z-scores (the standardised effect sizes of genetic variants on each trait) across each variable, for a given SNP, against its associated

LD score, multiplying the slope by the number of SNPs in the reference panel in the range of 5-50% MAF (B. K. Bulik-Sullivan *et al.*, 2015).

Finally, genetic correlation scores are produced through the normalisation of genetic covariance by SNP-based heritability. More detail regarding the LD score regression process can be found elsewhere (B. Bulik-Sullivan *et al.*, 2015).

To rigorously control for false positives, given the multiple comparisons made in the analysis, a Bonferroni correction was applied to the *p*-value threshold. Genetic correlations were considered statistically significant at a *p*-value < 0.05/49 (Bonferroni correction applied to account for the 49 genetic correlation pairs tested).

6.2.5 Co-localisation analysis (COLOC) analysis details

To follow up on the genetic correlation analysis, we performed genetic co-localisation for relevant pairs of traits with statistically significant positive genetic correlations, assuming no sample overlap between them.

To prepare for COLOC analysis, we clumped SNPs to identify genomic regions of interest which can be used between traits. Using PLINK, we performed clumping on senile, overall and diabetic cataract GWAS results, as at least one of these were used during the analysis. The clumping process applied a primary *p*-value threshold $p1 < 5\times10^{-8}$ to select lead SNPs and a secondary *p*-value threshold $p2 < 0.1$ for additional SNPs in the group. SNPs were grouped based on both an LD threshold ($r^2 = 0.001$) and the default PLINK clumping window of 250 kb, meaning that SNPs in weak LD with each other and located beyond this distance were excluded. The groups identified across GWA sets, according to these thresholds were used for the COLOC analysis.

By examining independent SNPs from one GWAS against another, we assessed whether the association signals for each trait aligned between two groups of SNPs. The co-localisation analysis was conducted using the COLOC package (Giambartolomei *et al.*, 2014). COLOC estimates five posterior probabilities (PP) for each SNP group, which indicates the probability of five different hypotheses: H0 (no association for either trait), H1 (association with trait 1 only), H2 (association with trait 2 only), H3 (distinct causal SNPs for each trait), and H4 (a single shared causal SNP for both traits). COLOC employs a Bayesian approach to calculate probabilities for the different hypotheses. The COLOC analysis helps to assess whether a shared causal variant is possible across two different traits.

A posterior probability for hypothesis 4 > 75% indicates a strong support of co-localisation within that given group (Giambartolomei *et al.*, 2014).

For relevant pairs, based on the results of the COLOC analysis, the SNP nexus tool, a functional annotation tool for genetic variants (https://www.snp-nexus.org/v4/), was used to identify relevant genes.

6.3 Genetic correlation results

	Senile cataract	Congenital cataract	Diabetic cataract	Overall cataract	Type 2 diabetes	Body Max Index	Asthma
Senile cataract		*-0.230*	***0.469***	***0.993***	*0.018*	*-0.128*	*0.092*
Congenital cataract	*0.110*		*-0.186*	*-0.111*	*0.071*	*-0.432*	*0.137*
Diabetic cataract	***1.20E-09***	*0.168*		***0.491***	***1.025***	*0.426*	***0.196***
Overall cataract	***0.000***	*0.300*	***3.61E-27***		***0.168***	*-0.045*	***0.142***
Type 2 diabetes	*0.707*	*0.521*	***3.17E-80***	***4.42E-06***		*0.095*	***0.253***
Body Max Index	*0.539*	*0.312*	*0.124*	*0.775*	*0.484*		*0.435*
Asthma	*0.192*	*0.323*	***0.019***	***0.010***	***4.45E-08***	*0.556*	

Figure 6.1: Genetic Correlation analysis using the LDSC tool between cataract subtypes and overall cataract with associated risk factors. On the top right is the genetic correlation score, green indicating a positive score and red indicating a negative score. The bottom right provides the p-value for each of the pairs. The graph is divided by grey cells were each of the pairs were tested against themselves and this provided a score of 1.0 (2-signficant figures). Bold values indicate p-value ≤ 0.05.

Genetic correlation results are shown in Figure 6.1, which presents a correlation matrix across the 49 tested genetic correlations using the LDSC package. Figure 6.1 highlights genetic correlations between cataract subtypes, marked by a yellow outline, and further shows correlations between overall cataract and related risk factors.

A statistically significant positive genetic correlation was found between senile cataract and diabetic cataract (r_g = 0.469, p = 1.2×10^{-9}), suggesting shared genetic factors that affect both traits. Similarly, overall cataract and T2D showed a statistically significant positive genetic correlation (r_g = 0.168, p = 4.4×10^{-6}), indicating that shared genetic variants impact both traits in the same direction.

While congenital cataract generally showed negative genetic correlations with other traits, such as senile cataract (r_g = -0.230, p = 0.110), diabetic cataract (r_g = -0.186, p = 0.168), and BMI (r_g = -0.432, p = 0.312), these results were not statistically significant. While a negative score, as observed in these results, suggest that shared genetic factors may impact these

traits in opposite directions, the lack of statistical significance observed suggests that these traits may actually be genetically independent or their genetic correlation is too small to identify with the data used.

Although asthma and overall cataract had a positive genetic correlation (r_g = 0.142, p = 0.010), this did not meet the Bonferroni-corrected significance threshold (p < 0.05/49). This was also observed between diabetic cataract which showed a positive genetic correlation with asthma (r_g = 0.196, p = 0.019). However, as these pairs reached a nominal level of significance at the conventional threshold (p < 0.05), they were included in the co-localisation analysis for further investigation.

Additional statistically significant positive genetic correlations were observed between other pairs of traits, such as cataract–senile cataract and diabetic cataract–T2D. However, these results more likely to be influenced by sample overlap and similarities in trait definitions across the investigated pairs.

6.4 Co-localisation results

Co-localisation analysis was conducted for the pairs: diabetic cataract–asthma, senile cataract–diabetic cataract, and cataract–asthma. These pairs were selected for co-localisation analysis based on the results of the genetic correlation analysis, due to the positive score and statistical significance observed.

The results of each analysis overall cataract – T2D, asthma – diabetic cataract, senile cataract – diabetic cataract and asthma – cataract can be found in Supplementary Tables 6.1, 6.2, 6.3, and 6.4 of the Appendix, respectively.

We identified several sites of strong co-localisation. In the analysis of overall cataract and T2D we identified five groups of SNPs had a Posterior Probability for Hypothesis (PPH) 4 ≥ 0.75, providing strong evidence that either the lead SNP or a linked SNP within these group is likely driving the shared genetic association across both traits. Section 6.5 outlines the respective genes linked to the lead SNPs per group, where information was available via SNP Nexus.

Two groups on chromosome 16 (SNPlist_2 and SNPlist_3) showed a PPH4 of 1, providing definitive evidence of shared genetic variants in these regions. Additionally, two groups on chromosome 6 (SNPlist_4 and SNPlist_6) had a PPH4 of 0.99, and one group on chromosome 9 (SNPlist_2) showed a PPH4 of 0.98. These results suggest the presence of shared genetic variants within these genomic regions, contributing to the observed genetic overlap between diabetes and the overall cataract definition.

These results are summarised in Supplementary Table 6.1.

As shown in Supplementary Table 6.2, for the asthma and diabetic cataract analysis, no groups of SNPs had a PPH4 ≥ 0.75. This suggests that, within the tested genomic regions, there is limited evidence for a shared causal variant influencing both traits. Instead, the highest scores were observed under PPH2, suggesting that these SNPs are more strongly associated with one trait, with limited evidence of shared genetic influence. Similar results were observed for senile and diabetic cataracts, where no groups had a PPH4 ≥ 0.75. Instead, the majority of groups showed a PPH1 ≥ 0.75, also indicating association with only one trait.

In the asthma and overall cataract analysis, a group of SNPs on chromosome 6 (SNPlist_4) showed a strong PPH4 of 0.96, indicating a strong likelihood of a shared causal variant driving the association between these traits. This result implies that shared genetic factors between asthma and cataract may play a significant role between the two traits.

6.5 SNP to Gene Analysis

The lead SNPs for the association between overall cataract and T2D are rs1364063 and rs6499270 on chromosome 16, rs9273529 and rs2857709 on chromosome 6, and rs10757274 on chromosome 9. For the asthma and cataract association on chromosome 6, the lead SNP is rs9273529.

Using the SNP Nexus tool, rs9273529 was identified as being linked to the *HLA-DQB1* gene. This gene, along with other human leukocyte antigen (*HLA*) genes, plays a crucial role in the immune system by recognising and differentiating between antigens. *HLA* genes are also associated with autoimmune and inflammatory diseases. (Crux and Elahi, 2017). On chromosome 16, rs6499270 is linked to *WWP2*. Finally, on chromosome 9, rs10757274 is linked to *CDKN2B-AS1*. SNP Nexus was unable to identify genes related to rs1364063 on chromosome 16 and rs2857709 on chromosome 6. These findings highlight potential genetic contributors to the shared biological mechanisms underlying cataract and T2D.

6.6 Discussion

In our investigation, we used publicly available GWAS data along with GWAS results generated from the UKB to explore potential shared genetic mechanisms across different cataract subtypes (diabetic cataract, senile cataract, and congenital cataract) and their associated risk factors (TD2, BMI, and asthma). Through LDSC genetic correlations and COLOC co-localisation analyses, we assessed potential genetic overlap between these traits. Our findings revealed several pairs with statistically significant positive genetic correlations, particularly between senile cataract and diabetic cataract (r_g = 0.469, p = 1.2×10^{-9}) and between overall cataract and T2D (r_g = 0.168, p = 4.4×10^{-6}). We also found weaker evidence of genetic correlation between asthma and overall cataract (r_g = 0.142, p = 0.010) and between

asthma and diabetic cataract (r_g = 0.196, p = 0.019). Our co-localisation analysis provided strong evidence of shared genetic variants between overall cataract and T2D, and between cataract and asthma at SNP rs9273529 (linked to *HLA-DQB1*), suggesting genetic overlap between these traits and shared underlying genetic mechanisms. Furthermore, between overall cataract and T2D, rs6499270 and rs10757274 were found to be linked to *WWP2* and *CDKN2B-AS1*, respectively.

T2D is a known risk factor for cataract and has been extensively investigated in both observational research and genetic analyses, where it is often used as a positive control to compare findings against the known causal association (Li, Wan and Zhao, 2014; Yuan, Wolk and Larsson, 2022). In our analysis, we observed a statistically significant positive genetic correlation between T2D and overall cataract (r_g = 0.168, p = 4.4×10^{-6}). While our overall cataract definition includes diabetic cataract cases (10.2% of overall cataract cases), as indicated by the reduced positive correlation, the genetic correlation may also reflect the broader genetic overlap encompassing non-diabetic cataract cases. These findings align with existing literature that has hypothesised shared genetic mechanisms underlying both T2D and cataract development. For instance, a population-based study in East Asian population also used genetic correlation to explore the relationship between cataract and T2D, finding strong statistical evidence for genetic overlap between the two traits (H. Zhang *et al.*, 2021). Our findings further support existing research on the underlying mechanisms connecting T2D and cataract, such as advanced glycation of lens proteins due to elevated glucose levels in individuals with T2D. Additional underlying biological pathways are supported by our investigation such as, the increased presence of free radicals in T2D and cataract patients reduces antioxidant activity in the lens, leading to greater oxidative stress and lens damage (Pollreisz and Schmidt-Erfurth, 2010). Statistically significant results were observed between senile and diabetic cataracts in the genetic correlation analysis (r_g = 0.469, p = 1.2×10^{-9}). However, the co-localisation findings indicate that genetic variants are predominantly associated with one trait rather than the other. This supports the hypothesis that diabetes may accelerate the aging process, contributing to the earlier development of cataracts (Mishra *et al.*, 2023).

Co-localisation was observed between asthma and cataract, with a shared lead SNP, rs9273529, and its associated SNP group showing a 96% posterior probability (PP) for hypothesis 4, suggesting that the two traits share a causal variant. While the relationship between asthma and cataract has been previously investigated (Li and Wang, 2022; Savran and Ulrik, 2023), understanding their shared genetic mechanisms remains to be incomplete. Our findings support some underlying genetic mechanisms previously reported, such as in the Blue Mountain Eye Study, where long-term use of inhaled and oral corticosteroids was linked

to an increased risk of cataract development, specifically the posterior subcapsular and nuclear subtypes (Wang *et al.*, 2009). This association is thought to arise from corticosteroid effects on lens receptors, ultimately leading to cataract formation. Both cataract and asthma have been linked to oxidative stress and inflammation, which may represent a shared biological pathway between the conditions, as suggested by our genetic correlation and co-localisation results (Michaeloudes *et al.*, 2022; Thompson *et al.*, 2022). Further investigation into the lead SNPs for each group was conducted using SNP Nexus to further understand the potential genetic mechanisms underlying the associations. For the shared lead SNPs between cataract and T2D, rs9273529 was identified in the *HLA-DQB1* region, rs6499270 was linked to the *WWP2* gene, and rs10757274 was associated with the *CDKN2B-AS1* gene. However, SNP Nexus was unable to identify genes related to rs1364063 on chromosome 16 and rs2857709 on chromosome 6.

Cyclin-dependent kinase inhibitor 2B antisense RNA 1 (*CDKN2B-AS1*) has previously been associated with conditions such as coronary heart disease, atherosclerosis, cancers as well as diabetes (Xiao *et al.*, 2021). *CDKN2B-AS1* is involved in mediating senescence, inflammation, and extracellular matrix (ECM) accumulation, processes that may play a significant role in the development of primary open-angle glaucoma (POAG). (Rathi *et al.*, 2020). Furthermore, *WWP2* has previously been identified as being associated with both cataract and T2D. *WWP2* plays a key role in various biological processes including immune response, apoptosis, and cell signalling, but its dysfunction can contribute to various conditions, particularly those influenced by oxidative stress (You *et al.*, 2023). As previously discussed, oxidative stress is a known contributor to cataract pathogenesis and also plays a critical role in the development and progression of T2D (Kaur *et al.*, 2012; Caturano *et al.*, 2023). Overall, findings from our genetic correlation and co-localisation analyses, highlight *CDKN2B-AS1* as a shared gene of interest in both cataract and T2D due to its potential role in these shared biological pathways. Further investigation is needed to determine whether these genes play a direct causal role in cataract formation and potential therapeutic targets.

The lead SNP identified in the asthma and cataract co-localisation analysis, rs9273529, is associated with the *HLA-DQB1* gene, which is commonly linked to allergic sensitisation, asthma, and immune function (Smit *et al.*, 2014). *HLA* genes including *HLA-DQB1*, have been previously reported to be associated with autoimmune diseases such as type 1 diabetes (Nguyen *et al.*, 2013). Additionally, *HLA-DQB1* has been implicated in various other inflammation-related conditions. This connection suggests a potential relevance to cataract formation, as chronic inflammation has been identified as a contributor to the development of cataract. Inflammatory processes can induce oxidative stress in the lens, leading to protein degradation and the development of lens opacities (Huang *et al.*, 2016, 2024).

Congenital cataract exhibited negative genetic correlations (between 0 and -1) with some traits, such as senile cataract (r_g = -0.230, p = 0.110), diabetic cataract (r_g = -0.186, p = 0.168), and BMI (r_g = -0.432, p = 0.312) and positive genetic correlation scores (between 0 and 1) with other traits such as asthma (r_g = 0.071, p = 0.521) and T2D (r_g = 0.138, p = 0.323), however, these results were not statistically significant. These findings suggest that shared genetic factors may have opposite effects on these traits. Unlike other cataract subtypes, which are influenced by environmental factors such as diabetes and aging, congenital cataract are primarily associated with inherited genetic variants. The observed inverse associations may indicate protective biological pathways that could be explored further with more robust congenital cataract data and improved statistical power.

However, our study has limitations that should be considered when interpreting the results. Due to computational restraints when using the LDSC package for genetic correlation, the optimal definition of diabetic cataract was not used in this analysis. Unlike other GWAS studies on diabetic cataracts that define cases and controls within a diabetic population, using individuals with diabetes as both cases (those with cataracts) and controls (those without), our definition included cataract cases identified through diagnostic and operational codes following an initial T2D diagnosis, representing incident cases of cataracts after diabetes onset (Chang *et al.*, 2016). This approach did not provide a control group limited to individuals with diabetes and thus limits the association we were able to gather during our GWA analysis for diabetic related cataract. Furthermore, diabetic cataract is a subset of the broader cataract group. The weaker genetic correlation score observed between diabetic cataract and the overall cataract may reflect genetic factors specific to diabetic cataract that are not shared across all cataract subtypes. For instance, the genetic correlation score between senile cataract and diabetic cataract is similarly moderate, suggesting distinct but overlapping genetic contributions.

Furthermore, a limitation of our study was the inability to differentiate between specific cataract subtypes, such as posterior subcapsular, cortical, and nuclear cataracts. The UKB data predominantly contains nuclear cataract cases, which is the primary subtype of age-related cataracts. However, research has suggested that specific risk factors are associated with particular subtypes; for example, the Singapore Malay Eye Study revealed that diabetes have distinct effects on different cataract subtypes such as cortical and posterior subcapsular with different magnitudes of effects (Tan *et al.*, 2020), thus it would be interesting to see if stronger genetic correlation or shared variants are present across different cataract subtypes. However, a significant challenge is the limited availability of large-scale datasets that differentiate between clinical cataract subtypes. As previously discussed, there is evidence of subtype-specific associations, and further investigation, particularly through genetic correlation

analyses, could improve our understanding of the underlying biological distinctions between these subtypes. As shown in Figure 6.1, two genetic correlation scores, congenital cataract with itself and T2D with diabetic cataract, were slightly inflated above 1. This inflation may be due to sample-related issues, such as the smaller sample sizes for these traits, and could be corrected by increasing our sample size.

While our study would have benefited from the use of Genome-wide Complex Trait Analysis (GCTA), which provides more accurate estimates of SNP-based heritability and genetic relationships between traits when compared to LDSC, its use was not feasible due to computational constraints. Instead, LDSC was chosen as it allowed the inclusion of externally published GWAS summary statistics, making it a more practical option for this analysis (Speed *et al.*, 2017).

Overall, our study highlights the potential genetic overlap between cataract subtypes and risk factors such as T2D and asthma. Our findings not only align with the existing literature but also expand on it by identifying specific genomic regions and genes that may underlie these shared mechanisms. Notably, overlapping genes identified in this investigation play key roles in inflammation and oxidative stress mechanisms. These results suggest that cataract may share broader biological pathways with metabolic and inflammatory diseases, reinforcing the importance of systemic approaches to cataract prevention and treatment.

Our study has demonstrated clear evidence of shared genetic mechanisms between cataract and traits such as T2D and asthma, including the identification of specific genes like *CDKN2B-AS1* and *WWP2* that may contribute to both cataract formation and other conditions. The results highlight the complex genetic architecture overlapping cataract subtypes and potential exploration of therapeutic targets. Future investigation should focus on refining cataract subtype definitions (posterior subcapsular, cortical and nuclear) and their genetic overlap with different risk factors. Overall, there is evidence of shared genetic mechanisms, such as the *CDKN2B-AS1* and *WWP2* genes in T2D and cataracts, which present promising opportunities for further investigation as potential therapeutic targets.

7 Discussion

During this investigation, genetic data was incorporated into several studies on cataract. The aims of this thesis, as outlined in Section 1.7, were to explore the overlapping genetic mechanisms among cataract subtypes, as well as the relationship between cataracts and well-known environmental risk factors. Additionally, this work aimed to identify potential modifiable risk factors to aid in cataract prevention and assess the efficacy of potential drug targets as alternative treatment options to surgery, thereby reducing the burden on healthcare systems.

In this section, the results will be critically evaluated in relation to the project's aims. This includes discussing the overarching strengths and limitations of the study, interpreting the implications of the findings, and proposing directions for future research to further our understanding of cataracts.

7.1 Summary of results

In Section 6, cataract subtypes were classified based on the disease's aetiological categories, including senile cataract, congenital cataract, and diabetic cataract. An overall cataract definition was also utilised to examine associations with environmental risk factors such as T2D, asthma, and BMI. The subdivision of cataract subtypes aimed to enhance understanding of the shared biological pathways across different forms of the disease. Additionally, the analysis explored how environmental factors, such as asthma, might provide insights into potential therapeutic targets that could also influence cataract development. To evaluate these relationships, genetic correlation and co-localisation analyses were conducted.

Overall, the results of this investigation highlighted an opposing effect of congenital cataracts compared to other cataract subtypes and BMI, however, as these results were not statistically significant their effect is distinguishable from 0 (null effect). Statistically significant genetic correlations were observed between certain pairs, including senile and diabetic cataracts, overall cataracts and T2D, asthma and overall cataracts, as well as diabetic cataracts and asthma. These findings were further examined using co-localisation analysis to better understand whether the observed genetic correlation may be driven by shared causal variants acting through specific biological pathways. This analysis identified genes of interest, such as *WWP2* and *CDKN2B-AS1*, which are shared genetic variants between cataracts and T2D. The results address the aim of exploring genetic overlap between cataract subtypes and clarify potential shared genetic variants with environmental factors.

The second aim of this project, identifying potential modifiable risk factors, was addressed in Section 4 and Section 5. Section 4 explored the potential causal relationship between vitamin

D levels, vitamin D deficiency, and cataract risk. This investigation incorporated both observational and MR analyses, using UKB data as well as publicly available GWAS data. While a correlative association between vitamin D levels and cataract risk was identified, no robust evidence was found to support a causal relationship between vitamin D levels, vitamin D deficiency, and cataract risk.

Section 5 further investigated a potential modifiable risk factor, alcohol consumption, this analysis investigated the relationship between alcohol consumption and cataract risk through observational and MR approaches. A one-sample MR analysis used the rs1229984 variant in the *ADH1B* gene as a genetic proxy for alcohol consumption to infer causality. This was supplemented by a two-sample MR analysis using publicly available GWA data for alcohol consumption. Additionally, a multivariable MR analysis was conducted to account for the potential mediating role of BMI in the relationship between alcohol consumption and cataract risk. While the observational results align with previous findings on the association between alcohol consumption and cataract, the genetic analyses found no evidence of a causal relationship.

The results presented in Sections 4 and 5 do not support the clinical efficacy of alcohol consumption modification or vitamin D supplementation as preventive measures for cataracts.

The third aim of this study was to explore potential alternative treatment options for cataracts. Specifically, the suitability of lanosterol as a drug target was investigated by assessing lanosterol-related genetic variants. Using a generated cataract GWAS, genetic variants within the lanosterol synthase gene region were examined. This was followed by a look-up analysis of previously published genetic associations with phytosterol-to-lanosterol ratios. Finally, a genetic risk score analysis was conducted to test the association between lanosterol, within the cholesterol synthesis pathway, and cataract risk. No statistically significant associations between SNPs in the lanosterol synthase gene region and cataract were identified. Furthermore, the look-up analysis and genetic risk score analysis provided no evidence of an association between lanosterol genetic variants and cataract risk.

7.2 Overall strengths and limitations

A significant majority of previous studies on cataracts relied on observational techniques. While these studies provided a strong foundation for further research, they lacked the ability to assess critical elements such as causality. Understanding causality is essential for gaining deeper insights into the biology of diseases, identifying preventive risk factors, and exploring potential drug targets. Furthermore, many observational studies were limited by comparatively small sample sizes.

