A causal association of genetically predicted serum 25-hydroxyvitamin D concentrations on the odds of tuberculosis disease: A Two-sample Mendelian Randomization Study

by Mercy Kimani

Submission date: 19-Nov-2020 04:10PM (UTC+0300)
Submission ID: 1451044427
File name: MSc_Project_Mercy_Kimani_signed.pdf (2.43M)
Word count: 15022
Character count: 85410

University of Nairobi

School of Mathematics

ISSN: 2410-1397

Master of Science in Biometry

A causal association of genetically predicted serum 25-hydroxyvitamin D concentrations on the odds of tuberculosis disease: A Two-sample Mendelian Randomization Study

Research Report in Mathematics, Number 48, 2020

Mercy W. Kimani

Nov 2020

Submitted to the School of Mathematics in partial fulfilment for a degree in Master of Science in Biometry

A causal association of genetically predicted serum 25-hydroxyvitamin D concentrations on the odds of tuberculosis disease: A Two-sample Mendelian Randomization Study

Research Report in Mathematics, Number 48, 2020

Mercy W. Kimani

School of Mathematics
College of Biological and Physical sciences
Chiromo, off Riverside Drive
30197-00100 Nairobi, Kenya

Master Thesis

Submitted to the School of Mathematics in partial fulfilment for a degree in Master of Science in Biometry

Submitted to: The Graduate School, University of Nairobi, Kenya

Abstract

Tuberculosis is a major cause of ill health and death. In 2018, an estimated 10 million people became ill and 1.2 million died from the disease. Vitamin D deficiency is also a major public health problem, affecting about a billion people globally. Studies have reported an association between vitamin D deficiency and an increased risk of tuberculosis. Some of the mechanisms underlying this association may include the role of vitamin D in both innate and adaptive immune functions. This study aimed to investigate whether there is a causal association between serum 25-hydroxyvitamin D (25(OH)D) levels and odds of tuberculosis disease in Asian and European populations using two-sample Mendelian randomization.

I used Mendelian randomization (MR), a technique that employs single nucleotide polymorphisms (genetic variants) as instrumental variables to estimate the causal effect of serum-25(OH)D (intervention) on the outcome of tuberculosis disease. I used genome-wide association studies (GWAS) summary statistics from the UK Biobank as the exposure data and from the International Tuberculosis Human Genetics Consortium (ITHGC) as the outcome data. Results from this study showed no evidence of a causal association of genetically predicted serum-25(OH)D concentrations on the odds of tuberculosis disease.

Master Thesis in Mathematics at the University of Nairobi, Kenya.
ISSN 2410-1397: Research Report in Mathematics

DISTRIBUTOR: School of Mathematics, University of Nairobi, Kenya

Declaration and Approval

I the undersigned declare that this dissertation is my original work and to the best of my knowledge, it has not been submitted in support of an award of a degree in any other university or institution of learning.

Signature Date

MERCY W. KIMANI
Reg No. I56/12265/2018

In my capacity as a supervisor of the candidate's dissertation, I certify that this dissertation has my approval for submission.

Signature Date

Dr Nelson Owuor Onyango
School of Mathematics,
University of Nairobi,
Box 30197, 00100 Nairobi, Kenya.
E-mail: onyango@uonbi.ac.ke

Sarah Atkinson 13th November 2020

Signature Date

Prof Sarah Atkinson
Department of Epidemiology and Demography
KEMRI-Wellcome Trust Research Programme,
Box 230, 80108 Kilifi, Kenya.
E-mail: satkinson@kemri-wellcome.org

Signature Date

Dr Tom Yates
Department of Infectious Diseases,
Imperial College London,
London SW7 2AZ, UK.
E-mail: t.yates@imperial.ac.uk

Dedication

This project is dedicated to my parents and siblings.

Acknowledgments

First, I thank God for his unwavering grace in making all this possible.

I acknowledge my supervisors, Prof. Sarah Atkinson and Dr. Tom Yates who conceptualized this project. Thank you for your unending support, insights and mentorship. I would also like to acknowledge Dr. Nelson Onyango for his counsel especially on the school requirements for this project. My heartfelt appreciation goes to Dr. Sarah's group and especially Reagan for their constructive assistance and discussions.

Many thanks to Dr. Ruth Mitchell for her insightful help on my analyses methods. I also appreciate the International Tuberculosis Host Genetics Consortium (ITHGC) for providing and allowing me to use their unpublished data.

I acknowledge IDeAL, KEMRI-Wellcome Trust for their financial support throughout this project. This work waswas supported through the DELTAS Africa Initiative [DEL-15-003]. The DELTAS Africa Initiative is an independent funding scheme of the African Academy of Sciences (AAS)'s Alliance for Accelerating Excellence in Science in Africa (AESA) and supported by the New Partnership for Africa's Development Planning and Coordinating Agency (NEPAD Agency) with funding from the Wellcome Trust [107769/Z/10/Z] and the UK government. The views expressed in this publication are those of the author(s) and not necessarily those of AAS, NEPAD Agency, Wellcome Trust or the UK government.

Finally, I thank my family, friends, course mates and colleagues who have given me emotional support and encouraged me when things were not working as expected especially during the pandemic.

Mercy W. Kimani

Nairobi, 2020.

Contents

Bibliography 36

Figures and Tables

Figures

Tables

1 Introduction

1.1 Background

Vitamin D deficiency and tuberculosis are both important global health concerns. Tuberculosis is a social and economic burden especially in South-East Asia, Africa, and the West Pacific, areas that account for 86-percent of cases(World Health Organization, 2020). An estimated 1.2 million people died from tuberculosis in 2019 making it one of the leading causes of death worldwide (World Health Organization, 2020). This is despite WHO reporting a two-percentage decline in the incidence of tuberculosis cases between 2018 and 2019 with an estimated ten million people becoming ill in 2019(World Health Organization, 2020). Vitamin D deficiency affects almost a billion people globally with the greatest prevalence observed in the Middle East and Asia, especially among women (Holick, 2007; Natasja & Lips, 2011). A recent study has also reported that Africa is not an exception with approximately one in every five people having insufficient levels.4 Vitamin D deficiency is associated with the risk of both skeletal and non-skeletal diseases (Mogire et al., 2020). Some of the skeletal diseases include rickets in children, osteomalacia in adults and accelerated osteoporosis in older people (Bouillon et al., 2019). A causal association with multiple sclerosis has also been reported (Rhead et al., 2016). Many epidemiological studies provide evidence of an increased risk of tuberculosis (TB) in individuals with vitamin D deficiency (S. J. Huang et al., 2017; Nnoaham & Clarke, 2008; Venturini et al., 2014).

Genetic variants are used as instrumental variables to estimate whether there is a causal effect of an exposure (serum-25(OH)D levels) on an outcome (odds of tuberculosis), a technique called Mendelian Randomization (MR). In MR, genetic variants mimic the random assignment of treatments as in randomized controlled trials. Since genotypes are randomly allocated at meiosis, associations of the genetic variants with risk of tuberculosis will not be constrained by unaccounted for or unmeasured environmental risk factors (confounders). Also, since disease outcome cannot modify genotype, bias from reverse causality will be avoided (Sheehan, Didelez, Burton, & Tobin, 2008; Burgess & Thompson, 2015).

1.2 Statement of Problem

Many epidemiological studies have reported an association between vitamin D deficiency and an increased risk of tuberculosis disease. Randomized controlled trials (RCT) in children and individuals living with the human immunodeficiency virus (HIV) have recently

reported no impact of vitamin D supplementation on the incidence of tuberculosis disease and infection. Further RCTs have been suggested and although they are the most robust means of evaluating potential interventions, they remain expensive and their capacity is limited. Mendelian randomization (MR) tests whether an exposure has a causal effect on an outcome by use of genetic variants as instrumental variables. Genetic variants are randomly assorted during meiosis, a process that mimics the random assignment of treatment in an RCT. This random allocation, as in a RCT, is independent of environmental factors meaning that exposure-outcome associations are unlikely to be a result of confounding by environmental exposures or reverse causation. This study seeks to explain whether serum vitamin D levels have a causal association with tuberculosis disease, which will also inform whether further RCTs of vitamin D for the prevention of tuberculosis are warranted.

1.3 Objectives

1.3.1 Overall objective

The general objective is to infer the causal effect of serum vitamin D levels on tuberculosis disease using two-sample Mendelian randomization.

1.3.2 Specific objectives

I. To determine the causal effect of vitamin D status on tuberculosis disease in a country-level analysis of different Asian and European countries.

II. To estimate the overall causal effect of vitamin D status on tuberculosis disease in a meta-analysis of the Asian and European country-level results.

1.4 Significance of the Study

Tuberculosis is a major cause of ill health and death. There is, therefore, a need for intervention studies aimed at the prevention and treatment of tuberculosis. Findings from this study will help inform whether vitamin D deficiency has a causal effect on tuberculosis. Since vitamin D deficiency can be readily corrected for example by supplementing food products such as flour, the results from this study might help in the prevention of tuberculosis. Results will also inform whether the suggestion of further RCTs of vitamin D for the prevention of tuberculosis is warranted.

2 Chapter 2: Literature Review

2.1 An introduction to Tuberculosis

Tuberculosis (TB) is an infectious disease caused by bacteria of the Mycobacterium tuberculosis complex (MTC) (Castets, Boisvert, Grumbach, Brunel, & Rist, 1968; Källenius et al., 1999; De Jong et al., 2009). Tuberculosis is thought to have spread to other parts of the world from Africa (Gutierrez et al., 2005). Ancient Mycobacterium tuberculosis (Mtb) strains, which originated from environmental mycobacteria, do not transmit in humans although they can still be isolated in immunocompromised individuals in some parts of East Africa (Supply et al., 2013; Delogu, Sali, & Fadda, 2013). Increase in population density especially in urban areas has led to the emergence of a more virulent and highly transmissible strain called modern Mtb that is responsible for most cases of TB (Gagneux, 2012; Wirth et al., 2008).

Mtb infection occurs when an individual inhales the aerosol droplets of tubercle bacilli. When these bacilli enter the alveoli of the lungs, they trigger innate immune responses (Urdahl, Shafiani, & Ernst, 2011). Failure of the innate immune response to eliminate the pathogen leads to its multiplication inside the macrophage and later spread to other tissues and organs via the bloodstream and lymphatic fluid (Wolf et al., 2008; Balasubramanian et al., 1996). This triggers the cell-mediated immune functions that control the replication of the bacteria by forming a granuloma that encapsulates the bacilli (Ottenhoff & Kaufmann, 2012). The granuloma later calcifies and resides in different organs and tissues in a dormant and metabolically inactive state called latent or dormant TB (Neyrolles et al., 2006; Hernandez-Pando et al., 2000; Feldman & Baggenstoss, 1938). When cell-mediated immunity weakens, the bacteria replicates actively leading to active TB (Gengenbacher & Kaufmann, 2012). Failure of the cell-mediated immune response to initially control the replication of the bacteria leads to the primary TB disease which is particularly seen in children (T. A. Thomas, 2017). Ingestion of unpasteurised milk is also a route for transmission of zoonotic tuberculosis which is caused by Mycobacterium bovis (De la Rua-Domenech, 2006). Tuberculosis manifests in different forms but the most common is pulmonary TB.

2.2 An overview of vitamin D

Vitamin D is a fat-soluble prohormone that exists in two forms, vitamin D3 (cholecalciferol), which is the animal form and vitamin D2 (ergocalciferol), which is the plant form. Vitamin D3 is primarily produced in the skin by ultraviolet B (UVB) irradiation on 7-

dehydrocholesterol, a precursor of cholesterol biosynthesis(Sylvia, DareV, Puneet, AdamJ, & LeilaJ, 2010). Some vitamin D is also obtained from the diet through absorption in the intestine, which is essential for individuals not exposed to the sun, and for at-risk groups including growing children, pregnant women and the elderly (Zhu et al., 2012; Yun et al., 2017; Schilling, 2012). Vitamin D3 is converted to 25-hydroxyvitamin D (25(OH)D) in the liver, which is the main circulating form of vitamin D and which is used to measure serum-25(OH)D levels. 25(OH)D is later converted into 1,25-dihydroxyvitamin D (1,25(OH)2D) in the kidney, which is the active vitamin D metabolite (Sylvia et al., 2010). Serum vitamin D levels are subject to several factors including genetics, skin pigmentation, season, latitude, use of sunscreens, clothing style, nutrition and supplements (Natasja & Lips, 2011). Low serum vitamin D levels have also been reported among individuals suffering from chronic kidney disease (Franca Gois, Wolley, Ranganathan, & Seguro, 2018).

2.3 Association between vitamin D and TB

2.3.1 Vitamin D supplementation and TB trials

There are two randomized controlled trials (RCTs) testing the effect of vitamin D supplementation on the incidence of tuberculosis disease/infection that has been published so far. One of the RCTs enrolled 8,851 (4,418 in the supplementation group; 4,433 in the placebo group) children aged between 6 and 13 years old who were followed up for three years in Mongolia. The inclusion criteria were a negative result for Mtb infection at baseline according to the QuantiFERON-TB Gold In-Tube assay [QFT, Qiagen]. After a 3-year follow-up, results from this study showed that vitamin D supplementation (weekly oral dose of 14,000 IU of vitamin D3 supplements) did not affect the prevention of tuberculosis infection (adjusted risk ratio [aRR] 1.10; 95% confidence interval [CI] 0.87-1.38) or tuberculosis disease (aRR 0.87; 95% CI 0.49-1.55) (Ganmaa et al., 2020).

The second RCT study enrolled 4,000 (2,001 in the supplementation group; 1,999 in the placebo) adults above 18 years and living with the human immune-deficiency virus (HIV) who were initiating antiretroviral therapy (ART) at the time of randomization. The inclusion criteria were: a serum-25(OH)D concentration of less than 30 ng/ml as quantified with a commercial enzyme immunoassay (Immunodiagnostics; Boldon, UK); non-pregnant women; participants not enrolled in another trial; and intention to stay in Dar es -Salaam, Tanzania for at least a year. TB was diagnosed using sputum smears that were stained using the Ziehl-Nielsen technique to examine for acid-fast bacilli (AFB) and chest X-ray. The cartridge-based Xpert MTB/RIF (Xpert) assay (GeneXpert System; CA, USA) was used to test participants who were sputum AFB negative but still suspected for TB. Supplementation involved a weekly oral dose of 50000 IU of vitamin D3 supplements for the first month followed by a daily oral dose of 2000 IU of vitamin D3 supplements. After 1-year follow-up, results from this study also found no impact of Vitamin D supplementation on the incidence of pulmonary tuberculosis (hazard ratio [HR] 0·78, 95% CI

0·54–1·13; p=0.19) or mortality (HR 1·04; 95% CI 0·85–1·25; p=0·78).(Sudfeld et al., 2020) They suggested that further RCTs might be warranted.

