Imperial College London

INTERPLAY BETWEEN VITAMIN D AND METABOLIC FACTORS IN COLORECTAL CANCER DEVELOPMENT: A MOLECULAR EPIDEMIOLOGY APPROACH

KAWTHAR AL-DABHANI

SUBMITTED FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF EPIDEMIOLOGY AND BIOSTATISTICS

SCHOOL OF PUBLIC HEALTH

IMPERIAL COLLEGE LONDON

DECLARATION OF ORIGINALITY

This is to certify that to the best of my knowledge, the content of this thesis is my own work. Any material which is not my own work has been properly acknowledged. This thesis has not been submitted for any other degrees.

COPYRIGHT DECLARATION

ACKNOWLEDGEMENT

I would like to thank God for guiding me onto this path, and providing me with the strength to face the challenges that came my way. I would also like to thank my family for supporting me throughout this journey, without their support I would not have been able to complete this research.

I would not be here if I was not introduced and directed to the field of epidemiology by Dr.Nasrin Maselli and Dr.Laith Abu-Raddad at Weill Cornell Medical College-Qatar, Dr.Thena and Dr.Enas at Qatar National Research Fund, and Dr. Halima Bensmail at Qatar Computing Research Institute.

To Qatar Foundation, Qatar National Research Fund, my sponsors, thank you fro funding my studies and for supporting my academic career. A special thanks to my coach and mentor Dr. Ayman Bassil, who saw the potential in me and encouraged me to switch from a managerial career path onto a scientific and research career path. I am grateful for his support and guidance. I would also like to thank Susi, the administrative wizard, who provided immeasurable support and is an amazing friend.

To all the PhD students in VC11/12: Javiera, Todd, Faridah, Jelena, Michael, Ville, Olivia, Saredo, Dipender, and Areti, thank you for all the good times we had in the office and outside. You all have made coming into the office very enjoyable and memorable.

And finally I would like to thank my supervisors Joanna Tzoulaki, Kostas Tsilidis and Marc Gunter for their constant support throughout my PhD.

ABSTRACT

Observational studies have reported that higher levels of serum vitamin D are associated with lower risk of colorectal cancer (CRC) and metabolic syndrome (MetS). Given the fact that MetS and its components are also associated with CRC, the potential causal pathways between vitamin D, MetS and its components, and CRC are not well understood. At the same time, observational studies have inherit limitations and can only assess association rather than causation. Alternatively, the Mendelian randomisation (MR) approach uses genetic variants as proxies for an environmental exposure and can be used to provide more robust evidence for potential causality.

In this thesis I used a variety of methods to study the interlinked effects of vitamin D, MetS, and CRC. Mediation analysis was used to assess whether the association between vitamin D and CRC was mediated by MetS in the EPIC cohort. I further assessed the potential causal association between vitamin D and CRC using both individual and summary-level data in EPIC, UK Biobank, the GECCO consortium and data from the SUNLIGHT consortium. Moreover, I assessed the direction of potential causal relationship between vitamin D and MetS components using summary-level data from genetic consortia.

Among the 2,300 participants in EPIC, MetS mediated ~18% of the association between vitamin D and CRC. No significant causal association between vitamin D and CRC was found for either the individual-level data in EPIC (OR: 1.03, 95%CI: 0.99 – 1.06), or for the larger studies using summary-level data in the UK Biobank (OR: 0.86, 95%CI: 0.68 – 1.08) or GECCO (OR: 0.92, 95%CI: 0.76 – 1.10). I also assessed the potential causal association between vitamin D and MetS components and found that

a 1 standard deviation decrease in the natural log transformed 25-hydroxyvitamin D (25(OH)D) was associated with a 4% increase in HbA1c levels. The results also showed that high BMI and low levels of HDL cholesterol reduced the levels of 25(OH)D, and that high levels of HbA1c and SBP increased 25(OH)D levels.

To conclude, no evidence for a causal association was found between vitamin D and CRC. Further research needs to be conducted to understand the inconsistency of results between observational and MR analyses of vitamin D and CRC. Moreover, evidence of causality was found between vitamin D and MetS components; however, present methods cannot reliably infer the directionality of these associations.

Table of Contents

TABLES AND FIGURES

ABBREVIATIONS

$1,25(OH)_2D$	1,25-dihydroxyvitamin D/Calcitriol
25(OH)D	25-hydroxyvitaminD/Calcidiol
2SLS	Two Stage Least Square
7-DHC	7-DeHydroCholesterol
AIDS	Acquired Immune Deficiency Syndrome
AICR	American Institute for Cancer Research
ANCOVA	Analysis of CoVariance
ATPIII	Adult Treatment Panel III
BMD	Bone Mineral Density
BMI	Body Mass Index
CAP	College of American Pathologist
CDK	Cyclin-Dependent Kinase
CHARGE	Cohorts for Heart & Aging Research in Genomic Epidemiology
CI	Confidence Interval
CIDR	Centre for Inherited Disease Research
COPD	Chronic Obstructive Pulmonary Disease
CRC	Colorectal Cancer
CRP	C-Reactive Protein
CV	Coefficient of Variability
DAG	Directed Acyclic Graph
DBP	Diastolic Blood Pressure
D-CarDia	Vitamin D and Coronary Artery Risk Development in Young Adults

DHCR7	7-DeHydroCholesterol Reductase
DNA	Deoxyribonucleic acid
EDTA	Ethylenediaminetetraacetic Acid
EPIC	European Prospective investigation into Cancer and Nutrition
ES	Effect Size
FAP	Familial Adenomatous Polyposis
FGF-23	Fibroblasts Growth Factor 23
GECCO	Genetics and Epidemiology of Colorectal Cancer Consortium
GIANT	Genetic Investigation of ANthropometric Traits
GLGC	Global Lipids Genetic Consortium
Global BPGen	Global Blood Pressure Genetics Consortium
GRASP	Genome-wide Repository of Associations between SNPs and Phenotypes
GV	Genetic Variant
GWAS	Genome-Wide Association Studies
HbA1c	Haemoglobin A1c/ glycated haemoglobin
HDL	High Density Lipoproteins
HMC	Hamad Medical Corporation
HMGCR	3-hydroxy-3-methylglutaryl coenzyme A reductase
HNPCC	Hereditary Non-Polyposis Colorectal Cancer
HOMA-IR	Homeostatic Model Assessment-Insulin Resistance
HR	Hazard Ratio
HRT	Hormone Replacement Therapy
HWE	Hardy-Weinberg Equilibrium
IARC	International Agency for Research on Cancer

IBD	Inflammatory Bowel disease
ICBP	International Consortium for Blood Pressure
ICD	International Classification of Diseases
IDF	International Diabetes Federation
IGF	Insulin-like Growth Factor
IL-6	InterLeukin-6
IOM	Institute of Medicine
IRR	Incidence rate ratio
IT	Information Technology
IU	International Units
IV	Instrumental Variants
IVW	Inverse variance weighted
LD	Linkage Disequilibrium
LDL	Low Density Lipoprotein
MAF	Minor Allele Frequency
MAGIC	Meta-Analyses of Glucose and Insulin-related traits Consortium
MDE	Minimum Detectable Effect
MED	Minimal Erythema Dose
MET	Metabolic Equivalents
MetS	Metabolic Syndrome
MMR	Mismatch Repair
MR	Mendelian randomisation
NGSP	National Glycohemoglobin Standardization Program
NHANES	National Health and Nutrition Examination Survey
OC	Oral Contraceptive

OR	Odds Ratio
PHESANT	PHEnome Scan ANalysis Tool
PTH	Parathyroid Hormone
QA	Quality Assurance
QBB	Qatar Biobank
QC	Quality Control
QSS	Qatar Stepwise Survey
RAS	Renin-Angiotensin System
RCT	Randomised Control Trial
RDA	Recommended Daily Allowance
RR	Relative Risk
RXR	Retinoid X Receptor
SBP	Systolic Blood Pressure
SD	Standard Deviation
SE	Standard Error
SMD	Standardised Mean Difference
SNP	Single Nucleotide Polymorphism
SPF	Sun Protection Factor
SREBP-2	sterol regulatory element-binding protein 2
SRR	Standardised Rate Ratio
SUNLIGHT	Study of Underlying Genetic Determinants of Vitamin D and Highly Related Traits
T2D	Type 2 Diabetes
TG	Triglyceride
TINIA	Turbidimetric INhibition ImmunoAssay

TNF-α	Tumour Necrosis Factor-alpha
UK	United Kingdom
US	United States
USES	US Endocrine Society
UV-B	Ultraviolet-B
UW GAC	University of Washington Genetic Analysis Centre
VDBP	Vitamin D Binding Protein
VDR	Vitamin D Receptor
VDRE	Vitamin D Responsive Elements
WCRF	World Cancer Research Fund
WHI	Women's Health Initiative
WHO	World Health Organization
WHR	Waist-to-Hip Ratio

CHAPTER 1 BACKGROUND

This chapter provides a background on the epidemiology and risk factors of colorectal cancer (CRC). These risk factors include age, obesity, physical activity, smoking, and alcohol consumption. Additionally, this chapter discusses the relationship between dietary factors, including vitamin D, with the risk of CRC. Furthermore, this chapter looks at the definition and epidemiology of metabolic syndrome (MetS), and the relationship between MetS and CRC.

Finally, the chapter provides background information on vitamin D production, sources, and biological roles as well as determinants of vitamin D. It also discuss the relationship between vitamin D and MetS and its individual components.

1.1 Colorectal cancer

1.1.1 Burden of colorectal cancer

CRC is the third most commonly diagnosed cancer in men after lung and prostate cancer and the second most common diagnosed form of cancer in women after breast cancer, with an estimated 1.85 million new cases and approximately 880 thousand deaths in 2018 (1). The cumulative risk for CRC is higher in men than in women.

The majority of CRC cases occur in developed regions such as North America, Australia, New Zealand, and Europe while the lowest incidence rates are found in Africa and South-Central Asia (1). The incidence of CRC is increasing rapidly, particularly in countries that have transitioned from a low-income to a high-income economy, such as Japan, Singapore, and Kuwait (2,3). In addition to the variation in the spread of CRC across countries, variation was also found within ethnicities in some Asian countries (2,4). A study by Arnold *et al.* assessed time trends in CRC cancer incidence and mortality for 184 countries using two different sources: GLOBOCAN database and the United Nation Development Programme (5). The highest rate of incidence of CRC in males were found in Slovakia, 61.6 per 100,000, while the lowest incidence of CRC was found in sub-Saharan Africa, approximately 1.5 per 100,000 (5). Incidence rate for females tended to be lower than males, however, the geographical patterns were similar between the sexes. The geographical pattern for mortality rates generally followed the pattern for incidence rate; however, based on temporal characteristics of incidence and mortality, three different patterns for incidence and mortality of CRC were identified (5). The first group included Eastern European countries, as well as Latin America and Asia. This group had an increase in

both incidence and mortality over a 10-year period, with similar trends for males and females. The second group consisted of several European countries, Canada and Singapore, where they reported an increase incidence rate of CRC, but a decrease rate of mortality. The third group reported a decrease in both incidence and mortality rate of CRC, which was found in the highest human development index countries such as Australia, New Zealand, USA, and Japan (5).

From the year 2012 to 2018, there was an increase in CRC diagnosis amongst Europeans from approximately 470,000 to approximately half a million new cases (1,6). On the other hand, a study in the US on 1,198,421 hospitalised patients with CRC investigated the trends of CRC hospitalisation by age and stage for CRC between the years 2002 and 2012 reported a decrease in the annual number of patients admitted into the hospital with a diagnosis of CRC from 2002 with 117,754 patients to 98,175 patients in 2012. However, this decrease was only found in colon cancer (93,588 in 2002 and 72,300 in 2012), not in rectal cancer (24,166 in 2002 and 25,875 in 2012) (7). The proportion of younger patients (under 40, between 40-50, and between 50-65) increased during the 10 years, while the proportion of older patients (65-80 and greater than 80 years) decreased over the 10 year period (7). This increase in younger patients was seen in both colon and rectal cancer, with a significantly higher proportion increase in rectal cancer compared to colon cancer (7).

Survival and mortality rates of individuals with CRC, are highly dependent on the stage at which the disease is diagnosed, and or delivery of treatment. This typically range from 90% 5-year survival rate for CRCs detected at the localized stage, 70% for regional, and 10% for individuals diagnosed with distant metastatic cancer (5). According to the study by Moghadamyeghaneh *et al.* 61% of the patients younger

than 50 years had stage III or IV colon cancer, compared to 47.5% of patients 50 years or older (7).

1.1.2 Stages of Colorectal cancer

The stage of CRC describes location and spread of cancer in the body. Knowing the stage of CRC, can help determine treatment and the patient's prognosis. There are three different classification of CRC stages. Dukes first proposed a classification for CRC staging according to the extent of spread. Cases "A" are patients in which the carcinoma is limited to the wall of the rectum with no extension into the extrarectal tissues and no metastases in the lymph nodes. Cases "B" are patients with carcinoma in which it has spread to the extrarectal tissues, but has not metastasised into the lymph nodes. Cases "C" are patients in which the carcinoma has metastases present in the lymph nodes (8). The modified Dukes classification categorises cases C into C1, tumour with lymph node involvement, but not apical node, and C2, tumour with lymph node involvement, including apical node. Furthermore, cases "D" are patients that develop distant metastasis (9). In 1954 Astler and Coller developed the Astler-Coller staging system which consists of 8 stages; stage A, the tumour is limited to the mucosa, B1, the tumour has invaded into the muscularis, B2, the tumour has invaded into the serosa, B3, the tumour had invaded into adjacent organs. For C1, C2, and C3, they are all relevant to the B category but with lymph node involvement. Stage D includes distant metastasis (9). The most used staging system is the TNM system, T for has the tumour grown into the wall of the colon and rectum and how many layers, N; has the tumour spread to the lymph nodes, if so where and how many, and M; has the tumour metastasised to other parts of the body, if so where and how much. There are 5 stages, starting from stage 0 to stage IV, in which the American Joint Committee on Cancer agreed to standardise the TNM staging system (10). **Table 1.1** reports the

classification for colorectal cancer using the TNM and the American Joint Committee on cancer staging system.