The UKB served as the primary data source for this investigation, offering high-quality phenotypic and genotypic information alongside a large sample size. The UKB dataset includes approximately 500,000 individuals, enabling robust analyses with both sample and genetic QC steps (Bycroft *et al.*, 2018). For example, the overall cataract GWAS utilised throughout this study included 45,449 cataract cases and 353,371 controls post QC, providing a substantial dataset for analysis. The phenotypic data available in the UKB also allowed for improvements on previously published cataract definitions. Within this study, cataracts were assessed using accurate operative and diagnostic ICD codes, avoiding the misclassification risks associated with self-reported data.

While our cataract definition makes effective use of diagnostic and operation codes, there are important limitations to consider. These codes are derived from hospital linkage data through the NHS such as hospital admissions and primary care (Sudlow *et al.*, 2015). The use of these records may lead to a degree of misclassification of controls. For example, some individuals with visually significant cataract may not appear in the data if they underwent surgery through a private institution, thus would not be captured in NHS records, or were unable access care through the NHS. In addition, some patients may choose to delay or avoid cataract surgery altogether, leading to potential misclassification of controls as a result of no physical record of cataract despite suffering from the disease. Misclassification can also arise when lens extraction is carried out as part of a separate ocular operation, including glaucoma or retinal detachment surgery, or when cataract develops secondary to prior ocular interventions. In these cases, surgery may be recorded but not necessarily be a direct proxy for primary age-related cataract. While these limitations should be acknowledged, the use of diagnostic and operation codes currently represents the most reliable method of capturing cataract cases in the UKB.

Additionally, new definitions for cataract subtypes, such as senile cataracts and diabetic cataracts, were generated. These definitions incorporated increased numbers of cases and controls and leveraged the REGENIE technique to account for relatedness in GWAS analyses (Mbatchou *et al.*, 2021), maximising the number of cases included.

The UKB also provided corresponding dates for operation and diagnostic codes for cataracts, enabling the assessment of cataract incidence in observational analyses. This data enhanced the quality of the analyses by allowing for more accurate understanding of disease progression and the relationship between risk factors and the risk of cataract.

However, data availability posed a limitation for some analyses. For instance, as outlined in Section 6, a congenital cataract GWAS definition was utilised. The UKB, however, recorded only 140 cases of congenital cataracts, resulting in weak statistical power for the GWAS.

Consequently, the genetic correlation and co-localisation analyses were limited in power when using the congenital cataract definition. To adequately assess congenital cataracts, a larger number of cases is necessary, with differentiation between subtypes such as congenital cataracts present at birth versus those developing in early childhood (within the first two years). Similarly, in Section 3, limited genetic data were available for lanosterol, which constrained the analysis and prevented a direct assessment of the effect of lanosterol on cataract risk. Moreover, the analysis was based on blood measurements of lanosterol, which may not accurately reflect its role within the lens structure, as the lens is distinct from blood tissues. Drug-based studies often rely on eQTL to evaluate causal effects (Vitali *et al.*, 2019). However, in this case, available eQTL data represented the gene expression of lanosterol in blood, rather than in lens tissue. Expanding eQTL data to include lens-specific gene expression would provide a more comprehensive basis for genetic analysis and improve the ability to assess the role of lanosterol in cataract development.

The MR-based analyses presented in Sections 4 and 5 assessed alcohol consumption and vitamin D levels, including deficiency, using approaches such as one-sample, two-sample, and multivariable MR models. However, both vitamin D and alcohol consumption have been previously reported to exhibit non-linear effects on various diseases (Dan *et al.*, 2022; Visontay *et al.*, 2022). In the vitamin D study, a non-linear MR analysis was included as additional supplementary, following the currently recommended approach for minimising bias arising from violations of the assumption of a homogeneous exposure effect across the population. However, due to methodological challenges, this analysis was not incorporated into the main results (Burgess *et al.*, 2023; Hamilton *et al.*, 2024). Whilst my work has not established evidence for non-linearity of causal effects, to thoroughly investigate the potential non-linear relationships between exposures such as alcohol consumption and vitamin D, more reliable and robust non-linear MR methods are required.

7.3 Recommendations for future research

From this series of analyses, several recommendations for future research have emerged. One of the key limitations of this study was the lack of available data on the clinical differentiation of cataract subtypes. To enable subtype-specific associations to be identified, future research should focus on follow-up investigations within the UKB cohort to classify existing cataract cases into nuclear, cortical, and posterior subcapsular subtypes. This would allow for more precise genetic associations to be identified, improving our understanding of subtype-specific risk factors and associated genetic variants to each.

Additionally, another data limitation was the low number of congenital cataract cases available for analysis. Future studies could expand the investigation of congenital cataract by increasing

case numbers through data linkage with other cohorts or by expanding the analysis to additional participants from other ethnicities where congenital cataract is more common. This would enable a more comprehensive assessment of the genetic and environmental determinants of congenital cataract, which remains a limited research area.

Furthermore, while GWAS has provided valuable insights into common variants associated with cataract, these loci typically explain only a small fraction of the overall genetic variation in the disease. Rare variants, which often have larger effect sizes, represent an important but underexplored component of cataract genetics. The inclusion of rare variants would further aid in the exploration of cataract subtypes such as congenital as well as non-congenital forms, such as age-related.

CNVs also represent an additional layer of genetic variation. To date, CNV analysis in cataract has been conducted primarily in congenital subtypes. With ongoing advances in whole-genome sequencing and CNV detection methods, it will become increasingly feasible to systematically evaluate CNVs in large population-based cohorts. This could clarify their contribution to age-related cataract and other subtypes, where the role of structural variation remains poorly understood. Integrating analyses of common SNPs, rare variants, and CNVs will therefore be essential to capture the genetic architecture of cataract and better understand its pathogenesis across different subtypes.

Age is the most significant non-modifiable risk factor for age-related cataract; the older an individual becomes, the more likely they are to develop cataract (see Section 1.4.1). In epidemiological studies, this strong age-dependence can be modelled using time-to-event (survival) analyses (Abd ElHafeez *et al.*, 2021). With the expansion of large-scale population cohorts linked to health records, it has become feasible to apply survival-based methods such as Cox proportional hazards regression in GWAS, rather than relying solely on traditional logistic or linear regression approaches. Survival analyses have the advantage of incorporating age at onset and time-to-event information, which may capture the underlying biology of cataract more effectively when compared to case-control designs. As computational methods and efficiency continue to improve, applying survival-based GWAS approaches could provide more accurate insights into the effect of genetic risk factors on cataract (Pedersen *et al.*, 2023). Furthermore, GWAS designs could also be strengthened by improved matching of older participants within the case and control groups. This would reduce potential age-related imbalances between cases and controls as result of the impact of aging on cataract development, thus minimising the effect of age distribution differences within the model and enhancing the interpretability of results.

While this investigation focused on two modifiable risk factors, vitamin D and alcohol consumption, future research should expand to include additional known risk factors for cataract, such as vitamin C and air pollution. The feasibility of such analyses is largely dependent on the availability of genetic instruments for these traits, but a more comprehensive understanding of these risk factors could help identify causally modifiable interventions for cataract prevention.

Furthermore, while this study investigated lanosterol as a potential drug target, which has shown promising pre-clinical findings, other potential targets have also been suggested to aid in protein degradation within a cataractous lens. One such candidate is 25-hydroxycholesterol, which may influence lens transparency. However, further genetic data and functional validation are required to assess whether there is genetic evidence supporting its potential for future human trials.

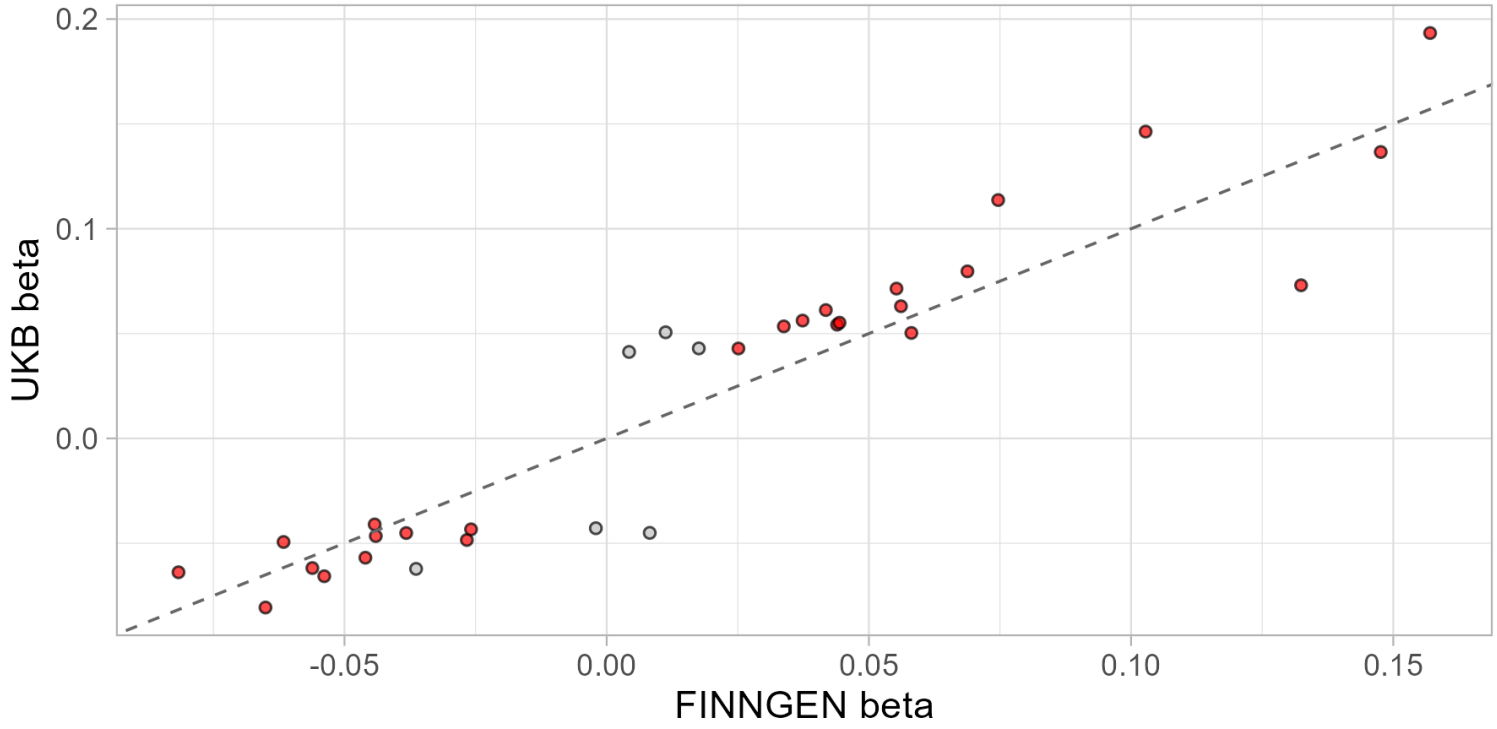
7.4 Implications of findings and conclusion

This investigation has provided valuable insights into the underlying genetic mechanisms of different cataract subtypes and identified potential biological pathways linking cataracts with type 2 diabetes. Additionally, commonly associated cataract risk factors, such as alcohol consumption and vitamin D deficiency, were evaluated for their causal relationship with cataract. The findings indicate no evidence that modifying these exposures would have a direct impact on cataract prevention or delay. A previously proposed drug target, lanosterol, was also assessed, and genetic data suggest that it is unlikely to be an effective therapeutic option. Collectively, this work has expanded the current understanding of cataract genetics, offering novel insights into its aetiology. By advancing knowledge of cataract genetics and aetiology, this research contributes to the foundations of precision medicine, providing a basis for future improvements in risk prediction, earlier detection, and the development of targeted preventive and therapeutic strategies. These findings not only contribute to the broader genetic epidemiology of cataracts but also provide a foundation for future research to further explore and build upon.

8 Appendix

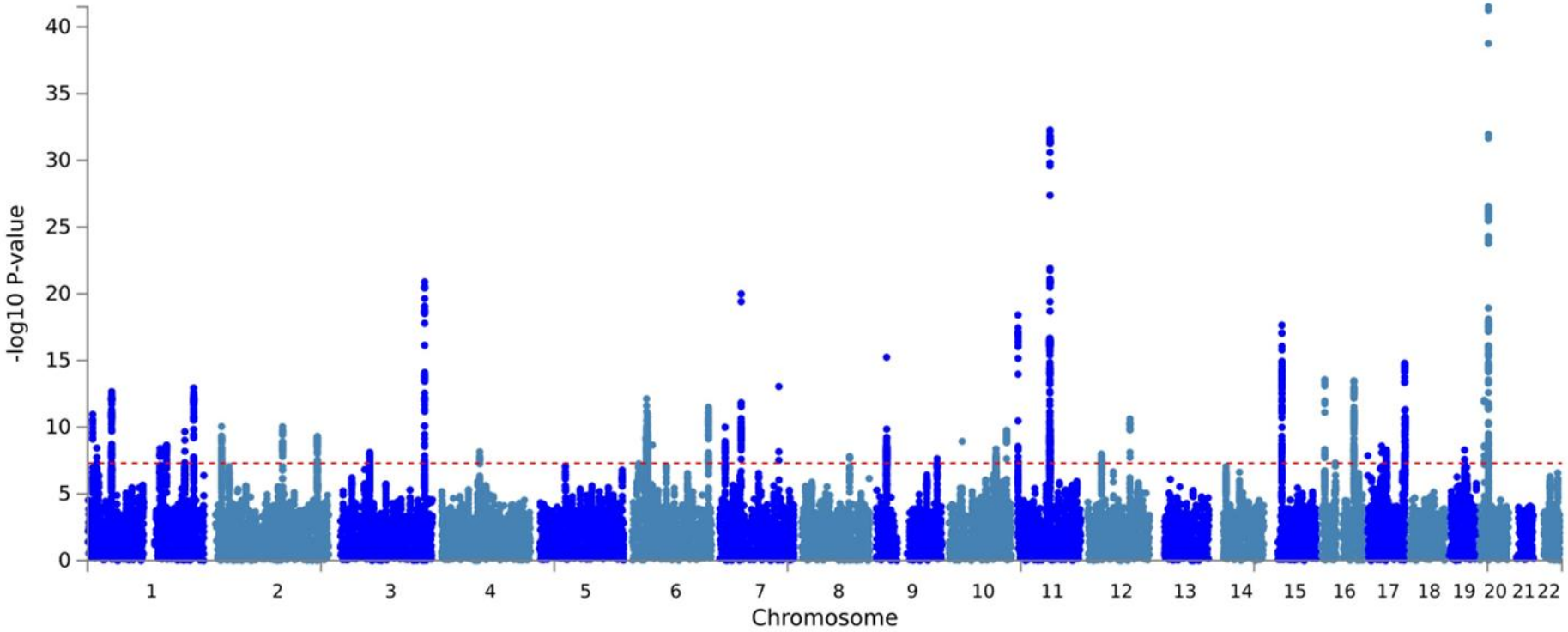
Supplementary Figure 3.1

Scatter plot comparing the beta values of independent SNPs found in UKB cataract GWAS results against those present in the FinnGen R9 cataract senile summary statistics. Red points indicate SNPs statistically significant across both UKB (at *p*-value ≤ $5x10^{-8}$) and FinnGen (*p*-value ≤ 0.05/ No. of independent SNPs in UKB cataract GWAS). Grey points indicate SNPs not statistically significant in FinnGen. The dashed grey line represents the identity line (slope = 1), indicating where SNP effect estimates are identical across both studies. Deviations from this line highlight where either study overestimates or underestimates the effect sizes. The correlation coefficient between effect estimates is r = 0.938.



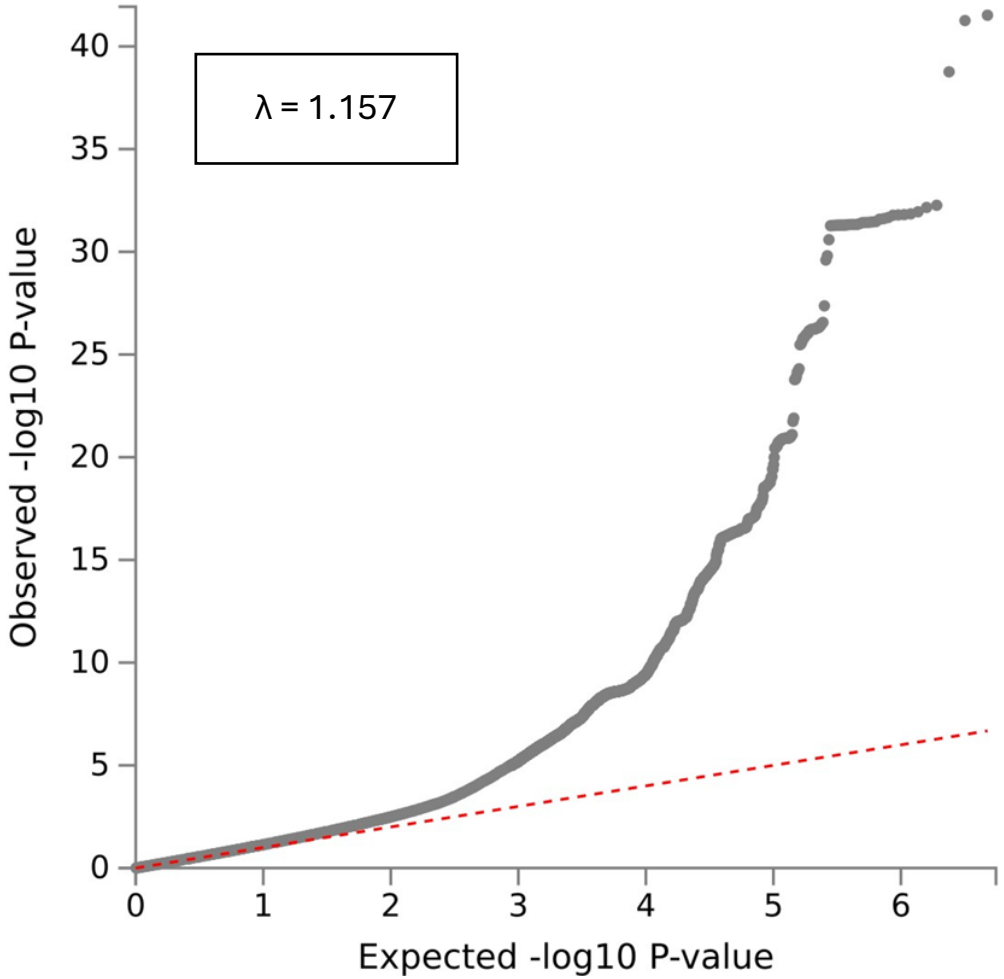
Supplementary Figure 3.2

Manhattan Plot (UKB GWAS summary statistics) produced using FUMA (https://fuma.ctglab.nl/). The dotted red line is set at conventional GWAS significant threshold of $5x10^{-8}$.



Supplementary Figure 3.3

QQ plot (UKB GWAS summary statistics), with genomic inflation factor for UKB cataract GWAS (λ = 1.157). Dotted red line indicates the null hypothesis of no association. Upward curve at the tail of the plot is indicative of true genetic associations amongst SNPs.



Supplementary Table 3.1

SNPs identified in the *LSS* gene region (expanded by 5Kb) for Approach 1.

SNP	CHR	BP	EA	OR	LOG.OR.SE	P	EAF
rs56293013	21	47603413	G	1.007	0.009	0.444	0.216
rs35960150	21	47603509	T	1.008	0.012	0.500	0.120
rs79044044	21	47603849	G	1.013	0.013	0.335	0.090
rs73144734	21	47603893	T	1.013	0.013	0.327	0.093
rs62215189	21	47603995	G	1.005	0.009	0.554	0.243
rs78331376	21	47604061	A	1.003	0.020	0.890	0.034
rs9647242	21	47604357	T	1.012	0.013	0.346	0.093
rs116170889	21	47604361	A	1.006	0.017	0.713	0.049
rs116442826	21	47604708	C	0.999	0.011	0.961	0.141
rs12482343	21	47604967	A	0.996	0.013	0.738	0.092
rs12482390	21	47605225	A	0.999	0.010	0.954	0.150
rs12482934	21	47605335	T	0.996	0.013	0.766	0.092
rs79937763	21	47605510	A	1.005	0.012	0.696	0.109
rs62212860	21	47605797	G	0.994	0.015	0.693	0.066
21:47606511_CA_C	21	47606511	C	1.015	0.007	0.044	0.531
21:47606751_TTTTG_T	21	47606751	T	1.015	0.007	0.043	0.533
rs118141253	21	47606769	C	1.017	0.020	0.391	0.038
rs8129267	21	47606797	A	0.985	0.015	0.315	0.931
rs8134131	21	47606805	T	0.985	0.015	0.304	0.931
rs148902926	21	47606935	G	0.999	0.011	0.914	0.121
rs183470806	21	47606955	T	0.968	0.035	0.356	0.013
rs142838362	21	47606966	T	0.961	0.021	0.060	0.034
rs146600734	21	47606986	C	1.005	0.017	0.775	0.053
rs182705598	21	47607000	C	0.930	0.031	0.018	0.017
rs570448320	21	47607056	T	0.979	0.035	0.546	0.012

rs9984805	21	47607057	G	1.006	0.008	0.453	0.332
rs28801979	21	47607088	A	1.005	0.017	0.775	0.050
rs118008360	21	47607364	T	1.060	0.039	0.134	0.010
rs555228834	21	47607713	ATT	1.003	0.009	0.720	0.217
rs11702775	21	47607847	C	1.003	0.009	0.767	0.218
rs114033373	21	47608207	T	1.008	0.017	0.633	0.049
rs2968	21	47608580	A	1.013	0.007	0.085	0.572
rs76408279	21	47608710	T	1.008	0.017	0.661	0.049
rs56173768	21	47608877	A	1.003	0.009	0.772	0.216
rs9717	21	47609510	C	1.014	0.008	0.083	0.642
rs118184861	21	47609578	G	1.012	0.031	0.699	0.014
rs914247	21	47609677	G	1.013	0.008	0.085	0.633
rs2330408	21	47610066	C	1.013	0.008	0.086	0.634
rs201035928	21	47610269	GGCATGGG GCTCCCT	0.998	0.015	0.896	0.061
rs73144744	21	47611285	C	1.014	0.011	0.202	0.144
rs2187118	21	47611310	G	1.012	0.008	0.106	0.573
rs2187119	21	47611358	G	1.003	0.009	0.748	0.217
rs76715041	21	47611545	A	1.001	0.023	0.966	0.026
rs17293705	21	47611799	A	1.035	0.027	0.201	0.019
rs2839139	21	47612166	C	1.013	0.008	0.090	0.634
rs116474329	21	47612424	A	1.008	0.017	0.661	0.049
rs2187120	21	47612466	G	1.013	0.008	0.077	0.573
rs2187121	21	47612471	C	1.013	0.008	0.077	0.573
rs146297977	21	47612512	CTG	1.012	0.008	0.111	0.627
rs73908575	21	47612512	G	1.042	0.027	0.130	0.020
rs3902367	21	47612867	G	1.014	0.008	0.082	0.642
rs372958158	21	47612941	C	0.987	0.020	0.513	0.037
rs561628345	21	47612986	A	0.960	0.027	0.131	0.024