2.3.2 Observational studies of the relationship between vitamin D and TB

The relationship between vitamin D and tuberculosis has been suggested since as early as the eighteenth century when fish oil was used to treat tuberculosis of the skin.(G. Rook, 1988). In the 1940s, Solaria rooms that concentrate the sun's rays were used to treat tuberculosis (Mehta, 1939). Many epidemiological studies and meta-analyses also provide evidence supporting an association between vitamin D levels and tuberculosis. The following are some examples of recent studies. Results from a nested case-control study, systematic review and individual-participant data meta-analyses of household contacts of TB patients in Peru reported that low serum vitamin D only predicts TB disease risk in a dose-dependent manner which strengthens a claim for a causal association. In this study, the meta-analysis results revealed that vitamin D deficiency (<50 nmol/L) was associated with increased risk of progression to TB disease before (odds ratio [OR] 1.49; 95% CI 1.07–2.07; p=0.02), and after adjusting for age, gender, BMI, and HIV status (adjusted odds ratio [aOR] 1.48; 95% CI 1.04–2.10; p=0.03). Vitamin D insufficiency (50–75 nmol/L) was also associated with progression to TB disease after adjusting for age, gender, BMI, and HIV status (aOR 1.33; 95% CI 1.00–1.78; p=0.05) (Aibana et al., 2019). In their meta-analysis, Huang and colleagues (S. J. Huang et al., 2017) found that vitamin D deficiency was associated with an increased risk of TB (OR 2.57; 95% CI 1.74-3.80; P=0.00001; $I^2 = 83$) with similar results being observed in a subgroup analysis of the Asian population (OR 2.62; 95% CI 1.63-4.23; P=0.0001; I2=71). They also found that vitamin D deficiency was associated with an increased risk of developing active TB in latent TB infected (LTBI) subjects (OR 4.26; 95% CI 2.48-7.30; P=0.00001; $I^2 = 48\%$). Additionally, they reported that vitamin D deficiency was associated with an increased risk of tuberculin skin test (TST) conversion (OR 3.99; 95% CI 1.88-8.45; P=0.0003; $I^2 = 0$). In a meta-analysis of 3,998 immune-suppressed individuals, Zeng and colleagues (Zeng et al., 2015) reported an increased risk of active TB in individuals with serum 25(OH)D levels 12.5 nmol/l (pooled OR 4.556; 95% CI 2.200–9.435; $I^2 = 11.9\%$; P<0.001) and within serum vitamin D levels between 13–25 nmol/L (pooled OR 3.797; 95% CI 1.935–7.405; $I^2 = 84.1\%$; P<0.001).

A cross-sectional study of household contacts in Mongolia reported that low levels of serum 25(OH)D were associated with latent tuberculosis infection (LTBI) (OR 0.92, 95% CI 0.88–0.97) after adjusting for age, sex, BMI, socio-economic, behavioural, and clinical characteristics (Gurjav et al., 2019). Additionally, a case-control study of prisoners in Brazil found that low serum vitamin D levels (<30 ng/ml) were associated with active TB in both univariate (OR 4.37; 95% CI 1.56–12.25) and multivariate analyses (aOR 3.77; 95% CI 1.04–13.64; after adjusting for drug use, previous incarceration and black race) (Macedo et al., 2018). Pre-antiretroviral therapy (pre-ART) vitamin D deficiency was also associated with the risk of incident TB (adjusted hazard ratio [aHR] 3.66, 95%

CI 1.16–11.51) in a multinational case-cohort study of HIV-infected adults initiating ART (Tenforde et al., 2017). In a Korean case-control study, vitamin D deficiency (<20 ng/ml) was also associated with a higher odds of TB (OR 2.64; 95% CI 1.33-5.22) with severe vitamin D deficiency (<10 ng/ml) being associated with an even higher odds of TB (OR 11.75; 95% CI 6.53-21.15) (Hong et al., 2014). Venturini and colleagues also reported that vitamin D deficiency (<25 nmol/L) was correlated with TB infection (risk ratio [RR] 1.61; 95% CI:1.086-2.388; P=0.018) with a higher correlation reported in active-TB compared to latent TB and controls (RR 4.587; 95% CI: 1.190-9.608; P<0.0001), in a case-control study of children in the United Kingdom and Italy (Venturini et al., 2014). Moreover, a prospective study of 572 TB contacts in Spain found that reduced vitamin D levels (<20 ng/ml) were associated with risk of TB incidence (HR 0.88; 95%CI 0.80–0.97) (Arnedo-Pena et al., 2015).

2.3.3 Mechanisms underlying an association between Vitamin D and TB

Vitamin D has long been known for its role in the skeletal system. Studies on the extra-skeletal functions of vitamin D began in the late 20th century. These roles include lipid metabolism, cell proliferation and differentiation (Holick, 2003b, 1995, 2003a; DeLuca, 2004), insulin resistance (Chiu, Chu, Go, & Saad, 2004; Scragg, Sowers, & Bell, 2004), endocrine functions (DeLuca, 2004), inflammation (Gysemans et al., 2005; Haddad Kashani et al., 2018; Hossein-Nezhad et al., 2013) and immune functions (Cantorna, Zhu, Froicu, & Wittke, 2004; Yang, Smith, Prahl, Luo, & DeLuca, 1993; Manolagas, Provvedini, & Tsoukas, 1985). The functions of vitamin D in the immune system span across both innate and adaptive host immune responses. 1,25-dihydroxyvitamin D induces production of the cathelicidin antimicrobial peptide (LL-37), in lymphocytes, neutrophils, human alveolar macrophages, and epithelial cells (Rivas-Santiago et al., 2008; Ashenafi et al., 2018). In vitro studies have shown that calcitriol (the active form of vitamin D, also known as 1,25-dihydroxy vitamin D), in these cells up-regulates the expression of LL-37 by the human cationic antimicrobial protein-18 (Hcap-18) gene (Martineau et al., 2007; Liu, Stenger, Tang, & Modlin, 2007; Rekha et al., 2015; Liu et al., 2006). LL-37 directly kills mycobacteria by osmotic lysis (Choi, Chow, & Mookherjee, 2012),66 and also induces autophagy, increases chemotaxis in macrophages, and regulates the production of chemokines and the secretion of cytokines (Skodric-Trifunovic et al., 2014). In vitro studies show that after infection, Toll-like receptors (TLR) are activated in human macrophages and monocytes (Liu et al., 2009). This increases the expression of vitamin D receptor and vitamin D-1-hydroxylase genes and therefore induces the antimicrobial peptide, cathelicidin which kills intracellular mycobacteria. When calcitriol binds to a membrane VDR on a Mtb infected macrophage, it induces a phagocyte superoxide burst which is mediated by a nicotinamide adenine dinucleotide phosphate (NADPH)-dependent phagocyte oxidase (Sly, Lopez, Nauseef, & Reiner, 2001; G. A. Rook et al., 1986). It also reduces the viability of MTB by enhancing phagolysosomal fusion in a phosphatidylinositol-3-kinases (PI3K)-dependent manner (Liu et al., 2009). Moreover, binding of 1,25-dihydroxyvitamin

D to the nuclear VDR upregulates innate functions such as induction of nitric oxide (NO) synthase and hence increased production of the toxic NO (Rockett et al., 1998).

2.4 GWAS studies

2.4.1 Genes associated with vitamin D status

Four genes associated with serum 25-hydroxyvitamin D (25(OH)D) concentrations have majorly been reported in different genome-wide association studies (GWAS). These genes are in the vitamin D synthesis and metabolism pathway (Figure 1). The group-specific component (GC) gene encodes the vitamin D binding protein (DBP), a protein that binds and transports vitamin D metabolites (Speeckaert, Huang, Delanghe, & Taes, 2006). The 7-dehydrocholesterol reductase (DHCR7) gene encodes an enzyme that converts 7-dehydrocholesterol to cholesterol using NADPH hence removing the substrate from the synthetic pathway of vitamin D3. The nicotinamide adenine dinucleotide synthetase-1 (NADSYN1) gene encodes an enzyme involved in the last step in nicotinamide adenine dinucleotide (NAD) biosynthesis (Hara et al., 2003; Waterham & Wanders, 2000). The DHCR7 and NADSYN1 genes are together described as the DHCR7/NADSYN1 gene locus. The cytochrome P450 2R1 (CYP2R1) gene encodes microsomal vitamin D 25-hydroxylase, an enzyme that catalyses the hydroxylation of vitamin D3 at carbon-25 to an active ligand for the vitamin D receptor (Cheng, Levine, Bell, Mangelsdorf, & Russell, 2004). The cytochrome P450 family 24 subfamily A member 1 (CYP24A1) gene encodes 24-hydroxylase, an enzyme that initiates the degradation of 1,25-dihydroxy vitamin D (1,25(OH)2D) and 25(OH)D by catalysing the hydroxylation reactions at the side-chain (Jones, Prosser, & Kaufmann, 2012).

2.4.2 GWAS of Vitamin D

Genome-wide association study (GWAS) is a method of analysing genetic variations that are related to a phenotype by comparing genotype frequencies in different individuals in an observational study. Several GWAS of serum 25-hydroxyvitamin D (25(OH)D) concentrations has been done in populations of different ethnicities. GWAS of 417,580 individuals of European ancestry from the UK Biobank (UKB) identified 143 loci that were associated with 25(OH)D concentration. 18,864 single nucleotide polymorphisms (SNPs) were genome-wide significant ($P < 510^{-8}$) with SNPs on the GC and CYP2R1 loci being the most associated ($P < 1.010^{-400}$) (Revez et al., 2020). A different GWAS of 401,460 individuals of European ancestry (white British) from the UKB reported 69 loci that were associated with 25(OH)D concentration. Among these were the four genes on the vitamin D metabolism and synthesis pathway. The study also identified 138 genome-wide significant and conditionally independent SNPs ($P < 6.610^{-9}$) (Manousaki et al., 2020). The SUNLIGHT (Study of Underlying Genetic Determinants of Vitamin D and Highly Related Traits) consortium GWAS of 79,366 European-ancestry individuals identified six

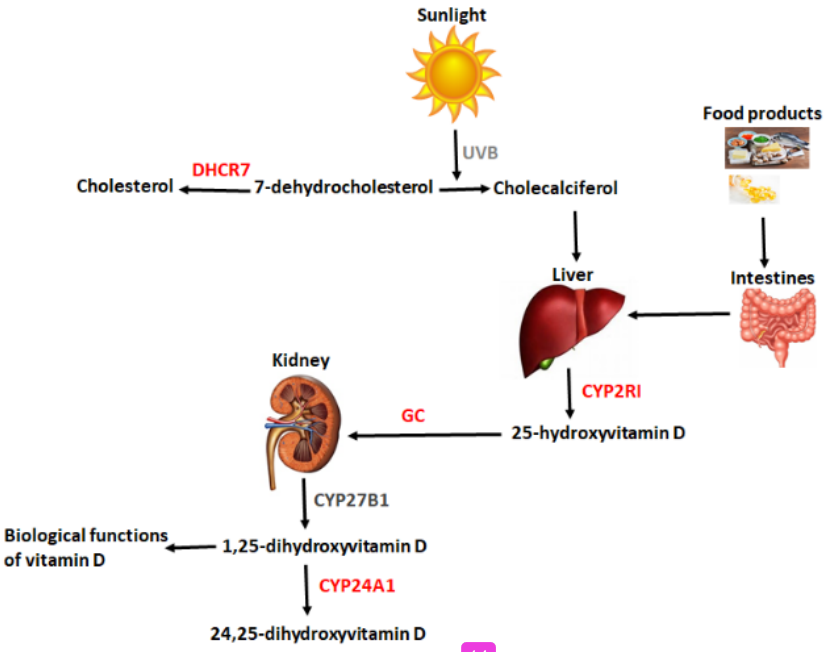

Figure 1. Genes associated with serum vitamin D levels

loci (GC, CYP2R1, DHCR7/NADSYN1, CYP24A1, Sec23 Homolog A, coat protein complex II (COPII) component [SEC23A], Amidohydrolase domain containing 1 [AMDHD1]) to be associated with 25(OH)D (Jiang et al., 2018). Four of these loci had been reported in a previous GWAS of about 30,000 individuals of European descent by the same consortium (GC, CYP2R1, CYP24A1, DHCR7/NADSYN1) (T. J. Wang et al., 2010). An earlier GWAS of 4501 individuals of European ancestry had identified SNPs in the GC, DHCR7/NADSYN1 and CYP2R1 genes that were associated with 25(OH)D concentrations (Ahn et al., 2010). GWAS of 1,829 US women reported 32 SNPs located in or near GC and CYP2R1 loci that had genome-wide significant associations ($P < 510^{-8}$) with 25(OH)D concentrations (O'Brien et al., 2018). GWAS of serum 25(OH)D on 3538 individuals from Asian (Indian) population in Punjabi reported six loci that were associated with 25(OH)D concentration with only one reaching the genome-wide significance (Forkhead Box A2 gene [FOXA2]: $p = 4.4710^{-9}$). The association between GC and CYP2R1 top variants with 25(OHD) concentrations in Europeans was also confirmed in the Punjabi sample (GC; $p = 0.007$ and CYP2R1; $p = 0.019$) (Sapkota et al., 2016).

2.4.3 Previous studies on vitamin D polymorphisms and TB

The genomic function of vitamin D is mediated by the vitamin D receptor (VDR). VDR is a nuclear hormone receptor, present on monocytes, activated T and B lymphocytes, and many other cells and tissues, that binds 1,25-dihydroxyvitamin D (1,25(OH)_2D) (Y. Wang, Zhu, & DeLuca, 2012). Once 1,25(OH)_2D binds to the nuclear VDR it reg-

ulates the transcription of vitamin D-responsive genes and also activates non-genomic signal transduction pathways in the cells (Bikle, 2014; Gil, Plaza-Diaz, & Mesa, 2018).

Currently, no study has tested the association between vitamin D polymorphisms on the GC, CYP2R1, CYP24A1 and DHCR7/NADSYN1 genes and the risk of TB. However, studies have found an association between VDR polymorphisms and TB. In a Taiwanese population, genotypes, AA of rs73126, GG of rs1544410 and A carriers of rs7041 were associated with susceptibility to TB (Lee et al., 2016). Three SNPs (rs11574143, rs11574079, and rs11168287) in the VDR gene were also associated with pulmonary tuberculosis in a Tibetan-Chinese population (Hu et al., 2016; Varahram et al., 2009). In a meta-analysis of 42 well-controlled studies, two variants (rs1544410 and rs731236) on the VDR gene were significantly associated with TB in the overall population and the south Asians in an ethnicity analysis (Xu & Shen, 2019). In the subgroup analyses, the rs7975232 variant was associated with extra-pulmonary TB while the rs2228570 variant was associated with both pulmonary and extra-pulmonary TB.

2.5 Mendelian Randomization

2.5.1 Describing the technique

Mendelian randomization uses genetic variants as instrumental variables in observational data to determine if an exposure (risk factor) causes a disease outcome (D. C. Thomas & Conti, 2004). Genetic variants proxy the differences in the risk factor. A genetic variant is a section of deoxyribonucleic acid (DNA), mostly a single nucleotide polymorphism (SNP) that differs between individuals. These genetic variants have to satisfy the instrumental variables assumptions which state that: (Greenland, 2000) (Figure 2)

16

I. The genetic variant must be associated with the exposure.

II. The genetic variant should not be associated with confounders of the exposure-outcome relationship.

III. The genetic variant should only be associated with the outcome through the exposure.

In MR, instrumental variables mimic the random assignment of treatments as in an experimental study (Figure 3) (Nitsch et al., 2006). Although randomized controlled trials (RCTs) are the most robust means of evaluating interventions, they are expensive and their capacity is likely to remain limited. The advantage of using MR is that the method reduces bias due to confounding and reverse causation. This is because genetic variants are randomly assorted during meiosis and are therefore independent of environmental

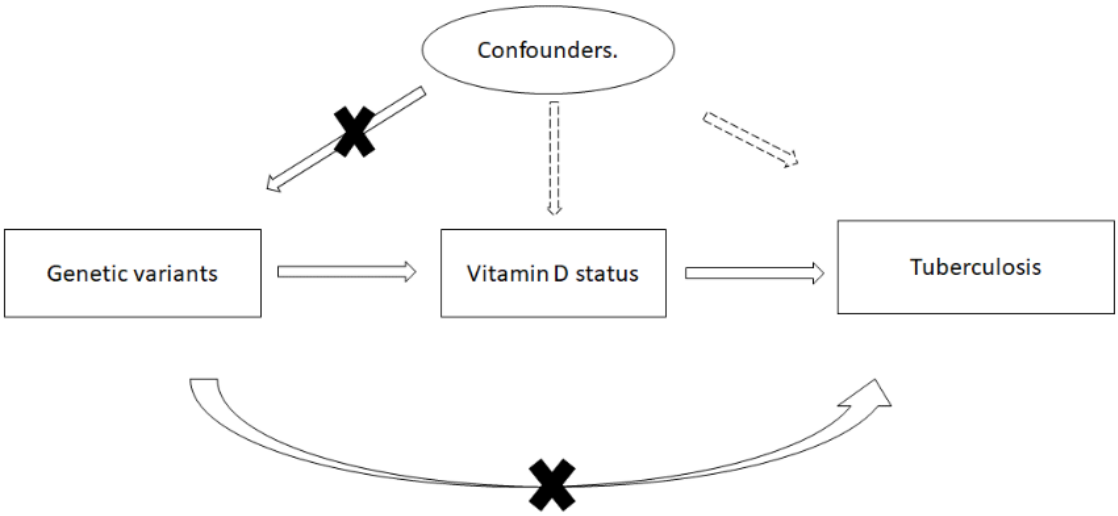

Figure 2. Assumptions of the genetic variants for Mendelian randomization

factors. MR has also been used to confirm the results of RCTs and to predict null results in RCTs costing upwards of USD 100 million (Labos, Brophy, Smith, Sniderman, & Thanassoulis, 2018; Yarmolinsky et al., 2018).