The World Health Organisation (WHO) classified histologically tumours of the colon and rectum into epithelial tumours and mesenchymal tumours. The defining feature of colorectal adenocarcinoma is the invasion through the muscularis mucosae into the submucosa. Mucinous adenocarcinoma is classified if greater than 50% of the lesion is composed of mucin, which contains malignant epithelium as acinar structures. This histopathological type consists of many high frequency micro-satellite instability carcinomas (11). The signet-ring cell carcinoma is defined by the presence of greater than 50% of the tumour cells with prominent intracytoplasmic mucin. The signet-ring cell fills the cytoplasm, displacing the nucleus. Signet-ring cells can occur in the mucin of mucinous adenocarcinoma or in minimal extracellular mucin (11). Adenosquamous carcinoma shows features of both squamous carcinoma and adenocarcinoma either as separate areas or mixed within the tumour (11). Medullary carcinoma is a rare variant characterised by sheets of malignant cells with vesicular nuclei, prominent nucleoli, and pink cytoplasm. Undifferentiated carcinoma are rare tumours that lack morphological evidence of differentiation beyond an epithelial tumour and have variable histological features (11).

Furthermore, the CRC subtyping consortium unified six independent molecular classification system into a single consensus system with four distant groups, based on gene expression data. The four groups are microsatellite instability immune (CMS1), which is characterised by hypermutation, unstable microsatellite, and strong immune activation, canonical (CMS2), characterised by marked WNT and MYC signalling activation, metabolic (CMS3), characterised by evident metabolic

dysregulation, and mesenchymal (CMS4), which is characterised by prominent TGF-β activation, stromal invasion, and angiogenesis (12).

1.1.3 Natural history of disease development

Colorectal carcinoma develops within pre-existing adenomas. Although the prevalence of adenomas are not known, only a minority of colorectal adenomas are needed to undergo malignant transformation into CRC (13). The National Polyp Study, a RCT investigating the effective surveillance of patients with colorectal adenomas found that patients with multiple adenomas and/or with villous adenoma were more likely to have at least one adenoma with high-grade dysplasia (13). Moreover, the larger the size of the adenoma, the higher the percentage of high-grade dysplasia. High-grade dysplasia in greater than or equal 1 adenoma was also found to increase with age. Severe dysplasia in adenomas is a histopathological marker for increased CRC risk (13).

There are a number of factors that play a role in the polyp to CRC sequence including gene mutations (microsatellite instability and aneuploidy) and epigenetic alterations (CpG Island Methylator Phenotype (CIMP) (14).

The polyp to cancer progression involves a step that initiates the formation of benign neoplasms, followed by the progression to a more histologically advanced neoplasm and then the transformation of the tumour to invasive carcinoma (14). Premalignant serrated polyps are associated with CIMP, which has a high frequency of aberrantly methylated CpG dinucleotides (14). While conventional tubular adenomas are more commonly initiated by biallelic inactivation of the APC tumour-suppressor gene, which displays chromosome instability, gaining or losing large portions of chromosomes (14). The most common form of genetic instability is chromosomal instability, which is found

in approximately 85% of CRCs (14). The varied methods to determine chromosomal instability means no agreed upon criteria to determine whether CRC displays chromosomal instability. Moreover, the mechanism that gives rise to chromosomal instability in tumour progression is poorly understood (14). However, there is some evidence that chromosomal instability promotes cancer progression by increasing clonal diversity (14).

Microsatellite instability, another form of genetic mutations, accounts for approximately 15% of CRCs (14). Microsatellite unstable CRC has been defined by the presence of approximately 30% unstable microsatellite loci from 5-10 loci selected at a National Cancer Institute consensus conference (14). Some evidence suggests that the loss of mismatch repair activity and the onset of microsatellite instability accelerates tumour progression (14). The mechanism of microsatellite unstable CRC involves the inactivation of the DNA mismatch repair genes by DNA methylation or by somatic mutation (14). Individuals with hereditary cancer syndrome almost always develop microsatellite unstable CRC, due to mutations in at least one of the mismatch repair genes (14).

Epigenetic instability in CRC is demonstrated by the hypermethylation of loci that contain CpG islands, as well as global DNA hypomethylation (14). CIMP is often defined as increased methylation of at least three loci from 5 gene-associated CpG islands (14). The mechanism that gives rise to CIMP is still under current investigation. However, there are some studies that showed that overexpression of the DNA methytranferases DNMT3B or DNMT1 has been correlated with the CIMP. Another potential mechanism is the inactivation of barriers that prevent the methylation of the normally unmethylated CpG islands. Alternatively, mutations in genes involved in

chromatin remodelling may mediate CIMP (14). These patterns of gene mutations and epigenetic alterations could be used to refine CRC screening approaches.

Table 1.1. TNM classification of Colorectal cancer and staging system

American Joint Committee on Cancer stage	T (Tumour)	N (Nodes)	M (Metastasis)
	TX: primary tumour cannot be evaluated	NX: Regional lymph nodes cannot be evaluated	
	T0: No tumour		
0	Tis: Carcinoma in situ		
I	T1-T2: Size and/or extent of primary tumour	N0: No cancer found in lymph nodes	
II	T3-T4: Size and/or extent of primary tumour		M0: No distant metastasis
III	T1-T4: Size and/or extent of primary tumour	N1-N2: Involvement of regional nodes	
IV	T1-T4: Size and/or extent of primary tumour	N3: Involvement of regional nodes	
	T1-T4: Size and/or extent of primary tumour	N1-N3: Involvement of regional nodes	M1: Distant metastasis

1.1.4 Modifiable and non-modifiable risk factors of CRC

CRC is widely believed to be primarily an environmentally-induced disease with approximately 70-80% of the variation in risk due to environmental factors (4). Among the various environmental factors, dietary components including red and processed meat, obesity, smoking and alcohol consumption, increase the risk of CRC (15–19). Higher levels of physical activity, circulating vitamin D, hormone replacement therapy (HRT), and aspirin use, have all been associated with lower CRC risk (20–23). The Nurses' Health Study of 121,701 female nurses reported that 37% of colon cancer cases that occurred in the cohort could have been avoided by behaviour modification (24). The factors that were significantly associated with colon cancer were physical activity, multivitamins, and calcium (24). Another study performed on men estimated that approximately a third to a half of colon cancer risk could have been avoided by behaviour modification (25). These modifiable risk factors include BMI, physical activity, alcohol consumption, smoking, red meat consumption, and folic acid supplementation (25). The remaining 20-25% of variation in CRC risk is probably due to familial cases or significant heritable components (26).

The section below provides an overview of the main risk factors for CRC that have been identified through a meta-analysis for CRC risk factors as well as the results from the 2018 report on diet, nutrition, and physical activity for CRC from the World Cancer Research Fund (WCRF) and the American Institute for Cancer Research (AICR)(19,27).

Modifiable risk factors

Body Mass Index, waist circumference, and waist-to-hip ratio

The WCRF and the AICR report on diet, nutrition and physical activity, provides one of the most comprehensive systematic reviews and meta-analysis on risk factors for CRC (19). Body fat, assessed by body mass index (BMI), is positively associated with risk of CRC with a clear dose-response relationship (19). The WCRF meta-analysed 38 prospective studies (71,089 cases) and reported a 5% increase in CRC risk per 5 kg/m^2 increase in BMI (Relative Risk (RR): 1.05, 95%CI: 1.03 – 1.07, I^2=74.2). This association was slightly stronger in men than in women (RR: 1.08, 95%CI: 1.04 – 1.11, I^2=83% and RR: 1.05, 95%CI: 1.02 – 1.08, I^2=83% respectively, per 5 kg/m^2 increase in BMI). These associations remained significant for CRC subtypes (per 5 kg/m^2 increase, RR: 1.02, 95%CI: 1.01 – 1.04, I^2=59% and RR: 1.07, 95%CI: 1.05 – 1.09, I^2=72% for rectal and colon cancer respectively).

Abdominal fat, measured as waist circumference or waist-to-hip ratio, also increased the risk of CRC (19). A dose-response meta-analysis of 8 studies (4,301 cases) for waist circumference and CRC risk reported a 2% increased risk in CRC per 10 cm increase in waist circumference (RR: 1.02, 95%CI: 1.01 – 1.03, I^2=0%). However, when stratified by sex, the risk of CRC was present amongst women only, (RR per 10 cm increase in waist circumference: 1.03, 95%CI: 1.02 – 1.04, I^2=0%) (19).

Body fat has been associated with higher levels of insulin, which can then promote cell growth and inhibit apoptosis increasing the risk of CRC (19). Moreover, obesity stimulates the release of inflammatory response, such as tumour necrosis factor α and interleukin 6, which stimulates the liver to produce C-reactive protein, increasing the risk of CRC (28,29).

Physical activity

The WCRF report stated that higher levels of physical activity was associated with a lower risk of CRC (19). A meta-analysis of 12 studies (8,396 cases) comparing the highest and lowest physical activity levels reported a 20% decrease in the risk of colon cancer (RR: 0.80, 95%CI: 0.72 – 0.88, I^2=39%). However, the meta-analysis limited to rectal cancer, which consisted of 9 studies with 2,326 cases, reported no association between physical activity and rectal cancer (RR: 1.04, 95%CI: 0.92 – 1.18, I^2=9%) (19). Physical activity has been associated with reduced body fatness, which has been reported to be associated with insulin resistance and inflammation (28,29). However, it remains unclear whether the association between physical activity and CRC that is not accompanied by weight loss has a significant impact on these pathways (19).

Cigarette smoking

Generally, there is consistent data regarding the association of long term smoking with CRC risk (30–33). A meta-analysis of 24 studies reported a 9% increased risk of colon cancer (standard rate ratio (SRR): 1.09, 95%CI: 1.01 – 1.18, I^2=45.6%) for ever smokers vs never smokers and a 24% increased risk of rectal cancer (SRR: 1.24, 95%CI: 1.16 – 1.39, I^2=40.4%) (34). Furthermore, the increased risk was similar between former and current smokers (SRR: 1.24 95%CI: 1.12 – 1.39, I^2=40.4% and SRR: 1.20, 95%CI: 1.11 – 1.30, I^2=0.0, for colon and rectal cancer respectively), suggesting that the effect of smoking persists for several years (34). Cigarette smoking has been shown that angiogenesis facilitate tumour growth, invasion, and metastasis and/or the suppression if cell-mediated immunity (35).

Alcohol consumption

Acetaldehyde in alcohol can be carcinogenic (36). Higher ethanol consumption can also induce oxidative stress by increasing the production of reactive oxygen species which are carcinogenic (19). The WCRF conducted a dose-dependent meta-analysis of 16 studies (15,896 cases) and found that CRC risk increased by 7% per 10 g/day of ethanol (RR: 1.07, 95%CI: 1.05 – 1.08, I^2=27.7%) (19). There was an effect modification for the risk of CRC with alcohol consumption between men and women (per 10 g/day, RR: 1.08, 95%CI: 1.06 – 1.09, I^2=0% and RR: 1.04, 95%CI: 1.00 – 1.07, I^2=44%, respectively). Moreover, the association remained consistent when stratified into colon and rectal cancer (per 10 g/day RR: 1.07, 95%CI: 1.05 – 1.09, I^2=34%, RR=1.08, 95%CI: 1.07 – 1.10, I^2=0%, respectively) (19).

Hormone Replacement Therapy

A meta-analysis of three randomised control trials (RCTs) (n=122 cases), 13 prospective cohort studies (n=3,713 cases) and 14 retrospective studies (n=2421 cases) reported that ever versus never use of hormone therapy reduced the risk of CRC by 16% (RR: 0.84, 95%CI: 0.81 – 0.88, I^2=3.3%) (37).

A meta-analysis of 19 studies on oral contraceptives reported a significant inverse association between ever versus never use with the risk of CRC (summary RR: 0.82 95%CI: 0.76 – 0.88, I^2=48.8%) (38).

Strong inverse relationships have also been reported between estradiol, estrone, and free estradiol with CRC risk in a Women's Health Initiative (WHI)-Clinical Trial nested case-control study with 401 CRC cases and 802 matched controls. After adjustments for insulin, free Insulin-like Growth Factor (IGF)-1, and C-Reactive Protein (CRP), the Odds Ratio (OR) was 0.58, 95%CI: 0.38 – 0.90, OR: 0.44, 95%CI: 0.28 – 0.68, and

OR: 0.43, 95%CI: 0.27 – 0.69 when comparing the fourth quartile to the first quartile for estradiol, estrone, and free estradiol respectively (39).

One mechanism explaining this inverse relationship is that oestrogens and progestins may reduce bile acid production, in which high levels of bile acid concentrations may enhance colon carcinogenesis. Another mechanism reported that HRT decreases both IGF and IGFBP-3 levels decreasing the risk of CRC (40).

However, the nature of this relationship is uncertain as a case-cohort study performed in the WHI-Observational study with 438 CRC cases and 816 controls, reported that high versus low levels of endogenous circulating oestradiol increased the risk of CRC after adjusting for obesity and other CRC risk factors (Hazard Ratio (HR): 1.53, 95%CI: 1.02 – 2.27) (41).

Diet and nutrition

Overall the report from the WCRF and the AICR concluded that there was strong evidence that high intakes of processed and red meat and alcoholic drinks, increased the risk of CRC, whereas intake of whole grains, fibre, dairy and calcium decreased the risk of CRC (19).

In detail, there was strong evidence that red and processed meat increased the risk of CRC (RR: 1.12, 95%CI: 1.04 – 1.21, I^2=70.2%) (19). However, separately, a dose-response meta-analysis of 8 studies reported no association between 100 g/day of red meat and CRC risk (RR: 1.12, 95%CI: 1.00 – 1.25, I^2=23.6%). While a dose-response meta-analysis of 10 studies on processed meat reported a 16% increased risk of CRC per 50 g/day of processed meat. When stratified by CRC subtype, rectal cancer reported no association with processed meat per 50 g/day (RR: 1.08, 95%CI: 1.00 –

1.18, I^2=0%), while colon cancer reported a 23% increased risk of CRC per 50 g/day of processed meat (RR: 1.23, 95%CI: 1.11 – 1.35, I^2=26%).

Red and processed meat are rich in haem iron and has been shown to promote colorectal tumorigenesis. There are two main mechanisms in which haem iron could cause CRC, haem iron can promote the formation of endogenous N-nitroso compounds, which have carcinogenic properties. Haem iron can also promote the production of free radicals and lipid peroxidation which is a risk factor for several diseases (42,43). Furthermore, red and processed meat are often cooked at high temperatures which produces heterocyclic amines and polycyclic aromatic hydrocarbons which have been associated with CRC risk. Processed meat usually has a higher fat content compared to red meat, which can stimulate tumorigenesis through the synthesis of bile acid (19).

There was also strong evidence that high intake of dietary fibre (RR: 0.91, 95%CI: 0.88 – 0.94, I^2=0.0% per 10 g/day), wholegrains (RR: 0.83, 95%CI: 0.78 – 0.89, I^2=18.2% per 90 g/day), dairy products (RR: 0.87, 95%CI: 0.83 – 0.90, I^2=18.4% per 400 g/day) and dietary calcium (RR: 0.94, 95%CI: 0.93 – 0.96, I^2=0.0% per 200 mg/day) decreased the risk of CRC (19).