rs2187122	21	47613004	C	1.014	0.008	0.079	0.639
rs10686231	21	47613160	GTGTA	1.006	0.013	0.648	0.104
rs58557992	21	47613161	T	1.004	0.011	0.710	0.129
21:47613285_ATG_A	21	47613285	A	1.011	0.015	0.484	0.071
rs144693442	21	47613749	A	1.006	0.017	0.716	0.049
rs79223895	21	47614169	C	0.999	0.015	0.930	0.061
rs2254522	21	47614443	G	1.005	0.010	0.572	0.180
rs2254524	21	47614469	C	1.014	0.008	0.079	0.642
rs35785446	21	47614553	A	1.007	0.017	0.686	0.049
rs2277824	21	47614660	T	1.005	0.011	0.666	0.130
rs9981910	21	47614975	T	1.005	0.011	0.679	0.130
rs76224955	21	47615038	A	0.998	0.022	0.912	0.028
rs12483507	21	47615403	C	0.992	0.026	0.764	0.022
rs146843341	21	47615891	A	0.998	0.022	0.921	0.032
rs7282841	21	47616071	C	1.013	0.008	0.087	0.633
rs2839140	21	47616080	A	1.013	0.007	0.086	0.572
rs76723404	21	47616737	G	0.998	0.015	0.913	0.061
rs9979525	21	47616818	C	1.013	0.008	0.087	0.634
rs55870069	21	47616850	A	1.016	0.013	0.216	0.095
rs78273090	21	47616905	A	0.999	0.015	0.955	0.061
rs9976233	21	47616913	A	1.013	0.007	0.085	0.572
rs11909555	21	47617489	G	1.005	0.010	0.582	0.180
rs10854480	21	47617810	C	0.987	0.008	0.098	0.363
21:47618181_TA_T	21	47618181	T	1.012	0.008	0.111	0.633
rs115295783	21	47618203	C	1.008	0.017	0.656	0.049
rs7280110	21	47619039	A	1.013	0.007	0.087	0.572
rs4819213	21	47619300	A	1.013	0.007	0.087	0.572
rs117106063	21	47619333	T	1.007	0.017	0.680	0.049
rs4819214	21	47619784	A	1.013	0.008	0.089	0.633
rs2839141	21	47620082	A	1.013	0.008	0.089	0.633

rs28560443	21	47621036	C	1.012	0.017	0.494	0.051
rs797014314	21	47621715	C	0.989	0.008	0.180	0.561
rs117393766	21	47621751	A	0.997	0.015	0.851	0.061
rs62212862	21	47621869	A	0.987	0.008	0.083	0.358
rs191009864	21	47622000	A	0.901	0.035	0.003	0.013
rs55689527	21	47622727	G	0.987	0.007	0.081	0.428
rs2839142	21	47622870	C	1.002	0.009	0.859	0.219
rs7282352	21	47622981	C	0.995	0.011	0.645	0.869
rs11702145	21	47623068	G	1.013	0.010	0.177	0.175
rs4818828	21	47623274	G	0.987	0.008	0.081	0.358
rs4819215	21	47623573	A	0.985	0.007	0.047	0.433
rs34105866	21	47623848	G	0.996	0.011	0.710	0.870
rs771351927	21	47624536	CAAAAA	0.989	0.016	0.488	0.058
rs114334512	21	47624548	G	1.019	0.018	0.304	0.044
rs2075906	21	47625544	T	0.996	0.011	0.713	0.870
rs2839143	21	47625658	C	1.013	0.010	0.179	0.175
rs76660727	21	47626104	A	0.992	0.015	0.592	0.066
rs16978976	21	47626728	T	1.015	0.013	0.231	0.096
rs117330398	21	47627090	A	1.011	0.032	0.737	0.014
rs2839144	21	47627245	A	0.986	0.008	0.076	0.366
rs140139047	21	47628082	T	1.068	0.038	0.079	0.011
rs73144751	21	47628375	T	1.003	0.009	0.779	0.217
rs11089053	21	47628711	G	0.996	0.011	0.735	0.871
rs148436940	21	47628715	C	0.995	0.036	0.882	0.012
rs78155037	21	47628894	C	0.991	0.015	0.557	0.067
rs2839145	21	47629268	A	0.989	0.008	0.161	0.586
rs138130258	21	47630129	A	0.992	0.015	0.589	0.066
rs2839146	21	47630550	T	0.987	0.008	0.084	0.358
rs4819216	21	47630862	A	0.987	0.008	0.105	0.359
rs2839147	21	47630951	C	0.994	0.011	0.620	0.873

rs140233907	21	47631165	C	1.005	0.030	0.859	0.017
rs2001809	21	47631199	T	0.981	0.008	0.015	0.357
rs2009213	21	47631245	G	1.000	0.009	0.977	0.758
rs78276120	21	47631784	G	0.992	0.015	0.587	0.068
rs370233614	21	47631961	A	1.018	0.018	0.315	0.044
21:47631963_TGG GCAGGGAG_T	21	47631963	T	0.991	0.008	0.213	0.409
rs117773065	21	47632067	G	0.991	0.015	0.555	0.068
rs2839148	21	47632276	A	1.012	0.014	0.390	0.078
rs2839149	21	47632580	T	1.003	0.011	0.783	0.120
rs2839151	21	47632995	C	1.015	0.013	0.239	0.096
rs9974665	21	47633789	G	1.011	0.008	0.157	0.394
rs11701000	21	47634477	A	1.002	0.009	0.863	0.221
rs4239841	21	47634499	C	0.884	0.073	0.092	0.997
rs914248	21	47634572	G	0.999	0.011	0.900	0.872
21:47634860_GT_G	21	47634860	G	1.011	0.014	0.416	0.077
rs536505576	21	47634862	A	1.011	0.014	0.416	0.077
rs73144753	21	47634915	A	1.001	0.009	0.873	0.221
rs34115287	21	47635176	C	1.016	0.010	0.106	0.174
rs60322177	21	47635545	A	1.021	0.018	0.257	0.044
rs2839152	21	47635577	T	0.976	0.016	0.117	0.067
rs9980968	21	47635627	G	1.010	0.008	0.176	0.396
rs11701729	21	47635713	A	0.994	0.008	0.460	0.722
rs2839153	21	47635856	G	0.999	0.011	0.894	0.872
rs73386515	21	47635883	T	1.020	0.018	0.276	0.044
rs76489504	21	47636557	A	1.011	0.014	0.431	0.077
rs141367193	21	47637197	ACTGTAGGT	0.990	0.008	0.197	0.335
rs2839154	21	47637760	T	1.007	0.007	0.365	0.444
rs61591722	21	47637767	C	1.004	0.011	0.734	0.121
rs34625510	21	47638019	T	1.012	0.014	0.408	0.077

rs202056351	21	47638053	AC	0.992	0.015	0.617	0.065
rs9984986	21	47638463	G	1.017	0.008	0.035	0.680
rs11702846	21	47638872	T	1.001	0.009	0.914	0.221
rs2277826	21	47639492	G	1.009	0.008	0.233	0.590
rs74328331	21	47639548	T	0.991	0.015	0.555	0.068
rs117110314	21	47639614	C	1.023	0.038	0.552	0.010
rs73144754	21	47639876	A	1.001	0.009	0.915	0.221
rs2839155	21	47639992	A	1.011	0.007	0.152	0.522
rs6518278	21	47640571	G	1.014	0.008	0.083	0.643
rs35679325	21	47640980	G	1.009	0.008	0.237	0.590
rs2839156	21	47641196	A	1.002	0.011	0.866	0.128
rs13049175	21	47641326	A	1.011	0.007	0.152	0.522
rs13046451	21	47641373	A	1.011	0.007	0.151	0.522
rs2839157	21	47641700	T	1.014	0.008	0.083	0.643
rs2839158	21	47641794	T	1.016	0.010	0.115	0.173
rs2280959	21	47641996	G	0.989	0.007	0.152	0.478
rs2280958	21	47642016	T	0.989	0.007	0.152	0.478
rs2280957	21	47642272	T	0.989	0.007	0.152	0.478
rs2280956	21	47642323	G	0.989	0.007	0.151	0.478
rs2280955	21	47642397	T	0.986	0.008	0.078	0.357
rs11558754	21	47642609	A	1.016	0.010	0.114	0.173
rs117806396	21	47642817	T	1.017	0.019	0.378	0.038
rs11909228	21	47642914	T	0.998	0.011	0.878	0.872
rs73144762	21	47643382	G	1.015	0.010	0.117	0.173
rs12151996	21	47643442	A	1.009	0.008	0.223	0.397
rs4819217	21	47644092	T	0.991	0.008	0.243	0.410
rs4819218	21	47644169	T	0.990	0.007	0.155	0.478
rs6518282	21	47644279	A	0.990	0.007	0.155	0.478
rs6518283	21	47644334	T	0.998	0.011	0.854	0.872
rs6518284	21	47644428	G	0.998	0.011	0.860	0.872

rs6518285	21	47644667	C	0.990	0.007	0.155	0.478
rs200769143	21	47644808	GA	0.993	0.015	0.619	0.068
rs139806871	21	47644999	C	0.992	0.015	0.605	0.068
rs12152059	21	47645087	G	1.039	0.027	0.157	0.020
rs148982990	21	47645196	C	0.941	0.029	0.040	0.018
rs13052806	21	47645670	G	0.990	0.007	0.162	0.478
rs56333186	21	47645920	A	1.017	0.013	0.181	0.096
rs9984242	21	47645970	C	0.997	0.007	0.693	0.506
rs13052767	21	47646041	T	0.967	0.033	0.309	0.014
rs116885460	21	47646896	T	0.993	0.015	0.613	0.068
rs999691	21	47647382	A	0.987	0.008	0.089	0.357
rs76428496	21	47647439	A	0.993	0.015	0.619	0.068
21:47648729_CCC CGCCCCT_C	21	47648729	C	0.986	0.017	0.398	0.947
rs915803	21	47648872	G	0.998	0.007	0.771	0.498
rs576793892	21	47649103	AGGGCGG	0.996	0.011	0.739	0.165
rs567709402	21	47649193	C	0.967	0.037	0.362	0.012
rs915804	21	47649802	C	0.996	0.007	0.547	0.565
rs73144764	21	47649917	T	0.999	0.009	0.934	0.219
21:47650214_GT_G	21	47650214	G	0.997	0.011	0.814	0.872
rs2298694	21	47650362	T	1.010	0.014	0.486	0.076
rs2839159	21	47651943	C	0.999	0.009	0.897	0.225
rs75300582	21	47652120	T	0.993	0.015	0.666	0.063
rs2839160	21	47652185	T	0.996	0.007	0.561	0.561
rs2839161	21	47652228	A	0.989	0.008	0.155	0.340
rs75159209	21	47652285	T	1.010	0.013	0.437	0.090
rs2839162	21	47652549	G	0.998	0.012	0.871	0.885
rs548916783	21	47652668	CA	1.016	0.012	0.193	0.112
rs79066226	21	47652906	T	0.993	0.015	0.666	0.063
rs17183473	21	47653345	G	0.993	0.015	0.666	0.063

rs8133857	21	47653422	C	1.000	0.011	0.997	0.881
21:47653623_CAG_C	21	47653623	C	0.997	0.014	0.850	0.073

Supplementary Table 3.2

Full results of Approach 2. List of SNPs present in cataract GWAS and statistically significant and independent in each phytosterol-to-lanosterol ratio summary statistic at *p*-value < $5x10^{-8}$ and r^2 threshold of < 0.1, respectively. Heatmap of SNP's phytosterol-to-lanosterol ratio summary statistic betas where green represents beta > 0 and red represents beta < 0. SNPs have been ordered with respect to *p*-value in Cataract GWAS.

					Phytosterol-to-Lanosterol Ratio Pairings and Corresponding Betas							
SNP	**CHR**	**Effect Allele**	**Beta in Cataract GWAS**	**P-value in Cataract GWAS**	brf_laf	brt_laf	caf_laf	cat_laf	sif_laf	sit_laf	stf_laf	stt_laf
rs612169	9	G	0.021	0.008	NA	NA	NA	0.047	NA	NA	NA	NA
rs550057	9	T	0.019	0.023	NA	NA	0.058	NA	0.060	0.058	NA	NA
rs17424122	2	A	-0.034	0.028	NA	NA	NA	0.120	NA	NA	NA	NA
rs111559090	2	A	-0.019	0.059	NA	NA	0.067	NA	NA	NA	0.060	NA
rs4076834	2	G	-0.028	0.066	NA	-0.232	NA	-0.210	NA	NA	NA	NA
rs60668987	2	A	-0.039	0.070	NA	NA	NA	NA	NA	NA	NA	-0.273
rs10208987	2	G	0.023	0.092	NA	NA	NA	NA	NA	-0.102	NA	NA
rs8302	2	C	-0.014	0.144	0.074	0.069	NA	0.064	0.079	NA	NA	NA
rs6735229	2	C	-0.010	0.167	-0.081	-0.082	-0.061	-0.075	-0.088	-0.098	-0.060	NA
rs77370416	2	C	0.023	0.197	0.133	0.157	NA	NA	NA	NA	NA	NA
rs13427362	2	G	0.017	0.246	NA	NA	NA	NA	-0.103	NA	NA	NA
rs3846662	5	G	-0.008	0.316	-0.050	-0.050	-0.047	-0.047	NA	NA	NA	NA
rs10205879	2	C	-0.008	0.353	NA	NA	NA	NA	NA	0.067	NA	NA
rs12916	5	C	-0.005	0.495	NA	NA	NA	NA	NA	NA	NA	-0.047
rs11057839	12	T	0.006	0.605	NA	NA	NA	NA	NA	0.070	NA	NA
rs67734975	2	G	-0.009	0.608	NA	NA	NA	NA	0.125	0.135	NA	NA
rs7598542	2	C	0.005	0.612	NA	NA	NA	NA	NA	NA	NA	-0.082
rs7599981	2	G	-0.004	0.660	NA	-0.056	NA	NA	NA	NA	NA	NA
rs10070119	5	T	-0.003	0.664	NA	NA	NA	NA	NA	-0.053	NA	NA
rs145288624	2	T	-0.005	0.718	NA	0.107	NA	NA	NA	0.116	NA	NA
rs7590687	2	C	0.005	0.749	NA	NA	NA	-0.134	NA	NA	NA	-0.097
rs138958276	2	A	0.004	0.814	NA	NA	NA	NA	NA	NA	NA	-0.123
rs140488605	2	T	-0.004	0.817	NA	NA	NA	0.149	NA	NA	NA	0.133

Supplementary Analysis 4.1

The “SUMnlmr” package was used to conduct our non-linear MR analysis (Mason and Burgess, 2022).

Non-Linear MR method

To explore the presence of a possible non-linear association between vitamin D levels and incident cataract, non-linear MR (NLMR) analyses using a fractional polynomial model were conducted. The population was stratified using three separate methods: residual, log-transforming the exposure prior to residual stratification, and double-ranked. The results of all three methods were triangulated to increase the reliability of our observation (Staley and Burgess, 2017; Burgess, 2023; Tian *et al.*, 2023).

The “SUMnlmr” package was used to conduct all non-linear MR analyses. Based on the widely used “nlmr” package from Staley and Burgess, but allowing for the implementation of the double-ranked stratification method from the Tian *et al.* “DRMR” package (Staley and Burgess, 2017; Mason and Burgess, 2022; Tian *et al.*, 2023)

For residual stratification methods, the UKBB sample was split into 10 strata by regressing vitamin D levels on the GRS and stratifying the data based on the residual variation. Using the double-ranked method, participants are first ranked into pre-strata according to their GRS. Participants within the pre-strata were then ranked according to their vitamin D levels.

GRS representing elevated vitamin D levels, described in the one-sample MR method, were used as the instrumental variable within the NLMR. For each stratum, the localised average causal effect (LACE) estimate was calculated. A meta-regression between the LACE estimates and mean vitamin D level for each stratum was performed by fitting fractional polynomial models with degrees 1 and 2, with the best fitting model selected. The best fitting fractional polynomial model was identified through a fractional polynomial degree test. Evidence that a fractional polynomial model with degree 2 was a better fit than a fractional polynomial model with degree 1 was indicated by p-value < 0.05 (Staley and Burgess, 2017). The analysis was controlled for sex, age, agexsex, age squared and the first 10 genetic PCs. Further details of the “SUMnlmr” package and fractional polynomial method can be obtained elsewhere (https://github.com/amymariemason/SUMnlmr) (Mason and Burgess, 2022).

Due to reported bias of NLMR methods, notably the widely used residual method, and to further assess the robustness of our NLMR, an additional negative control analysis was conducted between vitamin D levels and age (Smith, 2023). Age has previously been used as a negative control within NLMR analyses using vitamin D, with a null effect expected to be observed (Hamilton *et al.*, 2024). The analysis was controlled for sex and the first 10 genetic PCs.

Non-Linear MR results

The exposure-outcome relationship between vitamin D levels and cataract was investigated using a fractional polynomial model, as shown in Supplementary Analysis Figure 4.1.1, 4.1.2, and 4.1.3. For this analysis, cataract was defined by the UKBB incident cataract phenotype. Three different stratification methods were used: residual; log-transformed exposure residual; and double-ranked. The model produced results of three non-linearity tests: fractional polynomial non-linearity *p*-value; quadratic *p*-value; and Cochran Q *p*-value.

Residual

When testing for the model of best fit, there was no evidence to suggest that the fractional polynomial model of degree 2 was a better fit to a fractional polynomial model of degree 1 ($p = 0.073$). When testing for non-linearity, both quadratic ($p = 0.006$) and Cochran Q ($p = 0.044$) tests indicated some statistically significant evidence of a non-linear relationship between vitamin D levels and incident cataract. However, when evaluating the non-linear relationship between vitamin D levels and incident cataract using the fractional polynomial non-linearity test, there was no statistically significant evidence to support the presence of a non-linear relationship ($p = 0.078$).

Log-transformed residual

The fractional polynomial degree test found the best-fitting fractional polynomial model was degree 1 ($p = 0.156$). Again, there was some statistically significant evidence of a non-linear relationship (quadratic test *p*-value = 0.040). However, remaining non-linearity tests suggested no evidence of non-linear effects in the association between vitamin D and incident cataract (fractional polynomial test *p*-value = 0.421, Cochran Q test *p*-value = 0.641).

Double-ranked

The best-fitting fractional polynomial model had degree 1 (p = 0.637). All non-linearity tests found no evidence for a non-linear relationship between vitamin D and incident cataract (fractional polynomial test p-value = 0.489, quadratic test p-value = 0.448, Cochran Q test p-value = 0.325).

For full LACE estimates please see Supplementary Analysis Tables 4.1.1, 4.1.2, and 4.1.3.

Negative control analysis

As previously suggested, vitamin D and age should not present a non-linear relationship. However, this analysis presented statistically significant evidence for a non-linear relationship between vitamin D levels and age when using residual stratification (fractional polynomial test p-value = 8.28×10^{-5}, quadratic test p-value = 0.014, Cochran Q test p-value = 3.71×10^{-4}). Furthermore, a non-linear association was also established using a log-transformed exposure (quadratic test p-value = 0.035). However, when using the double-ranked stratification method, no non-linear relationship was observed between vitamin D levels and age, in line with expectations, suggesting the presence of bias within the residual and log-transformed residual fractional polynomial methods.

For full negative control LACE estimates and non-linearity tests please see Supplementary Analysis Tables 4.1.4, 4.1.5, 4.1.6, and 4.1.7.

Supplementary Analysis Table 4.1.1

Residual non-linear MR LACE estimates (10 strata).

Strata	Beta	SE	Lower CI (95%)	Upper CI (95%)	*p*-value
1	-0.014	0.009	-0.033	0.004	0.121
2	-0.009	0.006	-0.021	0.003	0.144
3	0.001	0.006	-0.011	0.013	0.819
4	-0.007	0.006	-0.019	0.005	0.260
5	0.008	0.006	-0.004	0.020	0.190
6	-0.001	0.006	-0.012	0.011	0.928
7	0.004	0.006	-0.008	0.015	0.532
8	-0.010	0.006	-0.022	0.002	0.099
9	0.009	0.006	-0.003	0.021	0.140
10	0.011	0.005	0.001	0.022	0.034

Supplementary Analysis Table 4.1.2

Log-transformed non-linear MR LACE estimates (10 strata).

Strata	Beta	SE	Lower CI (95%)	Upper CI (95%)	*p*-value
1	-0.428	0.334	-1.083	0.227	0.200
2	0.005	0.304	-0.591	0.602	0.986
3	-0.167	0.306	-0.768	0.434	0.586
4	0.020	0.310	-0.587	0.627	0.949
5	0.237	0.307	-0.364	0.839	0.440
6	0.061	0.305	-0.537	0.660	0.841
7	-0.120	0.305	-0.717	0.476	0.692
8	0.094	0.309	-0.511	0.699	0.760
9	0.233	0.312	-0.378	0.844	0.455
10	0.683	0.347	0.003	1.362	0.049

Supplementary Analysis Table 4.1.3

Double-ranked non-linear MR LACE estimates (10 strata).

Strata	Beta	SE	Lower CI (95%)	Upper CI (95%)	*p*-value
1	-0.016	0.018	-0.051	0.019	0.373
2	-0.012	0.012	-0.035	0.012	0.324
3	0.004	0.009	-0.015	0.022	0.699
4	0.006	0.008	-0.009	0.022	0.429
5	0.009	0.007	-0.004	0.022	0.161
6	0.003	0.006	-0.008	0.014	0.607
7	-0.010	0.005	-0.020	0.000	0.055
8	0.001	0.005	-0.008	0.010	0.811
9	0.006	0.004	-0.002	0.014	0.139
10	0.003	0.004	-0.004	0.010	0.387

Supplementary Analysis Table 4.1.4

Negative control residual non-linear MR LACE estimates (10 strata).

Strata	Beta	SE	Lower CI (95%)	Upper CI (95%)	*p*-value
1	0.074	0.022	0.032	0.117	0.001
2	0.042	0.014	0.015	0.068	0.002
3	0.013	0.013	-0.013	0.039	0.334
4	0.032	0.013	0.006	0.058	0.016
5	0.002	0.013	-0.024	0.027	0.907
6	-0.006	0.013	-0.031	0.020	0.675
7	-0.023	0.013	-0.048	0.003	0.082
8	-0.008	0.013	-0.034	0.018	0.555
9	-0.011	0.013	-0.037	0.014	0.383
10	0.021	0.012	-0.002	0.043	0.070

Supplementary Analysis Table 4.1.5

Negative control log-transformed non-linear MR LACE estimates (10 strata).

Strata	Beta	SE	Lower CI (95%)	Upper CI (95%)	*p*-value
1	1.673	0.768	0.166	3.179	0.030
2	1.107	0.662	-0.190	2.404	0.094
3	-0.709	0.678	-2.039	0.620	0.296
4	1.102	0.678	-0.227	2.432	0.104
5	-0.475	0.675	-1.797	0.848	0.482
6	-0.287	0.670	-1.600	1.027	0.669
7	-0.859	0.667	-2.166	0.448	0.198
8	-0.366	0.671	-1.681	0.948	0.585
9	-0.074	0.665	-1.378	1.230	0.911
10	0.309	0.705	-1.072	1.690	0.661

Supplementary Analysis Table 4.1.6

Negative control double-ranked non-linear MR LACE estimates (10 strata).