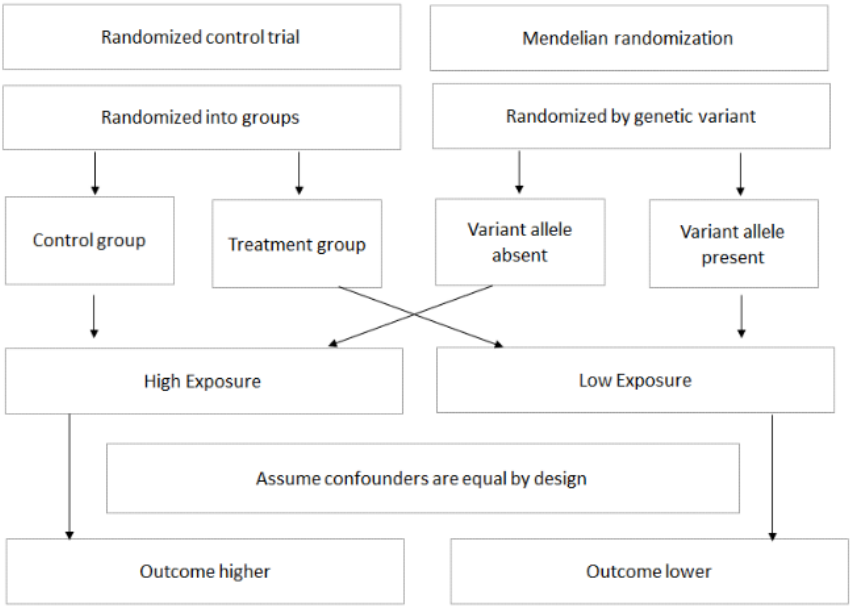

Figure 3. Comparison of Mendelian randomization and randomized controlled trial

Two sample-MR is an MR that uses summary statistics (that is beta-coefficients and standard errors) from two different, non-overlapping studies (GWAS of vitamin D and GWAS of TB in our study) to estimate whether a risk factor causally affects an outcome by estimating the IV-exposure and IV-outcome associations (Burgess et al., 2015). Summarized data are available from large consortia of genetic association studies where estimates

are obtained from regression models of either the exposure or the outcome on a genetic variant. These data are often from large sample sizes, which increases the power of the analysis. The data can also be easily accessed and this allows reproducibility of the analysis and hence ensures transparency.

2.5.2 Previous vitamin D MR studies

Many Mendelian randomization (MR) studies between genetically predicted serum-25(OH)D levels and different phenotypes have been conducted with some reporting a causal association. For example, vitamin D deficiency was found to have a causal relationship with increased all-cause mortality in two MR studies of 10,501 participants of European ancestry and 95,766 participants of Danish descent. These studies used genetic scores of SNPs on the DHCR7 and CYP2R1 loci to estimate the hazard ratio (HR) for mortality per 20 nmol/L decrease in genetically determined 25(OH)D levels. The HR estimates for the two genetic score in the European population study was 1.32 (95% CI: 0.84 to 2.20) and 1.35 (95%CI of 0.81 to 2.37) while that of the Danish population was 1.30 (95% CI: 1.051.61) (Aspelund et al., 2019; Afzal, Brøndum-Jacobsen, Bojesen, & Nordestgaard, 2014). The odds ratio (OR) of Alzheimer's disease (AD) per genetically-predicted one standard deviation increase in serum 25(OH)D was 0.86 (95% CI 0.78–0.94; $p = 0.002$) in a two-sample MR of 54,162 individuals of European ancestry (Larsson, Traylor, Markus, & Michaëlsson, 2018). A different MR of 54,162 individuals of European ancestry demonstrated that a one standard deviation decrease in natural log-transformed 25(OH)D increased the risk of AD (OR 1.25; 95% CI 1.03– 1.51; $p = 0.021$) (Mokry et al., 2016). Causal association between serum-25(OH)D and risk of ovarian cancer was reported in a two-sample MR study of 31,719 women of European ancestry with an OR of 1.17 (95% CI 0.76–1.78) per 20-nmol/L decrease in 25(OH)D levels (Ong et al., 2016). In a meta-analysis of ten studies of European and Asian populations (428,904 individuals), a 25 nmol/L increase in genetically determined 25(OH)D concentration was associated with a lower risk of diabetes (OR 0.86; 95% CI 0.77-0.97; $p = 0.01$) using a genetic score of DHCR7 and CYP2R1 synthesis SNPs (Lu et al., 2018). MR analyses of non-Hispanic whites (10,071 individuals) and Swedish population (12,097 participants) showed that increasing levels of 25(OH)D are associated with a decreased risk of multiple sclerosis (OR 0.79; 95% CI 0.64-0.99) and (OR 0.86; 95% CI 0.76–0.98) respectively. A meta-analysis of the two populations resulted in a combined OR of 0.85 (95% CI 0.76–0.94) (Rhead et al., 2016). One standard deviation decrease of genetically determined log-transformed 25(OH)D levels was also associated with an increased risk of multiple sclerosis (OR 2.02; 95% CI; 1.65–2.46; $p = 7.7210^{12}$) in an MR of 38,589 MS participants of European ancestry (Mokry et al., 2015). Genetically elevated levels of serum-25(OH)D was associated with a decreased risk of delirium (HR 0.74; 95% CI 0.62–0.87; $p = 0.0004$) in an MR of 313,121 participants of European descent from the UK biobank (Bowman et al., 2019).

3 Chapter 3: Methods

3.1 Data Sources for MR Analyses

I conducted a two-sample Mendelian randomization using summary statistics from non-overlapping Vitamin D and TB GWAS datasets. The Vitamin D GWAS datasets provided the exposure data while the TB GWAS datasets provided the outcome data. The flow diagram below shows the steps I took in conducting the MR (Figure 4).

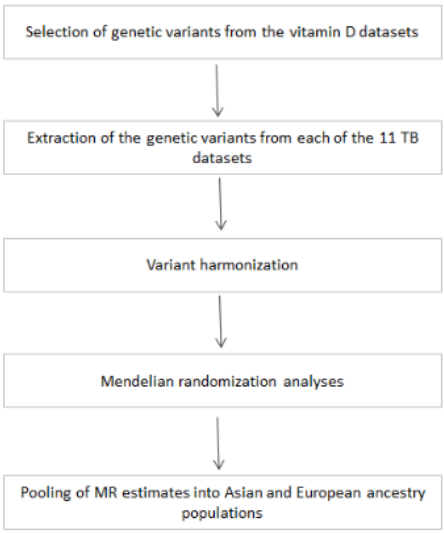

Figure 4. Flow diagram of analyses conducted.

3.1.1 Vitamin D GWAS Studies

I selected the instrumental variables for the European ancestry population, from the UK Biobank (UKB) vitamin D GWAS (Revez et al., 2020). The UKB is a population cohort of over 502,000 European individuals, aged between 40 and 69 years old who were recruited between 2006 and 2010 (Bycroft et al., 2018). Phenotypic, genotypic and clinical data were collected on these individuals. The North West Multicentre Research Ethics Service Committee approved the study and all participants signed informed consent. Individuals of European ancestry were identified using the 1000 Genome project reference cluster. Genome-wide genetic variation of the single nucleotide polymorphism (SNPs) was captured using the UK Biobank Axiom Array (Applied Biosystems) which contains 825,927 markers. The vitamin D GWAS study conducted a genome-wide association of 25-hydroxyvitamin D (25(OH)D) concentration in 417580 individuals from the UKB and combined them in a meta-analysis with results from a previous GWAS of 79,366 Euro-

peans (Revez et al., 2020; Jiang et al., 2018). 25(0H)D levels were measured using DiaSorin LIASON® chemiluminescent immunoassay which has a detection range of 10 nmol/L < 25(OH)D < 375 nmol/L (Revez et al., 2020).

To date, there is only one GWAS of serum 25(OH)D on individuals of Asian ancestry which recruited 3538 individuals from an Indian population in Punjabi (Sapkota et al., 2016). This study replicated an association between GC ($p = 0.007$) and CYP2R1 ($p = 0.019$) loci with 25(OHD) concentrations (T. J. Wang et al., 2010). However, due to the low number of study participants, I choose to use SNPs that were found to be associated with serum 25(OH)D levels in previous Asian MR studies (T. Huang et al., 2019; Lu et al., 2018). These MR studies genotyped 95,680 individuals randomly selected from the China Kadoorie Biobank (CKB) using a 384-SNP panel (Illumina GoldenGate). Specific SNPs (CYP2R1-rs10741657, DHCR7- rs12785878, GC-rs2282679, CYP24A1-rs6013897 identified from a previous genome-wide study in Europeans84 were included in the genotyping panel. SNPs in the DHCR7, CYP2R1, and GC loci were significantly associated with plasma 25(OH)D concentration while the CYP24A1 SNP had a consistent direction of effect with the European GWAS although the association was not significant (Lu et al., 2018).

3.1.2 TB GWAS Studies

The outcome data were summary statistics from eight individual tuberculosis GWAS studies across Asia and Europe from the International Tuberculosis Host Genetics Consortium (ITHGC) (Figure 5). This ongoing collaboration has collated a total of 15,573 TB cases and 24,843 controls (5309 cases and 8033 controls of Asian ancestry and 6739 cases and 13386 controls of European ancestry) genotyped data.

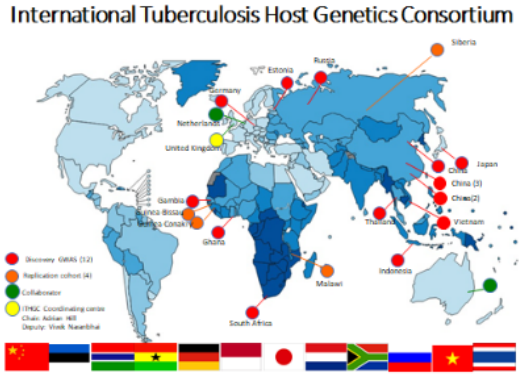

Figure 5. Map of the countries participating in the International Tuberculosis Host Genetics Consortium

3.2 MR Assumptions

Different assumptions have to be met while conducting a Mendelian randomization analysis. The first instrumental variable (IV) assumption is the assumption of relevance that states that the genetic variant should be associated with the serum-25(OH)D levels. The second IV assumption is the assumption of independence, which states that the variant should not be associated with any confounder of the serum-25(OH)D levels- TB association. The third IV assumption is the exclusion restriction, which states that the association between the variant and TB should only be through the serum-25(OH)D levels. These assumptions were addressed in the selection of genetic variants and also in the sensitivity analyses as described in the sections below.

3.3 Selection of genetic variants

The instrumental variables were single nucleotide polymorphisms (SNPs) associated with serum vitamin D levels (25(OH)D) at genome-wide significance ($p < 510^{-8}$) (Table 3). I chose eight SNPs with the lowest p-values based on reported GWAS associations and their use in previous MR studies (Aspelund et al., 2019; Larsson et al., 2018; Mokry et al., 2015).These SNPs are on the GC (rs2282679), CYP24A1 (rs17216707 and rs6013897), DHCR7/NADSYN1 (rs11234027 and rs12785878) and CYP2R1 (rs12794714 and rs10741657) genes. The genes encode proteins in the vitamin D synthesis and metabolic pathway and since they have biological relevance to exposure, they are more plausible to the assumption of relevance. To ensure that the SNPs were independent, I used the 1000 genomes project data as a reference cluster to estimate linkage disequilibrium between the SNPs. SNPs that had a coefficient of determination ($R^2 = 1$) were in linkage disequilibrium and therefore the SNP with the lowest p-value in their association with serum-25(OH)D were retained.

I systematically screened the selected variants in PhenoScanner `http://www.phenoscanner.medschl.cam.ac.uk/`, a database that gives the human genotype-phenotype association, to ensure the variants were not directly associated with any confounder of the exposure-outcome association in fulfilment of the MR assumption of independence. Further testing for the fulfilment of the instrumental variable assumptions was done in the sensitivity analyses discussed below.

I calculated the statistical power for the Mendelian randomization with a binary outcome using a web-based application `https://sb452.shinyapps.io/power/`. The input variables included: the TB data sample size; the ratio of TB cases to controls; coefficient of determination (R^2) of serum vitamin D levels on genetic variants (13%), as reported in the vitamin D GWAS (Revez et al., 2020); the change in outcome per standard error change in exposure as reported in an observational study (2.57) (S. J. Huang et al., 2017); and a type-I error rate of 0.05.

3.4 Mendelian Randomization Analyses

All Mendelian randomization (MR) tests were done in the statistical software R version 3.6.1 (Team et al., 2020), using the TwoSampleMR R Package `https://mrcieu.github.io/TwoSampleMR/authors.html` (Hemani et al., 2018).

3.4.1 Variant Harmonization

I performed variant harmonization to ensure that the effect allele of the selected variant on the serum-25(OH)D dataset is the same effect allele on the TB dataset (Table 4).