Dietary fibre and wholegrains fermented in the bowel can form butyrate, a fatty acid that has anti-proliferative effects. High fibre diets can also reduce insulin resistance, which can reduce the risk of CRC. Calcium has been suggested to reduce cell proliferation and promote cell differentiation, by influencing different cell signalling pathways (19). Calcium also binds to bile acid and free fatty acids reducing the effect on the colon and rectum (19).

Suggestive evidence indicated that low intake of non-starchy vegetables increase the risk of CRC (19). Eleven studies were included in the dose-response meta-analysis for non-starchy vegetables and the risk of CRC, (n=14,136 cases) and reported a 2% decreased risk of CRC per 100 g/day of non-starchy vegetables (RR: 0.98, 95%CI: 0.96 – 0.99, I^2=0%). When stratified by sex, no association was found in women per 100 g/day of non-starchy vegetable (RR: 0.99, 95%CI: 0.96 – 1.01, I^2=0%), while a 4% decreased risk in CRC was found in men (RR: 0.96, 95%CI: 0.93 – 0.99, I^2=33%) (19). Moreover, when stratified into CRC subtypes no association was found for rectal cancer (RR: 0.97, 95%CI: 0.95 – 0.99, I^2=0% and RR: 0.99, 95%CI: 0.96 – 1.02, I^2=0% for colon and rectal cancer respectively). Vegetables consists of anti-tumorigenic agents including dietary fibre, folic acids, vitamin C and selenium (19).

The WCRF also reported a suggestive inverse association between fish (RR: 0.89, 95%CI: 0.80 – 0.99 per 100 g/day) and foods containing vitamin C (RR: 0.94, 95%CI: 0.89 – 0.99 per 40 mg/day) with the risk of CRC (19). When stratified by either sex or CRC subtype, no association was found for women or for colon and rectal cancer subtypes with fish consumption. Fish consist of long-chain n-3 polyunsaturated fatty acids suppress inflammatory pathways reducing the risk of CRC (19).

Metabolic syndrome

A meta-analysis of 8 cohort studies, 2 nested case-controls, 4 case-controls and 1 RCT reported that MetS significantly increased the risk of CRC in men (RR: 1.33, 95%CI: 1.18 – 1.50, I^2=45%) and women (RR:1.41, 95%CI: 1.18 – 1.70, I^2=58%) (44). The association between MetS and CRC may be mediated by dysregulation of growth signals including insulin, IGF-1, and adiponectin, which can contribute to CRC-related processes (45).

Glycated haemoglobin

Glycated Haemoglobin (HbA1c) is a marker of hyperglycaemia and is formed when glucose binds to haemoglobin. It is commonly used in diagnosing diabetes. By measuring HbA1c, an average estimate of blood sugar levels over a period of weeks to months can be obtained.

A systematic review and meta-analysis of 8 studies involving 2,137 cases and 820,317 controls assessing the association between HbA1c and CRC reported a 22% increased risk of CRC for the highest versus the lowest categories of HbA1c (pooled RR: 1.22, 95%CI: 1.02 – 1.47, I^2=24.8%) (46). One study included in the meta-analysis mentioned above was a nested case control study performed on 1,026 incident CRC cases and 1,026 matched controls in the European Prospective Investigation into Cancer and Nutrition (EPIC). They reported a 10% increased risk in CRC per 10% increase in HbA1c percentage (OR: 1.10, 95%CI: 1.01 – 1.19) (47). However, after the exclusion of diabetic participants (n=211), the association between HbA1c and CRC was null (OR: 1.07, 95%CI: 0.90 – 1.27 per 10% increase in HbA1c percentage) (47). Insulin resistance and hyperglycaemia may increase the risk of CRC via inflammatory, oxidative stress, and proliferative pathways (46).

Hypertension

Research investigating the role of blood pressure on CRC is limited. A meta-analysis of 15 datasets with 8,880 cases of CRC reported that high blood pressure increased the risk of CRC compared to low levels of blood pressure (RR: 1.09 (95%CI: 1.01 – 1.18, I^2=45%) (44), in which 11 of the 15 datasets reported a non-significant positive association with CRC and the remaining four reported a non-significant negative

associations (44). Possible mechanisms between hypertension and CRC remain uncertain and further clarification needs to be done.

Dyslipidaemia

Dyslipidaemia is associated with inflammation, oxidative stress and insulin resistance which may enhance colorectal carcinogenesis (48). A meta-analysis of 9 prospective studies with a mean duration of follow up of 12 years, reported that high versus low concentration of serum triglycerides was associated with an 18% increased risk in CRC (SRR: 1.18, 95%CI: 1.04 – 1.34, I^2=47.8%) (48). A dose-response analysis was conducted on three studies and reported a 1% increase in risk of CRC per 50 mg/dL increment in triglycerides (SRR: 1.01, 95%CI: 1.00 – 1.03, I^2=0%) (48).

A meta-analysis of 6 prospective studies reported that high levels of high density lipoprotein (HDL)-cholesterol was not associated with CRC risk compared to low HDL cholesterol levels (SRR: 0.84, 95%CI: 0.69 – 1.02, I^2=42.5%) (48). Another meta-analysis of 6 studies (three of which are different from the above meta-analysis) also reported a non-significant association between low levels of HDL cholesterol with the risk of CRC compared to high levels of HDL cholesterol (SRR: 1.12, 95%CI: 0.98 – 1.27) (44).

Vitamin D

Although the main physiological role of vitamin D is to control calcium homeostasis, it has been hypothesised that the active metabolite of vitamin D, 1,25-dihydroxyvitamin D (1,25$(OH)_2$D), also known as calcitriol, controls cell growth, and improves expression of various genes regulating the normal structure and function of the colon crypt, as well as its function in apoptosis (22,49). It has also been hypothesised that

vitamin D plays a role in CRC development (22), with consistent associations between vitamin D levels and CRC in observational studies. A pooled analysis of 15 studies on serum 25-hydroxyvitamin D (25(OH)D) and the risk of CRC, reported a 33% lower risk of CRC comparing high to low levels of serum 25(OH)D (OR: 0.67, 95%CI: 0.59 – 0.76) (50).

Contrary to the observational evidence, intervention studies to date have failed to find protective relationships between vitamin D supplementation and CRC. For instance, one double-blind, placebo-controlled trial on 36,282 postmenopausal women, in which for an average of 7 years 18,176 women received 1,000 mg/day of calcium and 400 International Units (IU) of vitamin D_3 daily, while the remaining 18,106 received placebo, reported that supplementation of vitamin D and calcium had no effect on the incidence of CRC (HR: 1.08, 95%CI: 0.86 – 1.34) (51). It could be argued that the dosage of vitamin D given was below the recommendation levels set by the Institute of Medicine (IOM) of 600 – 800 IU per day for adults (52). Therefore, the efficacy of the dose may have been too weak. Moreover, participants were allowed to take their own vitamin D and calcium supplements, biasing the results towards the null. Another limitation was that the participants were not required to screen for CRC, which may have reduced the number of CRC cases detected.

A more recent double-blind, placebo-controlled RCT involved 2,303 healthy postmenopausal women aged 55 years and older. These participants were followed up for four years with 1,156 participants assigned to 2,000 IU/day of vitamin D_3 and 1,500 mg/day of calcium carbonate and the remaining 1,147 were assigned a placebo. This study investigated the association between vitamin D and calcium supplementation with all cancer (excluding non-melanoma skin cancers) and reported

no evidence of an association between vitamin D and all cancer (HR: 0.70, 95%CI: 0.47 – 1.02) (53). There were several limitations in this study as well, including short follow-up time, since the average latency period for developing CRC from adenomas is between 5 to 10 years (3). As with the previous RCT, participants in the placebo group were also allowed to take vitamin D supplements on their own, which may have biased the results towards the null. Moreover, baseline serum 25(OH)D was high in both the treatment and the placebo group (82.5 nmol/L and 81.7 nmol/L respectively). A systematic review and meta-analysis reported a U-shaped dose response association between 25(OH)D and CRC risk, with 137 nmol/L of 25(OH)D being the level at which the RR of CRC was the lowest at 0.65 compared to a RR of 1.0 at 30 nmol/L of serum 25(OH)D and a RR of ~0.8 for 200 nmol/L of serum 25(OH)D (54).

Non-modifiable risk factors

Age

The risk of CRC diagnosis increases after the age of 40 and rises sharply after the age of 50, with more than 90% of CRC cases occurring in those aged 50 or older (3). However, more recently studies also indicated an increase in CRC rates in younger individuals (3). A study by Davis *et al.* reported that CRC incidence in the US was higher among participants aged 20 to 49 in 2006 compared to 1987, with the most increase in CRC rate occurring in the 40 to 44 age group (55). The increase in CRC incidence in younger participants may be attributed to earlier diagnoses, which implies that CRC screening may need to begin at an earlier age.

Inflammatory Bowel disease

Inflammatory Bowel Disease (IBD) is used to describe two diseases, ulcerative colitis and Crohn's disease. Ulcerative colitis causes inflammation of the mucosa of the colon and rectum. While Crohn's disease causes inflammation of the full thickness of the bowel wall and may involve any part of the digestive tract from the mouth to the anus. Patients with ulcerative colitis are reported to have an increased risk for CRC malignancies by 4 to 20 fold (3,56). A meta-analysis of 12 studies reported that Crohn's disease was also associated with an increased risk of CRC (RR: 2.5, 95%CI: 1.3 – 4.7) (57).

Adenomatous polyps

Nearly 95% of sporadic CRC develops from adenomatous polyps, which are benign noncancerous growths (3). Individuals with a history of adenomas have a higher risk of developing CRC compared to individuals without a history of adenomas (3). The progression from benign adenoma to invasive carcinoma has been associated with a series of genetic mutations, including mutations in the *APC* and the *K-RAS* gene, causing activation of proto-oncogenes and loss of function of tumour suppressor genes (58). CRC is a result of a stepwise accumulation of multiple somatic mutation and therefore the development from adenomas to malignancies has a long latency of 5-10 years (3). Adenoma size is also relevant in CRC development, in which only 1% of adenoma less than 10 mm develop into CRC (59). While 50% of adenomas greater than 20 mm develop into CRC (59).

Inherited Genetic risk

Twin studies placed CRC second among the common cancers in terms of heritability after prostate cancer, in which 35% of the risk of CRC could be explained by heritable

factors (26,60). High penetrance mutations in genes including *APC*, deoxyribonucleic acid (DNA) mismatch repair genes (MMR), *AXIN2*, and *LKB1* account for 3 to 5% of CRC cases. Approximately 70-80% of these SNPs have also been associated with the risk of CRC in genome-wide association studies (GWAS) (28,61,62). The fraction of cases attributable to high-penetrance alleles is relatively low compared to the low-penetrance alleles (26).

The most common inherited conditions are Familial Adenomatous Polyposis (FAP) and Hereditary Non-Polyposis CRC (HNPCC), which mainly involve high penetrance mutations (3,26). However, they only correspond to approximately 5%of the total CRC burden (63). HNPCC involves cancer predisposition to seven other organs beside the colon and rectum including the endometrium, stomach, ovaries, small intestine, and brain. The genetic basis of HNPCC involves germline mutation in DNA MMR genes, which behave like tumour suppressors of *MLH1, MSH2, MHS6* and *PMS2* genes with a penetrance of approximately 80% for CRC (3,26,56,64).

FAP accounts for less than 1% of CRC cases and usually begins to appear in late childhood or adolescence and if untreated, one or more adenomas will possibly develop into adenocarcinoma. Therefore, the penetrance of this syndrome is approximately 100%. The gene responsible for this disease is the mutated form of *APC* gene, a tumour suppressor. Approximately 75-80% of individuals with *APC*-associated polyposis conditions have an affected parent (3,26,56). Individuals with previous adenomas are more likely to develop new adenomas. A study by the Netherlands Foundation for the Detection of Hereditary Tumours performed on a registry of families with HNPCC and FAP reported that 14 patients with adenoma and 162 patients without adenoma from the first colonoscopy were followed up with a

second colonoscopy. Results from the second colonoscopy reported that 3 of the 14 patients developed a new adenoma (21%, 95% Confidence Interval (CI): 4.7 – 50.8), compared to 8 of the 162 patients (5%, 95%CI: 1.6 – 8.3) (65).

The majority of CRC cases occur in individuals without a family history of the disease or predisposing illness. However 10-20% of CRC patients, have first-degree relatives that are affected, and for these family members the risk of CRC was approximately three times higher than that of the general population (3,56). Genetic association analyses have been used to identify predisposing CRC alleles using candidate gene approaches and genome-wide association studies.

1.1.5 Candidate gene studies for colorectal cancer

Candidate gene studies identifies risk factors associated with a particular disease (66). This is done by selecting candidate genes based on the relevance in the mechanism of the disease. The SNPs are selected based on functional consequence that affect gene regulation. Then the gene variant is verified for the disease association in observational studies (66). Previous studies have identified several genetic susceptibility variants for CRC. SNPs in the genes *ADH1C*, *APC*, *CCDN1*, *IL6*, *IL8*, *IRS1*, *MTHFR*, *PPARG*, *VDR* and *ARL11* were used as candidate genes for a study from the EPICOLON consortium (63). The expression of most of these genes are altered in CRC and are involved in processes for CRC. Four SNPs were significant in EPICOLON stage 1 (rs698 in *ADH1C*, rs1800795 in *IL6*, rs3803185 in *ARL11*, and rs2102302 in *GALNTL2*), but only rs3803185 was replicated in another independent CRC cohort (63). A more recent study on 55 CRC cases discovered two new genes *PTPN12* and *LRP6* that contribute to the susceptibility of CRC (67). Candidate gene

studies have high statistical power, however, one limitation to candidate gene analysis is the inability to discover new genes (68).

1.1.6 Genome-wide association studies for colorectal cancer

GWAS allow for the testing of more than a million single nucleotide polymorphisms (SNP) with a trait. A GWAS performed by Schumacher *et al.* combined data from four large consortia with 18,299 cases and 19,655 controls. Schumacher *et al*. replicated the results of 41 loci for CRC from previous studies and additionally found 6 novel loci that reached the genome-wide threshold of $5x10^{-8}$ (69). A more recent GWAS published earlier this year by Schmit *et al.* included genetic data from 53 observational studies and clinical trials, including the consortia from the GWAS mentioned above (n=163,315) (70). Schmit *et al.* identified 11 novel variants in the discovery GWAS on CRC with 9 variants independently replicated along with 70 variants that were previously published (70). Overall, the 76 variants (3 of the 70 previously known risk variants were excluded due to Minor Allele Frequency (MAF)<0.1 and 2 of the 11 novel variants were not replicated) that were found in this study explained 11.9% of the variance of CRC (70).