Strata	Beta	SE	Lower CI (95%)	Upper CI (95%)	*p*-value
1	0.046	0.038	-0.029	0.122	0.228
2	0.031	0.026	-0.019	0.081	0.227
3	0.009	0.020	-0.031	0.048	0.661
4	0.005	0.017	-0.028	0.037	0.784
5	0.001	0.014	-0.028	0.029	0.963
6	-0.017	0.013	-0.042	0.008	0.187
7	-0.003	0.011	-0.025	0.019	0.787
8	0.008	0.010	-0.012	0.028	0.425
9	-0.001	0.009	-0.018	0.017	0.941
10	-0.005	0.008	-0.021	0.010	0.522

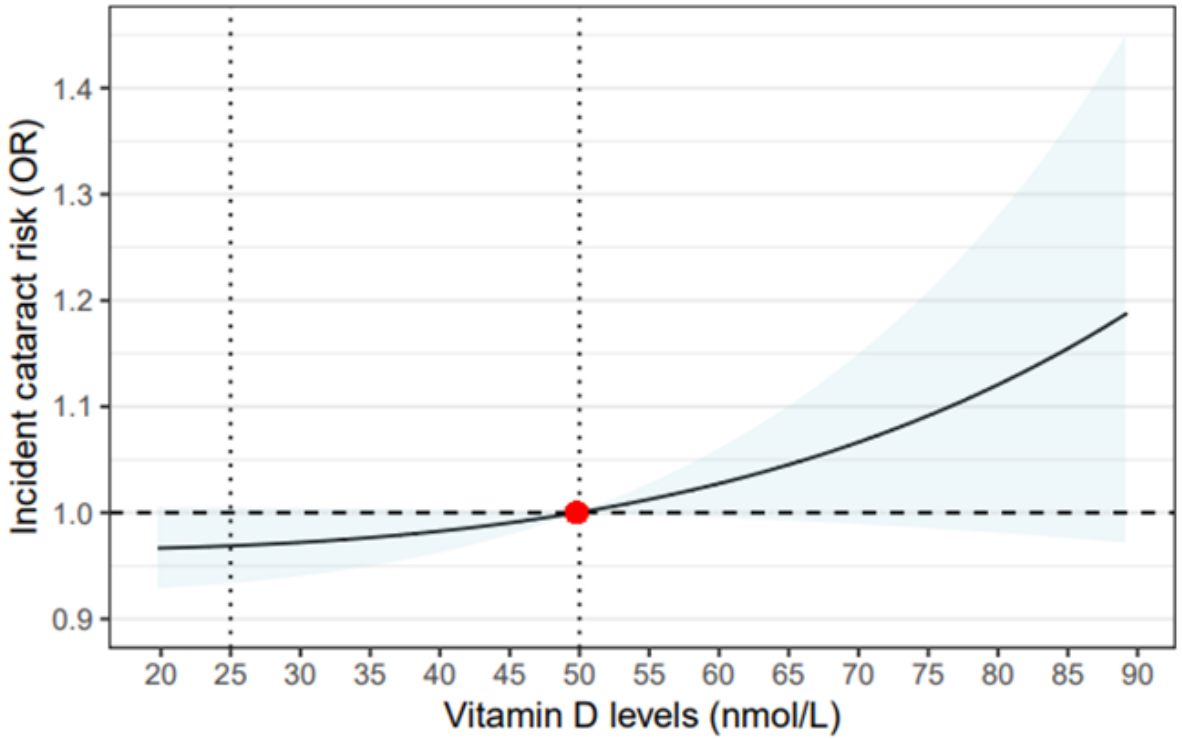
Supplementary Analysis Table 4.1.7

Negative control non-linear MR full non-linearity test *p*-values.

Non-linearity tests (*p*-value)	**Residual**	**Log-transformed residual**	**Double-ranked**
Fractional polynomial degree	0.236	0.068	0.366
Fractional polynomial non-linearity	8.28E-05	0.347	0.316
Quadratic	0.014	0.035	0.292
Cochran Q	3.71E-04	0.127	0.736

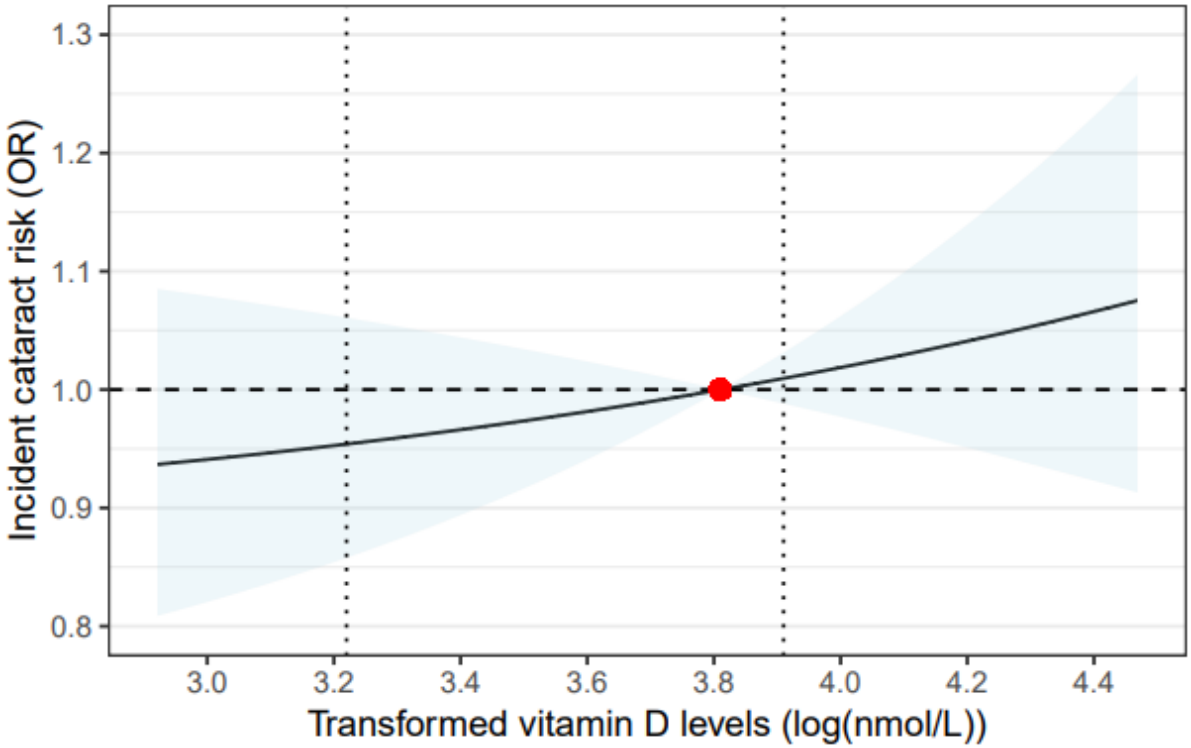
Supplementary Analysis Figure 4.1.1

Residual non-linear MR, incident cataract risk (OR) versus vitamin D levels (nmol/L).



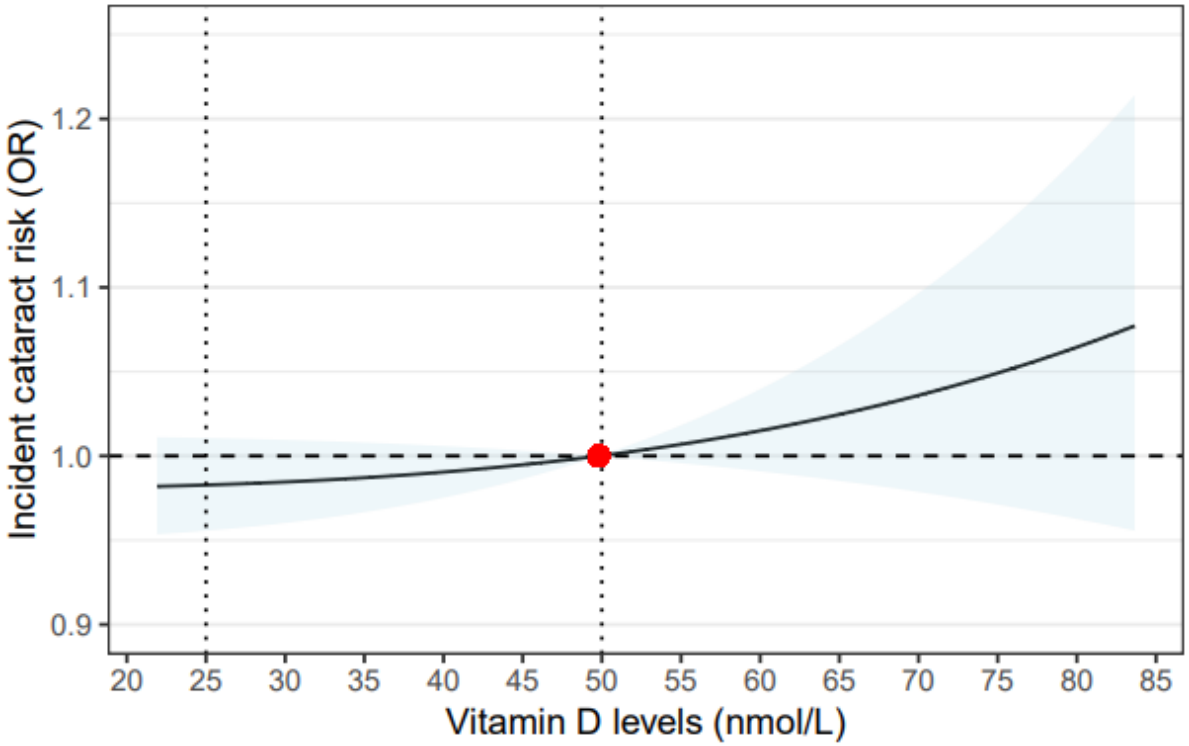
Supplementary Analysis Figure 4.1.2

Log-transformed non-linear MR, incident cataract risk (OR) versus vitamin D levels (log(nmol/L)).



Supplementary Analysis Figure 4.1.3

Double-ranked non-linear MR, incident cataract risk (OR) versus vitamin D levels (nmol/L).



Supplementary Table 4.1.1

Full MR results across all methods for vitamin D levels from SUNLIGHT consortium Jiang *et al.* (2018) and UKB cataract.

MR Method	No. of SNPs	Beta	SE	P	Lower CI (95%)	Upper CI (95%)	OR	OR Lower CI (95%)	OR Upper CI (95%)
MR Egger	7	0.233	0.128	0.127	-0.017	0.484	1.263	0.983	1.623
Weighted median	7	0.141	0.082	0.087	-0.021	0.302	1.151	0.980	1.353
Inverse variance weighted	7	0.115	0.075	0.125	-0.032	0.263	1.122	0.968	1.301
Simple mode	7	0.065	0.147	0.674	-0.224	0.354	1.067	0.800	1.425
Weighted mode	7	0.144	0.084	0.138	-0.021	0.310	1.155	0.979	1.363

Supplementary Table 4.1.2

Full MR results across all methods for vitamin D deficiency and UKB cataract.

MR Method	No. of SNPs	Beta	SE	P	Lower CI (95%)	Upper CI (95%)	OR	OR Lower CI (95%)	OR Upper CI (95%)
MR Egger	17	-0.038	0.022	0.104	-0.081	0.005	0.963	0.922	1.005
Weighted median	17	-0.028	0.018	0.119	-0.063	0.007	0.973	0.939	1.007
Inverse variance weighted	17	-0.014	0.014	0.344	-0.042	0.014	0.987	0.959	1.015
Simple mode	17	0.048	0.038	0.228	-0.027	0.123	1.049	0.973	1.130
Weighted mode	17	-0.025	0.017	0.151	-0.058	0.008	0.975	0.943	1.008

Supplementary Table 4.1.3

Full pleiotropy results using MR Egger intercept for vitamin D levels and deficiency with UKB cataract.

Exposure	MR Egger Intercept	SE	P
Vitamin D levels	-0.006	0.005	0.305
Vitamin D deficiency	0.005	0.004	0.163

Supplementary Table 4.2.1

Full MR results across all methods for vitamin D levels from SUNLIGHT consortium Jiang *et al.* (2018) and multi-ethnic cataract.

MR Method	No. of SNPs	Beta	SE	P	Lower CI (95%)	Upper CI (95%)	OR	OR Lower CI (95%)	OR Upper CI (95%)
MR Egger	7	0.094	0.114	0.445	-0.129	0.318	1.099	0.879	1.374
Weighted median	7	0.079	0.077	0.306	-0.072	0.230	1.082	0.931	1.258
Inverse variance weighted	7	0.093	0.067	0.165	-0.038	0.224	1.097	0.963	1.251
Simple mode	7	0.065	0.137	0.653	-0.204	0.334	1.067	0.816	1.396
Weighted mode	7	0.076	0.079	0.376	-0.079	0.231	1.079	0.924	1.259

Supplementary Table 4.2.2

Full MR results across all methods for vitamin D deficiency and multi-ethnic cataract.

MR Method	No. of SNPs	Beta	SE	P	Lower CI (95%)	Upper CI (95%)	OR	OR Lower CI (95%)	OR Upper CI (95%)
MR Egger	15	-0.028	0.019	0.180	-0.066	0.011	0.973	0.936	1.011
Weighted median	15	-0.015	0.016	0.332	-0.046	0.016	0.985	0.955	1.016
Inverse variance weighted	15	-0.012	0.013	0.361	-0.037	0.013	0.988	0.964	1.014
Simple mode	15	-0.003	0.029	0.929	-0.059	0.054	0.997	0.942	1.056
Weighted mode	15	-0.017	0.015	0.269	-0.046	0.012	0.983	0.955	1.012

Supplementary Table 4.2.3

Full pleiotropy results using MR Egger intercept for vitamin D levels and deficiency with multi-ethnic cataract.

Exposure	MR Egger Intercept	SE	P
Vitamin D levels	-0.0001	0.005	0.987
Vitamin D deficiency	0.003	0.003	0.299

Supplementary Table 4.3.1

Full MR results across all methods for vitamin D levels from Manousaki *et. al.* (2020) and UKB cataract.

MR Method	No. of SNPs	Beta	SE	P	Lower CI (95%)	Upper CI (95%)	OR	OR Lower CI (95%)	OR Upper CI (95%)
MR Egger	90	0.034	0.035	0.330	-0.034	0.102	1.035	0.966	1.108
Weighted median	90	0.073	0.035	0.038	0.004	0.142	1.076	1.004	1.153
Inverse variance weighted	90	0.013	0.027	0.617	-0.039	0.066	1.014	0.961	1.068
Simple mode	90	-0.051	0.076	0.506	-0.200	0.099	0.950	0.819	1.104
Weighted mode	90	0.035	0.028	0.208	-0.019	0.089	1.036	0.981	1.093

Supplementary Table 4.3.2

Full MR results across all methods for vitamin D levels from Manousaki *et. al.* (2020) and multi-ethnic cataract.

MR Method	No. of SNPs	Beta	SE	P	Lower CI (95%)	Upper CI (95%)	OR	OR Lower CI (95%)	OR Upper CI (95%)
MR Egger	62	0.066	0.039	0.097	-0.011	0.143	1.068	0.989	1.154
Weighted median	62	0.038	0.034	0.262	-0.029	0.105	1.039	0.972	1.111
Inverse variance weighted	62	0.026	0.031	0.396	-0.034	0.087	1.026	0.966	1.090
Simple mode	62	-0.067	0.078	0.391	-0.220	0.086	0.935	0.802	1.089
Weighted mode	62	0.045	0.026	0.095	-0.007	0.096	1.046	0.993	1.101

Supplementary Table 4.3.3

Full pleiotropy results using MR Egger intercept for vitamin D levels from Manousaki *et. al.* (2020) with UKB and multi-ethnic cataract.

Outcome	MR Egger Intercept	SE	P
UKB cataract	-0.001	0.001	0.352
Multi-ethnic cataract	-0.003	0.002	0.112

Supplementary Table 4.4.1

Full MR results across all methods for vitamin D levels from the SUNLIGHT consortium Jiang *et al.* (2018) (restricting to *DHCR7*, *CYP2R1*, *GC*, and *CYP24A1*) and UKB cataract.

MR Method	No. of SNPs	Beta	SE	P	Lower CI (95%)	Upper CI (95%)	OR	OR Lower CI (95%)	OR Upper CI (95%)
MR Egger	4	0.118	0.153	0.522	-0.182	0.417	1.125	0.834	1.518
Weighted median	4	0.153	0.083	0.063	-0.008	0.315	1.166	0.992	1.370
Inverse variance weighted	4	0.152	0.080	0.058	-0.005	0.308	1.164	0.995	1.361
Simple mode	4	0.150	0.115	0.284	-0.075	0.375	1.161	0.927	1.455
Weighted mode	4	0.152	0.087	0.177	-0.017	0.322	1.164	0.983	1.379

Supplementary Table 4.4.2

Full MR results across all methods for vitamin D levels from the SUNLIGHT consortium Jiang *et al.* (2018) (restricting to *DHCR7*, *CYP2R1*, *GC*, and *CYP24A1*) and multi-ethnic cataract.

MR Method	No. of SNPs	Beta	SE	P	Lower CI (95%)	Upper CI (95%)	OR	OR Lower CI (95%)	OR Upper CI (95%)
MR Egger	4	0.033	0.139	0.833	-0.238	0.305	1.034	0.788	1.356
Weighted median	4	0.082	0.078	0.295	-0.071	0.235	1.085	0.931	1.265
Inverse variance weighted	4	0.108	0.071	0.129	-0.031	0.247	1.114	0.969	1.280
Simple mode	4	0.071	0.131	0.624	-0.185	0.328	1.074	0.831	1.388
Weighted mode	4	0.076	0.079	0.404	-0.078	0.231	1.079	0.925	1.260

Supplementary Table 4.4.3

Full pleiotropy results using MR Egger intercept for vitamin D levels from the SUNLIGHT consortium Jiang *et al.* (2018) (restricting to *DHCR7*, *CYP2R1*, *GC*, and *CYP24A1*) with UKB and multi-ethnic cataract.

Outcome	MR Egger Intercept	SE	P
UKB cataract	0.002	0.008	0.818
Multi-ethnic cataract	0.005	0.008	0.595

Supplementary Table 5.1

Full MR results across all methods for alcohol consumption and UKB cataract.

MR Method	No. of SNPs	Beta	SE	P	Lower CI (95%)	Upper CI (95%)	OR	OR Lower CI (95%)	OR Upper CI (95%)
MR Egger	10	0.231	0.358	0.537	-0.470	0.932	1.259	0.625	2.539
Weighted median	10	0.234	0.160	0.143	-0.080	0.548	1.264	0.924	1.730
Inverse variance weighted	10	0.338	0.185	0.068	-0.024	0.701	1.403	0.976	2.016
Simple mode	10	0.080	0.270	0.774	-0.450	0.610	1.083	0.638	1.841
Weighted mode	10	0.206	0.180	0.282	-0.147	0.559	1.229	0.863	1.749

Supplementary Table 5.2

Full MR results across all methods for alcohol consumption and multi-ethnic cataract.

MR Method	No. of SNPs	Beta	SE	P	Lower CI (95%)	Upper CI (95%)	OR	OR Lower CI (95%)	OR Upper CI (95%)
MR Egger	10	0.111	0.290	0.713	-0.459	0.680	1.117	0.632	1.974
Weighted median	10	0.066	0.111	0.555	-0.152	0.283	1.068	0.859	1.327
Inverse variance weighted	10	0.104	0.175	0.552	-0.239	0.448	1.110	0.787	1.565
Simple mode	10	0.076	0.303	0.807	-0.517	0.669	1.079	0.596	1.953
Weighted mode	10	0.076	0.119	0.538	-0.157	0.309	1.079	0.855	1.362

Supplementary Table 6.1

Co-localisation results for overall cataract and type 2 diabetes showing the posterior probabilities for each hypothesis (0-4). No.SNPs represents SNPs present across both data sets, when SNPs were not present across each group they were removed.

SNP list	No. SNPs	PP H0	PP H1	PP H2	PP H3	PP H4
chr1_SNPlist_1	16	0%	0%	0%	100%	0%
chr1_SNPlist_2	27	0%	0%	100%	0%	0%
chr1_SNPlist_3	5	26%	0%	74%	0%	0%
chr1_SNPlist_4	13	1%	0%	98%	0%	1%
chr1_SNPlist_5	30	2%	0%	92%	0%	7%
chr10_SNPlist_1	25	0%	0%	8%	91%	1%
chr11_SNPlist_1	15	0%	0%	86%	0%	14%
chr11_SNPlist_2	9	0%	0%	99%	0%	1%
chr11_SNPlist_3	19	0%	0%	80%	1%	19%
chr12_SNPlist_1	13	1%	0%	98%	0%	1%
chr15_SNPlist_1	12	0%	0%	98%	0%	2%
chr15_SNPlist_2	4	0%	0%	94%	0%	6%
chr16_SNPlist_1	27	0%	0%	100%	0%	0%
chr16_SNPlist_2	26	0%	0%	0%	0%	100%
chr16_SNPlist_3	7	0%	0%	0%	0%	100%
chr16_SNPlist_4	11	27%	14%	0%	0%	59%
chr17_SNPlist_1	13	0%	0%	97%	1%	2%
chr17_SNPlist_2	1	100%	0%	0%	0%	0%
chr2_SNPlist_1	33	0%	0%	98%	0%	1%
chr2_SNPlist_2	16	0%	0%	98%	0%	1%
chr2_SNPlist_3	15	1%	0%	96%	0%	2%
chr20_SNPlist_1	25	0%	0%	99%	0%	1%
chr20_SNPlist_2	11	2%	0%	97%	0%	1%
chr3_SNPlist_1	11	0%	0%	98%	1%	2%
chr3_SNPlist_2	13	5%	0%	91%	1%	3%
chr4_SNPlist_1	8	65%	0%	35%	0%	0%
chr6_SNPlist_1	55	0%	0%	0%	29%	71%
chr6_SNPlist_2	37	0%	0%	0%	58%	42%
chr6_SNPlist_4	58	0%	0%	0%	1%	99%
chr6_SNPlist_5	1	99%	0%	1%	0%	0%
chr6_SNPlist_6	13	0%	0%	0%	1%	99%
chr6_SNPlist_9	13	1%	0%	94%	0%	5%
chr7_SNPlist_1	26	0%	0%	97%	0%	2%
chr7_SNPlist_2	8	99%	0%	1%	0%	0%
chr7_SNPlist_3	18	11%	0%	88%	0%	1%
chr9_SNPlist_1	17	1%	0%	95%	2%	1%
chr9_SNPlist_2	28	0%	0%	0%	2%	98%

Supplementary Table 6.2

Co-localisation results for asthma and diabetic cataract showing the posterior probabilities for each hypothesis (0-4). No.SNPs represents SNPs present across both data sets, when SNPs were not present across each group they were removed.