3.4.2 Primary Analysis

In the primary analysis, I used the inverse-variance weighted (IVW) method. Mathematically, we can characterise an instrumental variable (IV) as a random variable by assuming that we have an outcome Y that is a function of a measured exposure X and unmeasured confounders U. We express X as a function of the genetic variant G and the confounder U. We, therefore, rewrite the IV assumptions as:

I. G is associated with X
II. G is independent of U
III. G is independent of Y conditional on X and U

Let $\hat{\beta}_{Yj}$ be the genetic association with a standard error σ_{Yj} of an individual variant j= 1.....J with TB while $\hat{\beta}_{Xj}$ is the genetic association with a standard error σ_{Xj} of the variant with serum-25(OH)D concentrations. The genetic variants are not in linkage disequilibrium and are therefore assumed to be independently distributed. The genetic variants are also oriented such that the $\hat{\beta}_{Xj}$ estimates are all either positive or negative. The IVW estimate is calculated by a meta-analysis of each variant's ratio estimate. The ratio estimate of the j^{th} genetic variant is denoted as $\hat{\beta}_{IVj}$ and is calculated as : (Lawlor, Harbord, Sterne, Timpson, & Davey Smith, 2008)

$$\hat{\beta}_{IVj} = \frac{\hat{\beta}_{Yj}}{\hat{\beta}_{Xj}} \tag{1}$$

Since I was using non-overlapping datasets, I assumed that the correlation between $\hat{\beta}_{Yj}$ and $\hat{\beta}_{Xj}$ is zero (Pierce & Burgess, 2013), hence variance is:

$$Var(\hat{\beta}_{IVj}) = \frac{\sigma^2_{Yj}}{\hat{\beta}^2_{Xj}} \tag{2}$$

and the standard error is:

$$SE(\hat{\beta}_{IV\,j}) = \frac{\sigma_{Y\,j}}{\hat{\beta}_{X\,j}} \tag{3}$$

The IVW estimate is a weighted mean of the individual variant causal estimates which is calculated as:(Borenstein, Hedges, Higgins, & Rothstein, 2011)

$$\hat{\beta}_{IVW} = \frac{\sum_j W_j \hat{\beta}_{IV\,j}}{\sum_j W_j} \tag{4}$$

where

$$W_j = SE(\hat{\beta}_{IV\,j})^{-2} \tag{5}$$

are the inverse variance weights. The standard error (assuming a fixed effect model) is:

$$SE(\hat{\beta}_{IVW}) = \sqrt{\frac{1}{\sum_j \hat{\beta}_{X\,j}^2 \sigma_{Y\,j}^{-2}}} \tag{6}$$

I used the multiplicative random-effect IVW model. This model ensures that the relative weighting of each variant estimate does not change. The model also calculates the heterogeneity explained by each variant's causal estimate (Bowden et al., 2017). This model assumes that the $\hat{\beta}_{IV\,j}$ estimates are normally distributed with mean β and variance $\phi_M^2 \sigma_{IV\,j}^2$.

$$\hat{\beta}_{IV\,j} \sim N(\beta, \phi_M^2 \sigma_{IV\,j}^2) \tag{7}$$

This model allows the estimation of the residual standard error, which is equivalent to the heterogeneity parameter ϕ_M as part of the model. The point estimate from a multiplicative random-effects model is similar to the point estimate from a fixed-effect model (Thompson & Sharp, 1999). This model also assumes balanced pleiotropy (pleiotropic effects on TB has an equal possibility of either being positive or negative) to provide valid causal estimates (Bowden et al., 2017). I went ahead and performed sensitivity analyses to test how robust of our primary analysis results are. Analyses were done separately for each country and then pooled for each broad ethnicity.

3.4.3 Sensitivity Analysis

The robust analyses are meant to address horizontal pleiotropy (the MR assumption of exclusion restriction). They also allow the inference of valid causal effect although their assumptions are weaker compared to the primary analysis assumptions (Burgess et al., 2019). I used the weighted median regression and MR-Egger analyses for the sensitivity

analysis MR-Egger analysis is a weighted linear regression model of the $\hat{\beta}_{Yj}$ on the $\hat{\beta}_{Xj}$ where all the $\hat{\beta}_{Xj}$ associations are oriented to be all either positive or negative (Bowden, Davey Smith, & Burgess, 2015).

$$\hat{\beta}_{Yj} = \alpha_{0E} + \alpha_E \hat{\beta}_{Xj} \tag{8}$$

The intercept α_{0E} estimates the average horizontal pleiotropic effect (direct effect of the variants on TB that are not through serum-25(OH)D levels). The causal effect is the slope α_E estimated from a weighted regression of the genetic variant-TB association on the genetic variant-serum 25(OH)D levels association. The model is similar to that of IVW except that it includes an intercept term (α_E equals $\hat{\beta}_{IVW}$ when α_{0E} is zero.) When the intercept differs from zero, then there is horizontal pleiotropy and hence the IVW estimator is biased. The α_E estimate represents the true effect of causality Uuder the Instrument Strength Independent of Direct Effect (InSIDE) assumption. (Bowden et al., 2015). This assumption allows genetic variants to have pleiotropy independent of the variant-serum 25(OH)D levels associations and hence some heterogeneity in the causal estimates of individual variants may be observed. This heterogeneity does not invalidate the analysis since it is a consequence of weakening IV assumptions. The weighted median regression works under the majority valid assumption which states that 50% of the selected genetic variants are valid IVs (Bowden et al., 2015). The weighted median estimator is the median of a distribution having $\hat{\beta}_{IVj}$ as its

$$P_{IVj} = 100(S_j - \frac{W_j}{2}) \tag{9}$$

th percentile, where

$$S_j = \sum_{j=1}^{j} W_j \tag{10}$$

This method takes the variant-specific causal estimate and calculates the central tendency of the select estimates. The median of the distribution is not influenced by the outlying estimates.

3.4.4 Scatter Plot and Test for Heterogeneity

Using the Cochran's Q test, I did a heterogeneity test to establish whether the IV causal estimates vary based on the individual genetic variant (Bowden et al., 2019). This test relies on the assumption that all the genetic variants identify a similar causal parameter. H_0: IV causal estimates are homogeneous H_1: IV causal estimates are heterogeneous The statistic is calculated as:

$$Q = \sum_j W_j (\hat{\beta}_{IVj} - \hat{\beta}_{IVW})^2 \tag{11}$$

and has a chi-square distribution with J-1 degrees of freedom, where, J denotes the total genetic variants. Besides, I did an I-squared (I^2) heterogeneity test to measure the percentage of variability in the effect estimates. This is calculated as:

$$I^2 = (\frac{Q - df}{Q})x100\% \tag{12}$$

Where Q is the Cochran's Q statistic and df is the degree of freedom of the Q. This was also visualised using a scatter plot of the genetic association with TB against the genetic association with serum-25(OH)D levels, including their confidence intervals. Each point on the graph represents an individual genetic variant. Under the null hypothesis, these points should lie on a straight line through the origin. Variants that lay outside the line were tested for pleiotropy.

3.4.5 Funnel Plot and Test for Directional Pleiotropy

In addition to the MR-Egger test for directional pleiotropy, I also used funnel plots. A funnel plot is used to measure the precision of the IV estimates and to test for directional pleiotropy(Bowden et al., 2015). I plotted the reciprocal of the standard error of the IV estimates

$$\frac{1}{SE(\hat{\beta}_{IVj}} = \frac{\hat{\beta}_{Xj}}{\sigma_{Yj}} \tag{13}$$

against the IV estimates

$$\frac{\hat{\beta}_{Yj}}{\hat{\beta}_{Xj}} \tag{14}$$

Asymmetry in the funnel plot signifies horizontal pleiotropy. This indicates that the estimated causal effect of the risk factor on the outcome may be biased.

3.5 Meta-analyses

I pooled the country IVW estimates to represent two ethnic populations; Asian and European and conducted meta-analyses. I also pooled the IVW estimates for all the countries and carried out an overall meta-analysis. I used the I^2 heterogeneity test to measure the percentage of variability of the effect estimates. I also used the tau-squared (τ^2) which explains the between-study (between-country in this case) variance. I used both the random and fixed effect models to calculate the pooled effect estimate of serum 25(OH)D levels on tuberculosis. I used the metagen function in the meta package of R statistical software (Balduzzi, Rücker, & Schwarzer, 2019).

4 Chapter 4: Results

4.0.1 Extraction of the Instrumental Variables

I conducted Mendelian randomization (MR) analyses to determine the effect of genetically predicted serum 25-hydroxyvitamin D (25(OH)D) on the odds of tuberculosis disease. Instrumental variables (IV) were extracted from the UK Biobank vitamin D GWAS and were genetic variants (SNPs) that had a genome-wide association ($p = 510^{-8}$) with serum-25(OH)D77 (Table 3). SNPs on the same loci were found to be in linkage disequilibrium ($R^2 = 1$) after clumping to test for independence of the IVs. The SNP with the lowest p-value in their association with serum-25(OH)D were retained and I, therefore, remained with four IVs. The effect allele of all the SNPs were oriented such that the effect size estimates (beta) of the genetic association with serum-25(OH)D were all positive. I extracted the four IVs from the different outcome datasets. I did harmonization to ensure that the effect allele of an IV on the exposure dataset was the same effect allele on the outcome dataset (Table 4-5).

Table 1. Effects of genetic variants on serum 25(OH)D levels

SNP	Gene	EA/OA	EAF	N	Beta	SE	P-value	Var
rs10741657	CYP2R1	A/G	0.40	417580	0.09	0.002	$1.00e^{-300}$	0.0038
rs12785878	DHCR7/NADSYN1	T/G	0.79	417580	0.11	0.002	$1.00e^{-300}$	0.0040
rs2282679	GC	T/G	0.71	417580	0.19	0.002	$1.00e^{-300}$	0.0150
rs6013897	CYP24A1	T/A	0.80	417041	0.04	0.003	$4.25e^{-50}$	0.0005

EA, effect allele; OA, other allele; EAF, effect allele frequency; N, sample size; Beta, effect size of the serum-25(OH)D-increasing allele on rank-based inverse normal transformed (RINT) 25OHD levels; SE, standard error; Var, variance in 25(OH)D levels explained by each SNP

China1 data						
SNP	N	EA/OA	EAF	Beta	SE	P-value
rs1074165	1070	A/G	0.38	0.04	0.13	0.37
rs12785878	1070	T/G	0.48	-0.04	0.13	0.35
rs2282679	1070	T/G	0.67	0.06	0.13	0.33
rs6013897	1070	T/A	0.83	0.06	0.17	0.37
China2 data						
SNP	N	EA/OA	EAF	Beta	SE	P-value
rs10741657	2390	A/G	0.38	0.01	0.06	0.41

rs12785878	2390	T/G	0.52	-0.06	0.06	0.17
rs2282679	2390	T/G	0.69	0.13	0.06	0.02
rs6013897	2390	T/A	0.84	0.10	0.08	0.12
China3 data						
SNP	N	EA/OA	EAF	Beta	SE	P-value
rs10741657	2509	A/G	0.37	-0.15	0.06	0.01
rs12785878	2509	T/G	0.48	0.02	0.06	0.34
rs2282679	2509	T/G	0.69	0.09	0.06	0.07
rs6013897	2509	T/A	0.84	-0.02	0.08	0.37
Japan data						
SNP	N	EA/OA	EAF	Beta	SE	P-value
rs10741657	3950	A/G	0.36	-0.05	0.06	0.21
rs12785878	3950	T/G	0.33	-0.02	0.06	0.38
rs2282679	3950	T/G	0.73	-0.02	0.07	0.36
rs6013897	3950	T/A	0.91	0.10	0.12	0.18
Thailand data						
SNP	N	EA/OA	EAF	Beta	SE	P-value
rs10741657	1431	A/G	0.3	0.05	0.08	0.27
rs12785878	1431	T/G	0.3	-0.1	0.09	0.14
rs2282679	1431	T/G	0.75	0.15	0.12	0.1
rs6013897	1431	T/A	0.81	-0.11	0.1	0.15
Estonia data						
SNP	N	EA/OA	EAF	Beta	SE	P-value
rs10741657	7286	A/G	0.43	0.05	0.10	0.29
rs12785878	7286	T/G	0.65	0.18	0.10	0.03
rs2282679	7286	T/G	0.72	-0.03	0.10	0.38
rs6013897	7285	T/A	0.82	0.21	0.12	0.04
Germany data						
SNP	N	EA/OA	EAF	Beta	SE	P-value
rs10741657	919	A/G	0.39	0.09	0.11	0.21

rs12785878	919	T/G	0.68	0.04	0.11	0.37
rs2282679	919	T/G	0.73	0.04	0.12	0.37
rs6013897	919	T/A	0.81	-0.15	0.14	0.13
Russia data						
SNP	N	EA/OA	EAF	Beta	SE	P-value
rs10741657	11445	A/G	0.42	-0.03	0.03	0.13
rs12785878	11445	T/G	0.68	-0.05	0.03	0.04
rs2282679	11445	T/G	0.70	0.04	0.03	0.09
rs6013897	11445	T/A	0.83	-0.002	0.04	0.47

Table 2. Effect of the instrumental variables on the odds of tuberculosis disease in each study country

4.1 Mendelian Randomization Results

I had 100% statistical power in all the countries to detect a causal association between serum-25(OH)D concentrations and the odds of tuberculosis disease except in Estonia which had 99.8% power and Germany with 99.9% power. The primary estimates, which were calculated by the inverse variance weighted (IVW) method revealed a lack of causal association between genetically elevated serum-25(OH)D and the risk of TB disease ($p > 0.05$) (Table 6-7). The sensitivity analyses using MR-Egger and weighted median confirmed the lack of association (Table 8). These results were also visualised using forest plots (Figure 6).

China1 results				
SNP	Sample size	Beta	SE	P-value
rs10741657	1070	0.48	1.43	0.73
rs12785878	1070	-0.44	1.15	0.70
rs2282679	1070	0.30	0.70	0.67
rs6013897	1070	1.54	4.48	0.73
All - Inverse variance weighted	1070	0.18	0.55	0.74
All - MR Egger	1070	0.08	1.31	0.96
All - Weighted median	1070	0.24	0.57	0.70
China2 results				
SNP	Sample size	Beta	SE	P-value

rs10741657	2390	0.16	0.69	0.81
rs12785878	2390	-0.52	0.55	0.34
rs2282679	2390	0.67	0.33	0.04
rs6013897	2390	2.62	2.18	0.23
All - Inverse variance weighted	2390	0.35	0.33	0.28
All - MR Egger	2390	0.51	0.95	0.65
All - Weighted median	2390	0.43	0.30	0.15
China3 results				
SNP	Sample size	Beta	SE	P-value
rs10741657	2509	-1.66	0.69	0.02
rs12785878	2509	0.22	0.54	0.69
rs2282679	2509	0.49	0.34	0.15
rs6013897	2509	-0.74	2.18	0.73
All - Inverse variance weighted	2509	0.10	0.43	0.82
All - MR Egger	2509	1.17	0.95	0.34
All - Weighted median	2509	0.36	0.30	0.23
Japan results				
SNP	Sample size	Beta	SE	P-value
rs10741657	3950	-0.54	0.68	0.43
rs12785878	3950	-0.17	0.58	0.77
rs2282679	3950	-0.12	0.34	0.73
rs6013897	3950	2.60	2.86	0.37
All - Inverse variance weighted	3950	-0.17	0.27	0.53
All - MR Egger	3950	-0.33	0.70	0.68
All - Weighted median	3950	-0.14	0.29	0.62
Thailand results				
SNP	Sample size	Beta	SE	P-value
rs10741657	1431	0.56	0.93	0.55
rs12785878	1431	-0.88	0.80	0.27
rs2282679	1431	0.78	0.61	0.20

rs6013897	1431	-2.82	2.69	0.29
All - Inverse variance weighted	1431	0.18	0.50	0.72
All - MR Egger	1431	1.43	1.02	0.30
All - Weighted median	1431	0.62	0.50	0.22
Estonia results				
SNP	Sample size	Beta	SE	P-value
rs10741657	7286	0.58	1.05	0.58
rs12785878	7286	1.66	0.90	0.07
rs2282679	7286	-0.16	0.54	0.76
rs6013897	7286	5.62	3.26	0.08
All - Inverse variance weighted	7286	0.45	0.58	0.44
All - MR Egger	7286	-1.41	0.99	0.29
All - Weighted median	7286	0.22	0.48	0.65
Germany results				
SNP	Sample size	Beta	SE	P-value
rs10741657	919	0.98	1.23	0.42
rs12785878	919	0.34	1.02	0.74
rs2282679	919	0.20	0.61	0.74
rs6013897	919	-4.04	3.60	0.26
All - Inverse variance weighted	919	0.27	0.48	0.57
All - MR Egger	919	0.85	1.11	0.52
All - Weighted median	919	0.26	0.51	0.62
Russia results				
SNP	Sample size	Beta	SE	P-value
rs10741657	11445	-0.35	0.31	0.26
rs12785878	11445	-0.47	0.28	0.09
rs2282679	11445	0.22	0.16	0.18
rs6013897	11445	-0.07	1.02	0.95
All - Inverse variance weighted	11445	-0.02	0.18	0.90
All - MR Egger	11445	0.42	0.39	0.40