1.1.7 Screening

Decreasing CRC mortality rates has been observed in many countries and are likely due to increase in CRC screening programmes. The fact that CRC develops over several years provides a unique opportunity for early detection before invasive cancer or metastasis occurs. Early detection via screening has the potential to make drastic improvements in survival supported by evidence of a higher one-year net survival for patients diagnosed at stage I (localised) compared to stage IV (metastasis), 98% compared to 40% in England respectively (71). Although CRC screening can detect

and remove adenomatous polyps, and therefore prevent the development of CRC, the Cancer Prevention and Early Detection 2015-2016 report, indicated that only 58.6% of adults in the United States (US) age 50 and older were up to date with their screening in the past five years (49,72). The English National Health Service Bowel Cancer Screening Program detects CRC by testing individuals between the ages of 60-69 using either a guaiac faecal occult blood test or bowelscope screening, which is then followed by a colonoscopy for individuals with significant findings. This study reported that approximately 50-58% of the individuals in the UK underwent either the guaiac faecal occult blood test or bowelscope screening (73).

The next section discusses the details of the production, metabolism and regulation of vitamin D. Additionally it discusses the biological roles and sources of vitamin. It also provides a definition for vitamin D deficiency and further discuss the correlates and determinants of vitamin D and concludes with epidemiological studies that investigate the association of MetS and its components with vitamin D levels.

1.2 Vitamin D

1.2.1 Production, metabolism and regulation

Vitamin D, a fat soluble secosteroid hormone (a steroid with a broken ring), is acquired from three sources, from sun exposure, from diet, and supplements (74). Vitamin D can be obtained through consumption of fortified dairy products, cereals, and fish. There are two forms of vitamin D: vitamin D_3 which is obtained from animal sources (cholcalciferol) and sunlight exposure, and vitamin D_2 which is found in plants (ergocalciferol) (74). Most vitamin D in humans is obtained from sun exposure (vitamin D_3) (74).

Ultraviolet B (UV-B) radiation from the sun converts 7-dehydrocholesterol (7-DHC) that is found in the skin into pre-vitamin D_3, which is immediately induced thermally by isomerisation into vitamin D_3 (**Figure 1.1**). 7-DHC is also converted into cholesterol by the enzyme 7-dehydrocholesterol reductase (DHCR7). Since cholesterol is a precursor to vitamin D, inhibiting the synthesis of cholesterol, will increase the synthesis of vitamin D. Studies have shown that high levels of cholesterol causes DHCR7 degradation and increases the production of vitamin D (75). Vitamin D is then transported through the blood circulation by binding to the vitamin D binding protein (VDBP) into the liver, where vitamin D is converted into 25(OH)D or calcidiol by the enzyme 25-hydroxylase. 25(OH)D is considered the principal circulating form of vitamin D and is used as the most reliable biomarker for vitamin D status in epidemiological research (76). 25(OH)D, still bound to VDBP re-enters the circulation and undergoes a second hydroxylation by the enzyme 1-α-hydroxylase in the kidneys to produce the active form of vitamin D, $1,25(OH)_2D$, also known as calcitriol. This hydroxylation is performed by the gene *CYP27B1* and is found in the proximal renal

tubules. The half-life of 25(OH)D is between 10 days to 3 weeks compared to the half-life of $1,25(OH)_2D$ which is approximately 4-6 hours. The gold standard measurement for vitamin D is liquid chromatography coupled to tandem mass spectrometry (LC/TMS) (**Table 1.2**), which allows for the distinction of the two forms of vitamin D, vitamin D_2 and vitamin D_3 (77).

To express precision of immunoassay test results, two measurements of the Coefficient of Variability (CV) are reported: the inter-assay CV and the intra-assay CV. The inter-assay reports the consistency between plate to plate, while the intra-assay CV reports the consistency within the plate (78). Inter-assay CV less than 15% and intra-assay CV less than 10% are generally acceptable.

Table 1.2. Different measurements of vitamin D and advantages and disadvantages of each measurement

Serum 25(OH)D measurement	Advantages and disadvantages
chemiluminescent immunoassay (CLIA)	Distinguish between vitamin D_2 and vitamin D_3 Variable cross-reactivity for $24,25(OH)_2D_3$ and C3-epi-$25(OH)D_2$ Used in the Qatar Biobank
Enzyme-linked immunoassay (ELISA)	Underestimated $25(OH)D_2$ by ~25% Variable cross-reactivity for $24,25(OH)_2D_3$ and C3-epi-$25(OH)D_2$ Used in EPIC
High performance liquid chromatography (HPLC)	Distinguish between vitamin D_2 and vitamin D_3 Accuracy of method is difficult to assess Difficulty maintaining detectors
Non tandem mass spectrometry (LCMS-MS)	Measures simultaneously all species of 25-hydroxylated vitamin D as well as downstream dihydroxylated metabolites Has high analytical specificity and sensitivity and relatively short chromatography run time Instruments are costly and requires well-trained personnel for operation and maintenance

$1,25(OH)_2D$ binds to the vitamin D receptor (VDR) along with its heterodimer retinoid X receptor (RXR), which then binds to specific nucleotide sequence in the DNA known

as vitamin D responsive elements (VDRE). Transcription factors will then bind to this complex where it then regulates the transcription of genes involved in cell growth, differentiation and metastasis (79).

In addition to activating 25(OH)D by 1-α-hydroxylase into $1,25(OH)_2D$ in the kidney, 25(OH)D can also be converted into $24,25(OH)_2D$ by the enzyme *CYP24A1* into water soluble inactive forms that are excreted in the bile. $1,25(OH)_2D$ also stimulates its own destruction in the kidneys by the same 24-hydroxylase into $1,24,25(OH)_3D$ (74,76,80). Unlike the 25-hydroxylase, 1-α-hydroxylase is tightly regulated primarily by three hormones: the parathyroid hormone (PTH) levels, fibroblasts growth factor 23 (FGF-23), and $1,25(OH)_2D$ (81). Absorption of renal and intestinal calcium and phosphorous is increased in the presence of $1,25(OH)_2D$. It also induces the expression of 24-hydroxylase which catabolizes both 25(OH)D and $1,25(OH)_2D$. Low serum calcium and phosphate levels enhance the activity of PTH which stimulates the transcription of 1-α-hydroxylase. Concurrently, high levels of $1,25(OH)_2D$ suppresses PTH production at the level of transcription and stimulates production of FGF-23 in the bone. As the level of FGF-23 increases, the expression of 1-α-hydroxylase is supressed (74). Therefore, calcitriol enhances the efficiency of calcium and phosphorous absorption along the intestines (82).

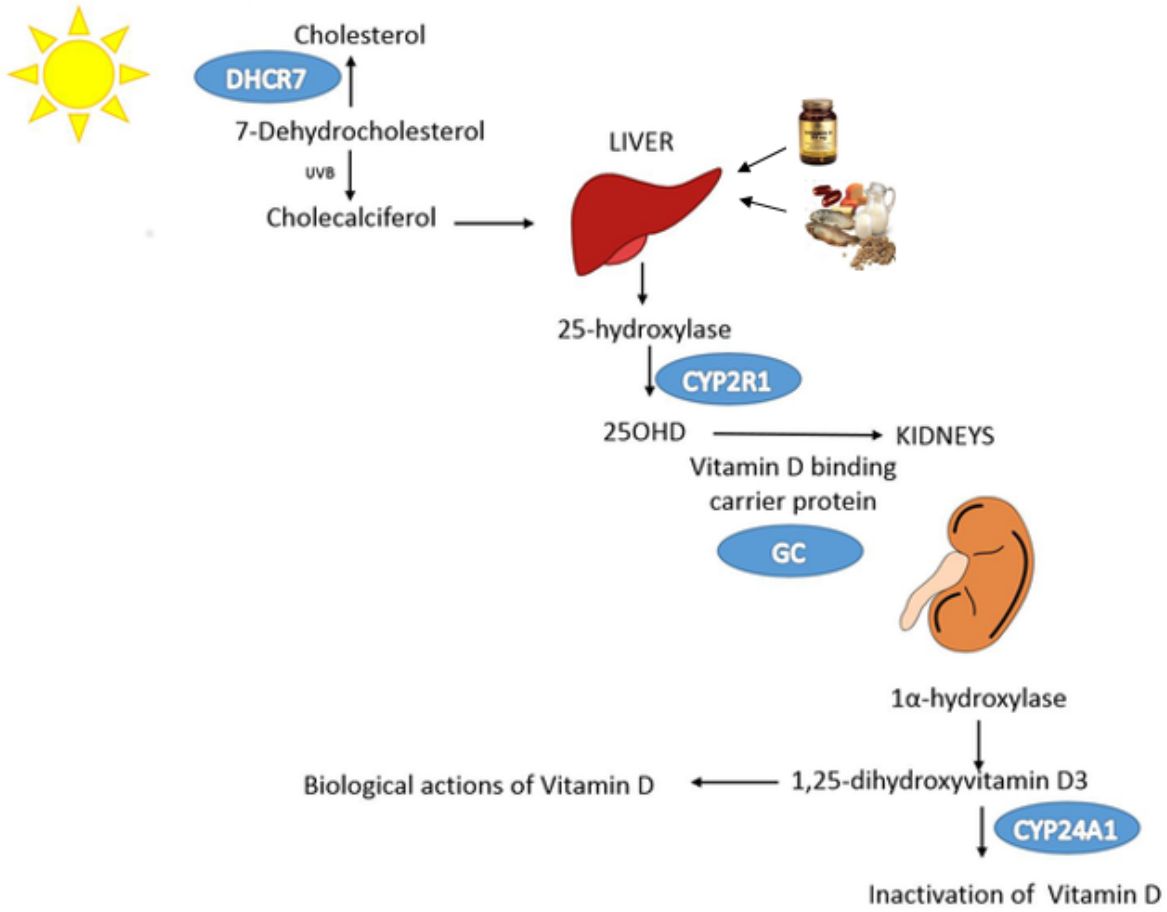


Figure 1.1 Vitamin D production and metabolism

Source reproduced from Mokry *et al.* (83) with permission under the terms of the Creative Commons Attribution License. The images of vitamin D supplements, food product and two arrows pointing in the direction of the liver have been added to the figure.

Although the kidney is the main source of 1-α-hydroxylase, epithelial cells in the skin, lungs, breast, intestine, prostate, and colon also express 1-α-hydroxylase (81,84). Similarly, brain, prostate, breast and colon tissues have VDRs and respond to $1,25(OH)_2D$. VDRs are also expressed in β-cells, skeletal muscles, immune system and in adipose tissues (85,86). A recent study by Gallone *et al.*, reported a total of 43,332 VDR-binding variants, in which these variants were associated with 17 disease traits, mainly immune and inflammatory disorders including Graves' disease, Crohn's disease, irritable bowel syndrome and Type I Diabetes (87). Although the physiology

of vitamin D in these organs remains unknown, their main biological function has been identified as anti-proliferative (82).

1.2.2 Sources of Vitamin D

UV-B Irradiation

Synthesis of vitamin D in the skin from sun exposure is the most important source of vitamin D. The time of day, season, latitude, and skin pigmentation are factors that need to be considered to determine how much vitamin D is produced during sun exposure. Exposure of the arms and legs to 0.5 minimal erythemal dose (25% to 50% of the time it would take to develop a sunburn) is equivalent to consuming 3,000 IU of vitamin D_3, this is dependent on the intensity and UV-B irradiation (74,76).

Food sources

Very few foods contain natural vitamin D. Fatty fish (such as salmon, tuna, and mackerel) and dry shiitake mushroom consists of the highest amounts of natural vitamin D found in foods (**Table 1.3**) (74,81). Beef liver, cheese and egg yolks contain small amount of vitamin D (82). Consuming foods that have been fortified with vitamin D increases vitamin D intake. In the United Kingdom (UK), vitamin D is only added to milk and margarine, while in the US, vitamin D is added to cereals, flour, bread, milk, and milk products (88,89), which may be one reason individuals from the UK have lower vitamin D levels compared to individuals from the US (90).

Table 1.3 Dietary and supplemental sources of Vitamin D

Source	Vitamin D content
Natural Sources	
Salmon	
Fresh, wild (3.5 oz)	~600 – 1000 IU of vitamin D_3
Fresh, farmed (3.5 oz)	~100 – 250 IU of vitamin D_3 or D_2
Canned (3.5 oz)	~300 IU of vitamin D_3
Sardines, canned (3.5 oz)	
Mackerel, canned (3.5 oz)	~250 IU of vitamin D_3
Tuna, canned (3.6 oz)	~230 IU of vitamin D_3
Cod liver oil (1 tsp)	~400 – 1000 IU of vitamin D_3
Shiitake mushrooms	
Fresh (3.5 oz)	~100 IU of vitamin D_2
Sundried (3.5 oz)	~1600 IU of vitamin D_2
Egg yolk	~20 IU of vitamin D_3 or D_2
Exposure to sunlight, UVB radiation (0.5 minimal erythemal dose)	~3000 IU of vitamin D_3
Fortified foods	
Fortified milk	~100 IU/8 oz, usually vitamin D_3
Fortified orange juice	~100 IU/8 oz vitamin D_3
Infant formulas	~100 IU/8 oz vitamin D_3
Fortified yogurts	~100 IU/8 oz, usually vitamin D_3
Fortified butter	~50 IU/3.5 oz, usually vitamin D_3
Fortified margarine	~430 IU/3.5 oz, usually vitamin D_3
Fortified cheese	~100 IU/3 oz, usually vitamin D_3
Fortified breakfast cereals	~100 IU/serving, usually vitamin D_3
Supplements	
Prescription	
Vitamin D_2 (ergocalciferol)	50,000 IU/capsule
Drisdol (Vitamin D_2) liquid supplements	8000 IU/ml
Over the counter	
Multivitamin	400 IU vitamin D, D_2, or D_3
Vitamin D_3	400, 800, 1,000, and 2,000 IU

Source: Reproduced with permission from Holick *et al.* (74), Copyright Massachusetts Medical Society.

1.2.3 Vitamin D supplementation and recommended vitamin D intake

According to the IOM, the recommended dietary allowance (RDA) for vitamin D represents a daily intake that is sufficient to maintain bone health and normal calcium metabolism in healthy individuals. This value varies based on age and circumstances. Supplementation with 400 IU of vitamin D, approximately raises serum 25(OH)D levels to 45 nmol/L (91). Approximately 600 to 800 IU daily supplementation of vitamin D is recommended for most of the population according to the IOM. However, a range of studies have reported the optimal vitamin D dosage for fracture prevention is between 800 to 1600 IU (92). On the other hand, the US Endocrine Society (USES) recommended daily supplementation of between 1,500-2,000 IU to maintain musculoskeletal health (93).