SNP list	No. SNPs	PP H0	PP H1	PP H2	PP H3	PP H4
chr1_SNPlist_1	27	82%	0%	18%	0%	0%
chr10_SNPlist_1	23	0%	0%	99%	0%	1%
chr16_SNPlist_1	15	1%	0%	98%	0%	2%
chr19_SNPlist_1	59	1%	0%	97%	0%	2%
chr19_SNPlist_2	1	5%	0%	95%	0%	1%
chr4_SNPlist_1	11	0%	0%	99%	0%	0%
chr6_SNPlist_1	332	0%	0%	0%	84%	16%
chr6_SNPlist_2	234	0%	0%	87%	9%	4%
chr6_SNPlist_4	247	0%	0%	13%	18%	69%

Supplementary Table 6.3

Co-localisation results for senile cataract and diabetic cataract showing the posterior probabilities for each hypothesis (0-4). No.SNPs represents SNPs present across both data sets, when SNPs were not present across each group they were removed.

SNP list	No. SNPs	PP H0	PP H1	PP H2	PP H3	PP H4
chr1_SNPlist_1	60	0%	79%	0%	0%	21%
chr1_SNPlist_2	20	2%	96%	0%	0%	2%
chr10_SNPlist_1	46	2%	96%	0%	0%	2%
chr11_SNPlist_1	35	0%	67%	0%	0%	33%
chr11_SNPlist_2	39	0%	40%	0%	0%	60%
chr11_SNPlist_3	28	0%	96%	0%	0%	4%
chr15_SNPlist_1	29	0%	60%	0%	1%	38%
chr2_SNPlist_1	25	1%	97%	0%	0%	2%
chr2_SNPlist_2	55	4%	89%	0%	0%	7%
chr20_SNPlist_1	31	0%	83%	0%	0%	17%
chr3_SNPlist_1	23	0%	98%	0%	0%	2%
chr7_SNPlist_1	28	0%	94%	0%	0%	6%
chr7_SNPlist_2	25	1%	98%	0%	0%	1%
chr7_SNPlist_3	7	0%	83%	0%	0%	17%

Supplementary Table 6.4

Co-localisation results for asthma and overall cataract showing the posterior probabilities for each hypothesis (0-4). No.SNPs represents SNPs present across both data sets, when SNPs were not present across each group they were removed.

SNP list	No. SNPs	PP H0	PP H1	PP H2	PP H3	PP H4
chr1_SNPlist_1	19	0%	0%	99%	0%	1%
chr1_SNPlist_2	24	90%	0%	10%	0%	0%
chr1_SNPlist_3	8	38%	0%	61%	0%	1%
chr1_SNPlist_4	15	99%	0%	1%	0%	0%
chr1_SNPlist_5	32	2%	0%	97%	0%	1%
chr10_SNPlist_1	30	0%	0%	99%	0%	1%
chr11_SNPlist_1	21	0%	0%	99%	0%	1%
chr11_SNPlist_2	10	0%	0%	98%	0%	2%
chr11_SNPlist_3	25	0%	0%	98%	0%	1%
chr12_SNPlist_1	16	1%	0%	96%	0%	2%
chr15_SNPlist_1	13	0%	0%	95%	0%	5%
chr15_SNPlist_2	1	100%	0%	0%	0%	0%
chr16_SNPlist_1	30	0%	0%	99%	0%	1%
chr16_SNPlist_2	32	0%	0%	99%	0%	1%
chr16_SNPlist_3	14	4%	0%	95%	0%	1%
chr16_SNPlist_4	28	7%	0%	89%	0%	4%
chr17_SNPlist_1	39	2%	0%	97%	0%	1%
chr17_SNPlist_2	20	0%	0%	99%	0%	1%
chr2_SNPlist_1	33	0%	0%	96%	0%	4%
chr2_SNPlist_2	13	93%	0%	7%	0%	0%
chr2_SNPlist_3	23	82%	0%	18%	0%	0%
chr20_SNPlist_1	28	0%	0%	99%	0%	1%
chr20_SNPlist_2	21	99%	0%	1%	0%	0%
chr3_SNPlist_1	16	100%	0%	0%	0%	0%
chr3_SNPlist_2	21	5%	0%	94%	0%	1%
chr4_SNPlist_1	9	1%	0%	97%	0%	2%
chr6_SNPlist_1	361	0%	0%	51%	47%	2%
chr6_SNPlist_2	93	0%	0%	35%	44%	21%
chr6_SNPlist_4	354	0%	0%	0%	4%	96%
chr6_SNPlist_5	126	5%	1%	30%	6%	58%
chr6_SNPlist_6	46	0%	0%	22%	13%	65%
chr6_SNPlist_7	141	1%	0%	94%	2%	3%
chr6_SNPlist_9	15	3%	0%	96%	0%	1%
chr7_SNPlist_1	27	98%	0%	2%	0%	0%
chr7_SNPlist_2	9	0%	0%	98%	0%	2%
chr7_SNPlist_3	23	11%	0%	88%	0%	1%
chr9_SNPlist_1	23	64%	0%	35%	0%	1%
chr9_SNPlist_2	18	3%	0%	95%	0%	1%

9 Bibliography

Abd ElHafeez, S. *et al.* (2021) 'Methods to Analyze Time-to-Event Data: The Cox Regression Analysis', *Oxidative Medicine and Cellular Longevity*, 2021, p. 1302811. Available at: https://doi.org/10.1155/2021/1302811.

Abdellah, M.M. *et al.* (2019) 'Association of Serum 25-Hydroxyl Vitamin D Deficiency and Age-Related Cataract: A Case-Control Study', *Journal of Ophthalmology*, 2019, p. e9312929. Available at: https://doi.org/10.1155/2019/9312929.

Abu-Amero, K.K., Kondkar, A.A. and Chalam, K.V. (2016) 'Resveratrol and Ophthalmic Diseases', *Nutrients*, 8(4), p. 200. Available at: https://doi.org/10.3390/nu8040200.

Agha, M. and Agha, R. (2017) 'The rising prevalence of obesity: part A: impact on public health', *International Journal of Surgery. Oncology*, 2(7), p. e17. Available at: https://doi.org/10.1097/IJ9.0000000000000017.

Aina, F.O. *et al.* (2006) 'Hormone replacement therapy and cataract: a population-based case-control study', *Eye (London, England)*, 20(4), pp. 417–422. Available at: https://doi.org/10.1038/sj.eye.6701877.

Albert, D.M. and Gamm, D.M. (2024) *Cataract | Causes, Symptoms & Treatment | Britannica*. Available at: https://www.britannica.com/science/cataract-eye (Accessed: 21 December 2024).

Allen, N. *et al.* (2012) 'UK Biobank: Current status and what it means for epidemiology', *Health Policy and Technology*, 1(3), pp. 123–126. Available at: https://doi.org/10.1016/j.hlpt.2012.07.003.

Altshuler, D., Donnelly, P., and The International HapMap Consortium (2005) 'A haplotype map of the human genome', *Nature*, 437(7063), pp. 1299–1320. Available at: https://doi.org/10.1038/nature04226.

American Diabetes Association (2010) 'Diagnosis and Classification of Diabetes Mellitus', *Diabetes Care*, 33(Suppl 1), pp. S62–S69. Available at: https://doi.org/10.2337/dc10-S062.

Amin, H.A. and Drenos, F. (2021) 'No evidence that vitamin D is able to prevent or affect the severity of COVID-19 in individuals with European ancestry: a Mendelian randomisation study of open data', *BMJ Nutrition, Prevention & Health*, p. bmjnph. Available at: https://doi.org/10.1136/bmjnph-2020-000151.

Ang, M.J. and Afshari, N.A. (2021) 'Cataract and systemic disease: A review', *Clinical & Experimental Ophthalmology*, 49(2), pp. 118–127. Available at: https://doi.org/10.1111/ceo.13892.

Armstrong, R.A. (2014) 'When to use the Bonferroni correction', *Ophthalmic & Physiological Optics: The Journal of the British College of Ophthalmic Opticians (Optometrists)*, 34(5), pp. 502–508. Available at: https://doi.org/10.1111/opo.12131.

Arranz, S. *et al.* (2012) 'Wine, Beer, Alcohol and Polyphenols on Cardiovascular Disease and Cancer', *Nutrients*, 4(7), pp. 759–781. Available at: https://doi.org/10.3390/nu4070759.

Awidi, A.A. *et al.* (2024) 'The Effect of Racial, Ethnic, and Socioeconomic Differences on Visual Impairment before Cataract Surgery', *Ophthalmology*, 132(1), pp. 98–107. Available at: https://doi.org/10.1016/j.ophtha.2024.07.021.

Balashova, L.M. *et al.* (2018) 'Lanomax as a Drug in Cataract Treatment: A Case Study', *Biophysics*, 63(4), pp. 655–661. Available at: https://doi.org/10.1134/S0006350918040024.

Beard, E. *et al.* (2019) 'Associations between socio-economic factors and alcohol consumption: A population survey of adults in England', *PLoS ONE*, 14(2), p. e0209442. Available at: https://doi.org/10.1371/journal.pone.0209442.

Becker, C. *et al.* (2018) 'Cataract in patients with diabetes mellitus—incidence rates in the UK and risk factors', *Eye*, 32(6), pp. 1028–1035. Available at: https://doi.org/10.1038/s41433-017-0003-1.

Beebe, D.C., Holekamp, N.M. and Shui, Y.-B. (2010) 'Oxidative Damage and the Prevention of Age-Related Cataracts', *Ophthalmic Research*, 44(3), pp. 155–165. Available at: https://doi.org/10.1159/000316481.

Behndig, A. *et al.* (2023) 'Annual Report 2022 Based on Data from the European Registry of Quality Outcomes for Cataract and Refractive Surgery'.

Bell, S.J. *et al.* (2020) 'Congenital cataract: a guide to genetic and clinical management', *Therapeutic Advances in Rare Disease*, 1, p. 2633004020938061. Available at: https://doi.org/10.1177/2633004020938061.

Benjamin, E., Burns, B.S. and Proctor, W.R. (2013) 'Tobacco Smoke Exposure Greatly Increases Alcohol Consumption in C57BL/6 Mice', *Alcoholism, clinical and experimental research*, 37(0 1), pp. E364–E372. Available at: https://doi.org/10.1111/j.1530-0277.2012.01911.x.

Berkowitz, S.T. *et al.* (2024) 'Ophthalmology Workforce Projections in the United States, 2020 to 2035', *Ophthalmology*, 131(2), pp. 133–139. Available at: https://doi.org/10.1016/j.ophtha.2023.09.018.

Berulava, T. and Horsthemke, B. (2010) 'The obesity-associated SNPs in intron 1 of the FTO gene affect primary transcript levels', *European Journal of Human Genetics*, 18(9), pp. 1054–1056. Available at: https://doi.org/10.1038/ejhg.2010.71.

Bhargava, R. *et al.* (2015) 'Phacoemulsification versus small incision cataract surgery in patients with uveitis', *International Journal of Ophthalmology*, 8(5), pp. 965–970. Available at: https://doi.org/10.3980/j.issn.2222-3959.2015.05.20.

Bhatti, S.N. *et al.* (2020) 'Exploration of Alcohol Consumption Behaviours and Health-Related Influencing Factors of Young Adults in the UK', *International Journal of Environmental Research and Public Health*, 17(17), p. 6282. Available at: https://doi.org/10.3390/ijerph17176282.

Bhide, A., Shah, P.S. and Acharya, G. (2018) 'A simplified guide to randomized controlled trials', *Acta Obstetricia et Gynecologica Scandinavica*, 97(4), pp. 380–387. Available at: https://doi.org/10.1111/aogs.13309.

Bixler, J.E. (2019) 'Cataracts and Their Treatment in People with Diabetes', in *Prevention and Management of Diabetes-Related Eye Disease*. Arlington (VA): American Diabetes

Association (ADA Clinical Compendia Series). Available at: http://www.ncbi.nlm.nih.gov/books/NBK544518/ (Accessed: 16 March 2025).

Black, R.J. *et al.* (2016) 'The Association between Systemic Glucocorticoid Use and the Risk of Cataract and Glaucoma in Patients with Rheumatoid Arthritis: A Systematic Review and Meta-Analysis', *PLoS ONE*, 11(11), p. e0166468. Available at: https://doi.org/10.1371/journal.pone.0166468.

Blair, K. *et al.* (2024) 'Amblyopia', in *StatPearls*. Treasure Island (FL): StatPearls Publishing. Available at: http://www.ncbi.nlm.nih.gov/books/NBK430890/ (Accessed: 21 December 2024).

Boehm, F.J. and Zhou, X. (2022) 'Statistical methods for Mendelian randomization in genome-wide association studies: A review', *Computational and Structural Biotechnology Journal*, 20, p. 2338. Available at: https://doi.org/10.1016/j.csbj.2022.05.015.

Bourne, R. *et al.* (2006) 'Outcomes of cataract surgery in Pakistan: results from The Pakistan National Blindness and Visual Impairment Survey', *The British Journal of Ophthalmology*, 91(4), pp. 420–426. Available at: https://doi.org/10.1136/bjo.2006.106724.

Bourne, R.R.A. *et al.* (2003) 'Outcomes of cataract surgery in Bangladesh: results from a population based nationwide survey', *The British Journal of Ophthalmology*, 87(7), pp. 813–819.

Bowden, J. *et al.* (2016) 'Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator', *Genetic Epidemiology*, 40(4), pp. 304–314. Available at: https://doi.org/10.1002/gepi.21965.

Bro, T. *et al.* (2023) 'Two point four million cataract surgeries: 30 years with the Swedish National Cataract Register, 1992-2021', *Journal of Cataract & Refractive Surgery*, 49(8), p. 879. Available at: https://doi.org/10.1097/j.jcrs.0000000000001209.

Brown, C.J. and Akaichi, F. (2015) 'Vitamin D deficiency and posterior subcapsular cataract', *Clinical Ophthalmology*, 9, pp. 1093–1098. Available at: https://doi.org/10.2147/OPTH.S84790.

Bulik-Sullivan, B. *et al.* (2015) 'An atlas of genetic correlations across human diseases and traits', *Nature Genetics*, 47(11), pp. 1236–1241. Available at: https://doi.org/10.1038/ng.3406.

Bulik-Sullivan, B.K. *et al.* (2015) 'LD Score regression distinguishes confounding from polygenicity in genome-wide association studies', *Nature Genetics*, 47(3), pp. 291–295. Available at: https://doi.org/10.1038/ng.3211.

Burgess, S. *et al.* (2020) 'A robust and efficient method for Mendelian randomization with hundreds of genetic variants', *Nature Communications*, 11(1), p. 376. Available at: https://doi.org/10.1038/s41467-019-14156-4.

Burgess, S. *et al.* (2023) 'Guidelines for performing Mendelian randomization investigations: update for summer 2023', *Wellcome Open Research*, 4, p. 186. Available at: https://doi.org/10.12688/wellcomeopenres.15555.3.

Burgess, S. (2023) 'Violation of the Constant Genetic Effect Assumption Can Result in Biased Estimates for Non-Linear Mendelian Randomization', *Human Heredity*, 88(1), pp. 79–90. Available at: https://doi.org/10.1159/000531659.

Burgess, S., Butterworth, A. and Thompson, S.G. (2013) 'Mendelian Randomization Analysis With Multiple Genetic Variants Using Summarized Data', *Genetic Epidemiology*, 37(7), pp. 658–665. Available at: https://doi.org/10.1002/gepi.21758.

Burgess, S. and Thompson, S.G. (2017) 'Interpreting findings from Mendelian randomization using the MR-Egger method', *European Journal of Epidemiology*, 32(5), pp. 377–389. Available at: https://doi.org/10.1007/s10654-017-0255-x.

Bycroft, C. *et al.* (2018) 'The UK Biobank resource with deep phenotyping and genomic data', *Nature*, 562(7726), pp. 203–209. Available at: https://doi.org/10.1038/s41586-018-0579-z.

Cashman, K.D. *et al.* (2016) 'Vitamin D deficiency in Europe: pandemic?', *The American Journal of Clinical Nutrition*, 103(4), pp. 1033–1044. Available at: https://doi.org/10.3945/ajcn.115.120873.

Caturano, A. *et al.* (2023) 'Oxidative Stress in Type 2 Diabetes: Impacts from Pathogenesis to Lifestyle Modifications', *Current Issues in Molecular Biology*, 45(8), pp. 6651–6666. Available at: https://doi.org/10.3390/cimb45080420.

Chakraborty, R., Read, S.A. and Vincent, S.J. (2020) 'Understanding Myopia: Pathogenesis and Mechanisms', in M. Ang and T.Y. Wong (eds) *Updates on Myopia: A Clinical Perspective*. Singapore: Springer, pp. 65–94. Available at: https://doi.org/10.1007/978-981-13-8491-2_4.

Chan, E., Mahroo, O.A.R. and Spalton, D.J. (2010) 'Complications of cataract surgery', *Clinical & Experimental Optometry*, 93(6), pp. 379–389. Available at: https://doi.org/10.1111/j.1444-0938.2010.00516.x.

Chan, H.-N. *et al.* (2022) 'Vitamin D and Ocular Diseases: A Systematic Review', *International Journal of Molecular Sciences*, 23(8), p. 4226. Available at: https://doi.org/10.3390/ijms23084226.

Chang, C. *et al.* (2016) 'A Genome-Wide Association Study Provides New Evidence That CACNA1C Gene is Associated With Diabetic Cataract', *Investigative Ophthalmology & Visual Science*, 57(4), pp. 2246–2250. Available at: https://doi.org/10.1167/iovs.16-19332.

Chang, C.C. *et al.* (2015) 'Second-generation PLINK: rising to the challenge of larger and richer datasets', *GigaScience*, 4(1), pp. s13742-015-0047–8. Available at: https://doi.org/10.1186/s13742-015-0047-8.

Chang, S.-W. and Lee, H.-C. (2019) 'Vitamin D and health - The missing vitamin in humans', *Pediatrics & Neonatology*, 60(3), pp. 237–244. Available at: https://doi.org/10.1016/j.pedneo.2019.04.007.

Chang, Y. -m. *et al.* (2009) 'Sun exposure and melanoma risk at different latitudes: a pooled analysis of 5700 cases and 7216 controls', *International Journal of Epidemiology*, 38(3), pp. 814–830. Available at: https://doi.org/10.1093/ije/dyp166.

CHEC (2024) *CHEC launches premium NHS EDoF lens solution to cataract patients*, *CHEC*. Available at: https://chec.uk/chec-launches-premium-nhs-edof-lens-solution-to-cataract-patients/ (Accessed: 8 January 2025).

Chen, X. *et al.* (2020) 'Prevalence and risk factors on age-related cataract and surgery in adults over 50 years old in Binhu District, Wuxi, China', *International Journal of Ophthalmology*, 13(3), pp. 445–451. Available at: https://doi.org/10.18240/ijo.2020.03.12.

Chen, Xinyi *et al.* (2021) 'Cataract: Advances in surgery and whether surgery remains the only treatment in future', *Advances in Ophthalmology Practice and Research*, 1(1), p. 100008. Available at: https://doi.org/10.1016/j.aopr.2021.100008.

Chen, X.-J. *et al.* (2018) 'Lanosterol and 25-hydroxycholesterol dissociate crystallin aggregates isolated from cataractous human lens via different mechanisms', *Biochemical and Biophysical Research Communications*, 506(4), pp. 868–873. Available at: https://doi.org/10.1016/j.bbrc.2018.10.175.

Choquet, H. *et al.* (2021) 'A large multiethnic GWAS meta-analysis of cataract identifies new risk loci and sex-specific effects', *Nature Communications*, 12(1), p. 3595. Available at: https://doi.org/10.1038/s41467-021-23873-8.

Chu, L. *et al.* (2020) 'Vibration of effects in epidemiologic studies of alcohol consumption and breast cancer risk', *International Journal of Epidemiology*, 49(2), pp. 608–618. Available at: https://doi.org/10.1093/ije/dyz271.

Chua, J. *et al.* (2015) 'Ancestry, Socioeconomic Status, and Age-Related Cataract in Asians: The Singapore Epidemiology of Eye Diseases Study', *Ophthalmology*, 122(11), pp. 2169–2178. Available at: https://doi.org/10.1016/j.ophtha.2015.06.052.

Chua, S.Y. *et al.* (2021) 'Alcohol Consumption and Incident Cataract Surgery in Two Large UK Cohorts', *Ophthalmology*, 128(6), pp. 837–847. Available at: https://doi.org/10.1016/j.ophtha.2021.02.007.

Chua, S.Y.L. *et al.* (2021) 'The Association of Ambient Air Pollution With Cataract Surgery in UK Biobank Participants: Prospective Cohort Study', *Investigative Ophthalmology & Visual Science*, 62(15), p. 7. Available at: https://doi.org/10.1167/iovs.62.15.7.

Chugh, P.K. and Dabas, A. (2021) 'Price Dispersion of Vitamin D Supplements Over Time: An Initiative for Prescriber Education', *Indian Journal of Endocrinology and Metabolism*, 25(2), pp. 142–147. Available at: https://doi.org/10.4103/ijem.ijem_159_21.

Cicinelli, M.V. *et al.* (2023) 'Cataracts', *The Lancet*, 401(10374), pp. 377–389. Available at: https://doi.org/10.1016/S0140-6736(22)01839-6.

Coleman, J.R.I. *et al.* (2016) 'Quality control, imputation and analysis of genome-wide genotyping data from the Illumina HumanCoreExome microarray', *Briefings in Functional Genomics*, 15(4), pp. 298–304. Available at: https://doi.org/10.1093/bfgp/elv037.

Cook, C. (2010) 'Mode of administration bias', *The Journal of Manual & Manipulative Therapy*, 18(2), pp. 61–63. Available at: https://doi.org/10.1179/106698110X12640740712617.

Costenbader, K.H. (2022) 'Vitamin D and fish oil supplements and risk of autoimmune disease', *BMJ*, 376, p. o243. Available at: https://doi.org/10.1136/bmj.o243.

Crux, N.B. and Elahi, S. (2017) 'Human Leukocyte Antigen (HLA) and Immune Regulation: How Do Classical and Non-Classical HLA Alleles Modulate Immune Response to Human Immunodeficiency Virus and Hepatitis C Virus Infections?', *Frontiers in Immunology*, 8, p. 832. Available at: https://doi.org/10.3389/fimmu.2017.00832.

Cui, A. *et al.* (2023) ‘Global and regional prevalence of vitamin D deficiency in population-based studies from 2000 to 2022: A pooled analysis of 7.9 million participants’, *Frontiers in Nutrition*, 10. Available at: https://www.frontiersin.org/articles/10.3389/fnut.2023.1070808 (Accessed: 10 September 2023).

Dan, L. *et al.* (2022) ‘Nonlinear Association between Serum 25-Hydroxyvitamin D and All-Cause Mortality in Adults with Inflammatory Bowel Disease in a Prospective Cohort Study’, *The Journal of Nutrition*, 152(9), pp. 2125–2134. Available at: https://doi.org/10.1093/jn/nxac148.

Danysh, B.P. and Duncan, M.K. (2009) ‘The Lens Capsule’, *Experimental eye research*, 88(2), pp. 151–164. Available at: https://doi.org/10.1016/j.exer.2008.08.002.

Daszynski, D.M. *et al.* (2019) ‘Failure of Oxysterols Such as Lanosterol to Restore Lens Clarity from Cataracts’, *Scientific Reports*, 9(1), p. 8459. Available at: https://doi.org/10.1038/s41598-019-44676-4.

Davies, N.M., Holmes, M.V. and Smith, G.D. (2018) ‘Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians’, *BMJ*, 362, p. k601. Available at: https://doi.org/10.1136/bmj.k601.