All - Weighted median	11445	0.04	0.15	0.77

Table 3. Mendelian randomization results for the effect of genetically elevated serum-25(OH)D on the odds of tuberculosis disease

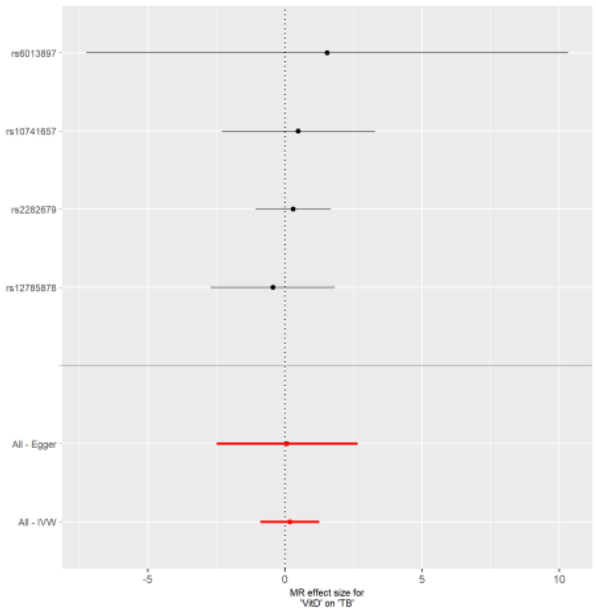

Figure 6. China 1

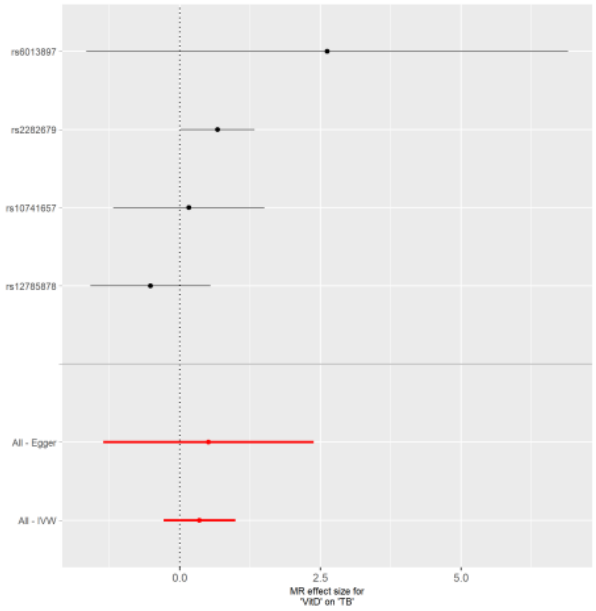

Figure 7. China 2

Figure 8. Forest plots results for the effect of genetically predicted serum-25(OH)D on the odds of tuberculosis disease

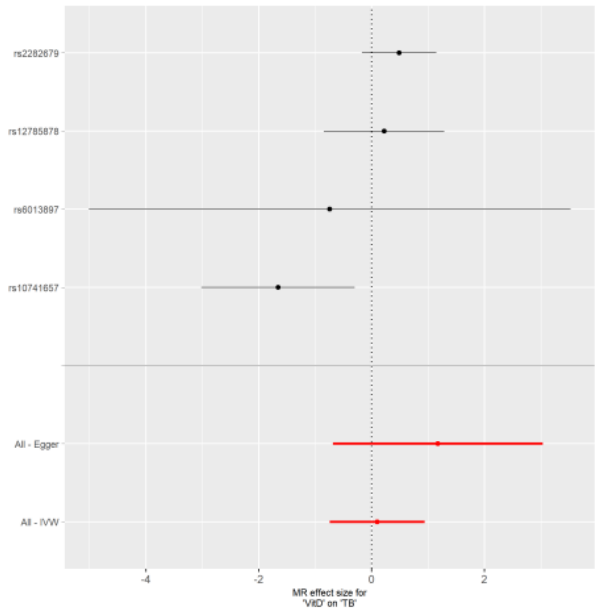

Figure 8. China 3

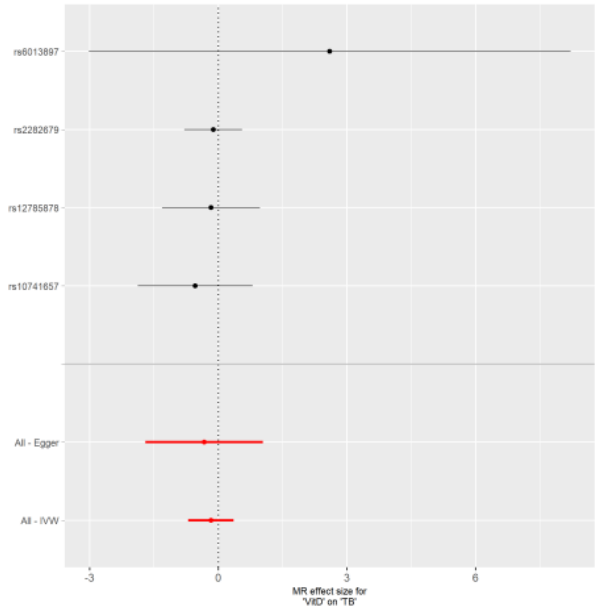

Figure 9. Japan

Figure 10. Forest plots results for the effect of genetically predicted serum-25(OH)D on the odds of tuberculosis disease

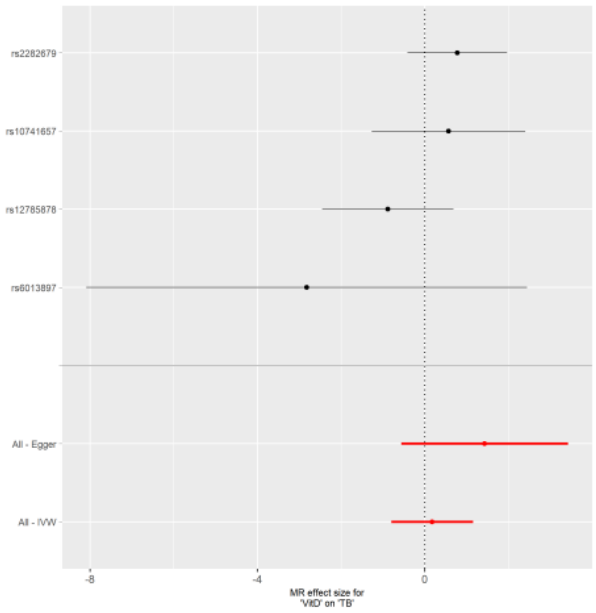

Figure 10. Thailand

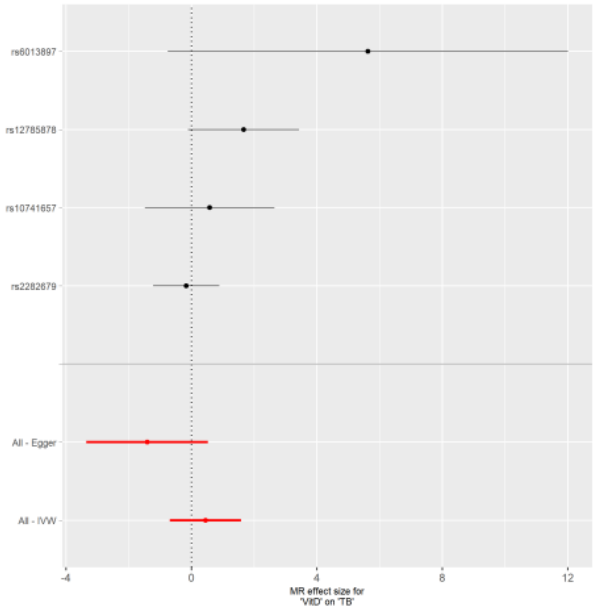

Figure 11. Estonia

Figure 12. Forest plots results for the effect of genetically predicted serum-25(OH)D on the odds of tuberculosis disease

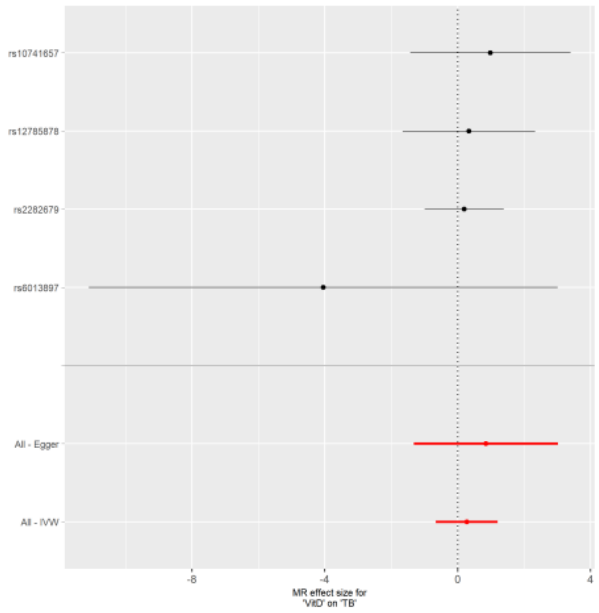

Figure 12. Germany

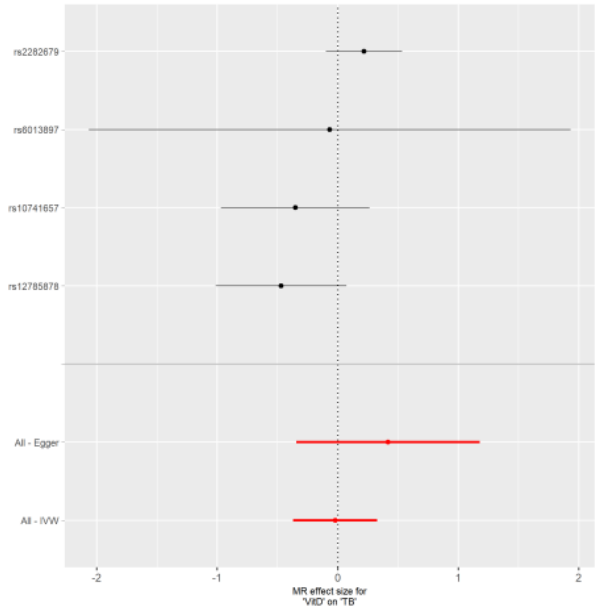

Figure 13. Russia

Figure 14. Forest plots results for the effect of genetically predicted serum-25(OH)D on the odds of tuberculosis disease

4.1.1 Heterogeneity and Pleiotropy Results

Heterogeneity was initially assessed using a scatter plot of genetic association with tuberculosis against the genetic association with serum-25(OH)D concentrations (Figure 7). Most of the genetic variants (points on the graph) in the different countries lied away

from the straight line through the origin and were therefore tested for pleiotropy. Further tests for heterogeneity of the IV estimates based on individual genetic variant were done using the Cochrane's Q test under the null hypothesis of homogeneity. Results for the Cochrane's Q test showed a lack of heterogeneity ($P \geq 0.05$) among the genetic variants and therefore confirms that our select IVs were valid (Table 6). The amount of heterogeneity was expressed using the I-squared statistic. An I-squared below 50% represents homogeneity and only one of the country had slight heterogeneity (China 3; $I^2 = 62\%$) (Table 9). The MR-Egger intercepts are the average pleiotropic effect of a genetic variant. Under the null hypothesis of no pleiotropy, the intercepts should not differ from zero. Results of the test for directional pleiotropy showed no evidence of pleiotropy ($P \geq 0.05$) (Table 9). These results further validated our choice of genetic variants.

Table 4. Heterogeneity and Pleiotropy results

Country	Q	Q-df	Q-pval	I^2	egger-intercept	egger-se	egger-pval
China1	0.46	3	0.93	0	0.01	0.16	0.94
China2	4.64	3	0.20	35	-0.02	0.12	0.87
China3	7.97	3	0.05	62	-0.15	0.12	0.34
Japan	1.25	3	0.74	0	0.02	0.09	0.83
Thailand	4.14	3	0.25	28	-0.16	0.11	0.31
Estonia	5.65	3	0.13	47	0.25	0.12	0.17
Germany	1.79	3	0.62	0	-0.08	0.14	0.62
Russia	5.96	3	0.11	50	-0.06	0.05	0.34

Q.

Cochrane's Q statistic for heterogeneity; Q_{df}, $Q statistic degrees of freedom$; I^2, test for heterogeneity; Egger, $intercept, MR-Egger measure of pleiotropy$; $Egger_{se}, MR-Eggerstandarderror$; $Egger_{p}val, MR-Eggerpvalue$

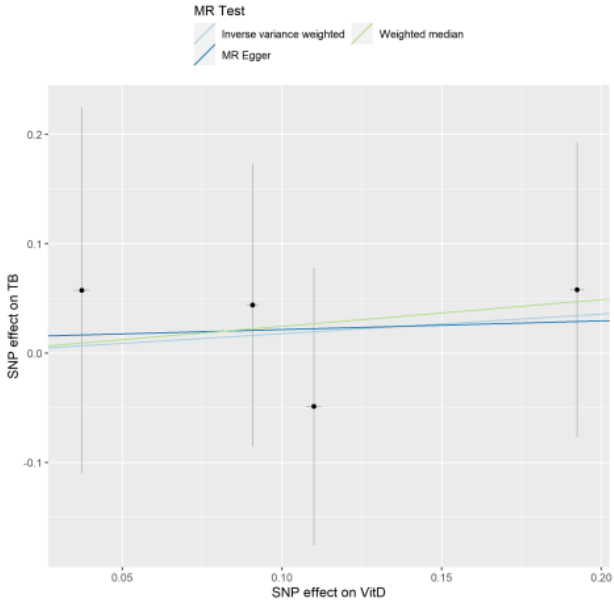

Figure 15. China 1

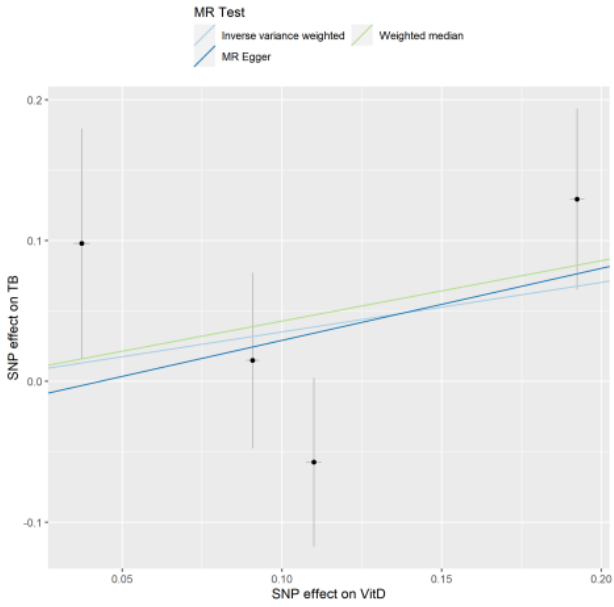

Figure 16. China 2

Figure 17. Forest plots results for the effect of genetically predicted serum-25(OH)D on the odds of tuberculosis disease