One study reported that some individuals require more vitamin D than others to reach a given concentration of serum 25(OH)D (91). Children and adults that suffer from obesity, tend to sequester vitamin D due to body fat, and require 2 to 5 times more vitamin D to maintain sufficient levels of serum 25(OH)D (76). Moreover, individuals taking anti-seizure medication, acquired immune deficiency syndrome (AIDS) medication, and glucocorticoids also require more vitamin D to maintain sufficient levels of serum 25(OH)D (76).

Precisely defining vitamin D deficiency on the basis of 25(OH)D values is still a matter of debate (94). The optimal levels of serum 25(OH)D concentration can be defined from biological effects. The most common definition for optimal 25(OH)D concentration depends on the concentration at which it supresses the PTH to its minimum to prevent bone loss. However, estimates of optimal serum 25(OH)D criteria by this definition is very wide; where it ranges between 30 to 100 nmol/L (95–97). There are other

definitions for optimal vitamin D which include the level of 25(OH)D associated with the highest bone mineral density (BMD), greatest calcium absorption, reduced rates of bone loss and reduced fracture rates (96).

Only a few studies measured optimal serum 25(OH)D in regard to factors other than skeletal health; one of which reported the optimal serum 25(OH)D concentration for CRC prevention being ≥90 nmol/L (95). Another study reported the optimal 25(OH)D concentration in relation to maintaining Homeostatic Model Assessment-Insulin Resistance (HOMA-IR) at healthy levels was 29 nmol/L (76,98).

1.2.4 Deficiency and Toxicity

Vitamin D deficiency is defined by the USES guidelines as less than 50 nmol/L in circulation, while individuals with 50 to 75 nmol/L are considered insufficient, and individuals with greater than 75 nmol/L have sufficient levels of serum 25(OH)D (**Table 1.4**) (99). As mentioned previously, vitamin D deficiency has been shown to be associated with many diseases (49,100).

On the other hand, vitamin D intoxication is extremely rare and can be caused by ingestion of high doses of vitamin D, and is characterised by hypercalcemia, hypercalciuria, and hyperphosphatemia, which are responsible for soft tissue and vascular calcification and nephrolithiasis. Doses of more than 50,000 IU/day raises the serum 25(OH)D levels to more than 375 nmol/L. Due to the tight regulation of the $1,25(OH)_2D$ in the kidney, as the level of serum 25(OH)D increases, 1-α-hydroxylase decreases due to the degrading enzyme, 24-hydroxylase. Thereby, keeping the concentration of $1,25(OH)_2D$ tightly regulated. However, as 25(OH)D concentration continues to rise, a limit is reached where $1,25(OH)_2D$ can no longer be regulated appropriately, which results in hypercalcemia and hyperphosphatemia. The symptoms

include nausea, dehydration, and lethargy (74,101). Supplementation of vitamin D_2 may have a greater risk of toxicity since it has a weaker affinity to VDBP in which it produces higher levels of free 25(OH)D_2 and 1,25$(OH)_2$D (102).

Table 1.4. Serum vitamin D cut-off points for vitamin D deficiency, inadequacy, adequacy, and toxicity according to the US Endocrine society and the Institute of Medicine

	US Endocrine Society	Institute of Medicine
Vitamin D deficient	<50 nmol/L	<30 nmol/L
Inadequate	50 – 75 nmol/L	30 – 50 nmol/L
Adequate	75 – 250 nmol/L	>50 nmol/L
Toxic	> 250 nmol/L	

1.2.5 Biological roles of vitamin D

Skeletal function

One of the main physiological functions of vitamin D is to maintain calcium and phosphorous levels to sustain a variety of metabolic functions and bone metabolism. Without vitamin D, only 10-15% of dietary calcium and approximately 60% of phosphorous is absorbed. Adequate amount of vitamin D is necessary to prevent rickets and osteomalacia (103). Low levels of 25(OH)D is associated with increased parathyroid hormone, which causes an increase in bone turnover, resulting in excess release of calcium from the bone, increasing the risk of osteoporosis and fractures (74).

Non-skeletal functions

Recently, there have been an increasing number of studies recognising non-skeletal effects of vitamin D, including cancers, diabetes, cardiovascular diseases and immune diseases (74).

Cancer

Observational studies have reported that higher levels of serum 25(OH)D were related to reduced incidence of many cancers (104). An umbrella review of meta-analyses of observational studies showed that high levels of 25(OH)D, compared to low levels of 25(OH)D, significantly reduced the risk of breast cancer and CRC (104). Colon, breast, and prostate tissues express 1- α-hydroxylase to produce $1,25(OH)_2D$ and have been hypothesised to prevent cancer by inducing cellular maturation, inducing apoptosis, and inhibiting angiogenesis. Once $1,25(OH)_2D$ completes its task in these cells, it initiates its own destruction by binding to the enzyme 24-hydroxylase, which will prevent it from entering the circulation that will influence calcium metabolism. Additionally, $1,25(OH)_2D$ has also been hypothesised to enhance the expression of P21 and P27, which are cyclin-dependent kinase (CDK) inhibitors in the G1 checkpoint of the cell cycle, to control cellular proliferation (76).

Autoimmune disease

There is a potential role of vitamin D in the pathogenesis of specific autoimmune diseases, including Type 1 Diabetes, multiple sclerosis, and Crohn's disease. In autoimmune disease, monocytes behave abnormally by over-secreting inflammatory chemicals, tumour necrosis factor alpha (TNF-α) and interleukin 6 (IL-6), which up-regulates the release of other inflammatory chemicals causing damage of healthy

tissues and the nervous system. Vitamin D suppresses the release of inflammatory cytokines of IL-6 and TNF-α in autoimmune conditions by monocytes (105).

Cardiovascular disease and Type 2 Diabetes Mellitus

In addition to contributing to autoimmune diseases, overproduction of inflammatory cytokines, caused by vitamin D deficiency may also contribute to the development and progression of heart failure. Moreover, low levels of $1,25(OH)_2D$ also results in elevated levels of PTH, which in turn has been associated with hypertension, cardiac arrhythmias, increased coronary calcification, and insulin resistance. However, the cause and effect mechanism of these relationships remain unclear. Vitamin D affects cardiac function by acting as a negative regulator of RAS, which helps regulate electrolyte and volume homeostasis. Disruption of the VDR gene leads to excessive RAS stimulation, which in turn is associated with cardiac hypertrophy and elevated blood pressure (106,107).

The presence of VDRs and VDBPs in pancreatic tissues suggests a role of vitamin D in insulin secretion. Vitamin D is essential for normal insulin release in response to glucose through the regulation of calcium. Vitamin D deficiency is associated with decreased pancreatic insulin secretion. One mechanism of action of vitamin D on insulin secretion involves the β-cell calcium-dependent endopeptidases, which converts proinsulin into insulin. Vitamin D may also act directly to induce β-cell insulin secretion by increasing the intracellular calcium concentration via non-selective voltage-dependent calcium channels (108,109).

An umbrella review of vitamin D reported a significant reduction in T2D comparing the top category of 25(OH)D levels to the bottom category (RR: 0.63, 95%CI: 0.56 – 0.69,

I^2=1%) (104). This study also reported a statistically significant reduction in cardiovascular disease comparing the top category of 25(OH)D levels to the bottom category (RR: 0.67, 95%CI: 0.55 – 0.82, I^2=74%) (104).

Most cardiovascular cells express 1-α-hydroxylase, enabling local synthesis of $1,25(OH)_2D$ and 24-hydroxylase. High levels of vitamin D may decrease blood pressure by many mechanisms including the RAS, PTH levels, and inflammation (104,106).

There are many other biological roles vitamin D has on the body including lung diseases (asthma, cystic fibrosis, chronic obstructive pulmonary disease (COPD), interstitial lung disease, and respiratory infections), adverse pregnancy outcomes, neurologic disorders, cognitive disorders (Alzheimer's disease and depression), and infectious diseases (76,104,110).

1.2.6 Epidemiology of Vitamin D Deficiency

Europe

Vitamin D deficiency is highly prevalent worldwide and has many potential health consequences (76). In 2014, 7.5% of children between the ages of 1.5 – 3 years in the UK were severely deficient in vitamin D (defined as less than 25 nmol/L, according to the USES). Moreover, according to the National Diet and Nutrition Survey, approximately 24.4% of girls aged 11 to 18 years, 16.9% of adult men, and 24.1% of adult women aged greater than 64 were also vitamin D deficient year round (111). However, these proportions increased during the winter months (~40% for adolescents and adults) and decreased during the summer months (13.4% and 8.4% for adolescents and adults respectively). In Germany, approximately 57% of both

males and females aged between 18 and 79 were vitamin D deficient (< 30 nmol/L). In Northern Europe (Denmark, Finland, Ireland, and Poland), nearly 67% of older men and 92% of adolescents had less than 50 nmol/L of serum 25(OH)D (112).

Middle East

Despite the long hours of sunlight in Qatar and surrounding regions, vitamin D deficiency has been shown to be highly prevalent in this region (113–118). For example, a Kuwaiti and an Emirati study reported that approximately 98% and 83% (respectively) of the participants had serum 25(OH)D levels less than 50 nmol/L (116,117). Approximately 46.2% of adolescents in Iran had less than 20 nmol/L of serum 25(OH)D levels (119) and around 81% of Saudi Arabian adolescent girls had serum 25(OH)D levels less than 25 nmol/L (120). Studies thus far conducted in Qatar on vitamin D were specific to a certain population (elderly and health professionals), in which these studies reported high prevalence of vitamin D deficiency (114,118).

1.2.7 Genetics of vitamin D

Twin and family-based studies suggest that genetics contribute to the variability of vitamin D. Such studies have the advantage of assessing relative effects of genetic and environmental factors on traits because they allow the estimation of genetic, common environmental and unshared environmental effects (121). Studies have reported that variability in serum 25(OH)D concentrations is mainly explained by additive genetic effect (122,123). A study on white male twins in the US reported only strong genetic influence in serum 25(OH)D during winter (November-March) (70%) and not summer (April-October) (0%) (121). On the contrary, a Swedish study performed on 204 adult male and female twins, reported that 48% of the total variance of serum 25(OH)D was explained by genetic variance in the summer (May-October),

while 0% of total variance was explained in the winter (November-April) (124). These differences in serum 25(OH)D levels in different seasons could probably explain the different results found in vitamin D studies (104).

Three genome-wide association studies on adult Europeans have identified SNPs associated with 25(OH)D concentrations (125–127). The latest study was published in 2018 by the SUNLIGHT consortium on 79,366 individuals. Overall these studies have highlighted 6 loci in association with vitamin D levels. The polymorphisms were located near the *GC, DHCR7, CYP24A1, AMDHD1, SEC23A,* and *CYP2R1* genes. *GC* encodes for VDBP that transports vitamin D to various tissues in the body. *DHCR7* is involved in the conversion of 7-dehydrocholestrol into pre-vitamin D, *CYP24A1* encodes 24-hydroxylase which initiates degradation of 25(OH)D and 1,25$(OH)_2$D, and *CYP2R1* encodes the enzyme that converts vitamin D into 25(OH)D (125). rs10745742 is located near the gene *AMDHD1* on chromosome 12, and rs8018720 is located near the gene *SEC23A* on chromosome 14. Little is known of the functions of these two genes for vitamin D.

The percentage of residual variance of circulating 25(OH)D explained by the SNPs in the gene regions found in the Jiang *et al.* study was 2.7%, after adjusting for known 25(OH)D covariates (128).

1.2.8 Environmental Risk factors of vitamin D deficiency

Although a large proportion of serum 25(OH)D variance is due to genetic heritability, approximately 57% of variance of serum 25(OH)D is due to common and unshared environmental variance (122). A review by Holick *et al.* reported that sunlight, age, sex, body size, season, ethnicity/skin pigmentation, sunscreen usage, and diet affect the levels of vitamin D (74). Moreover a literature review of determinants of vitamin D

levels found that HRT, oral contraceptive usage, menopausal status, smoking, physical activity, and cholesterol levels all affect the level of vitamin D concentrations (99–105). Below I will further discuss the relationship between these factors with vitamin D.

Age

According to two studies performed in the NHANES III population, the mean level of serum 25(OH)D decreased with age, regardless of season and gender (136,137). The first NHANES III study was performed on 18,875 participants aged ≥12 years. The mean level of serum 25(OH)D in winter and at lower latitudes for men aged >80 was 68.7 nmol/L compared to 78.6 nmol/L for men between the ages of 12 to 19. Older women (>80 years) also had lower levels of serum 25(OH)D in winter and at lower latitudes compared to younger women (12 – 19 years) (59.6 nmol/L and 64.9 nmol/L respectively) (136). Mean 25(OH)D remained higher in both young males and females compared to older males and females in the summer and at higher latitudes. One limitation to this study is that individuals that were taking vitamin D supplements were not excluded and thereby the levels of 25(OH)D could have been higher than individuals that did not take supplements. The second NHANES III study was performed on 15,390 adults (≥18 years) and excluded participants that were taking vitamin D supplementation and the conclusion remained consistent, older (>60 years) men and women had lower serum 25(OH)D (75.4 nmol/L and 64.5 nmol/L respectively) compared to younger (18-39 years) men and women (81.4 nmol/L and 77.0 nmol/L respectively) (137).

Sex

There are contradicting reports on the association between vitamin D levels and sex. A systematic review and meta-analysis on 82 studies of worldwide vitamin D status reported no significant sex-related differences in 25(OH)D (effect size (ES): 53.3 nmol/L 95%CI: 48.7 – 57.9, I^2=99.3% and ES: 54.2 nmol/L, 95%CI: 50.6 – 57.8, I^2=99.3 for females and males respectively) in Europe (138). Similar results were found for North American and the Asia/Pacific region. A nationwide cohort study by Chan *et al.* on blacks and non-Hispanic whites also reported that sex was not a significant predictor for serum 25(OH)D in either ethnic groups (139). These two studies observed substantial heterogeneity between the studies possibly due to unmeasured factors influencing vitamin D status.

While a more recent systematic review on 107 studies of vitamin D status among 630,093 individuals in southern European countries stratified by sex and age group, suggested that women had lower mean serum 25(OH)D levels compared to men (140). This might have been due to differences in lifestyle and personal characteristics of the individuals in this study.

Physical activity

Several studies have shown that physical activity is associated with serum vitamin D levels (141–144). A prospective study of approximately 10,500 adults (>35 years) from the NHANES III reported a positive association between physical activity (that meets the American College of Sports Medicine criteria) and serum 25(OH)D concentration (p-value trend: <0.01) (142). One limitation of this study was that physical activity was self-reported and therefore, misclassification could have occurred. A cross-sectional study on 559 adolescents between the ages 14-18 years measured physical activity using an accelerometer for one week and reported a positive correlation between

vigorous physical activity and serum 25(OH)D levels (partial correlation: 0.13, p-value: <0.01) (145). However, moderate physical activity showed no correlation with serum 25(OH)D (partial correlation: 0.06, p-value: 0.19) (145).