Davis, G. (2016) ‘The Evolution of Cataract Surgery’, *Missouri Medicine*, 113(1), pp. 58–62.

Day, A.C. *et al.* (2015) ‘The Royal College of Ophthalmologists’ National Ophthalmology Database study of cataract surgery: report 1, visual outcomes and complications’, *Eye*, 29(4), pp. 552–560. Available at: https://doi.org/10.1038/eye.2015.3.

Dayimu, A. (2024) ‘forestploter: Create a Flexible Forest Plot’. Available at: https://cran.r-project.org/web/packages/forestploter/index.html (Accessed: 23 July 2024).

Delcourt, C. *et al.* (2014) ‘Lifetime Exposure to Ambient Ultraviolet Radiation and the Risk for Cataract Extraction and Age-Related Macular Degeneration: The Alienor Study’, *Investigative Ophthalmology & Visual Science*, 55(11), pp. 7619–7627. Available at: https://doi.org/10.1167/iovs.14-14471.

Delgado-Rodríguez, M. and Llorca, J. (2004) ‘Bias’, *Journal of Epidemiology & Community Health*, 58(8), pp. 635–641. Available at: https://doi.org/10.1136/jech.2003.008466.

Diabetes UK (2017) *The future of diabetes*, *Diabetes UK*. Available at: https://www.diabetes.org.uk/support-us/campaign/other-campaigns/the-future-of-diabetes (Accessed: 21 December 2024).

Diaz-Torres, S. *et al.* (2024) ‘Uncovering genetic loci and biological pathways associated with age-related cataracts through GWAS meta-analysis’, *Nature Communications*, 15(1), p. 9116. Available at: https://doi.org/10.1038/s41467-024-53212-6.

Diego Zamora-de la Cruz *et al.* (2016) *Management of Traumatic Cataract*, *American Academy of Ophthalmology*. Available at: https://www.aao.org/eyenet/article/management-of-traumatic-cataract (Accessed: 29 November 2021).

Donachie, P.H.J. *et al.* (2024) *National Ophthalmology Database Audit - Seventh Annual Report of the National Cataract Audit - National Cataract Audit for the 2022 NHS Year: 01 April 2022 to 31 March 2023*. 7. The Royal College of Ophthalmologists, p. 124. Available at: https://nodaudit.org.uk/sites/default/files/2024-05/NOD%20Cataract%20Audit%207th%20Annual%20Report%202024_1.pdf.

D'Orazio, J. *et al.* (2013) 'UV Radiation and the Skin', *International Journal of Molecular Sciences*, 14(6), pp. 12222–12248. Available at: https://doi.org/10.3390/ijms140612222.

Drack, A.V. (2005) 'Pediatric Ophthalmology*', in D.A. Palay and J.H. Krachmer (eds) *Primary Care Ophthalmology (Second Edition)*. Philadelphia: Mosby, pp. 229–273. Available at: https://doi.org/10.1016/B978-032303316-9.50019-X.

Du, Y. *et al.* (2018) 'Traumatic Cataract in Children in Eastern China: Shanghai Pediatric Cataract Study', *Scientific Reports*, 8(1), p. 2588. Available at: https://doi.org/10.1038/s41598-018-20982-1.

Dubois, V.D. and Bastawrous, A. (2017) 'N-acetylcarnosine (NAC) drops for age-related cataract', *The Cochrane Database of Systematic Reviews*, 2017(2), p. CD009493. Available at: https://doi.org/10.1002/14651858.CD009493.pub2.

Ericson-Neilsen, W. and Kaye, A.D. (2014) 'Steroids: Pharmacology, Complications, and Practice Delivery Issues', *The Ochsner Journal*, 14(2), pp. 203–207.

Falkowska, M. *et al.* (2023) 'Influence of Diet, Dietary Products and Vitamins on Age-Related Cataract Incidence: A Systematic Review', *Nutrients*, 15(21), p. 4585. Available at: https://doi.org/10.3390/nu15214585.

Fan, H. *et al.* (2023) 'Fruit and vegetable intake and the risk of cataract: insights from the UK Biobank study', *Eye*, 37(15), pp. 3234–3242. Available at: https://doi.org/10.1038/s41433-023-02498-9.

Fan, X. *et al.* (2020) 'Vitamin C is a source of oxoaldehyde and glycative stress in age-related cataract and neurodegenerative diseases', *Aging Cell*, 19(7), p. e13176. Available at: https://doi.org/10.1111/acel.13176.

Ference, B.A. *et al.* (2019) 'Mendelian Randomization Study of *ACLY* and Cardiovascular Disease', *New England Journal of Medicine*, 380(11), pp. 1033–1042. Available at: https://doi.org/10.1056/NEJMoa1806747.

Fernández, J. *et al.* (2018) 'From Presbyopia to Cataracts: A Critical Review on Dysfunctional Lens Syndrome', *Journal of Ophthalmology*, 2018(1), p. 4318405. Available at: https://doi.org/10.1155/2018/4318405.

Finan, C. *et al.* (2017) 'The druggable genome and support for target identification and validation in drug development', *Science translational medicine*, 9(383), p. eaag1166. Available at: https://doi.org/10.1126/scitranslmed.aag1166.

Flaxman, S.R. *et al.* (2017) 'Global causes of blindness and distance vision impairment 1990-2020: a systematic review and meta-analysis', *The Lancet. Global Health*, 5(12), pp. e1221–e1234. Available at: https://doi.org/10.1016/S2214-109X(17)30393-5.

Fong, K.C. (2008) '1 - The ageing lens and classification of cataracts', in R. Malhotra (ed.) *Cataract*. Edinburgh: Butterworth-Heinemann, pp. 1–15. Available at: https://doi.org/10.1016/B978-0-08-044977-7.50006-9.

Fox, J.C. *et al.* (2024) 'Identification of pathogenic genetic variants in patients with acquired early-onset bilateral cataracts using next-generation sequencing', *Journal of American Association for Pediatric Ophthalmology and Strabismus*, 28(1), p. 103808. Available at: https://doi.org/10.1016/j.jaapos.2023.11.011.

Frampton, G. *et al.* (2014) 'The clinical effectiveness and cost-effectiveness of second-eye cataract surgery: a systematic review and economic evaluation', *Health Technology Assessment (Winchester, England)*, 18(68), pp. 1–205, v–vi. Available at: https://doi.org/10.3310/hta18680.

Friligkou, E. *et al.* (2024) 'Gene discovery and biological insights into anxiety disorders from a large-scale multi-ancestry genome-wide association study', *Nature Genetics* [Preprint]. Available at: https://doi.org/10.1038/s41588-024-01908-2.

Fry, A. *et al.* (2017) 'Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants With Those of the General Population', *American Journal of Epidemiology*, 186(9), pp. 1026–1034. Available at: https://doi.org/10.1093/aje/kwx246.

Fukai, K. *et al.* (2022) 'Alcohol use patterns and risk of incident cataract surgery: a large scale case–control study in Japan', *Scientific Reports*, 12, p. 20142. Available at: https://doi.org/10.1038/s41598-022-24465-2.

GBD 2016 Alcohol Collaborators (2018) 'Alcohol use and burden for 195 countries and territories, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016', *Lancet (London, England)*, 392(10152), pp. 1015–1035. Available at: https://doi.org/10.1016/S0140-6736(18)31310-2.

Giambartolomei, C. *et al.* (2014) 'Bayesian test for colocalisation between pairs of genetic association studies using summary statistics', *PLoS genetics*, 10(5), p. e1004383. Available at: https://doi.org/10.1371/journal.pgen.1004383.

Gil, Á., Plaza-Diaz, J. and Mesa, M.D. (2018) 'Vitamin D: Classic and Novel Actions', *Annals of Nutrition and Metabolism*, 72(2), pp. 87–95. Available at: https://doi.org/10.1159/000486536.

Glass, T.A. *et al.* (2013) 'Causal Inference in Public Health', *Annual review of public health*, 34, pp. 61–75. Available at: https://doi.org/10.1146/annurev-publhealth-031811-124606.

Gogate, P. (2010) 'Comparison of various techniques for cataract surgery, their efficacy, safety, and cost', *Oman Journal of Ophthalmology*, 3(3), pp. 105–106. Available at: https://doi.org/10.4103/0974-620X.71880.

Gong, Y. *et al.* (2015) 'Different Amounts of Alcohol Consumption and Cataract: A Meta-analysis', *Optometry and Vision Science*, 92(4), p. 471. Available at: https://doi.org/10.1097/OPX.0000000000000558.

Goyal, M.M. *et al.* (2009) 'Effect of vitamin C and E activity on surgically removed cataractous human lens epithelium cells', *Indian Journal of Clinical Biochemistry*, 24(4), pp. 375–380. Available at: https://doi.org/10.1007/s12291-009-0068-0.

Graw, J. (2019) 'Mouse models for microphthalmia, anophthalmia and cataracts', *Human Genetics*, 138(8), pp. 1007–1018. Available at: https://doi.org/10.1007/s00439-019-01995-w.

Hamilton, F.W. *et al.* (2024) 'Non-linear Mendelian randomization: detection of biases using negative controls with a focus on BMI, Vitamin D and LDL cholesterol', *European Journal of Epidemiology*, 39(5), pp. 451–465. Available at: https://doi.org/10.1007/s10654-024-01113-9.

Hammond, C.J. *et al.* (2001) 'The Heritability of Age-Related Cortical Cataract: The Twin Eye Study', *Investigative Ophthalmology & Visual Science*, 42(3), pp. 601–605.

Han, X. *et al.* (2020) 'Are smoking intensity and cessation related to cataract surgical risk in diabetic patients? Findings from the 45 and Up Study', *Eye*, 34(2), pp. 383–391. Available at: https://doi.org/10.1038/s41433-019-0550-8.

Harahap, J. and Rania, R. (2019) 'Cataracts Risk Factors and Comparison of Blood Glucose Levels in Diabetic and Non-Diabetic Patients towards the Occurrence of Cataracts', *Open Access Macedonian Journal of Medical Sciences*, 7(20), p. 3359. Available at: https://doi.org/10.3889/oamjms.2019.422.

Hariton, E. and Locascio, J.J. (2018) 'Randomised controlled trials—the gold standard for effectiveness research', *BJOG : an international journal of obstetrics and gynaecology*, 125(13), p. 1716. Available at: https://doi.org/10.1111/1471-0528.15199.

Hashemi, H. *et al.* (2020) 'Global and regional prevalence of age-related cataract: a comprehensive systematic review and meta-analysis', *Eye*, 34(8), pp. 1357–1370. Available at: https://doi.org/10.1038/s41433-020-0806-3.

Hashimi, M. *et al.* (2024) 'Using genetics to investigate the association between lanosterol and cataract', *Frontiers in Genetics*, 15. Available at: https://doi.org/10.3389/fgene.2024.1231521.

Hejtmancik, J.F. (2008) 'Congenital Cataracts and their Molecular Genetics', *Seminars in cell & developmental biology*, 19(2), pp. 134–149. Available at: https://doi.org/10.1016/j.semcdb.2007.10.003.

Hemani, G. *et al.* (2018) 'The MR-Base platform supports systematic causal inference across the human phenome', *eLife*. Edited by R. Loos, 7, p. e34408. Available at: https://doi.org/10.7554/eLife.34408.

Ho, F.K. *et al.* (2020) 'Modifiable and non-modifiable risk factors for COVID-19, and comparison to risk factors for influenza and pneumonia: results from a UK Biobank prospective cohort study', *BMJ Open*, 10(11), p. e040402. Available at: https://doi.org/10.1136/bmjopen-2020-040402.

Holden, B.A. *et al.* (2008) 'Global vision impairment due to uncorrected presbyopia', *Archives of Ophthalmology (Chicago, Ill.: 1960)*, 126(12), pp. 1731–1739. Available at: https://doi.org/10.1001/archopht.126.12.1731.

Holick, M.F. (2007) 'Vitamin D Deficiency', *New England Journal of Medicine*, 357(3), pp. 266–281. Available at: https://doi.org/10.1056/NEJMra070553.

Holmes, M.V. *et al.* (2014) 'Association between alcohol and cardiovascular disease: Mendelian randomisation analysis based on individual participant data', *BMJ*, 349, p. g4164. Available at: https://doi.org/10.1136/bmj.g4164.

Horwitz, J. *et al.* (1999) 'Lens α-crystallin: Function and structure', *Eye*, 13(3), pp. 403–408. Available at: https://doi.org/10.1038/eye.1999.114.

Howell, A.E. *et al.* (2018) 'Use of Mendelian Randomization for Identifying Risk Factors for Brain Tumors', *Frontiers in Genetics*, 9, p. 525. Available at: https://doi.org/10.3389/fgene.2018.00525.

Huang, C., Chen, W. and Wang, X. (2023) ‘Studies on the fat mass and obesity-associated (FTO) gene and its impact on obesity-associated diseases’, *Genes & Diseases*, 10(6), pp. 2351–2365. Available at: https://doi.org/10.1016/j.gendis.2022.04.014.

Huang, J. *et al.* (2015) ‘Improved imputation of low-frequency and rare variants using the UK10K haplotype reference panel’, *Nature Communications*, 6(1), p. 8111. Available at: https://doi.org/10.1038/ncomms9111.

Huang, J. *et al.* (2016) ‘Association between the HLA-DQB1 polymorphisms and the susceptibility of chronic hepatitis B: A comprehensive meta-analysis’, *Biomedical Reports*, 4(5), pp. 557–566. Available at: https://doi.org/10.3892/br.2016.632.

Huang, J. *et al.* (2024) ‘Association between systemic immune-inflammation index and cataract among outpatient US adults’, *Frontiers in Medicine*, 11. Available at: https://doi.org/10.3389/fmed.2024.1469200.

Huff, T., Mahabadi, N. and Tadi, P. (2024) ‘Neuroanatomy, Visual Cortex’, in *StatPearls*. Treasure Island (FL): StatPearls Publishing. Available at: http://www.ncbi.nlm.nih.gov/books/NBK482504/ (Accessed: 21 December 2024).

Hugosson, M. and Ekström, C. (2020) ‘Prevalence and risk factors for age-related cataract in Sweden’, *Upsala Journal of Medical Sciences*, 125(4), pp. 311–315. Available at: https://doi.org/10.1080/03009734.2020.1802375.

Hwang, J.L. and Weiss, R.E. (2014) ‘Steroid-induced diabetes: a clinical and molecular approach to understanding and treatment’, *Diabetes/metabolism research and reviews*, 30(2), pp. 96–102. Available at: https://doi.org/10.1002/dmrr.2486.

Hwang, L.-D. *et al.* (2019) ‘Using a two-sample Mendelian randomization design to investigate a possible causal effect of maternal lipid concentrations on offspring birth weight’, *International Journal of Epidemiology*, 48(5), pp. 1457–1467. Available at: https://doi.org/10.1093/ije/dyz160.

Igo, R.P., Kinzy, T.G. and Bailey, J.N.C. (2019) ‘Genetic Risk Scores’, *Current protocols in human genetics*, 104(1), p. e95. Available at: https://doi.org/10.1002/cphg.95.

Infinite Eyecare (2020) ‘Basic Eye Anatomy’, *Infinite Eyecare: High Quality Vision Service | Millwoods Edmonton*, 29 January. Available at: https://infiniteeyecare.ca/basic-eye-anatomy-and-how-it-works/ (Accessed: 21 December 2024).

Ingold, N., Amin, H.A. and Drenos, F. (2019) ‘Alcohol causes an increased risk of head and neck but not breast cancer in individuals from the UK Biobank study: A Mendelian randomisation analysis’. medRxiv, p. 19002832. Available at: https://doi.org/10.1101/19002832.

Iqbal, A.M. and Jamal, S.F. (2023) ‘Essential Hypertension’, in *StatPearls [Internet]*. Treasure Island (FL): StatPearls Publishing. Available at: https://www.ncbi.nlm.nih.gov/sites/books/NBK539859/ (Accessed: 21 December 2024).

Janiaud, P. *et al.* (2021) ‘Validity of observational evidence on putative risk and protective factors: appraisal of 3744 meta-analyses on 57 topics’, *BMC Medicine*, 19, p. 157. Available at: https://doi.org/10.1186/s12916-021-02020-6.

Jee, D. and Kim, E.C. (2015) ‘Association between serum 25-hydroxyvitamin D levels and age-related cataracts’, *Journal of Cataract and Refractive Surgery*, 41(8), pp. 1705–1715. Available at: https://doi.org/10.1016/j.jcrs.2014.12.052.

Jiang, H. *et al.* (2019) ‘Dietary vitamin and carotenoid intake and risk of age-related cataract’, *The American Journal of Clinical Nutrition*, 109(1), pp. 43–54. Available at: https://doi.org/10.1093/ajcn/nqy270.

Jiang, J. *et al.* (2013) ‘Copy number variations of DNA repair genes and the age-related cataract: Jiangsu Eye Study’, *Investigative Ophthalmology & Visual Science*, 54(2), pp. 932–938. Available at: https://doi.org/10.1167/iovs.12-10948.

Jiang, X. *et al.* (2018) ‘Genome-wide association study in 79,366 European-ancestry individuals informs the genetic architecture of 25-hydroxyvitamin D levels’, *Nature Communications*, 9(1), p. 260. Available at: https://doi.org/10.1038/s41467-017-02662-2.

Johar, S.R.K. *et al.* (2004) ‘Epidemiology based etiological study of pediatric cataract in western India’, *Indian Journal of Medical Sciences*, 58(3), pp. 115–121.

Kabbani, J. *et al.* (2024) ‘A survey of intraocular lens preferences of UK refractive surgeons for cataract surgery and refractive lens exchange’, *BMC Ophthalmology*, 24(1), p. 397. Available at: https://doi.org/10.1186/s12886-024-03639-8.

Kanclerz, P., Hecht, I. and Tuuminen, R. (2023) ‘Is Occasional Alcohol Drinking and Smoking Related to the Development of Age-Related Cataract?’, *Investigative Ophthalmology & Visual Science*, 64(13), p. 8. Available at: https://doi.org/10.1167/iovs.64.13.8.

Kanthan, G.L. *et al.* (2010) ‘Alcohol consumption and the long-term incidence of cataract and cataract surgery: the Blue Mountains Eye Study’, *American Journal of Ophthalmology*, 150(3), pp. 434-440.e1. Available at: https://doi.org/10.1016/j.ajo.2010.04.020.

Karimi, S., Arabi, A. and Shahraki, T. (2021) ‘Alcohol and the Eye’, *Journal of Ophthalmic & Vision Research*, 16(2), pp. 260–270. Available at: https://doi.org/10.18502/jovr.v16i2.9089.

Katre, D. *et al.* (2022) ‘The Prevalence of Cataract in Children’, *Cureus*, 14(10). Available at: https://doi.org/10.7759/cureus.30135.

Kaur, J. *et al.* (2012) ‘The Oxidative Stress in Cataract Patients’, *Journal of Clinical and Diagnostic Research : JCDR*, 6(10), pp. 1629–1632. Available at: https://doi.org/10.7860/JCDR/2012/4856.2626.

Kendall, J.M. (2003) ‘Designing a research project: randomised controlled trials and their principles’, *Emergency medicine journal: EMJ*, 20(2), pp. 164–168. Available at: https://doi.org/10.1136/emj.20.2.164.

Khairallah, M. *et al.* (2015) ‘Number of People Blind or Visually Impaired by Cataract Worldwide and in World Regions, 1990 to 2010’, *Investigative Ophthalmology & Visual Science*, 56(11), pp. 6762–6769. Available at: https://doi.org/10.1167/iovs.15-17201.

Khan, J. and Shaw, S. (2023) ‘Risk of cataract and glaucoma among older persons with diabetes in India: a cross-sectional study based on LASI, Wave-1’, *Scientific Reports*, 13(1), p. 11973. Available at: https://doi.org/10.1038/s41598-023-38229-z.

Khudayar, M. *et al.* (2022) 'The Association Between Deficiency of Vitamin D and Diabetes Mellitus Type 2 (DMT2)', *Cureus*, 14(2), p. e22221. Available at: https://doi.org/10.7759/cureus.22221.

Kim, S.-J. *et al.* (2023) 'Strabismus and Nystagmus in Patients With Pediatric Cataracts: Study Using Insurance Claims Data', *American Journal of Ophthalmology*, 248, pp. 87–95. Available at: https://doi.org/10.1016/j.ajo.2022.11.014.

Kisic, B. *et al.* (2012) 'Antioxidant Capacity of Lenses with Age-Related Cataract', *Oxidative Medicine and Cellular Longevity*, 2012(1), p. 467130. Available at: https://doi.org/10.1155/2012/467130.

Kiziltoprak, H. *et al.* (2019) 'Cataract in diabetes mellitus', *World Journal of Diabetes*, 10(3), pp. 140–153. Available at: https://doi.org/10.4239/wjd.v10.i3.140.

Klein, B.E. *et al.* (1995) 'Older-onset diabetes and lens opacities. The Beaver Dam Eye Study', *Ophthalmic Epidemiology*, 2(1), pp. 49–55. Available at: https://doi.org/10.3109/09286589509071451.

Klein, B.E., Klein, R. and Moss, S.E. (1995) 'Incidence of cataract surgery in the Wisconsin Epidemiologic Study of Diabetic Retinopathy', *American Journal of Ophthalmology*, 119(3), pp. 295–300. Available at: https://doi.org/10.1016/s0002-9394(14)71170-5.

Kurki, M.I. *et al.* (2023) 'FinnGen provides genetic insights from a well-phenotyped isolated population', *Nature*, 613(7944), pp. 508–518. Available at: https://doi.org/10.1038/s41586-022-05473-8.

Lai, K. *et al.* (2013) 'The Effects of Postmenopausal Hormone Use on Cataract: A Meta-Analysis', *PLoS ONE*, 8(10), p. e78647. Available at: https://doi.org/10.1371/journal.pone.0078647.

Lawlor, D.A. (2016) 'Commentary: Two-sample Mendelian randomization: opportunities and challenges', *International Journal of Epidemiology*, 45(3), pp. 908–915. Available at: https://doi.org/10.1093/ije/dyw127.

Lee, D.-S. *et al.* (2015) 'The Gender-Dependent Association between Obesity and Age-Related Cataracts in Middle-Aged Korean Adults', *PLOS ONE*, 10(5), p. e0124262. Available at: https://doi.org/10.1371/journal.pone.0124262.

LESH (2017) 'Cataract Statistics & Resources | LESH', 6 June. Available at: https://www.lasereyesurgeryhub.co.uk/data/cataract-statistics/ (Accessed: 21 December 2024).

Li, L., Wan, X. and Zhao, G. (2014) 'Meta-analysis of the risk of cataract in type 2 diabetes', *BMC Ophthalmology*, 14(1), p. 94. Available at: https://doi.org/10.1186/1471-2415-14-94.

Li, W. and Wang, B. (2022) 'Cross-sectional study of the association between asthma and cataract among 40 years and older in the USA', *BMC Ophthalmology*, 22(1), p. 340. Available at: https://doi.org/10.1186/s12886-022-02564-y.

Liang, T.Z. and Chao, J.H. (2023) 'Inhaled Corticosteroids', in *StatPearls*. Treasure Island (FL): StatPearls Publishing. Available at: http://www.ncbi.nlm.nih.gov/books/NBK470556/ (Accessed: 23 March 2025).