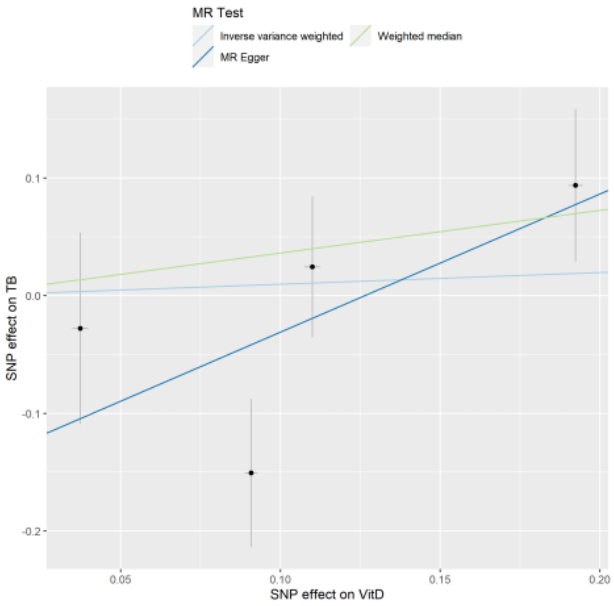

Figure 17. China 3

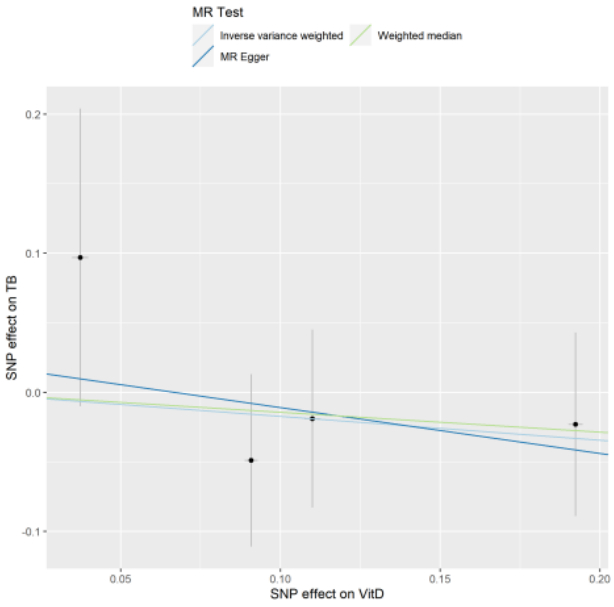

Figure 18. Japan

Figure 19. Forest plots results for the effect of genetically predicted serum-25(OH)D on the odds of tuberculosis disease

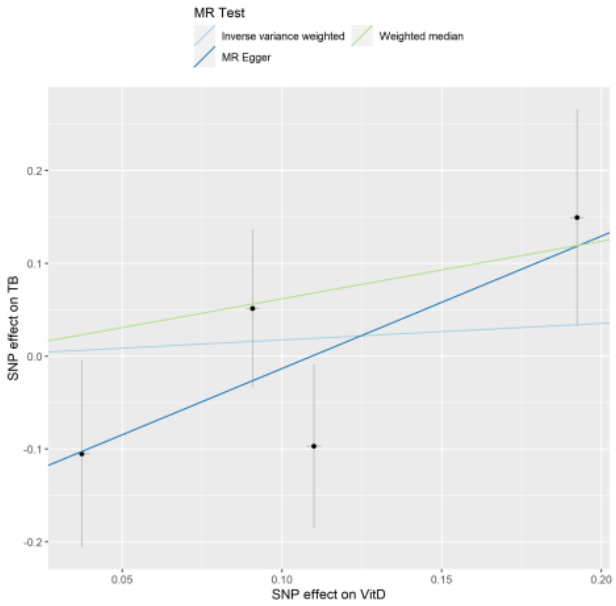

Figure 19. Thailand

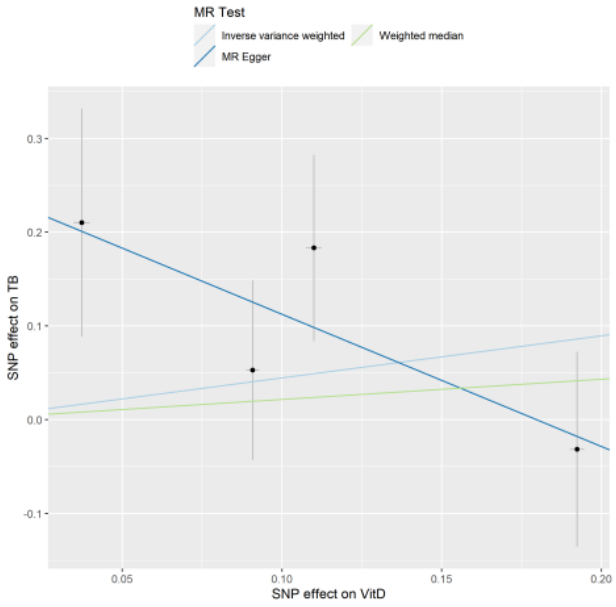

Figure 20. Estonia

Figure 21. Forest plots results for the effect of genetically predicted serum-25(OH)D on the odds of tuberculosis disease

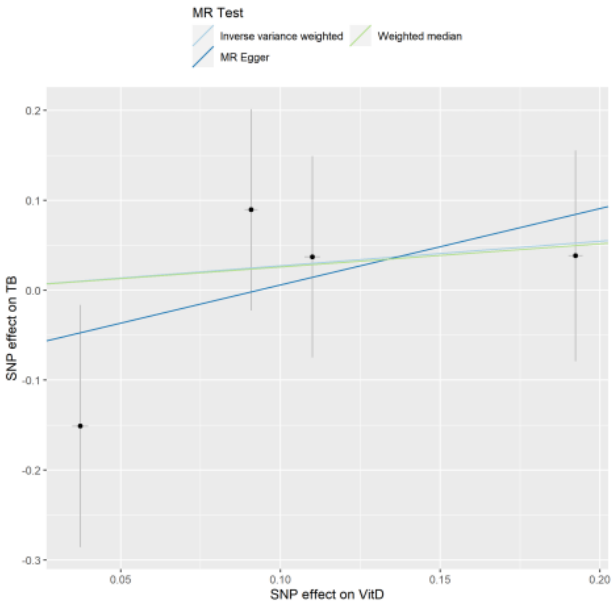

Figure 21. Germany

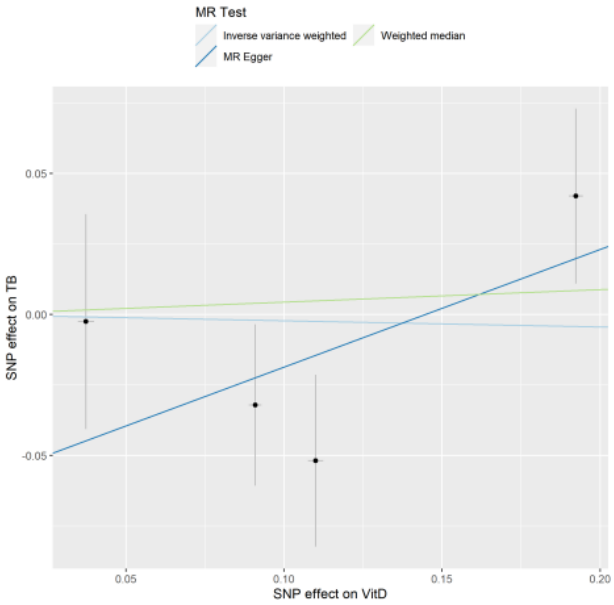

Figure 22. Russia

Figure 23. Forest plots results for the effect of genetically predicted serum-25(OH)D on the odds of tuberculosis disease

4.2 Meta-analyses

Using the Mendelian randomization inverse variance weighted (MR-IVW) results, I performed meta-analyses using the inverse variance weighted (IVW) - fixed effect model to test the overall association between genetically elevated serum-25(OH)D levels and the

odds of tuberculosis disease. I also tested the association within the Asian population and European population by doing meta-analyses of the MR-IVW estimates of the respective countries.

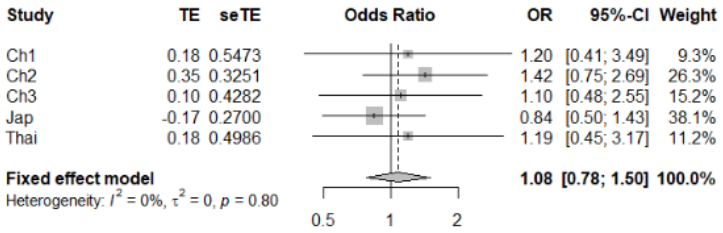

Figure 24. Meta-analysis of the MR-IVW estimates of the Asian countries

TE, Treatment effect which is the MR-IVW estimate of the odds of tuberculosis disease per standard deviation increase in genetically determined serum-25(OH)D levels; seTE, standard error of the treatment effect; OR, Odds ratio, 95%-CI, 95% confidence interval; I^2, I-squared test for heterogeneity; τ^2, tau-squared test for heterogeneity; p, p value; Ch1, China1 data; Ch2, China 2 data; Ch3, China 3 data; Jap, Japan data; Thai, Thailand data

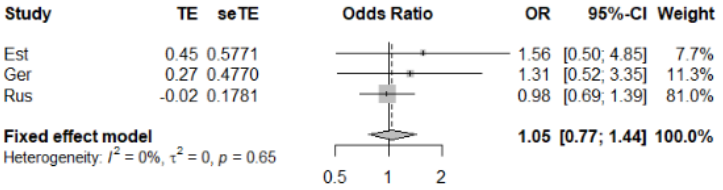

Figure 25. Meta-analysis of the MR-IVW estimates of the European countries

TE, Treatment effect which is the MR-IVW estimate of the odds of tuberculosis disease per standard deviation increase in genetically determined serum-25(OH)D levels; seTE, standard error of the treatment effect; OR, Odds ratio, 95%-CI, 95% confidence interval; I^2, I-squared test for heterogeneity; τ^2, tau-squared test for heterogeneity; p, p value; Est, Estonia data; Ger, Germany data; Rus, Russia data

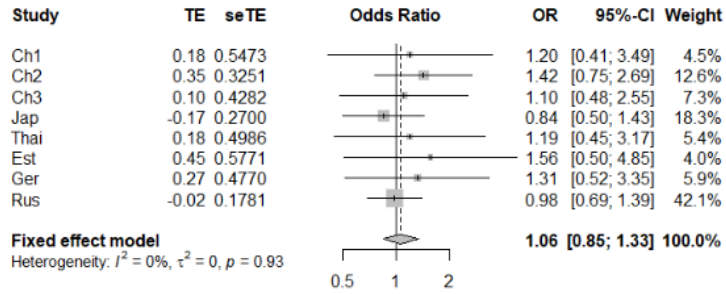

Figure 26. Meta-analysis of the MR-IVW estimates of both Asian and European countries

TE, Treatment effect which is the MR-IVW estimate of the odds of tuberculosis disease per standard deviation increase in genetically determined serum-25(OH)D levels; seTE, standard error of the treatment effect; OR, Odds ratio, 95%-CI, 95% confidence interval; I^2, I-squared test for heterogeneity; τ^2, tau-squared test for heterogeneity; p, p value; Ch1, China1 data; Ch2, China 2 data; Ch3, China 3 data; Jap, Japan data; Thai, Thailand data; Est, Estonia data; Ger, Germany data; Rus, Russia data

Meta-analyses of the IVW estimates showed no association between genetically elevated serum 25(OH)D levels and the odds of TB overall (odds ratio [OR]=1.08, 95% confidence interval [CI]=0.85-1.33, p=0.93) or in the two ethnic populations, European (OR=1.05, 95% CI=0.77-1.44, p=0.65) and Asian (OR=1.08, 95% CI=0.78-1.50, p=0.80) (Figure 8-10). I used the fixed-effect model within the meta-analysis since the IVW estimates had no evidence of both between and within-study heterogeneity ($I^2 = 0\%$, $\tau^2 = 0$).

5 Chapter 5: Discussion

Finding from our study suggest that genetically elevated serum-25(OH)D concentration were not causally associated with the odds of tuberculosis disease. The results are consistent with results from recently published randomized controlled trials (RCTs) testing the effect of vitamin D supplementation on the incidence of tuberculosis disease/infection. Sudfeld and colleagues found no impact of Vitamin D supplementation on mortality or the incidence of pulmonary tuberculosis in a large randomized controlled trial (RCT) enrolling adults living with HIV in Tanzania (Sudfeld et al., 2020). Vitamin D supplementation also showed no effect in the prevention of tuberculosis infection and disease in an RCT that enrolled children in Mongolia (Ganmaa et al., 2020). The difference in our study compared to the RCTs was the potential to capture all the sources of vitamin D which includes exposure to sunlight, diet and supplementation by use of genetic variants that affect serum-25(OH)D rather than supplementation alone as in RCTs. Besides, lifelong exposure to vitamin D is tested rather than the time-limited supplementation. Nonetheless, RCTs remain the gold standard and are the most robust means of evaluating potential interventions. The strengths of this study include the lack of potential confounding by environmental factors. These could be a possible explanation of the reported association between low vitamin D and the risk of TB reported in many observational studies. Since the outcome of disease does not affect the genetic make-up of an individual, the study also eliminated possible effects of reverse causation which might be found in observational studies. The study also had enough power to detect the causal association between serum-25(OH)D levels and TB disease. In addition, there were no violations of the assumptions of Mendelian randomization (MR) as far as they could be tested. Potential limitations of this study just like any other MR study is that I could not fully rule out the effect of possible genetic pleiotropy and linkage disequilibrium (LD) between the genetic variants that I used as instrumental variables. However, findings from the pleiotropy and independence tests showed no evidence of horizontal pleiotropy or LD to support the possibility. The cohorts in this study included individual of European and Asian ancestry and therefore these results may not apply to other ethnic populations. In conclusion, genetically elevated serum-25(OH)D does not has a causal association with tuberculosis disease.

5.1 Future Research

I hope to study the effect of genetically determined serum-25(0H)D on the risk of tuberculosis in the African population in the near future.

Bibliography

Afzal, S., Brøndum-Jacobsen, P., Bojesen, S. E., & Nordestgaard, B. G. (2014). Genetically low vitamin d concentrations and increased mortality: Mendelian randomisation analysis in three large cohorts. *Bmj*, *349*, g6330.

Ahn, J., Yu, K., Stolzenberg-Solomon, R., Simon, K. C., McCullough, M. L., Gallicchio, L., … others (2010). Genome-wide association study of circulating vitamin d levels. *Human molecular genetics*, *19*(13), 2739–2745.

Aibana, O., Huang, C.-C., Aboud, S., Arnedo-Pena, A., Becerra, M. C., Bellido-Blasco, J. B., … others (2019). Vitamin d status and risk of incident tuberculosis disease: A nested case-control study, systematic review, and individual-participant data meta-analysis. *PLoS medicine*, *16*(9), e1002907.

Arnedo-Pena, A., Juan-Cerdán, J., Romeu-García, A., García-Ferrer, D., Holguín-Gómez, R., Iborra-Millet, J., … Roach-Poblete, F. (2015). Vitamin d status and incidence of tuberculosis among contacts of pulmonary tuberculosis patients. *The International Journal of Tuberculosis and Lung Disease*, *19*(1), 65–69.

Ashenafi, S., Mazurek, J., Rehn, A., Lemma, B., Aderaye, G., Bekele, A., … others (2018). Vitamin d3 status and the association with human cathelicidin expression in patients with different clinical forms of active tuberculosis. *Nutrients*, *10*(6), 721.