Oral contraceptives and Hormone replacement therapy consumptions

Studies have suggested that the use of oestrogen in HRT and in oral contraceptives may increase the levels of serum 25(OH)D. Women consuming oral contraceptives have been shown to have higher levels of 25(OH)D compared to women that do not (129,146–148). Moreover, women taking HRT had increased the levels of $1,25(OH)_2D$ in several studies (135,149,150). No significant association was found with 25(OH)D (p-value> 0.05) (135,150).

Statins

Statins are drugs that lower the production of cholesterol (151). A prospective cohort study of 91 hyperlipidemic patients were given 10 – 20 mg doses of rosuvastatin for 8 weeks as primary or secondary prevention, according to NCEP ATP III guideline. Serum 25(OH)D measurements were taken before treatment and after treatment. A significant decrease was found for cholesterol, triglyceride levels, and LDL-C. Furthermore, a significant increase was found in serum 25(OH)D after 8 weeks of rosuvastatin treatment (152). A RCT on 134 hyperlipidemic patients that were not on lipid lowering medication, compared the influence of two statin drugs, rosuvastatin and fluvastatin on the levels of 25(OH)D reported similar results (153). Patients were randomised in a 1:1 ratio to rosuvastatin 10 mg or fluvastatin 80 mg for 6 months. A significant increase in 25(OH)D was found in patients that were given the rosuvastatin

treatment (from 11.8 to 35.2 ng/mL, p-value< 0.001). No significant change was found for fluvastatin treatment (from 9.6 to 10.2 ng/mL, p-value: 0.56) (153).

Smoking

Several studies conducted in Europe and the US reported that smokers had significantly lower levels of serum 25(OH)D compared to non-smokers (144,154–157). The mechanism as to how smoking affects vitamin D metabolism remains unclear, however, it has been shown to be associated with bone loss in smokers (158). On the contrary, a Norwegian study on 7,161 participants with measured serum 25(OH)D reported that smokers had higher serum levels of 25(OH)D (72 nmol/L) compared to non-smokers (52.3 nmol/L) (159). This difference may have been due to the overestimation of serum 25(OH)D by 15-20% in smokers compared to non-smokers due to the immunoassay they used to measure 25(OH)D levels, which was not found using other immunological or LC-MS-MS methods (159). Therefore, the results are not as reliable for smokers compared to non-smokers.

Alcohol

Observational studies have reported an inverse association between alcohol consumption and vitamin D levels (142,160). A RCT on 2,753 vitamin D sufficient participants using a 2x2 factorial design of 1,000 IU/day supplementation of vitamin D_3, 1,200 mg/day of calcium carbonate, both, or placebo in the US and Puerto Rico, reported that the percent increase of serum 25(OH)D was higher in participants that drank less than 1 g/day of alcohol compared to participants that drank more than 30 g/day after one year of 1,000 IU of vitamin D supplementation (p-value: 0.04) (157).

On the other hand, some studies have shown that with increased consumption of alcohol, there was a decreased risk of vitamin D deficiency (141,144). The study by Tonnesen *et al.* recruited 700 participants and reported a 32% decreased risk in vitamin D deficiency (<25 nmol/L) for individuals that drank 5 or more units of alcohol in the last 7 days compared to non-drinkers (RR: 0.68, 95%CI: 0.47 – 0.97) (141). One hypothesis is that ethanol increases both liver size and liver blood flow, which in turn increases the transformation of vitamin D_3 into 25(OH)D, as demonstrated in an in vivo study using a rat model (161).

Ethnicity/skin pigmentation

A systematic review of 12 experimental studies of vitamin D production following exposure to UV radiation for different skin pigmentations concluded that skin pigmentation influenced vitamin D production in 7 of the 12 studies, while the remaining 5 studies did not show an association (162). They have concluded that pigmented skin reduced the UV-induced production of 25(OH)D in the blood compared to fair skin (162).

A cross-sectional study performed on 1,530 Hispanic and African American individuals in three different locations in the US (Colorado, California and Texas) showed that the levels of 25(OH)D were highest in Hispanics residing in Colorado (45.7 nmol/L) compared to Hispanics in Texas (36.5 nmol/L), and the lowest levels of 25(OH)D were found in African Americans residing in California (27.5 nmol/L) (133). An ecological meta-analysis by Hagenau *et al.* reported that Caucasians (68 nmol/L) had on average higher levels of serum 25(OH)D compared to non-Caucasians (47 nmol/L) (163). These studies all suggests that serum vitamin D is associated with skin pigmentation.

Sunscreen usage/skin coverage

There have been conflicting results with regards to the effect sunscreen has on serum vitamin D levels. A review of previous studies discussed both sides of the conflict (164) and concluded that although sunscreens can block nearly all the UV rays, (instead it allows a fraction of the UV-B to be transmitted), these conflicting results may be due to the amount of sunscreen applied and duration of those individuals that apply sunscreen remain in the sun (164). A RCT on 37 volunteers that measured the effect of thickness of sunscreen with the recommended amount of 2 $mgcm^{-2}$ reported an inverse association with serum 25(OH)D (mean increase in 25(OH)D: 6.40 nmol/L) (165). However, lesser thickness (<2mg cm^{-2}) of sunscreen reported significant increase levels of 25(OH)D (mean increase in 25(OH)D: >10.20 nmol/L) (165).

Skin coverage also plays an important role in reducing sunlight penetrance to the skin. A study performed on 146 Jordanians between the ages 18 and 45, investigated three groups of individuals based on skin coverage. Group 1 (n=20) were women wearing Western-type dress (more skin revealed), group 2 (n=80) were women covered except for face and hands, and group 3 (n=23) were women that were completely covered; including hands and face (the remaining 22 participants were men). The results showed a significantly lower serum 25(OH)D concentration in women that covered up completely (24.30 nmol/L in the summer and 22.70 nmol/L in the winter) compared to women that wore more Western-type dresses in both the summer and winter seasons (36.70 nmol/L in the summer and 30.90 nmol/L in the winter, p-value <0.05 for both seasons) (166).

A study on 21 healthy Caucasian adults aged 23-47 years, indicated that participants that were exposed to UVB on the whole body and the upper body had a significant

increase of vitamin D (increase per dose 1 minimal erythema dose (MED): 0.45 nmol/L, 95%CI: 0.16 – 0.20, for total body exposure), while participants with only face and hands exposed had no significant increase (increase per dose 1 MED: 0.03 nmol/L, 95%CI: -0.01 – 0.04) (167).

Season/latitude

Several studies from around the world reported higher levels of serum 25(OH)D during the summer months and the lowest during autumn and winter months (141,168–170). Higher latitude, which has been associated with lower amounts of sunlight, have been shown to be associated with a higher risk of vitamin D deficiency (157,160). However, a systematic review of 195 studies of vitamin D status worldwide, observed differences in serum 25(OH)D levels by region with North America (higher latitude) (overall 25(OH)D: 68.73 nmol/L, 95%CI: 63.71 – 73.75, I^2=98.9%) having the highest levels of 25(OH)D compared to Middle East/Africa (lower latitude) (overall 25(OH)D: 49.05 nmol/L, 95%CI: 40.61 – 57.48, I^2=99.4%) (138). There were a few limitations to this study including the high level of heterogeneity among studies, as well the possibility of publication bias (138). Conversely, the ecological meta-analysis performed by Hagenau *et al.* reported no overall effect of latitude on serum 25(OH)D levels (β: -0.03 nmol/L per 1 degree north or south from the equator, p-value: 0.80) (163). However, when stratified by ethnicity, Caucasians had a significantly lower serum 25(OH)D with increase latitude (β: -0.69 nmol/L per 1 degree, p-value: 0.02), while non-Caucasians reported no association with latitude (β: 0.03 nmol/L per 1 degree, p-value: 0.90)(163), which may be due to skin pigmentation.

Metabolic syndrome

There have been an increasing number of studies that have investigated the effects of vitamin D on the risk of MetS with most studies indicating an inverse association between vitamin D and the risk of MetS (171–173). A meta-analysis of four prospective cohort studies with 6,554 cases of MetS concluded that vitamin D was inversely associated with MetS with a RR of 0.86 (95%CI: 0.80 – 0.92, I^2=0%) (174).

Adiposity

A meta-analysis of 23 observational studies investigating the association between obesity and the risk of vitamin D deficiency reported that vitamin D deficiency was 35% higher in overweight and obese participants compared to the normal weight participants (RR: 1.35, 95%CI: 1.21 – 1.50, I^2=87.3%), irrespective of age and latitude(175). However, this study failed to indicate whether vitamin D deficiency leads to obesity or vice versa.

A bi-directional MR study which infers directionality of causality was conducted on the vitamin D and Coronary Artery Risk Development in Young Adults (D-CarDia) study, which includes 21 cohort studies consisting of 42,024 individuals. 12 BMI-related SNPs and four vitamin D-related SNPs were selected to conduct the MR analysis. This study demonstrated that higher BMI led to lower 25(OH)D (β: -0.42, 95%CI: -0.71 – -0.31), and not vice versa (176).

RCTs are the gold standard for inferring causality as well as for untangling the direction of the association. A systematic review and meta-analysis of 25-hydroxyvitamin D and calcium supplementation studies on 42,430 participants with a median treatment duration of 12 months reported no effect of vitamin D supplementation on adiposity measurements (mean difference for BMI: -0.06, 95%CI: -0.14 – 0.03, p-value: 0.20)

(177). Furthermore, a dose-response analysis was performed for doses of vitamin D_3 <1,000 IU/day, 1,000- <2,000 IU/day, 2,000 - <4,000 IU/day, and >4,000 IU/day and reported no effect of vitamin D in any of the dosage groups with adiposity measurements (p-value> 0.05) (177). These results could be explained as there being no biological effect of vitamin D supplementation on adiposity. It could also be due to the heterogeneity of the studies in which they differed in terms of methodology.

There are several possible mechanisms for lower vitamin D concentrations in obese individuals including lower dietary intake of vitamin D rich foods, less exposure of skin to the sun. Elevated PTH levels with low levels of serum 25(OH)D might affect calcium influx into adipose cells and impede body fat loss, and the sequestration of 25(OH)D in adipose tissues (178,179).

Type 2 diabetes and HbA1c

An umbrella review of meta-analyses of observation studies on vitamin D reported a 37% lower risk of T2D comparing the top category of 25(HO)D to the bottom category (RR: 0.63, 95%CI: 0.56 – 0.69, I^2=1%) (180).

A MR study was done to assess the causal association between vitamin D and HbA1c levels. The association between the four vitamin D-associated SNPs with HbA1c was obtained from the Meta-Analyses of Glucose and Insulin-related traits Consortium (MAGIC). The association between vitamin D-associated SNPs with 25(OH)D concentrations was obtained from two studies, Ely and EPIC-Norfolk. Results from this study showed no causal effect between 25(OH)D concentration and HbA1c per 25(OH)D-lowering allele (β:0.01, 95%CI: -0.04 – 0.05) (181).

A systematic review and meta-analysis of 23 RCT studies analysing the effects of vitamin D supplementation on HbA1c reported reduced HbA1c levels with vitamin D supplementation among T2D patients (standardised mean difference (SMD): -0.25, 95%CI: -0.45 – 0.05, I^2=75.5%) (182).

Several mechanisms have been proposed to explain the impact vitamin D has on insulin resistance. Vitamin D can affect insulin action directly by stimulating the expression of insulin receptor, thereby enhancing insulin responsiveness for glucose transport, or indirectly by influencing β-cell insulin secretion through a rise in cytosolic calcium concentration via non-selective voltage dependent calcium channels. Changes in calcium in primary insulin target tissues may contribute to peripheral insulin resistance leading to decreased glucose transporter-4 activity (80,109,183–185).

Blood pressure

A systematic review and meta-analysis on 11 prospective studies with 283,537 participants and 55,816 incident cases of hypertension reported a significant inverse association between circulating 25(OH)D levels and the risk of hypertension when comparing the top to the bottom third category of baseline 25(OH)D (RR: 0.70, 95%CI: 0.58 – 0.86) (186).

Meanwhile a MR study using data from the D-CarDia collaboration, International Consortium for Blood Pressure (ICBP), Cohorts for Heart & Aging Research in Genomic Epidemiology (CHARGE) and Global Blood Pressure Genetics Consortium (Global BPGen) were used to assess the causal relationship between vitamin D and hypertension risk. The results from this study reported an inverse causal effect

between 25(OH)D and the risk of hypertension (OR: 0.92, 95%CI: 0.87 – 0.97) per 10% increase in 25(OH)D concentration. A 10% increase in 25(OH)D concentration was associated with 0.29 mmHg lower DBP (95%CI: 0·07 – 0.52). However, no causal effect was found between 25(OH)D with SBP (β: -0.37, 95%CI: -0.73 – 0.003) (180).

On the other hand, a systematic review and meta-analysis of 46 RCTs on the association between vitamin D supplementation (<1600 IU/day) with the risk of hypertension showed no statistically significant effect on systolic blood pressure (SBP) (difference in mean: 0.00, 95%CI: -0.80 – 0.80, I^2=21%) or diastolic blood pressure (DBP) (difference in mean: -0.10, 95%CI: -0.60 – 0.50, I^2=20%) in participants taking vitamin D supplements compared to placebo (187).

The renin-angiotensin system plays an important role in the regulation of blood pressure, in which excess activity of the RAS can lead to hypertension. Based on the inverse relationship between $1,25(OH)_2D$ and plasma renin activity, one mechanism linking vitamin D to hypertension is its role as a suppressor of renin biosynthesis to regulate of the RAS (188).

Lipid profile

A NHANES III study on 8,421 men and women aged ≥20 years, reported no association between serum 25(OH)D and low HDL cholesterol (OR: 0.92, 95%CI: 0.72 – 1.17) when comparing the fifth highest category to the baseline category (173). Another study reported an inverse association between both total vitamin D (p-trend <0.0001) and dietary vitamin D (p-trend: 0.05) with low HDL cholesterol. An inverse association was found between dietary vitamin D only with hypertriglyceridemia (p-trend: 0.03) (189). No information was available for the association between serum 25(OH)D concentration and lipids.

Two MR studies were found for the potential causal association between vitamin D with HDL cholesterol and triglycerides. A MR study (n=10,601) on vitamin D and cardiovascular diseases reported higher HDL cholesterol and lower triglyceride levels per doubling of vitamin D (relative difference %: 23.8, 95%CI: 3.0 – 48.6 and relative difference %: -30.5, 95%CI: -51.3 – -0.8, respectively) when using filaggrin genotype as the instrumental variable (IV) for vitamin D status (190). Studies have shown that loss-of-function mutations in the filaggrin gene increase 25(OH)D concentration by 10% (190).