Lindblad, B.E. *et al.* (2007) 'Alcohol consumption and risk of cataract extraction: a prospective cohort study of women', *Ophthalmology*, 114(4), pp. 680–685. Available at: https://doi.org/10.1016/j.ophtha.2006.07.046.

Linetsky, M. *et al.* (2014) 'UVA Light-excited Kynurenines Oxidize Ascorbate and Modify Lens Proteins through the Formation of Advanced Glycation End Products', *The Journal of Biological Chemistry*, 289(24), pp. 17111–17123. Available at: https://doi.org/10.1074/jbc.M114.554410.

Liou, S.-Y. (2014) 'Use of GWAS for drug discovery and development', *Yakugaku Zasshi: Journal of the Pharmaceutical Society of Japan*, 134(4), pp. 485–490. Available at: https://doi.org/10.1248/yakushi.13-00248-4.

Liu, A. and Manche, E.E. (2011) 'Bilateral posterior subcapsular cataracts associated with long-term intranasal steroid use', *Journal of Cataract & Refractive Surgery*, 37(8), pp. 1555–1558. Available at: https://doi.org/10.1016/j.jcrs.2011.05.020.

Liu, Y.-C. *et al.* (2017) 'Cataracts', *The Lancet*, 390(10094), pp. 600–612. Available at: https://doi.org/10.1016/S0140-6736(17)30544-5.

Locke, A.E. *et al.* (2015) 'Genetic studies of body mass index yield new insights for obesity biology', *Nature*, 518(7538), pp. 197–206. Available at: https://doi.org/10.1038/nature14177.

Lou, L. *et al.* (2018) 'Association of Sex With the Global Burden of Cataract', *JAMA Ophthalmology*, 136(2), pp. 116–121. Available at: https://doi.org/10.1001/jamaophthalmol.2017.5668.

Machiela, M.J. and Chanock, S.J. (2015) 'LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants', *Bioinformatics (Oxford, England)*, 31(21), pp. 3555–3557. Available at: https://doi.org/10.1093/bioinformatics/btv402.

Mailhot, G. and White, J.H. (2020) 'Vitamin D and Immunity in Infants and Children', *Nutrients*, 12(5), p. 1233. Available at: https://doi.org/10.3390/nu12051233.

Mailu, E.W. *et al.* (2020) 'Factors associated with the uptake of cataract surgery and interventions to improve uptake in low- and middle-income countries: A systematic review', *PLoS ONE*, 15(7), p. e0235699. Available at: https://doi.org/10.1371/journal.pone.0235699.

Makley, L.N. *et al.* (2015) 'Pharmacological chaperone for α-crystallin partially restores transparency in cataract models', *Science (New York, N.Y.)*, 350(6261), pp. 674–677. Available at: https://doi.org/10.1126/science.aac9145.

Malhotra, R. *et al.* (2014) 'Comparison of small incision cataract surgery with conventional extra capsular cataract surgery: an evaluation from resource poor setting in India', *International Journal of Clinical Trials*, 1(1), pp. 10–13. Available at: https://doi.org/10.5455/2349-3259.ijct20140503.

Manousaki, D. *et al.* (2017) 'Low-Frequency Synonymous Coding Variation in CYP2R1 Has Large Effects on Vitamin D Levels and Risk of Multiple Sclerosis', *American Journal of Human Genetics*, 101(2), pp. 227–238. Available at: https://doi.org/10.1016/j.ajhg.2017.06.014.

Manousaki, D. *et al.* (2020) 'Genome-wide Association Study for Vitamin D Levels Reveals 69 Independent Loci', *American Journal of Human Genetics*, 106(3), pp. 327–337. Available at: https://doi.org/10.1016/j.ajhg.2020.01.017.

Mason, A.M. and Burgess, S. (2022) 'Software Application Profile: SUMnlmr, an R package that facilitates flexible and reproducible non-linear Mendelian randomization analyses', *International Journal of Epidemiology*, 51(6), pp. 2014–2019. Available at: https://doi.org/10.1093/ije/dyac150.

Mbatchou, J. *et al.* (2021) 'Computationally efficient whole-genome regression for quantitative and binary traits', *Nature Genetics*, 53(7), pp. 1097–1103. Available at: https://doi.org/10.1038/s41588-021-00870-7.

McCarthy, S. *et al.* (2016) 'A reference panel of 64,976 haplotypes for genotype imputation', *Nature Genetics*, 48(10), pp. 1279–1283. Available at: https://doi.org/10.1038/ng.3643.

Meader, N. *et al.* (2016) 'A systematic review on the clustering and co-occurrence of multiple risk behaviours', *BMC Public Health*, 16(1), p. 657. Available at: https://doi.org/10.1186/s12889-016-3373-6.

Medeiros, S. (2016) *Dysfunctional Lens Syndrome, a New Way to Educate Patients*, *American Academy of Ophthalmology*. Available at: https://www.aao.org/eyenet/academy-live/detail/dysfunctional-lens-syndrome-educate-patients (Accessed: 21 December 2024).

Memon, A. *et al.* (2016) 'Age-related cataract and its types in patients with and without type 2 diabetes mellitus: A Hospital-based comparative study', *JPMA. The Journal of the Pakistan Medical Association*, 66, pp. 1272–1276.

Michaeloudes, C. *et al.* (2022) 'Molecular mechanisms of oxidative stress in asthma', *Molecular Aspects of Medicine*, 85, p. 101026. Available at: https://doi.org/10.1016/j.mam.2021.101026.

Mishra, D. *et al.* (2023) 'Enzymatic and biochemical properties of lens in age-related cataract versus diabetic cataract: A narrative review', *Indian Journal of Ophthalmology*, 71(6), p. 2379. Available at: https://doi.org/10.4103/ijo.IJO_1784_22.

Miyashita, H. *et al.* (2019) 'Association between ultraviolet radiation exposure dose and cataract in Han people living in China and Taiwan: A cross-sectional study', *PLOS ONE*, 14(4), p. e0215338. Available at: https://doi.org/10.1371/journal.pone.0215338.

Mohammadi, S.-F. *et al.* (2017) 'Obesity and Density of the Crystalline Lens: Revisiting a Growing Dilemma', *Biomedicine Hub*, 2(1), pp. 1–8. Available at: https://doi.org/10.1159/000454979.

Molday, R.S. and Moritz, O.L. (2015) 'Photoreceptors at a glance', *Journal of Cell Science*, 128(22), pp. 4039–4045. Available at: https://doi.org/10.1242/jcs.175687.

Molnar, K.S. *et al.* (2019) 'Mechanism of Action of VP1-001 in cryAB(R120G)-Associated and Age-Related Cataracts', *Investigative Ophthalmology & Visual Science*, 60(10), pp. 3320–3331. Available at: https://doi.org/10.1167/iovs.18-25647.

Monti, S. *et al.* (2018) 'Randomized controlled trials and real-world data: differences and similarities to untangle literature data', *Rheumatology*, 57(Supplement_7), pp. vii54–vii58. Available at: https://doi.org/10.1093/rheumatology/key109.

Moreau, K.L. and King, J.A. (2012) 'Protein Misfolding and Aggregation in Cataract Disease and Prospects for Prevention', *Trends in molecular medicine*, 18(5), pp. 273–282. Available at: https://doi.org/10.1016/j.molmed.2012.03.005.

Morris, A.P. *et al.* (2012) 'Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes', *Nature genetics*, 44(9), p. 981. Available at: https://doi.org/10.1038/ng.2383.

Motlagh, M. and Geetha, R. (2022) 'Physiology, Accommodation', in *StatPearls*. Treasure Island (FL): StatPearls Publishing. Available at: http://www.ncbi.nlm.nih.gov/books/NBK542189/ (Accessed: 21 December 2024).

Mylona, I. *et al.* (2019) 'Hypertension is the Prominent Risk Factor in Cataract Patients', *Medicina*, 55(8), p. 430. Available at: https://doi.org/10.3390/medicina55080430.

Naeem, M. *et al.* (2012) 'Cataract: trends in surgical procedures and visual outcomes; a study in a tertiary care hospital', *J Pak Med Assoc*, 62(3), p. 4.

Nair, R. and Maseeh, A. (2012) 'Vitamin D: The "sunshine" vitamin', *Journal of Pharmacology & Pharmacotherapeutics*, 3(2), pp. 118–126. Available at: https://doi.org/10.4103/0976-500X.95506.

NEI (2024) *Cataracts | National Eye Institute*, *National Eye Institute*. Available at: https://www.nei.nih.gov/learn-about-eye-health/eye-conditions-and-diseases/cataracts (Accessed: 8 January 2025).

Nelson, M.R. *et al.* (2015) 'The support of human genetic evidence for approved drug indications', *Nature Genetics*, 47(8), pp. 856–860. Available at: https://doi.org/10.1038/ng.3314.

Nguyen, C. *et al.* (2013) 'Definition of High-Risk Type 1 Diabetes HLA-DR and HLA-DQ Types Using Only Three Single Nucleotide Polymorphisms', *Diabetes*, 62(6), pp. 2135–2140. Available at: https://doi.org/10.2337/db12-1398.

NHS England (2023) *NHS England » 2023-25 NHS Payment Scheme (amended)*. Available at: https://www.england.nhs.uk/publication/2023-25-nhs-payment-scheme/ (Accessed: 8 January 2025).

Niazi, S. *et al.* (2023) 'Association between obesity and age-related cataract: an updated systematic review and dose-response meta-analysis of prospective cohort studies', *Frontiers in Nutrition*, 10, p. 1215212. Available at: https://doi.org/10.3389/fnut.2023.1215212.

NICE (2022) *Vitamin D deficiency in adults | Health topics A to Z | CKS | NICE*. Available at: https://cks.nice.org.uk/topics/vitamin-d-deficiency-in-adults/ (Accessed: 27 December 2021).

Nirmalan, P.K. *et al.* (2004) 'Risk factors for age related cataract in a rural population of southern India: the Aravind Comprehensive Eye Study', *British Journal of Ophthalmology*, 88(8), pp. 989–994. Available at: https://doi.org/10.1136/bjo.2003.038380.

Nizami, A.A. and Gulani, A.C. (2021) 'Cataract', in *StatPearls*. Treasure Island (FL): StatPearls Publishing. Available at: http://www.ncbi.nlm.nih.gov/books/NBK539699/ (Accessed: 3 December 2021).

Öktem, Ç. and Aslan, F. (2021) 'Vitamin D Levels in Young Adult Cataract Patients: A Case-Control Study', *Ophthalmic Research*, 64(1), pp. 116–120. Available at: https://doi.org/10.1159/000509602.

Palacios, C. and Gonzalez, L. (2014) 'Is vitamin D deficiency a major global public health problem?', *The Journal of steroid biochemistry and molecular biology*, 144PA, pp. 138–145. Available at: https://doi.org/10.1016/j.jsbmb.2013.11.003.

Pan, C.-W. *et al.* (2013) 'Myopia and Age-Related Cataract: A Systematic Review and Meta-analysis', *American Journal of Ophthalmology*, 156(5), pp. 1021-1033.e1. Available at: https://doi.org/10.1016/j.ajo.2013.06.005.

Pan, C.-W. and Lin, Y. (2014) 'Overweight, obesity, and age-related cataract: a meta-analysis', *Optometry and Vision Science: Official Publication of the American Academy of Optometry*, 91(5), pp. 478–483. Available at: https://doi.org/10.1097/OPX.0000000000000243.

Park, S. *et al.* (2013) 'Association Between Cataract and the Degree of Obesity', *Optometry and Vision Science*, 90(9), pp. 1019–1027. Available at: https://doi.org/10.1097/OPX.0b013e31829cae62.

Pastor-Valero, M. *et al.* (2007) 'Years of sunlight exposure and cataract: a case-control study in a Mediterranean population', *BMC Ophthalmology*, 7(1), p. 18. Available at: https://doi.org/10.1186/1471-2415-7-18.

Patel, R. *et al.* (2020) 'Systemic adverse effects from inhaled corticosteroid use in asthma: a systematic review', *BMJ Open Respiratory Research*, 7(1), p. e000756. Available at: https://doi.org/10.1136/bmjresp-2020-000756.

Patnaik, J.L. *et al.* (2024) 'Prevalence of cortical cataracts by sex, race, and ethnicity in a Colorado cohort', *Journal of Cataract and Refractive Surgery*, 50(3), pp. 301–303. Available at: https://doi.org/10.1097/j.jcrs.0000000000001329.

Pedersen, E.M. *et al.* (2023) 'ADuLT: An efficient and robust time-to-event GWAS', *Nature Communications*, 14(1), p. 5553. Available at: https://doi.org/10.1038/s41467-023-41210-z.

Pe'er, I. *et al.* (2008) 'Estimation of the multiple testing burden for genomewide association studies of nearly all common variants', *Genetic Epidemiology*, 32(4), pp. 381–385. Available at: https://doi.org/10.1002/gepi.20303.

Perlman, R.L. (2016) 'Mouse models of human disease', *Evolution, Medicine, and Public Health*, 2016(1), pp. 170–176. Available at: https://doi.org/10.1093/emph/eow014.

Petticrew, M. and Roberts, H. (2003) 'Evidence, hierarchies, and typologies: horses for courses', *Journal of Epidemiology & Community Health*, 57(7), pp. 527–529. Available at: https://doi.org/10.1136/jech.57.7.527.

Piano, M.R. (2017) 'Alcohol's Effects on the Cardiovascular System', *Alcohol Research : Current Reviews*, 38(2), pp. 219–241.

Plunk, A.D. *et al.* (2014) 'Alcohol Consumption, Heavy Drinking and Mortality: Re-Thinking the J-Shaped Curve', *Alcoholism, clinical and experimental research*, 38(2), pp. 471–478. Available at: https://doi.org/10.1111/acer.12250.

Pollreisz, A. and Schmidt-Erfurth, U. (2010) 'Diabetic Cataract—Pathogenesis, Epidemiology and Treatment', *Journal of Ophthalmology*, 2010(1), p. 608751. Available at: https://doi.org/10.1155/2010/608751.

Pös, O. *et al.* (2021) 'DNA copy number variation: Main characteristics, evolutionary significance, and pathological aspects', *Biomedical Journal*, 44(5), pp. 548–559. Available at: https://doi.org/10.1016/j.bj.2021.02.003.

Praveen, M.R. *et al.* (2010) 'A study to explore the risk factors for the early onset of cataract in India', *Eye*, 24(4), pp. 686–694. Available at: https://doi.org/10.1038/eye.2009.137.

Prickett, C.D. *et al.* (2004) 'Alcohol: Friend or Foe? Alcoholic Beverage Hormesis for Cataract and Atherosclerosis is Related to Plasma Antioxidant Activity', *Nonlinearity in Biology, Toxicology, Medicine*, 2(4), pp. 353–370. Available at: https://doi.org/10.1080/15401420490900272.

Pruim, R.J. *et al.* (2010) 'LocusZoom: regional visualization of genome-wide association scan results', *Bioinformatics*, 26(18), pp. 2336–2337. Available at: https://doi.org/10.1093/bioinformatics/btq419.

PubChem (2025) *LSS - lanosterol synthase (human)*. Available at: https://pubchem.ncbi.nlm.nih.gov/gene/LSS/human (Accessed: 8 January 2025).

Purola, P.K.M. *et al.* (2022) 'Prevalence and 11-Year Incidence of Cataract and Cataract Surgery and the Effects of Socio-Demographic and Lifestyle Factors', *Clinical Ophthalmology (Auckland, N.Z.)*, 16, pp. 1183–1195. Available at: https://doi.org/10.2147/OPTH.S355191.

Qi, L.-B. *et al.* (2016) 'Cataract-causing mutation S228P promotes βB1-crystallin aggregation and degradation by separating two interacting loops in C-terminal domain', *Protein & Cell*, 7(7), pp. 501–515. Available at: https://doi.org/10.1007/s13238-016-0284-3.

R Core Team (2021) *R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.* Available at: https://www.r-project.org/ (Accessed: 3 October 2022).

Rahi, J.S., Dezateux, C., and British Congenital Cataract Interest Group (2001) 'Measuring and interpreting the incidence of congenital ocular anomalies: lessons from a national study of congenital cataract in the UK', *Investigative Ophthalmology & Visual Science*, 42(7), pp. 1444–1448.

Rahman, S.T. *et al.* (2023) 'Vitamin D Supplementation and the Incidence of Cataract Surgery in Older Australian Adults', *Ophthalmology*, 130(3), pp. 313–323. Available at: https://doi.org/10.1016/j.ophtha.2022.09.015.

Raissy, H.H. *et al.* (2010) 'Risk of Cataracts in the Childhood Asthma Management Program Cohort', *The Journal of allergy and clinical immunology*, 126(2), pp. 389-392.e4. Available at: https://doi.org/10.1016/j.jaci.2010.05.007.

Rathi, S. *et al.* (2020) 'Molecular Genetics and Functional Analysis Implicate CDKN2BAS1-CDKN2B Involvement in POAG Pathogenesis', *Cells*, 9(9), p. 1934. Available at: https://doi.org/10.3390/cells9091934.

Rautiainen, S. *et al.* (2010) 'Vitamin C supplements and the risk of age-related cataract: a population-based prospective cohort study in women', *The American Journal of Clinical Nutrition*, 91(2), pp. 487–493. Available at: https://doi.org/10.3945/ajcn.2009.28528.

Ravindran, R.D. *et al.* (2011) 'Inverse Association of Vitamin C with Cataract in Older People in India', *Ophthalmology*, 118(10), pp. 1958-1965.e2. Available at: https://doi.org/10.1016/j.ophtha.2011.03.016.

Ravindran, R.D. *et al.* (2021) 'Seven-year trends in cataract surgery indications and quality of outcomes at Aravind Eye Hospitals, India', *Eye*, 35(7), pp. 1895–1903. Available at: https://doi.org/10.1038/s41433-020-0954-5.

Rechsteiner, D. *et al.* (2021) 'Genetic Analysis in a Swiss Cohort of Bilateral Congenital Cataract', *JAMA Ophthalmology*, 139(7), pp. 691–700. Available at: https://doi.org/10.1001/jamaophthalmol.2021.0385.

Reddy, P.Y., Giridharan, N.V. and Reddy, G.B. (2012) 'Activation of sorbitol pathway in metabolic syndrome and increased susceptibility to cataract in Wistar-Obese rats', *Molecular Vision*, 18, pp. 495–503.

Rees, J.L. (2004) 'The Genetics of Sun Sensitivity in Humans', *American Journal of Human Genetics*, 75(5), pp. 739–751.

Reyes, L.P. *et al.* (2023) 'Expression of oxysterols in human lenses: Implications of the sterol pathway in age-related cataracts', *The Journal of Steroid Biochemistry and Molecular Biology*, 225, p. 106200. Available at: https://doi.org/10.1016/j.jsbmb.2022.106200.

Richter, G.M. *et al.* (2012) 'Risk Factors for Cortical, Nuclear, Posterior Subcapsular, and Mixed Lens Opacities: The Los Angeles Latino Eye Study', *Ophthalmology*, 119(3), pp. 547–554. Available at: https://doi.org/10.1016/j.ophtha.2011.09.005.

Ritter, L.L. *et al.* (1993) 'Alcohol Use and Lens Opacities in the Beaver Dam Eye Study', *Archives of Ophthalmology*, 111(1), pp. 113–117. Available at: https://doi.org/10.1001/archopht.1993.01090010117037.

Roberts, J.E. (2011) 'Ultraviolet radiation as a risk factor for cataract and macular degeneration', *Eye & Contact Lens*, 37(4), pp. 246–249. Available at: https://doi.org/10.1097/ICL.0b013e31821cbcc9.

Rothberg, A.E. *et al.* (2023) 'Obesity Management for the Treatment of Type 2 Diabetes', in J.M. Lawrence et al. (eds) *Diabetes in America*. Bethesda (MD): National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK). Available at: http://www.ncbi.nlm.nih.gov/books/NBK603314/ (Accessed: 23 March 2025).

Rowan, S. *et al.* (2021) 'Aged Nrf2-Null Mice Develop All Major Types of Age-Related Cataracts', *Investigative Ophthalmology & Visual Science*, 62(15), p. 10. Available at: https://doi.org/10.1167/iovs.62.15.10.

Rowe, N.G. *et al.* (2000) 'Diabetes, fasting blood glucose and age-related cataract: the Blue Mountains Eye Study', *Ophthalmic Epidemiology*, 7(2), pp. 103–114.

Sabhapandit, D.S. (2019) 'Nuclear Cataracts: Symptoms, Causes, and Treatment', *Neoretina Blog*, 30 December. Available at: https://neoretina.com/blog/nuclear-cataracts-symptoms-causes-treatment/ (Accessed: 21 December 2024).

Sarkar, D. *et al.* (2023) ‘Age-related cataract - Prevalence, epidemiological pattern and emerging risk factors in a cross-sectional study from Central India’, *Indian Journal of Ophthalmology*, 71(5), pp. 1905–1912. Available at: https://doi.org/10.4103/ijo.IJO_2020_22.

Saunders, G.R.B. *et al.* (2022) ‘Genetic diversity fuels gene discovery for tobacco and alcohol use’, *Nature*, 612(7941), pp. 720–724. Available at: https://doi.org/10.1038/s41586-022-05477-4.

Savran, O. and Ulrik, C.S. (2023) ‘Inhaled Corticosteroid Exposure and Risk of Cataract in Patients with Asthma and COPD: A Systematic Review and Meta-Analysis’, *Journal of Ophthalmology*, 2023, p. 8209978. Available at: https://doi.org/10.1155/2023/8209978.

Scarborough, P. *et al.* (2011) ‘The economic burden of ill health due to diet, physical inactivity, smoking, alcohol and obesity in the UK: an update to 2006-07 NHS costs’, *Journal of Public Health*, 33(4), pp. 527–535. Available at: https://doi.org/10.1093/pubmed/fdr033.

Schoeler, T. *et al.* (2023) ‘Participation bias in the UK Biobank distorts genetic associations and downstream analyses’, *Nature Human Behaviour*, 7(7), pp. 1216–1227. Available at: https://doi.org/10.1038/s41562-023-01579-9.

Scholz, M. *et al.* (2022) ‘Genome-wide meta-analysis of phytosterols reveals five novel loci and a detrimental effect on coronary atherosclerosis’, *Nature Communications*, 13(1), p. 143. Available at: https://doi.org/10.1038/s41467-021-27706-6.

Sekhon, R.K., Rocha Cabrero, F. and Deibel, J.P. (2023) ‘Nystagmus Types’, in *StatPearls*. Treasure Island (FL): StatPearls Publishing. Available at: http://www.ncbi.nlm.nih.gov/books/NBK539711/ (Accessed: 21 December 2024).

Self, J.E. *et al.* (2020) ‘Cataract management in children: a review of the literature and current practice across five large UK centres’, *Eye*, 34(12), pp. 2197–2218. Available at: https://doi.org/10.1038/s41433-020-1115-6.

Seral-Cortes, M. *et al.* (2021) ‘Development of a Genetic Risk Score to predict the risk of overweight and obesity in European adolescents from the HELENA study’, *Scientific Reports*, 11(1), p. 3067. Available at: https://doi.org/10.1038/s41598-021-82712-4.

Shanmugam, P.M. *et al.* (2015) ‘Effect of lanosterol on human cataract nucleus’, *Indian Journal of Ophthalmology*, 63(12), pp. 888–890. Available at: https://doi.org/10.4103/0301-4738.176040.