Aspelund, T., Grübler, M. R., Smith, A. V., Gudmundsson, E. F., Keppel, M., Cotch, M. F., … others (2019). Effect of genetically low 25-hydroxyvitamin d on mortality risk: Mendelian randomization analysis in 3 large european cohorts. *Nutrients*, *11*(1), 74.

Balasubramanian, V., Pavelka, M. S., Bardarov, S. S., Martin, J., Weisbrod, T. R., McAdam, R. A., … Jacobs, W. R. (1996). Allelic exchange in mycobacterium tuberculosis with long linear recombination substrates. *Journal of bacteriology*, *178*(1), 273–279.

Balduzzi, S., Rücker, G., & Schwarzer, G. (2019). How to perform a meta-analysis with r: a practical tutorial. *Evidence-based mental health*, *22*(4), 153–160.

Bikle, D. D. (2014). Vitamin d metabolism, mechanism of action, and clinical applications. *Chemistry & biology*, *21*(3), 319–329.

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2011). *Introduction to meta-analysis*. John Wiley & Sons.

Bouillon, R., Marcocci, C., Carmeliet, G., Bikle, D., White, J. H., Dawson-Hughes, B., … others (2019). Skeletal and extraskeletal actions of vitamin d: current evidence and outstanding questions. *Endocrine reviews*, *40*(4), 1109–1151.

Bowden, J., Davey Smith, G., & Burgess, S. (2015). Mendelian randomization with invalid instruments: effect estimation and bias detection through egger regression. *International journal of epidemiology*, *44*(2), 512–525.

Bowden, J., Del Greco M, F., Minelli, C., Davey Smith, G., Sheehan, N., & Thompson, J.

(2017). A framework for the investigation of pleiotropy in two-sample summary data mendelian randomization. *Statistics in medicine*, *36*(11), 1783–1802.

Bowden, J., Del Greco M, F., Minelli, C., Zhao, Q., Lawlor, D. A., Sheehan, N. A., … Davey Smith, G. (2019). Improving the accuracy of two-sample summary-data mendelian randomization: moving beyond the nome assumption. *International journal of epidemiology*, *48*(3), 728–742.

Bowman, K., Jones, L., Pilling, L. C., Delgado, J., Kuchel, G. A., Ferrucci, L., … Melzer, D. (2019). Vitamin d levels and risk of delirium: A mendelian randomization study in the uk biobank. *Neurology*, *92*(12), e1387–e1394.

Burgess, S., Scott, R. A., Timpson, N. J., Smith, G. D., Thompson, S. G., Consortium, E.-I., et al. (2015). Using published data in mendelian randomization: a blueprint for efficient identification of causal risk factors. *European journal of epidemiology*, *30*(7), 543–552.

Burgess, S., Smith, G. D., Davies, N. M., Dudbridge, F., Gill, D., Glymour, M. M., … others (2019). Guidelines for performing mendelian randomization investigations. *Wellcome Open Research*, *4*.

Burgess, S., & Thompson, S. G. (2015). *Mendelian randomization: methods for using genetic variants in causal estimation*. CRC Press.

Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L. T., Sharp, K., … others (2018). The uk biobank resource with deep phenotyping and genomic data. *Nature*, *562*(7726), 203–209.

Cantorna, M. T., Zhu, Y., Froicu, M., & Wittke, A. (2004). Vitamin d status, 1, 25-dihydroxyvitamin d3, and the immune system. *The American journal of clinical nutrition*, *80*(6), 1717S–1720S.

Castets, M., Boisvert, H., Grumbach, F., Brunel, M., & Rist, N. (1968). Tuberculosis bacilli of the african type: preliminary note. *Revue de tuberculose et de pneumologie*, *32*(2), 179.

Cheng, J. B., Levine, M. A., Bell, N. H., Mangelsdorf, D. J., & Russell, D. W. (2004). Genetic evidence that the human cyp2r1 enzyme is a key vitamin d 25-hydroxylase. *Proceedings of the National Academy of Sciences*, *101*(20), 7711–7715.

Chiu, K. C., Chu, A., Go, V. L. W., & Saad, M. F. (2004). Hypovitaminosis d is associated with insulin resistance and β cell dysfunction. *The American journal of clinical nutrition*, *79*(5), 820–825.

Choi, K.-Y., Chow, L. N., & Mookherjee, N. (2012). Cationic host defence peptides: multifaceted role in immune modulation and inflammation. *Journal of innate immunity*, *4*(4), 361–370.

De Jong, B. C., Antonio, M., Awine, T., Ogungbemi, K., De Jong, Y. P., Gagneux, S., … others (2009). Use of spoligotyping and large sequence polymorphisms to study the population structure of the mycobacterium tuberculosis complex in a cohort study of consecutive smear-positive tuberculosis cases in the gambia. *Journal of clinical microbiology*, *47*(4), 994–1001.

De la Rua-Domenech, R. (2006). Human mycobacterium bovis infection in the united

kingdom: incidence, risks, control measures and review of the zoonotic aspects of bovine tuberculosis. *Tuberculosis*, *86*(2), 77–109.

Delogu, G., Sali, M., & Fadda, G. (2013). The biology of mycobacterium tuberculosis infection. *Mediterranean journal of hematology and infectious diseases*, *5*(1).

DeLuca, H. F. (2004). Overview of general physiologic features and functions of vitamin d. *The American journal of clinical nutrition*, *80*(6), 1689S–1696S.

Feldman, W. H., & Baggenstoss, A. H. (1938). The residual infectivity of the primary complex of tuberculosis. *The American journal of pathology*, *14*(4), 473.

Franca Gois, P. H., Wolley, M., Ranganathan, D., & Seguro, A. C. (2018). Vitamin d deficiency in chronic kidney disease: Recent evidence and controversies. *International journal of environmental research and public health*, *15*(8), 1773.

Gagneux, S. (2012). Host–pathogen coevolution in human tuberculosis. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1590), 850–859.

Ganmaa, D., Uyanga, B., Zhou, X., Gantsetseg, G., Delgerekh, B., Enkhmaa, D., … others (2020). Vitamin d supplements for prevention of tuberculosis infection and disease. *New England Journal of Medicine*, *383*(4), 359–368.

Gengenbacher, M., & Kaufmann, S. H. (2012). Mycobacterium tuberculosis: success through dormancy. *FEMS microbiology reviews*, *36*(3), 514–532.

Gil, A., Plaza-Diaz, J., & Mesa, M. D. (2018). Vitamin d: classic and novel actions. *Annals of Nutrition and Metabolism*, *72*(2), 87–95.

Greenland, S. (2000). An introduction to instrumental variables for epidemiologists. *International journal of epidemiology*, *29*(4), 722–729.

Gurjav, U., Ankhbat, M., Ganbaatar, G., Batjarga, K., Ochirbat, B., Baigal, D., … others (2019). Vitamin d deficiency is associated with tuberculosis infection among household contacts in ulaanbaatar, mongolia. *The International Journal of Tuberculosis and Lung Disease*, *23*(8), 919–923.

Gutierrez, M. C., Brisse, S., Brosch, R., Fabre, M., Omaïs, B., Marmiesse, M., … Vincent, V. (2005). Ancient origin and gene mosaicism of the progenitor of mycobacterium tuberculosis. *PLoS Pathog*, *1*(1), e5.

Gysemans, C. A., Cardozo, A. K., Callewaert, H., Giulietti, A., Hulshagen, L., Bouillon, R., … Mathieu, C. (2005). 1, 25-dihydroxyvitamin d3 modulates expression of chemokines and cytokines in pancreatic islets: implications for prevention of diabetes in nonobese diabetic mice. *Endocrinology*, *146*(4), 1956–1964.

Haddad Kashani, H., Seyed Hosseini, E., Nikzad, H., Soleimani, A., Soleimani, M., Tamadon, M. R., … Asemi, Z. (2018). The effects of vitamin d supplementation on signaling pathway of inflammation and oxidative stress in diabetic hemodialysis: a randomized, double-blind, placebo-controlled trial. *Frontiers in pharmacology*, *9*, 50.

Hara, N., Yamada, K., Terashima, M., Osago, H., Shimoyama, M., & Tsuchiya, M. (2003). Molecular identification of human glutamine-and ammonia-dependent nad synthetases carbon-nitrogen hydrolase domain confers glutamine dependency. *Journal of Biological Chemistry*, *278*(13), 10914–10921.

Hemani, G., Zheng, J., Elsworth, B., Wade, K. H., Haberland, V., Baird, D., … others (2018). The mr-base platform supports systematic causal inference across the human phenome. *Elife*, *7*, e34408.

Hernandez-Pando, R., Jeyanathan, M., Mengistu, G., Aguilar, D., Orozco, H., Harboe, M., … Bjune, G. (2000). Persistence of dna from mycobacterium tuberculosis in superficially normal lung tissue during latent infection. *The Lancet*, *356*(9248), 2133–2138.

Holick, M. F. (1995). Noncalcemic actions of 1, 25-dihydroxyvitamin d3 and clinical applications. *Bone*, *17*(2), S107–S111.

Holick, M. F. (2003a). Evolution and function of vitamin d. In *Vitamin d analogs in cancer prevention and therapy* (pp. 3–28). Springer.

Holick, M. F. (2003b). Vitamin d: A millenium perspective. *Journal of cellular biochemistry*, *88*(2), 296–307.

Holick, M. F. (2007). Vitamin d deficiency. *New England Journal of Medicine*, *357*(3), 266–281.

Hong, J., Kim, S., Chung, K., Kim, E., Jung, J., Park, M., … Kang, Y. A. (2014). Association between vitamin d deficiency and tuberculosis in a korean population. *The International journal of tuberculosis and lung disease*, *18*(1), 73–78.

Hossein-Nezhad, A., Mirzaei, K., Keshavarz, S., Ansar, H., Saboori, S., & Tootee, A. (2013). Evidences of dual role of vitamin d through cellular energy homeostasis and inflammation pathway in risk of cancer in obese subjects. *Minerva medica*, *104*(3), 295.

Hu, Q., Chen, Z., Liang, G., Mo, F., Zhang, H., Xu, S., … Jin, T. (2016). Vitamin d receptor gene associations with pulmonary tuberculosis in a tibetan chinese population. *BMC infectious diseases*, *16*(1), 469.

Huang, S. J., Wang, X.-H., Liu, Z.-D., Cao, W.-L., Han, Y., Ma, A.-G., & Xu, S.-F. (2017). Vitamin d deficiency and the risk of tuberculosis: a meta-analysis. *Drug design, development and therapy*, *11*, 91.

Huang, T., Afzal, S., Yu, C., Guo, Y., Bian, Z., Yang, L., … others (2019). Vitamin d and cause-specific vascular disease and mortality: a mendelian randomisation study involving 99,012 chinese and 106,911 european adults. *BMC medicine*, *17*(1), 160.

Jiang, X., O'Reilly, P. F., Aschard, H., Hsu, Y.-H., Richards, J. B., Dupuis, J., … others (2018). Genome-wide association study in 79,366 european-ancestry individuals informs the genetic architecture of 25-hydroxyvitamin d levels. *Nature communications*, *9*(1), 1–12.

Jones, G., Prosser, D. E., & Kaufmann, M. (2012). 25-hydroxyvitamin d-24-hydroxylase (cyp24a1): its important role in the degradation of vitamin d. *Archives of biochemistry and biophysics*, *523*(1), 9–18.

Källenius, G., Koivula, T., Ghebremichael, S., Hoffner, S. E., Norberg, R., Svensson, E., … Svenson, S. B. (1999). Evolution and clonal traits of mycobacterium tuberculosis complex in guinea-bissau. *Journal of clinical microbiology*, *37*(12), 3872–3878.

Labos, C., Brophy, J. M., Smith, G. D., Sniderman, A. D., & Thanassoulis, G. (2018). Evaluation of the pleiotropic effects of statins: a reanalysis of the randomized trial evi-

dence using egger regression—brief report. *Arteriosclerosis, thrombosis, and vascular biology*, *38*(1), 262–265.

Larsson, S. C., Traylor, M., Markus, H. S., & Michaëlsson, K. (2018). Serum parathyroid hormone, 25-hydroxyvitamin d, and risk of alzheimer's disease: A mendelian randomization study. *Nutrients*, *10*(9), 1243.

Lawlor, D. A., Harbord, R. M., Sterne, J. A., Timpson, N., & Davey Smith, G. (2008). Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. *Statistics in medicine*, *27*(8), 1133–1163.

Lee, S.-W., Chuang, T.-Y., Huang, H.-H., Liu, C.-W., Kao, Y.-H., & Wu, L. S.-H. (2016). Vdr and vdbp genes polymorphisms associated with susceptibility to tuberculosis in a han taiwanese population. *Journal of Microbiology, Immunology and Infection*, *49*(5), 783–787.

Liu, P. T., Schenk, M., Walker, V. P., Dempsey, P. W., Kanchanapoomi, M., Wheelwright, M., … others (2009). Convergence of il-1β and vdr activation pathways in human tlr2/1-induced antimicrobial responses. *PloS one*, *4*(6), e5810.

Liu, P. T., Stenger, S., Li, H., Wenzel, L., Tan, B. H., Krutzik, S. R., … others (2006). Toll-like receptor triggering of a vitamin d-mediated human antimicrobial response. *Science*, *311*(5768), 1770–1773.

Liu, P. T., Stenger, S., Tang, D. H., & Modlin, R. L. (2007). Cutting edge: vitamin d-mediated human antimicrobial activity against mycobacterium tuberculosis is dependent on the induction of cathelicidin. *The journal of immunology*, *179*(4), 2060–2063.

Lu, L., Bennett, D. A., Millwood, I. Y., Parish, S., McCarthy, M. I., Mahajan, A., … others (2018). Association of vitamin d with risk of type 2 diabetes: a mendelian randomisation study in european and chinese adults. *PLoS medicine*, *15*(5), e1002566.

Maceda, E. B., Goncalves, C. C., Andrews, J. R., Ko, A. I., Yeckel, C. W., & Croda, J. (2018). Serum vitamin d levels and risk of prevalent tuberculosis, incident tuberculosis and tuberculin skin test conversion among prisoners. *Scientific reports*, *8*(1), 1–9.

Manolagas, S. C., Provvedini, D. M., & Tsoukas, C. D. (1985). Interactions of 1, 25-dihydroxyvitamin d3 and the immune system. *Molecular and cellular endocrinology*, *43*(2-3), 113–122.

Manousaki, D., Mitchell, R., Dudding, T., Haworth, S., Harroud, A., Forgetta, V., … others (2020). Genome-wide association study for vitamin d levels reveals 69 independent loci. *The American Journal of Human Genetics*, *106*(3), 327–337.

Martineau, A. R., Wilkinson, K. A., Newton, S. M., Floto, R. A., Norman, A. W., Skolimowska, K., … others (2007). Ifn-γ-and tnf-independent vitamin d-inducible human suppression of mycobacteria: the role of cathelicidin ll-37. *The Journal of Immunology*, *178*(11), 7190–7198.

Mehta, P. (1939). Basic principles of solarium treatment. *The Indian Medical Gazette*, *74*(5), 291.

Mogire, R. M., Mutua, A., Kimita, W., Kamau, A., Bejon, P., Pettifor, J. M., … Atkinson, S. H. (2020). Prevalence of vitamin d deficiency in africa: a systematic review and meta-analysis. *The Lancet Global Health*, *8*(1), e134–e142.

Mokry, L. E., Ross, S., Ahmad, O. S., Forgetta, V., Smith, G. D., Leong, A., … Richards, J. B. (2015). Vitamin d and risk of multiple sclerosis: a mendelian randomization study. *PLoS Med*, *12*(8), e1001866.