A bi-directional MR investigating the causal relationship between elevated cholesterol levels and vitamin D, reported that a 50% decrease in plasma 25(OH)D levels reported a potential causal association with lower HDL cholesterol levels (change in HDL cholesterol %: -6.0, 95%CI: -10.0 – -2.3), which supported their observational study results (change in HDL cholesterol %: -1.1, 95%CI: -1.7 – -0.5) (191). In the reverse direction, halving of HDL cholesterol levels increased plasma 25(OH)D levels (change in 25(OH)D %: 20.0, 95%CI: 7.4 – 34.0) which contradicted their observational results (change in 25(OH)D %: -1.5, 95%CI: -2.2 – -0.7) (191).

A meta-analysis of 8 RCTs investigating the association between serum 25(OH)D and lipid profile showed no statistically significant relationship between serum 25(OH)D with either HDL cholesterol (mean difference: -0.14, 95%CI: -0.99 – 0.71, I^2= 16%) and triglycerides (mean difference: -1.92, 95%CI: -7.72 – 3.88, I^2= 46%) (192).

There are several mechanisms that might explain the association between vitamin D and lipids. Vitamin D deficiency significantly increases 3-hydroxy-3-methylglutaryl coenzyme A reductase (HMGCR) expression which is involved in cholesterol uptake, which in turn is regulated by sterol regulatory element-binding protein 2 (SREBP-2)

(193). This suggests that vitamin D deficiency induces an increase in cholesterol levels by promoting cholesterol synthesis. Knocking down insulin-induced gene (Insig-1/2) in rats blocked 1,25$(OH)_2$D activation of SREBP-2 and increased the expression of HMGCR and cholesterol levels (193). Additionally, the knock down of VDRs in rats blocked 1,25$(OH)_2$D stimulating the decrease of Insig-2 expression, decreasing SREBP-2 and increasing HMGCR which increases cholesterol synthesis (193).

1.3 Metabolic syndrome

In the previous sections of this chapter, a relationship between vitamin D and CRC was found with MetS. One of the aims of this thesis is to investigate whether MetS mediates the relationship between vitamin D and CRC. Mediation analysis will be further discussed in the next section. This section will go into further detail on the definition, the epidemiology and the components of MetS.

1.3.1 Definition of Metabolic syndrome

The term MetS is use to describe a constellation of risk factors including abdominal obesity, high glucose, high triglyceride, and low high-density lipoprotein cholesterol levels, and hypertension that increase the risk of developing cardiovascular disease, T2D and other complex disease (194). MetS is also known as syndrome X, the insulin resistance syndrome, and the deadly quartet. Several definitions of MetS have been proposed, including those by the WHO in 1998 in which they focused on the diagnosis and classification of T2D, which provoked discussion from the European Group for the Study of Insulin Resistance (EGIR) and the American Association of Clinical Endocrinologists (AACE) in 2003. The National Cholesterol Education Program Adult Treatment Panel III (NCEP/ATPIII) was published in the US in 2001 to guide therapy for LDL-cholesterol and coronary heart disease (195). This was followed by a similar definition by the International Diabetes Federation in 2005 (196). The US NCEP/ATPIII includes any three or more of the abnormalities mentioned above (195). While the International Diabetes Federation (IDF) defined MetS as central obesity plus two of the four abnormalities (39). The cut-off points for each of the abnormalities differ between each of the definitions and can be found in **Table 1.5**.

Table 1.5. WHO, EGIR, NCEP-ATP III and IDF definitions of metabolic syndrome

	WHO	EGIR	NCEP-ATP III	IDF
Criteria	T2D or impaired glucose tolerance, or insulin resistance, plus ≥2 of the following	Hyperinsulinaemia Plus ≥2 of the following	Any ≥3 of the following	Central obesity Plus ≥2 of the following
Central obesity	BMI > 30kg/m^2 or WHR > 0.9 (M) or > 0.85 (F)	WC ≥94cm (M) WC ≥80cm (F)	WC ≥102cm (M) WC ≥88cm (F)	WC-ethnic specific or BMI > 30kg/m^2
Dyslipidaemia	TG ≥150 mg/dL or HDL-C < 35mg/dL(M) <39 mg/dL (F)	TG ≥177 mg/dL or HDL-C < 39 mg/dL	TG ≥150 mg/dL or medication HDL-C < 40mg/dL(M) <50 mg/dL (F) or medication	TG ≥150 mg/dL or medication HDL-C < 40mg/dL(M) <50 mg/dL (F) or medication
Blood pressure	≥140/90 mmHg	≥140/90 mmHg or medication	SBP ≥ 130mmHg or DBP ≥85mmHg or medication	SBP ≥ 130mmHg or DBP ≥85mmHg or medication
Other	Microalbuminuria: Albumin excretion ≥ 20µg/min		Fasting plasma glucose ≥100 mg/dL or medicaiton	Fasting plasma glucose ≥100 mg/dL or previously diagnosed T2D

WHO: World Health Organisation, EGIR: European Group for the Study of Insulin Resistance, NCEP-ATP III: The National Cholesterol Education Program Adult Treatment Panel III, IDF: International Diabetes Federation, T2D: Type 2 Diabetes, BMI: body mass index, WHR: waist to hip ratio, WC: waist circumference, TG: triglyceride, HDL-C: high density lipoprotein-cholesterol, SBP: systolic blood pressure, DBP: diastolic blood pressure
Source reproduced from Ngai *et al.*(197) with permission under the terms of the Creative Commons Attribution License

1.3.2 Epidemiology of Metabolic syndrome

The reported prevalence of MetS varies depending on the definition used, and ethnic background of studies populations. It has been estimated that around 15-30% of the world's adult population have MetS and most of these are residing in developed countries (198), with the lowest prevalence of MetS in France at 7% and the highest in the US (199). A study comparing the effect of ethnic origin on MetS in the US reported the prevalence of MetS was lowest for non-Hispanic whites and African Americans (23.8% and 21.6% respectively) and were highest for Mexican Americans

(31.9%) (200). Moreover, the prevalence of MetS is age-dependent. The prevalence of MetS in the US increases from 7% in participants aged 20-29 years to 44% in participants aged 60-69 (200).

A recent study from the National Health and Nutrition Examination Survey (NHANES) on adults aged 18 and older, reported that MetS increased from 25% during the period of 1988 to 1994 to 34% between 2007 to 2012, with similar increases for both men and women separately (201). The Diabetes Epidemiology: Collaborative Analysis of Diagnostic Criteria in Europe (DECODE) study included 9 populations studies performed in Finland, the Netherlands, the UK, Sweden, Poland, and Italy reported that 41% of men and 38% of women had MetS at baseline at ages 47 to 71 years (202).

The high prevalence of MetS has also been shown in the States of the Gulf Cooperative Council which include Kuwait, Oman, Qatar, Bahrain, Saudi Arabia and the United Arab Emirates. The prevalence of MetS in men ranged between 20.7% - 37.2% according to the ATP III definition and between 29.6% - 36.2% according to the IDF definition. The prevalence of MetS was higher in women, ranging from 32.1% - 42.7% according to the ATP III definition and between 36.1% and 45.9% according to the IDF definition (203).

1.3.3 Components of Metabolic Syndrome

Obesity

Obesity is mainly driven by unhealthy diet choices as well as physical inactivity. The adipocytes in the adipose tissues can become enlarged leading to reduced blood supply and hypoxia due to nutrient excess. Hypoxia then results in inflammation in

the adipose tissue that is associated with obesity related comorbidities (204). Although obesity is a main risk factor for MetS, not everyone that is obese will develop insulin resistance. Approximately 30-40% of the variance of BMI is explained by genetic factors. GWAS provided insight into the genetics of BMI and reported more than 700 genetic variants to be significantly associated with BMI. However, only a small proportion (5%) of these genetic variants explained the variance of BMI (205).

Insulin Resistance

Insulin resistance individuals have abnormal levels of fasting glucose and/or hyperglycaemia, or reduction in insulin. In these individuals, the pancreatic beta cells need to secrete more insulin to overcome hyperglycaemia. Over time the pancreatic beta cells produce insufficient amount of insulin leading to hyperglycaemia and T2D (199,204).

Dyslipidaemia

Dyslipidaemia consists of a spectrum of abnormalities which include an increase in lipoproteins, elevated triglyceride levels, increased LDL-C, and low levels of HDL-C. Impaired insulin signalling increases free fatty acid levels, which serves as a substrate for the synthesis of triglycerides in the liver (204). Free fatty acids also increase the levels of very low density lipoprotein (VLDL) production. Insulin regulates the clearance of VLDL and therefore individuals with insulin resistance have an increase in VLDL production and a decrease in VLDL clearance (204).

Hypertension

Hyperglycaemia and hyperinsulinemia activate the RAS by increasing the expression of angiotensinogen, which may contribute to the development of hypertension in insulin resistant individuals. Insulin resistance and hyperinsulinemia can lead to the activation of the sympathetic nervous system, which leads the kidney to increase sodium reabsorption, increasing cardiac output, and arteries constrict resulting in hypertension (204).

1.4 Mediation Analysis

In epidemiological studies it is often necessary to disentangle the pathways that link exposure to outcome. As reported in the previous sections, vitamin D, MetS and CRC are all associated with each other. However, it is difficult to understand the causal sequence of events between these factors. Methods such as mediation analysis and Mendelian randomisation can help understand the casual sequence of events. In this section mediation analysis will be discussed and Mendelian randomisation will be discussed in detail in the next section.

Mediation analysis is used to assess to what extent the exposure has an effect on the outcome or if there is some other factor on the pathway between the exposure and outcome (mediator) that explains an effect on the outcome. Mediation analysis is a hypothesis regarding causal network, where the mediator may be on the causal pathway between the exposure of interest and the outcome of interest. **Figure 1.2** shows a DAG of MetS and its components mediating the relationship between vitamin D and CRC.

1.4.1 Assumptions for Mediation Analysis

There are four steps to conduct a mediation analysis according to Baron and Kenny: (1) The exposure is correlated with the outcome; (2) The exposure is correlated with the mediator; (3) The mediator is correlated with the outcome controlling for the exposure; and (4) The effect of the exposure on the outcome controlling for the mediator should be zero.

If these conditions are met, then MetS completely mediates the vitamin D to CRC relationship. If the first three steps are met and not the fourth, then partial mediation is indicated.

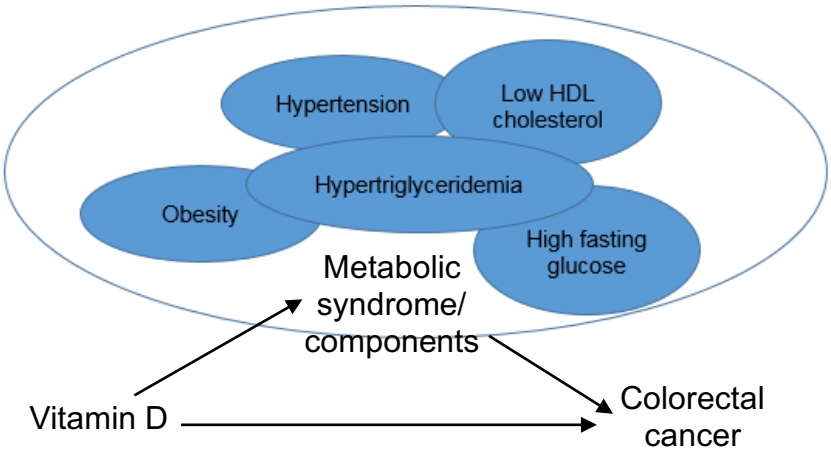


Figure 1.2. Directed acyclic graph of mediation where metabolic syndrome/components are mediators for the vitamin D to CRC association

1.4.2 Methods of Mediation

Several methods have been described for mediation analysis, including the difference of coefficient method, the product of coefficient method, and the causal method.

Difference of coefficient method: two regression models were conducted, one with the mediator and the other without the mediator. The beta values and SEs for each were obtained. If the ratio of the difference between the beta without mediator and beta with mediator divided by the SE of both models lies between the t-distribution of a large sample (between -1.96 – 1.96; p-value> 0.05) there is no mediation. Further details on calculations can be found in the Appendix B.

Product of coefficient method: Two regression models were conducted for this method. The regression of the mediator to the exposure model and the regression of the

exposure to the outcome adjusted for the mediator model. The coefficient of the exposure in the first regression was multiplied by the coefficient of the mediator in the second regression and divided by the SE of both models. This was then compared with the z-distribution for significance. More information on SE calculations can be found in Appendix B.

Counterfactual method/Causal inference:

In the counterfactual/causal method vitamin D was dichotomized into greater than or equal 50 nmol/L or less than 50 nmol/L, as this was defined to be vitamin D deficient according to the USES (206). The direct natural effect, measures the effect of the exposure on the outcome that is not through the mediator, the indirect natural effect, measures the effect of the exposure on the outcome that operates through the mediator, and the total effect, the product of the direct and indirect effect, were all calculated. Percent of mediation was calculated by dividing the beta estimates of the natural indirect effect by the total effect. **Equation 1** shows the calculation of direct and indirect effects.

$$E(M|X=x, C=c) = \beta_0 + \beta_1 x + \beta`_2 c$$

$$E(Y|X=x, M=m, C=c) = \theta_0 + \theta_1 x + \theta_2 m + \theta`_4 c \quad (1)$$

$$\text{The indirect effect} = \beta_1 \theta_2$$

$$\text{The direct effect} = \theta_1$$

Where M is the mediator, X is the exposure of interest, Y is the outcome of interest and C is the confounders. The direct effect is θ_1, which is the effect estimate for the association between exposure and outcome controlling for the mediator. While the indirect effect is the product of β1, which is the effect estimate for the association

between the exposure and mediator, and θ_2, which is the effect estimate for the association between the mediator and outcome controlling for the exposure.

1.5 Mendelian Randomisation

The observational associations presented above between vitamin D, MetS and its components and CRC suffer from several limitations which do not allow causal inference. These shortcomings of observational studies include confounding by lifestyle and socioeconomic status or other unknown factors which cannot always be measured and taken care of in observational research. Observational studies also suffer from reverse causation bias in which underlying disease already influences the levels of risk factor under investigation. In order to determine causes of a disease, the optimal way to address this is to conduct a RCT (207,208). While RCTs are considered the gold standard for disease aetiology, they have several limitations. RCTs can be expensive if the outcome is rare and requires a long duration of follow-up. Targeted treatment that has an effect only on the risk factor may not be available and some risk factors cannot be randomly allocated (for example smoking) due to ethical reasons. Moreover, compliance is often poor in RCTs and generalisability needs to be examined since RCTs are mostly conducted on a specific group of people (207,208). A number of RCTs to date have failed to confirm the protective effect of vitamin D supplementation and CRC risk seen in observational studies (51–53).