Sharma, A.K. *et al.* (2016) ‘Visual Outcome of Traumatic Cataract at a Tertiary Eye Care Centre in North India: A Prospective Study’, *Journal of Clinical and Diagnostic Research : JCDR*, 10(1), pp. NC05–NC08. Available at: https://doi.org/10.7860/JCDR/2016/17216.7049.

Sheeladevi, S. *et al.* (2016) ‘Global prevalence of childhood cataract: a systematic review’, *Eye*, 30(9), pp. 1160–1169. Available at: https://doi.org/10.1038/eye.2016.156.

Shiels, A. and Hejtmancik, J.F. (2015) ‘Molecular Genetics of Cataract’, *Progress in molecular biology and translational science*, 134, pp. 203–218. Available at: https://doi.org/10.1016/bs.pmbts.2015.05.004.

Shiels, A. and Hejtmancik, J.F. (2019) 'Biology of Cataracts and Opportunities for Treatment', *Annual review of vision science*, 5, pp. 123–149. Available at: https://doi.org/10.1146/annurev-vision-091517-034346.

Shu, Y. *et al.* (2023) 'Changing Trends in the Disease Burden of Cataract and Forecasted Trends in China and Globally from 1990 to 2030', *Clinical Epidemiology*, 15, pp. 525–534. Available at: https://doi.org/10.2147/CLEP.S404049.

Singh, P. and Tripathy, K. (2023) 'Presbyopia', in *StatPearls [Internet]*. StatPearls Publishing. Available at: https://www.ncbi.nlm.nih.gov/sites/books/NBK560568/ (Accessed: 21 December 2024).

Sizar, O. *et al.* (2023) 'Vitamin D Deficiency', in *StatPearls*. Treasure Island (FL): StatPearls Publishing. Available at: http://www.ncbi.nlm.nih.gov/books/NBK532266/ (Accessed: 21 December 2024).

Skelly, A.C., Dettori, J.R. and Brodt, E.D. (2012) 'Assessing bias: the importance of considering confounding', *Evidence-Based Spine-Care Journal*, 3(1), pp. 9–12. Available at: https://doi.org/10.1055/s-0031-1298595.

Slatkin, M. (2008) 'Linkage disequilibrium — understanding the evolutionary past and mapping the medical future', *Nature reviews. Genetics*, 9(6), p. 477. Available at: https://doi.org/10.1038/nrg2361.

Slob, E.A.W. and Burgess, S. (2020) 'A comparison of robust Mendelian randomization methods using summary data', *Genetic Epidemiology*, 44(4), pp. 313–329. Available at: https://doi.org/10.1002/gepi.22295.

Smit, L.A.M. *et al.* (2014) 'Human leukocyte antigen class II variants and adult-onset asthma: does occupational allergen exposure play a role?', *European Respiratory Journal*, 44(5), pp. 1234–1242. Available at: https://doi.org/10.1183/09031936.00068014.

Smith, G.D. (2023) 'Mendelian randomisation and vitamin D: the importance of model assumptions', *The Lancet Diabetes & Endocrinology*, 11(1), p. 14. Available at: https://doi.org/10.1016/S2213-8587(22)00345-X.

Sobczyk, M.K. *et al.* (2023) 'Systematic comparison of Mendelian randomisation studies and randomised controlled trials using electronic databases', *BMJ Open*, 13(9), p. e072087. Available at: https://doi.org/10.1136/bmjopen-2023-072087.

Sofianopoulou, E. *et al.* (2024) 'Estimating dose-response relationships for vitamin D with coronary heart disease, stroke, and all-cause mortality: observational and Mendelian randomisation analyses', *The Lancet Diabetes & Endocrinology*, 12(1), pp. e2–e11. Available at: https://doi.org/10.1016/S2213-8587(23)00287-5.

Speed, D. *et al.* (2017) 'Re-evaluation of SNP heritability in complex human traits', *Nature genetics*, 49(7), pp. 986–992. Available at: https://doi.org/10.1038/ng.3865.

Spence, J.D. (2015) 'Statins and Cataracts: Reverse Causality?', *Canadian Journal of Cardiology*, 31(5), p. 691.e11. Available at: https://doi.org/10.1016/j.cjca.2014.12.018.

Spieth, P.M. *et al.* (2016) 'Randomized controlled trials – a matter of design', *Neuropsychiatric Disease and Treatment*, 12, pp. 1341–1349. Available at: https://doi.org/10.2147/NDT.S101938.

Spreafico, R. *et al.* (2020) ‘Advances in Genomics for Drug Development’, *Genes*, 11(8), p. 942. Available at: https://doi.org/10.3390/genes11080942.

Staley, J.R. and Burgess, S. (2017) ‘Semiparametric methods for estimation of a nonlinear exposure-outcome relationship using instrumental variables with application to Mendelian randomization’, *Genetic Epidemiology*, 41(4), pp. 341–352. Available at: https://doi.org/10.1002/gepi.22041.

Sterrantino, A.F. (2024) ‘Observational studies: practical tips for avoiding common statistical pitfalls’, *The Lancet Regional Health - Southeast Asia*, 25. Available at: https://doi.org/10.1016/j.lansea.2024.100415.

Storey, P. *et al.* (2013) ‘Racial Differences in Lens Opacity Incidence and Progression: The Salisbury Eye Evaluation (SEE) Study’, *Investigative Ophthalmology & Visual Science*, 54(4), pp. 3010–3018. Available at: https://doi.org/10.1167/iovs.12-11412.

Su, D. *et al.* (2012) ‘A novel mutation in CRYAA is associated with autosomal dominant suture cataracts in a Chinese family’, *Molecular Vision*, 18, pp. 3057–3063.

Sudlow, C. *et al.* (2015) ‘UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age’, *PLOS Medicine*, 12(3), p. e1001779. Available at: https://doi.org/10.1371/journal.pmed.1001779.

Swerdlow, D.I. *et al.* (2015) ‘HMG-coenzyme A reductase inhibition, type 2 diabetes, and bodyweight: evidence from genetic analysis and randomised trials’, *The Lancet*, 385(9965), pp. 351–361. Available at: https://doi.org/10.1016/S0140-6736(14)61183-1.

Tam, V. *et al.* (2019) ‘Benefits and limitations of genome-wide association studies’, *Nature Reviews. Genetics*, 20(8), pp. 467–484. Available at: https://doi.org/10.1038/s41576-019-0127-1.

Tan, A.G. *et al.* (2019) ‘Evaluating the associations between obesity and age-related cataract: a Mendelian randomization study’, *The American Journal of Clinical Nutrition*, 110(4), pp. 969–976. Available at: https://doi.org/10.1093/ajcn/nqz167.

Tan, A.G. *et al.* (2020) ‘Incidence, progression and risk factors of age-related cataract in Malays: The Singapore Malay Eye Study’, *Clinical & Experimental Ophthalmology*, 48(5), pp. 580–592. Available at: https://doi.org/10.1111/ceo.13757.

Taylor, D. (1998) ‘Congenital cataract: The history, the nature and the practice’, *Eye*, 12(1), pp. 9–36. Available at: https://doi.org/10.1038/eye.1998.5.

Thompson, B. *et al.* (2022) ‘Oxidative stress induces inflammation of lens cells and triggers immune surveillance of ocular tissues’, *Chemico-Biological Interactions*, 355, p. 109804. Available at: https://doi.org/10.1016/j.cbi.2022.109804.

Tian, H. *et al.* (2023) ‘Relaxing parametric assumptions for non-linear Mendelian randomization using a doubly-ranked stratification method’, *PLOS Genetics*, 19(6), p. e1010823. Available at: https://doi.org/10.1371/journal.pgen.1010823.

Tiwari, A. and Balasundaram, P. (2023) ‘Public Health Considerations Regarding Obesity’, in *StatPearls*. Treasure Island (FL): StatPearls Publishing. Available at: http://www.ncbi.nlm.nih.gov/books/NBK572122/ (Accessed: 21 December 2024).

Tsai, M.-K., Gao, W. and Wen, C.-P. (2023) 'The relationship between alcohol consumption and health: J-shaped or less is more?', *BMC Medicine*, 21, p. 228. Available at: https://doi.org/10.1186/s12916-023-02911-w.

Uffelmann, E. *et al.* (2021) 'Genome-wide association studies', *Nature Reviews Methods Primers*, 1(1), pp. 1–21. Available at: https://doi.org/10.1038/s43586-021-00056-9.

UK Biobank (2007) 'UK Biobank: Protocol for a large-scale prospective epidemiological resource'. Available at: https://www.ukbiobank.ac.uk/media/gnkeyh2q/study-rationale.pdf (Accessed: 6 November 2024).

UK Biobank (2021) *Baseline assessment*, *UK Biobank*. Available at: https://www.ukbiobank.ac.uk/enable-your-research/about-our-data/baseline-assessment (Accessed: 8 January 2025).

Valenzuela, C.V. *et al.* (2019) 'Intranasal corticosteroids do not lead to ocular changes: a systematic review and meta-analysis', *The Laryngoscope*, 129(1), pp. 6–12. Available at: https://doi.org/10.1002/lary.27209.

Valero, M.P. *et al.* (2002) 'Vitamin C Is Associated with Reduced Risk of Cataract in a Mediterranean Population', *The Journal of Nutrition*, 132(6), pp. 1299–1306. Available at: https://doi.org/10.1093/jn/132.6.1299.

Van Staa, T.P. *et al.* (2000) 'Use of oral corticosteroids and risk of fractures', *Journal of Bone and Mineral Research: The Official Journal of the American Society for Bone and Mineral Research*, 15(6), pp. 993–1000. Available at: https://doi.org/10.1359/jbmr.2000.15.6.993.

Vance, M.C., Caverly, T.J. and Hayward, R.A. (2020) 'Underappreciated Bias Created by Measurement Error in Risk Factor Assessment—A Case Study of No Safe Level of Alcohol Consumption', *JAMA Internal Medicine*, 180(3), pp. 459–461. Available at: https://doi.org/10.1001/jamainternmed.2019.6116.

Vashist, P. *et al.* (2020) 'Association of cataract and sun exposure in geographically diverse populations of India: The CASE study. First Report of the ICMR-EYE SEE Study Group', *PLOS ONE*, 15(1), p. e0227868. Available at: https://doi.org/10.1371/journal.pone.0227868.

Veerbeek, M.A. *et al.* (2019) 'Differences in alcohol use between younger and older people: Results from a general population study', *Drug and Alcohol Dependence*, 202, pp. 18–23. Available at: https://doi.org/10.1016/j.drugalcdep.2019.04.023.

Vijayalakshmi, P. and Njambi, L. (2016) 'Postoperative care for paediatric cataract patients', *Community Eye Health*, 29(94), pp. 32–34.

Villanueva, R.A.M. and Chen, Z.J. (2019) 'ggplot2: Elegant Graphics for Data Analysis (2nd ed.)', *Measurement: Interdisciplinary Research and Perspectives*, 17(3), pp. 160–167. Available at: https://doi.org/10.1080/15366367.2019.1565254.

Visontay, R. *et al.* (2022) 'Are there non-linear relationships between alcohol consumption and long-term health?: a systematic review of observational studies employing approaches to improve causal inference', *BMC Medical Research Methodology*, 22, p. 16. Available at: https://doi.org/10.1186/s12874-021-01486-5.

Vitali, F. *et al.* (2019) 'Precision drug repurposing via convergent eQTL-based molecules and pathway targeting independent disease-associated polymorphisms', *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing*, 24, pp. 308–319.

Wain, L.V. *et al.* (2015) '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), pp. 769–781. Available at: https://doi.org/10.1016/S2213-2600(15)00283-0.

Waljee, A.K. *et al.* (2017) 'Short term use of oral corticosteroids and related harms among adults in the United States: population based cohort study', *BMJ*, 357, p. j1415. Available at: https://doi.org/10.1136/bmj.j1415.

Wang, C.-H. and Xin, Z.-K. (2024) 'Causal association between 25-hydroxyvitamin D status and cataract development: A two-sample Mendelian randomization study', *World Journal of Clinical Cases*, 12(16), pp. 2789–2795. Available at: https://doi.org/10.12998/wjcc.v12.i16.2789.

Wang, J.J. *et al.* (2009) 'Use of Inhaled and Oral Corticosteroids and the Long-term Risk of Cataract', *Ophthalmology*, 116(4), pp. 652–657. Available at: https://doi.org/10.1016/j.ophtha.2008.12.001.

Wang, K. *et al.* (2022) 'Oxysterol Compounds in Mouse Mutant αA- and αB-Crystallin Lenses Can Improve the Optical Properties of the Lens', *Investigative Opthalmology & Visual Science*, 63(5), p. 15. Available at: https://doi.org/10.1167/iovs.63.5.15.

Wang, T.J. *et al.* (2010) 'Common genetic determinants of vitamin D insufficiency: a genome-wide association study', *The Lancet*, 376(9736), pp. 180–188. Available at: https://doi.org/10.1016/S0140-6736(10)60588-0.

Wang, W. and Zhang, X. (2014) 'Alcohol Intake and the Risk of Age-Related Cataracts: A Meta-Analysis of Prospective Cohort Studies', *PLoS ONE*, 9(9), p. e107820. Available at: https://doi.org/10.1371/journal.pone.0107820.

Waring IV, G.O. (2020) 'Dysfunctional Lens Syndrome', *CRSToday*, March. Available at: https://crstoday.com/articles/2020-mar/dysfunctional-lens-syndrome (Accessed: 21 December 2024).

Watford, J. (2020) 'Guide to Cataract Surgery In The UK | LESH', 1 October. Available at: https://www.lasereyesurgeryhub.co.uk/cataract-surgery/ (Accessed: 21 December 2024).

Wei, L. *et al.* (2016) 'Association of vitamin C with the risk of age-related cataract: a meta-analysis', *Acta Ophthalmologica*, 94(3), pp. e170–e176. Available at: https://doi.org/10.1111/aos.12688.

West, R. (2017) 'Tobacco smoking: Health impact, prevalence, correlates and interventions', *Psychology & Health*, 32(8), pp. 1018–1036. Available at: https://doi.org/10.1080/08870446.2017.1325890.

Whitworth, J.A. *et al.* (1989) 'The hypertensive effect of synthetic glucocorticoids in man: role of sodium and volume', *Journal of Hypertension*, 7(7), pp. 537–549. Available at: https://doi.org/10.1097/00004872-198907000-00005.

Willoughby, C.E. *et al.* (2010) 'Anatomy and physiology of the human eye: effects of mucopolysaccharidoses disease on structure and function – a review', *Clinical &*

Experimental Ophthalmology, 38(s1), pp. 2–11. Available at: https://doi.org/10.1111/j.1442-9071.2010.02363.x.

Wise, S.J. *et al.* (2014) 'Statin use and risk for cataract: a nested case-control study of 2 populations in Canada and the United States', *The Canadian Journal of Cardiology*, 30(12), pp. 1613–1619. Available at: https://doi.org/10.1016/j.cjca.2014.08.020.

Wistow, G. (2012) 'The human crystallin gene families', *Human Genomics*, 6(1), p. 26. Available at: https://doi.org/10.1186/1479-7364-6-26.

Wolf, N. *et al.* (2005) 'Age-related cataract progression in five mouse models for anti-oxidant protection or hormonal influence', *Experimental Eye Research*, 81(3), pp. 276–285. Available at: https://doi.org/10.1016/j.exer.2005.01.024.

Wood, A.R. *et al.* (2016) 'Variants in the FTO and CDKAL1 loci have recessive effects on risk of obesity and type 2 diabetes, respectively', *Diabetologia*, 59, p. 1214. Available at: https://doi.org/10.1007/s00125-016-3908-5.

World Health Organisation (2023) *Hypertension*. Available at: https://www.who.int/news-room/fact-sheets/detail/hypertension (Accessed: 21 December 2024).

World Health Organisation (2024) *Diabetes*. Available at: https://www.who.int/news-room/fact-sheets/detail/diabetes (Accessed: 21 December 2024).

Xiao, M. *et al.* (2021) 'CDKN2B-AS1 participates in high glucose-induced apoptosis and fibrosis via NOTCH2 through functioning as a miR-98-5p decoy in human podocytes and renal tubular cells', *Diabetology & Metabolic Syndrome*, 13(1), p. 107. Available at: https://doi.org/10.1186/s13098-021-00725-5.

Xu, J. *et al.* (2020) 'Advances in pharmacotherapy of cataracts', *Annals of Translational Medicine*, 8(22), p. 1552. Available at: https://doi.org/10.21037/atm-20-1960.

Xu, L., You, Q.S. and Jonas, J.B. (2009) 'Prevalence of alcohol consumption and risk of ocular diseases in a general population: the Beijing Eye Study', *Ophthalmology*, 116(10), pp. 1872–1879. Available at: https://doi.org/10.1016/j.ophtha.2009.04.014.

Xue, A. *et al.* (2018) 'Genome-wide association analyses identify 143 risk variants and putative regulatory mechanisms for type 2 diabetes', *Nature Communications*, 9(1), p. 2941. Available at: https://doi.org/10.1038/s41467-018-04951-w.

Yang, T.-L. *et al.* (2009) 'HMGA2 Is Confirmed To Be Associated with Human Adult Height', *Annals of human genetics*, 74(1), p. 11. Available at: https://doi.org/10.1111/j.1469-1809.2009.00555.x.

Yavorska, O.O. and Burgess, S. (2017) 'MendelianRandomization: an R package for performing Mendelian randomization analyses using summarized data', *International Journal of Epidemiology*, 46(6), pp. 1734–1739. Available at: https://doi.org/10.1093/ije/dyx034.

Ye, J. *et al.* (2012) 'Smoking and Risk of Age-Related Cataract: A Meta-Analysis', *Investigative Ophthalmology & Visual Science*, 53(7), pp. 3885–3895. Available at: https://doi.org/10.1167/iovs.12-9820.

Ye, J. *et al.* (2014) 'Body Mass Index and Risk of Age-Related Cataract: A Meta-Analysis of Prospective Cohort Studies', *PLoS ONE*, 9(2). Available at: https://doi.org/10.1371/journal.pone.0089923.

Yi, J. *et al.* (2011) 'Epidemiology and molecular genetics of congenital cataracts', *International Journal of Ophthalmology*, 4(4), p. 422. Available at: https://doi.org/10.3980/j.issn.2222-3959.2011.04.20.

Yonova-Doing, E. *et al.* (2016) 'Genetic and Dietary Factors Influencing the Progression of Nuclear Cataract', *Ophthalmology*, 123(6), pp. 1237–1244. Available at: https://doi.org/10.1016/j.ophtha.2016.01.036.

Yonova-Doing, E. *et al.* (2020) 'Common variants in SOX-2 and congenital cataract genes contribute to age-related nuclear cataract', *Communications Biology*, 3(1), pp. 1–8. Available at: https://doi.org/10.1038/s42003-020-01421-2.

You, S. *et al.* (2023) 'Down-regulation of WWP2 aggravates Type 2 diabetes mellitus-induced vascular endothelial injury through modulating ubiquitination and degradation of DDX3X', *Cardiovascular Diabetology*, 22(1), p. 107. Available at: https://doi.org/10.1186/s12933-023-01818-3.

Younan, C. *et al.* (2002) 'Myopia and incident cataract and cataract surgery: the blue mountains eye study', *Investigative Ophthalmology & Visual Science*, 43(12), pp. 3625–3632.

Yu, X. *et al.* (2014) 'Hypertension and Risk of Cataract: A Meta-Analysis', *PLoS ONE*, 9(12), p. e114012. Available at: https://doi.org/10.1371/journal.pone.0114012.

Yuan, S., Wolk, A. and Larsson, S.C. (2022) 'Metabolic and lifestyle factors in relation to senile cataract: a Mendelian randomization study', *Scientific Reports*, 12(1), p. 409. Available at: https://doi.org/10.1038/s41598-021-04515-x.

Yuan, Z. *et al.* (2020) 'Testing and controlling for horizontal pleiotropy with probabilistic Mendelian randomization in transcriptome-wide association studies', *Nature Communications*, 11(1), p. 3861. Available at: https://doi.org/10.1038/s41467-020-17668-6.

Zabor, E.C., Kaizer, A.M. and Hobbs, B.P. (2020) 'Randomized Controlled Trials', *Chest*, 158(1 Suppl), pp. S79–S87. Available at: https://doi.org/10.1016/j.chest.2020.03.013.

Zarulli, V. *et al.* (2018) 'Women live longer than men even during severe famines and epidemics', *Proceedings of the National Academy of Sciences of the United States of America*, 115(4), pp. E832–E840. Available at: https://doi.org/10.1073/pnas.1701535115.

Zetterberg, M. and Celojevic, D. (2015) 'Gender and Cataract – The Role of Estrogen', *Current Eye Research*, 40(2), pp. 176–190. Available at: https://doi.org/10.3109/02713683.2014.898774.

Zhang, H. *et al.* (2021) 'Mendelian randomization study reveals a population-specific putative causal effect of type 2 diabetes in risk of cataract', *International Journal of Epidemiology*, 50(6), pp. 2024–2037. Available at: https://doi.org/10.1093/ije/dyab175.

Zhang, Jun *et al.* (2008) 'Targeted knockout of the mouse betaB2-crystallin gene (Crybb2) induces age-related cataract', *Investigative Ophthalmology & Visual Science*, 49(12), pp. 5476–5483. Available at: https://doi.org/10.1167/iovs.08-2179.

Zhang, X. *et al.* (2021) ‘Alcohol consumption and risk of cardiovascular disease, cancer and mortality: a prospective cohort study’, *Nutrition Journal*, 20(1), p. 13. Available at: https://doi.org/10.1186/s12937-021-00671-y.

Zhao, L. *et al.* (2015) ‘Lanosterol reverses protein aggregation in cataracts’, *Nature*, 523(7562), pp. 607–611. Available at: https://doi.org/10.1038/nature14650.

Zhao, M. *et al.* (2021) ‘Defect of LSS Disrupts Lens Development in Cataractogenesis’, *Frontiers in Cell and Developmental Biology*, 9, p. 788422. Available at: https://doi.org/10.3389/fcell.2021.788422.

Zheng, J. *et al.* (2017) ‘Recent Developments in Mendelian Randomization Studies’, *Current Epidemiology Reports*, 4(4), pp. 330–345. Available at: https://doi.org/10.1007/s40471-017-0128-6.

Zhou, A., Selvanayagam, J.B. and Hyppönen, E. (2022) ‘Non-linear Mendelian randomization analyses support a role for vitamin D deficiency in cardiovascular disease risk’, *European Heart Journal*, 43(18), pp. 1731–1739. Available at: https://doi.org/10.1093/eurheartj/ehab809.

Zhou, W. *et al.* (2018) ‘Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies’, *Nature genetics*, 50(9), pp. 1335–1341. Available at: https://doi.org/10.1038/s41588-018-0184-y.

Zhu, Z. *et al.* (2019) ‘Shared Genetics of Asthma and Mental Health Disorders: A Large-Scale Genome-Wide Cross-Trait Analysis’, *European Respiratory Journal* [Preprint]. Available at: https://doi.org/10.1183/13993003.01507-2019.

Zou, X. *et al.* (2020) ‘The Polymorphism rs2968 of LSS Gene Confers Susceptibility to Age-Related Cataract’, *DNA and Cell Biology*, 39(11), pp. 1970–1975. Available at: https://doi.org/10.1089/dna.2020.5872.