Mokry, L. E., Ross, S., Morris, J. A., Manousaki, D., Forgetta, V., & Richards, J. B. (2016). Genetically decreased vitamin d and risk of alzheimer disease. *Neurology*, *87*(24), 2567–2574.

Natasja, M., & Lips, P. (2011). Worldwide vitamin d status. *Best Pract Res Clin Endocrinol Metab*, *25*, 671–680.

Neyrolles, O., Hernández-Pando, R., Pietri-Rouxel, F., Fornès, P., Tailleux, L., Payán, J. A. B., … others (2006). Is adipose tissue a place for mycobacterium tuberculosis persistence? *PloS one*, *1*(1), e43.

Nitsch, D., Molokhia, M., Smeeth, L., DeStavola, B. L., Whittaker, J. C., & Leon, D. A. (2006). Limits to causal inference based on mendelian randomization: a comparison with randomized controlled trials. *American journal of epidemiology*, *163*(5), 397–403.

Nnoaham, K. E., & Clarke, A. (2008). Low serum vitamin d levels and tuberculosis: a systematic review and meta-analysis. *International journal of epidemiology*, *37*(1), 113–119.

O'Brien, K. M., Sandler, D. P., Shi, M., Harmon, Q. E., Taylor, J. A., & Weinberg, C. R. (2018). Genome-wide association study of serum 25-hydroxyvitamin d in us women. *Frontiers in genetics*, *9*, 67.

Ong, J.-S., Cuellar-Partida, G., Lu, Y., Study, A. O. C., Fasching, P. A., Hein, A., … others (2016). Association of vitamin d levels and risk of ovarian cancer: a mendelian randomization study. *International journal of epidemiology*, *45*(5), 1619–1630.

Ottenhoff, T. H., & Kaufmann, S. H. (2012). Vaccines against tuberculosis: where are we and where do we need to go? *PLoS Pathog*, *8*(5), e1002607.

Pierce, B. L., & Burgess, S. (2013). Efficient design for mendelian randomization studies: subsample and 2-sample instrumental variable estimators. *American journal of epidemiology*, *178*(7), 1177–1184.

Rekha, R. S., Rao Muvva, S. J., Wan, M., Raqib, R., Bergman, P., Brighenti, S., … Agerberth, B. (2015). Phenylbutyrate induces ll-37-dependent autophagy and intracellular killing of mycobacterium tuberculosis in human macrophages. *Autophagy*, *11*(9), 1688–1699.

Revez, J. A., Lin, T., Qiao, Z., Xue, A., Holtz, Y., Zhu, Z., … others (2020). Genome-wide association study identifies 143 loci associated with 25 hydroxyvitamin d concentration. *Nature communications*, *11*(1), 1–12.

Rhead, B., Bäärnhielm, M., Gianfrancesco, M., Mok, A., Shao, X., Quach, H., … others (2016). Mendelian randomization shows a causal effect of low vitamin d on multiple sclerosis risk. *Neurology Genetics*, *2*(5).

Rivas-Santiago, B., Hernandez-Pando, R., Carranza, C., Juarez, E., Contreras, J. L., Aguilar-Leon, D., … Sada, E. (2008). Expression of cathelicidin ll-37 during mycobacterium tuberculosis infection in human alveolar macrophages, monocytes, neutrophils,

and epithelial cells. *Infection and immunity*, *76*(3), 935–941.

Rockett, K. A., Brookes, R., Udalova, I., Vidal, V., Hill, A. V., & Kwiatkowski, D. (1998). 1, 25-dihydroxyvitamin d3 induces nitric oxide synthase and suppresses growth of mycobacterium tuberculosis in a human macrophage-like cell line. *Infection and immunity*, *66*(11), 5314–5321.

Rook, G. (1988). The role of vitamin d in tuberculosis. *American Journal of Respiratory and Critical Care Medicine*, *138*(4), 768–770.

Rook, G. A., Steele, J., Fraher, L., Barker, S., Karmali, R., O'riordan, J., & Stanford, J. (1986). Vitamin d3, gamma interferon, and control of proliferation of mycobacterium tuberculosis by human monocytes. *Immunology*, *57*(1), 159.

Sapkota, B. R., Hopkins, R., Bjonnes, A., Ralhan, S., Wander, G. S., Mehra, N. K., … Sanghera, D. K. (2016). Genome-wide association study of 25 (oh) vitamin d concentrations in punjabi sikhs: Results of the asian indian diabetic heart study. *The Journal of steroid biochemistry and molecular biology*, *158*, 149–156.

Schilling, S. (2012). Epidemic vitamin d deficiency among patients in an elderly care rehabilitation facility. *Deutsches Ärzteblatt International*, *109*(3), 33.

Scragg, R., Sowers, M., & Bell, C. (2004). Serum 25-hydroxyvitamin d, diabetes, and ethnicity in the third national health and nutrition examination survey. *Diabetes care*, *27*(12), 2813–2818.

Sheehan, N. A., Didelez, V., Burton, P. R., & Tobin, M. D. (2008). Mendelian randomisation and causal inference in observational epidemiology. *PLoS Med*, *5*(8), e177.

Skodric-Trifunovic, V., Blanka, A., Stjepanovic, M. I., Ignjatovic, S., Mihailovic-Vucinic, V., Sumarac, Z., … Ilic, K. (2014). The health benefits of vitamin d relevant for tuberculosis. *Journal of Medical Biochemistry*, *33*(4), 301.

Sly, L. M., Lopez, M., Nauseef, W. M., & Reiner, N. E. (2001). 1α, 25-dihydroxyvitamin d3-induced monocyte antimycobacterial activity is regulated by phosphatidylinositol 3-kinase and mediated by the nadph-dependent phagocyte oxidase. *Journal of Biological Chemistry*, *276*(38), 35482–35493.

Speeckaert, M., Huang, G., Delanghe, J. R., & Taes, Y. E. (2006). Biological and clinical aspects of the vitamin d binding protein (gc-globulin) and its polymorphism. *Clinica chimica acta*, *372*(1-2), 33–42.

Sudfeld, C. R., Mugusi, F., Muhihi, A., Aboud, S., Nagu, T. J., Ulenga, N., … Fawzi, W. W. (2020). Efficacy of vitamin d3 supplementation for the prevention of pulmonary tuberculosis and mortality in hiv: a randomised, double-blind, placebo-controlled trial. *The Lancet HIV*, *7*(7), e463–e471.

Supply, P., Marceau, M., Mangenot, S., Roche, D., Rouanet, C., Khanna, V., … others (2013). Genomic analysis of smooth tubercle bacilli provides insights into ancestry and pathoadaptation of mycobacterium tuberculosis. *Nature genetics*, *45*(2), 172–179.

Sylvia, C., DareV, A., Puneet, D., AdamJ, F., & LeilaJ, M. (2010). Vitamin d: Metabolism. *Endocrinology and Metabolism Clinics of North America*, 243–253.

Team, R. C., et al. (2020). *R: A language and environment for statistical computing*. Vienna,

Austria.

Tenforde, M. W., Yadav, A., Dowdy, D. W., Gupte, N., Shivakoti, R., Yang, W.-T., … others (2017). Vitamin a and d deficiencies associated with incident tuberculosis in hiv-infected patients initiating antiretroviral therapy in multinational case-cohort study. *Journal of acquired immune deficiency syndromes (1999)*, *75*(3), e71.

Thomas, D. C., & Conti, D. V. (2004). Commentary: the concept of 'mendelian randomization'. *International journal of epidemiology*, *33*(1), 21–25.

Thomas, T. A. (2017). Tuberculosis in children. *Pediatric Clinics*, *64*(4), 893–909.

Thompson, S. G., & Sharp, S. J. (1999). Explaining heterogeneity in meta-analysis: a comparison of methods. *Statistics in medicine*, *18*(20), 2693–2708.

Urdahl, K., Shafiani, S., & Ernst, J. (2011). Initiation and regulation of t-cell responses in tuberculosis. *Mucosal immunology*, *4*(3), 288–293.

Varahram, M., Farnia, P., ANOUSHEH, S., Kazampour, M., MIRZA, M., SEYF, S., … Velayati, A. A. (2009). The vdr and tnf-α gene polymorphisms in iranian tuberculosis patients: The study on host susceptibility.

Venturini, E., Facchini, L., Martinez-Alier, N., Novelli, V., Galli, L., de Martino, M., & Chiappini, E. (2014). Vitamin d and tuberculosis: a multicenter study in children. *BMC infectious diseases*, *14*(1), 652.

Wang, T. J., Zhang, F., Richards, J. B., Kestenbaum, B., Van Meurs, J. B., Berry, D., … others (2010). Common genetic determinants of vitamin d insufficiency: a genome-wide association study. *The Lancet*, *376*(9736), 180–188.

Wang, Y., Zhu, J., & DeLuca, H. F. (2012). Where is the vitamin d receptor? *Archives of biochemistry and biophysics*, *523*(1), 123–133.

Waterham, H. R., & Wanders, R. J. (2000). Biochemical and genetic aspects of 7-dehydrocholesterol reductase and smith-lemli-opitz syndrome. *Biochimica et Biophysica Acta (BBA)-Molecular and Cell Biology of Lipids*, *1529*(1-3), 340–356.

Wirth, T., Hildebrand, F., Allix-Béguec, C., Wölbeling, F., Kubica, T., Kremer, K., … others (2008). Origin, spread and demography of the mycobacterium tuberculosis complex. *PLoS Pathog*, *4*(9), e1000160.

Wolf, A. J., Desvignes, L., Linas, B., Banaiee, N., Tamura, T., Takatsu, K., & Ernst, J. D. (2008). Initiation of the adaptive immune response to mycobacterium tuberculosis depends on antigen production in the local lymph node, not the lungs. *The Journal of experimental medicine*, *205*(1), 105–115.

World Health Organization, W. (2020). *Global tuberculosis report 2020*. World Health Organization.

Xu, X., & Shen, M. (2019). Associations between vitamin d receptor genetic variants and tuberculosis: a meta-analysis. *Innate immunity*, *25*(5), 305–313.

Yang, S., Smith, C., Prahl, J. M., Luo, X., & DeLuca, H. F. (1993). Vitamin d deficiency suppresses cell-mediated immunity in vivo. *Archives of biochemistry and biophysics*, *303*(1), 98–106.

Yarmolinsky, J., Bonilla, C., Haycock, P. C., Langdon, R. J., Lotta, L. A., Langenberg, C., … others (2018). Circulating selenium and prostate cancer risk: a mendelian random-

ization analysis. *JNCI: Journal of the National Cancer Institute*, *110*(9), 1035–1038.

Yun, C., Chen, J., He, Y., Mao, D., Wang, R., Zhang, Y., … Yang, X. (2017). Vitamin d deficiency prevalence and risk factors among pregnant chinese women. *Public health nutrition*, *20*(10), 1746–1754.

Zeng, J., Wu, G., Yang, W., Gu, X., Liang, W., Yao, Y., & Song, Y. (2015). A serum vitamin d level< 25nmol/l pose high tuberculosis risk: a meta-analysis. *PloS one*, *10*(5), e0126014.

Zhu, Z., Zhan, J., Shao, J., Chen, W., Chen, L., Li, W., … Zhao, Z. (2012). High prevalence of vitamin d deficiency among children aged 1 month to 16 years in hangzhou, china. *BMC public health*, *12*(1), 126.

A causal association of genetically predicted serum 25-hydroxyvitamin D concentrations on the odds of tuberculosis disease: A Two-sample Mendelian Randomization Study

ORIGINALITY REPORT

14%	12%	13%	5%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to University of Nairobi Student Paper	1%
2	pt.scribd.com Internet Source	1%
3	journals.plos.org Internet Source	1%
4	www.researchsquare.com Internet Source	1%
5	epdf.tips Internet Source	1%
6	www.dovepress.com Internet Source	1%
7	Jing Dong, Puya Gharahkhani, Wong-Ho Chow, Marilie D. Gammon et al. "No Association Between Vitamin D Status and Risk of Barrett's Esophagus or Esophageal Adenocarcinoma: A Mendelian Randomization Study", Clinical	1%

	Gastroenterology and Hepatology, 2019 Publication	
8	research-information.bristol.ac.uk Internet Source	<1%
9	researchonline.lshtm.ac.uk Internet Source	<1%
10	www.repository.cam.ac.uk Internet Source	<1%
11	worldwidescience.org Internet Source	<1%
12	Submitted to University of Oxford Student Paper	<1%
13	arxiv.org Internet Source	<1%
14	Xia Jiang, Douglas P. Kiel, Peter Kraft. "The genetics of vitamin D", Bone, 2019 Publication	<1%
15	bmcinfectdis.biomedcentral.com Internet Source	<1%
16	Mingzhou Fu, Kelly M. Bakulski, Cesar Higgins, Erin B. Ware. "Mendelian randomization of dyslipidemia on cognitive impairment among older Americans", Cold Spring Harbor Laboratory, 2020 Publication	<1%

17	Omowunmi Aibana, Chuan-Chin Huang, Said Aboud, Alberto Arnedo-Pena et al. "Vitamin D status and risk of incident tuberculosis disease: a systematic review and individual participant data meta-analysis:", Cold Spring Harbor Laboratory, 2018 Publication	<1%
18	academic.oup.com Internet Source	<1%
19	www.coursehero.com Internet Source	<1%
20	ijponline.biomedcentral.com Internet Source	<1%
21	open.uct.ac.za Internet Source	<1%
22	"Handbook of Statistical Genomics", Wiley, 2019 Publication	<1%
23	Vitamin D, 2010. Publication	<1%
24	whatisvitamind3.com Internet Source	<1%
25	Bowden, Jack, George Davey Smith, Philip C. Haycock, and Stephen Burgess. "Consistent Estimation in Mendelian Randomization with	<1%

Some Invalid Instruments Using a Weighted Median Estimator", Genetic Epidemiology, 2016.
Publication

26	Jue-Sheng Ong, Puya Gharahkhani, Jiyuan An, Matthew H Law, David C Whiteman, Rachel E Neale, Stuart MacGregor. "Vitamin D and overall cancer risk and cancer mortality: a Mendelian randomization study", Human Molecular Genetics, 2018 Publication	<1%
27	Stephen Burgess, George Davey Smith, Neil M. Davies, Frank Dudbridge et al. "Guidelines for performing Mendelian randomization investigations", Wellcome Open Research, 2020 Publication	<1%
28	www.rj-robbins.com Internet Source	<1%
29	"Extraskeletal Effects of Vitamin D", Springer Science and Business Media LLC, 2018 Publication	<1%
30	Submitted to University of Birmingham Student Paper	<1%
31	aasldpubs.onlinelibrary.wiley.com Internet Source	<1%
32	Feng-Xiao Bu, Laura Armas, Joan Lappe, Yu Zhou, Guimin Gao, Hong-Wei Wang, Robert	<1%

Recker, Lan-Juan Zhao. "Comprehensive association analysis of nine candidate genes with serum 25-hydroxy vitamin D levels among healthy Caucasian subjects", Human Genetics, 2010
Publication

33	profiles.stanford.edu Internet Source	<1%
34	James R Staley, Stephen Burgess. "Semiparametric methods for estimation of a non-linear exposure–outcome relationship using instrumental variables with application to Mendelian randomization", Cold Spring Harbor Laboratory, 2017 Publication	<1%
35	onlinelibrary.wiley.com Internet Source	<1%
36	mafiadoc.com Internet Source	<1%
37	"Genetic Epidemiology", Springer Science and Business Media LLC, 2018 Publication	<1%

Exclude quotes	On	Exclude matches	< 19 words
Exclude bibliography	On		