1.5.1 Mendelian randomisation definition

MR analysis uses genetic variants to infer causal effects between risk factors and disease outcomes. The name MR comes from the two laws of Mendelian inheritance: the law of segregation and the law of independent assortment meaning that the inheritance of one trait is independent to the inheritance of other traits (the traits are

unlinked or distantly linked) (208). This implies that there is random allocation of group of alleles (similar to a RCT) which then overcomes the limitation of confounding mainly due to socioeconomic status and lifestyle factors). MR does not suffer from reverse causation since the disease cannot change an individual's genotype (207,208). **Figure 1.3** below shows the similarities between RCT and MR.

In MR, genetic variants are used to form subgroups analogous to RCT. The subgroups differ only in the exposure of interest and the difference in outcome is due to the causal effect of the exposure on the outcome (208). The genetic variants of an individual are not randomly assigned from our parents, however, the distribution of genetic variants in the population can be considered random with respect to the lifestyle and socioeconomic factors as long as mating is done at random. This can be tested by performing a Hardy-Weinberg Equilibrium (HWE) test to determine if the frequency of homozygotes and heterozygotes is similar to what is expected (208).

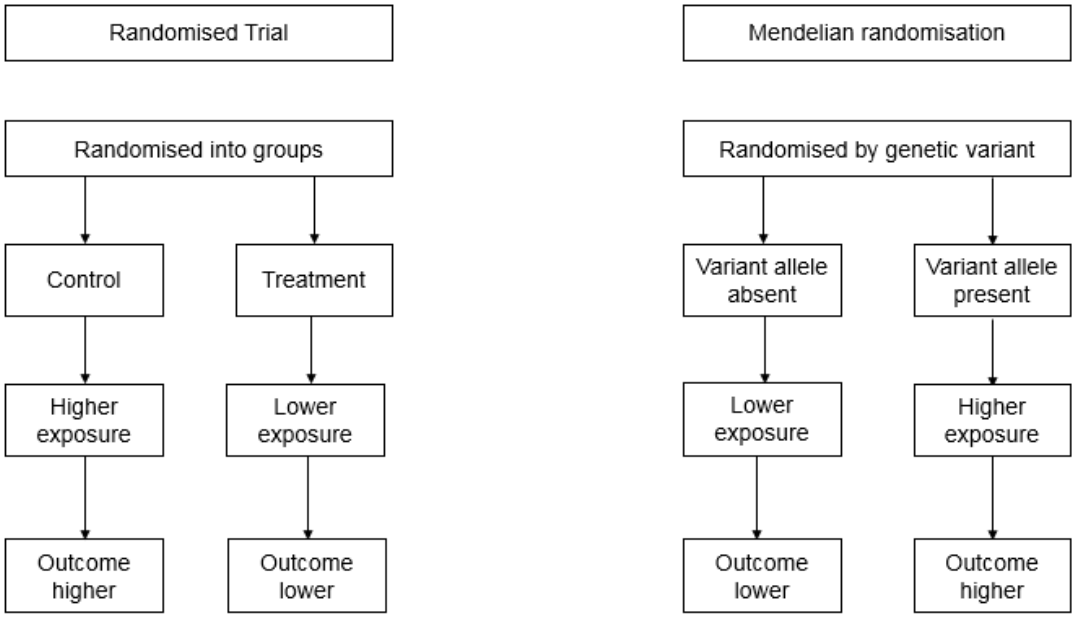


Figure 1.3. Comparison of a randomised controlled trial and Mendelian Randomisation. Reprinted from The Lancet, Vol. number: 366, Author(s): Aroon Hingorani and Steve Humphries, Title of article: Nature's randomised trials, Pages No.: 1906 - 1908, Copyright (2005), with permission from Elsevier (209).

1.5.2 Assumptions for causal inference

To perform a MR analysis, three conditions for the genetic variant (GV) to be an instrumental variable (IV) must be met:

1) The GV must be associated with the exposure of interest,
2) The GV is independent of the confounding factors that confound the association between the exposure and the outcome,
3) The GV is associated with the outcome only through the exposure (207).

This means that a genetic variant that is associated with only the exposure of interest and not with other risk factors that will affect the outcome must be found. It must be associated with the outcome only through the exposure (**Figure 1.4**).

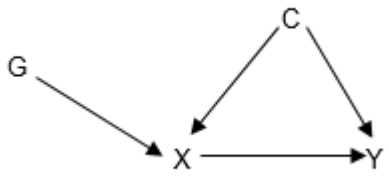


Figure 1.4. Directed acyclic graph (DAG) for the instrumental variables. G, instrumental variant; X, exposure; Y, outcome; C, confounder

There are three possibilities by which the IV assumptions can be violated. Firstly, pleiotropy, this refers to the GV being associated with multiple risk factors. This violates the second and/or third assumption making the genetic variant invalid. Horizontal pleiotropy occurs when the IV influences the exposure and the outcome through independent pathways (210). However, if the IV is associated with a trait which then influences another trait, this is considered vertical pleiotropy and does not violate the IV assumptions. Pleiotropy can be overcome by using IV located in genes that have only one known biological function.

The second category in which IV assumptions can be violated is non-Mendelian inheritance, in which genes do not follow Mendel's law of independent assortment. Linkage disequilibrium (LD) is when genetic variants are so close on the chromosome that they are inherited together thus not independent. LD can violate the third IV assumption if there is LD between the IV and a polymorphism that is associated with the outcome, in which the causal association between the exposure and the outcome will be confounded **Figure 1.5** (207). Another non-Mendelian inheritance violation is effect modification, in which there is an interaction between the effect of the exposure with a covariate which varies the causal effect across the strata.

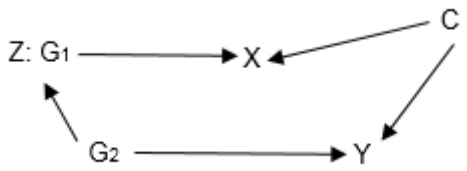


Figure 1.5. Linkage disequilibrium that leads to violation of the instrumental variable analysis. Z:G_1: genetic variant that is being used as the instrumental variable; G_2: genetic variant in linkage disequilibrium with $G_1$and related to Y ; X: exposure of interest; Y: outcome of interest; C: confounders

Source from Authors: Debbie A. Lawlor, Roger M. Harbord, Jonathan A. C. Sterne, et al., Title of article: Mendelian randomisation: Using genes as instruments for making causal inferences in epidemiology, Publishers: John Wily and Sons, Copyright © 2007 John Wiley & Sons, Ltd (207).

The third category is population effects. Population stratification occurs when the population under study can be divided into subpopulations. An association can be found between the IV and the exposure that could be due to the difference in subpopulation and not the IV, which would violate the IV assumptions. This can be controlled by restricting the study population, or by adjusting for genetic principal components in GWAS studies (208).

One other limitation to MR analysis is the usage of weak instruments which would provide little to no information. Normally, imprecise estimates of a causal effect would be expected with weak instruments, however, it has been recently realised that IV estimation performs badly with very weak instruments and can cause biased estimates with narrow confidence intervals (207). F-statistics greater than 10 are considered sufficient strength for the instruments (207).

1.5.3 Methods for instrumental variable analysis

There are many methods for IV analysis including the ratio of coefficient method (Wald ratio), two stage least square method (2SLS), inverse-variance weighted method (IVW), likelihood based methods, and semi parametric method. In this thesis I used the 2SLS, the IVW, and the likelihood based method. The 2SLS method consists of two regression stages: in the first stage, the exposure is regressed on the IV and in the second stage, the outcome is regressed on the fitted values of the first regression (208). This method was used for 1-sample MR analysis, where the association between IV and exposure and the association between IV and outcome are in the same dataset. The IVW method and the likelihood based method were both used in 2-sample MR analyses, where I obtained the summary statistics of the association between the IV and exposure and the association between the IV and the outcome from independent sources. The IVW method combines the ratio estimates from the individual variants using inverse-variance weights in a fixed effect meta-analysis model (**Equation 2**) (211). **Equation 2** shows the causal effect of the exposure on the outcome, where X_k is the beta estimate, regression coefficient, from the association between the IV and the exposure, Y_k is the beta estimate from association between

the IV and the outcome, and σ_{Yk} is the standard error (SE) of the association between the IV and the outcome.

$$\beta_{IVW} = \frac{\sum_k X_k Y_k \sigma^{-2} Y_k}{\sum_k X^2{}_k \sigma^{-2} Y_k} \qquad se(\beta_{IVW}) \sqrt{\frac{1}{\sum_k X^2{}_k \sigma^{-2} Y_k}} \tag{2}$$

The likelihood based model is constructed by assuming a linear relationship between the exposure and the outcome and a bivariate normal distribution for the genetic associations estimates with the exposure and the outcome (211).

1.6 Cohorts

1.6.1 Qatar Biobank

Data cohort

The aim of the QBB is to collect health, lifestyle and biological information from 60,000 men and women residing in Qatar and to follow them over the long term to record any incident health conditions. This will then be used to further investigate the genetic, lifestyle, and environmental risk factors, and their interactions for major chronic diseases affecting the population residing in Qatar. QBB was set up by the Qatar Foundation and the Supreme Council of Health in collaboration with Imperial College London.

As of the 16th of April 2017, 7,627 individuals participated in the QBB. In the current analysis, data from 1,205 QBB participants were available with 702 females and 503 males. The reason for the lower number of individuals included in this analysis is because the dataset was created in 2015; at that time approximately 1,200 participants were recruited. Participants who were either Qatari or non-Qataris who were long-term residents of Qatar (that have lived there for over 15 years) were included in the cohort. All participants gave informed consent. Participation was voluntary and registration was conducted via online, through the QBB website and telephone bookings. Participants completed questionnaires on health and lifestyle, including questions on socio-demographic factors, smoking habits, occupation, mobile phone use, and physical activity. All participants also completed a food frequency questionnaire on diet at the QBB centre, which asked participants how often they consumed various foods and beverages. The nurse would also record information on previous and

current health problems, family history of illnesses, medication use and in women, questions regarding reproductive factors.

Measured anthropometric measurements and body composition data were also obtained. Strength, vision, respiratory function, fitness, blood pressure, electrocardiogram, arterial stiffness and carotid ultrasound were also measured. Biological samples, which include ten tubes of blood (~60 ml), a urine sample and a saliva sample, were collected from each participant.

Preparation of blood samples, saliva, and urine

A proportion of the blood was used to measure 66 clinical biomarkers (including bone and joint markers, coagulation tests, diabetes related tests, differential white cell count, full blood count, sex steroid hormones, lipid profile, minerals, and vitamins) routinely for all participants (**Table 1.6**). Haematology and blood biochemistry were analysed by the laboratories at Hamad Medical Corporation (HMC).

Urine, saliva, and the remainder of the blood samples were divided into aliquots and stored into 2 dimensional barcode labelled microtubes at -80°C. The ethylenediaminetetraacetic acid (EDTA; used as an anticoagulant) blood samples were centrifuged to separate blood into its three layers: plasma; which consists of protein, hormone, minerals and salts, buffy coat (leucocytes), and erythrocytes. The buffy coat layer was aspirated using a hand pipette and pipetted into a 2-dimensional barcode labelled microtube -80°C.

Table 1.6. The 66 clinical biomarkers routinely measured in the Qatar Biobank

Group	Variable	Group	Variable
Bone and joint markers	Calcium	Sex steroid hormones	Estradiol
	Phosphorus		Sex hormone binding globulin
	Uric acid		Testosterone
	Vitamin D	Inflammation/Autoimmune	Rheumatoid factor
Coagulation tests	Activated partial thromboplastin time		C-Reactive protein
	Fibrinogen level	Lipid profile	Cholesterol
	International normalised ratio		High density lipoprotein
	Prothrombin time		Low density lipoprotein
Diabetes related tests	C-peptide		Triglycerides
	Glucose	Liver function tests	Albumin
	Glycated Haemoglobin A1c %		Alkaline phosphatase
	Insulin		Alanine transaminase
Differential white cell count	Basophil		Aspartate transaminase
	Basophil %		Gamma glutamyl transferase
	Eosinophils		Total bilirubin
	Eosinophils %		Total protein
	Lymphocytes	Minerals	Iron
	Lymphocytes %		Ferritin
	Monocyte		Magnesium
	Monocyte %		Total iron binding capacity
	Neutrophils	Muscle markers	Creatine kinase
	Neutrophils %		Myoglobin
	White blood cell	Thyroid function tests	Free triiodothyronine
Electrolytes and renal function tests	Chloride		Free thyroxine
	Serum creatinine		Thyroid stimulating hormone
	Bicarbonate	Vitamins	Vitamin B12
	Potassium		Folate serum
	Sodium	Other tests	Homocysteine

Group	Variable	Group	Variable
	Urea nitrogen		N-terminal brain-type natriuretic peptide
Full blood count	Haematocrit		
	Haemoglobin		
	Mean corpuscular haemoglobin		
	Mean corpuscular HGB concentration		
	Mean corpuscular volume		
	Mean platelet volume		
	Platelets		
	Red blood cell		

Source reproduced from Al Kuwari *et al.*(212) with permission under the terms of the Creative Commons Attribution License

Exposure and outcome assessments

Vitamin D assessment

25(OH)D was analysed in one lab at HMC, which is College of American Pathologist (CAP) accredited, using a LIAISON® 25 OH Vitamin D TOTAL Assay, a chemiluminescent immunoassay, where serum 25(OH)D_2 and 25(OH)D_3 were measured. The inter-assay CV for this methodology was 10.6% and the inter-assay CV was 5.4%. The lowest limit of detection was 10 nmol/L. This methodology was able to distinguish between vitamin D_2 and vitamin D_3, and the cross-reactivity with 3-epi-25(OH)D_3, to avoid overestimation, was small (<2%). In this study vitamin D deficiency was defined by the USES guidelines as less than 50 nmol/L in circulation, while individuals with 50-75 nmol/L were considered insufficient, and individuals with greater than 75 nmol/L were considered sufficient (99).

Anthropometric measures assessment

Standing height, weight and body composition, waist and hip measurements were all measured by a trained nurse. Participants wore gowns or light-weight clothing, and shoes and socks were removed. Height was measured with their feet flat on the surface of the base of the plate of the Seca-274 Stadiometer instrument. Feet were placed together and heels against the back of the plate with the participants' back as straight as possible against the rod with arms hanging by their sides. The Stadiometer head plate was brought down on top of their head (without any head gear, including caps, turbans, and hair styles). This data was then automatically imported and stored on the database. If the reading did not automatically import, the data was added manually.

Weight and body composition was measured using the Tanita BC-418 MA instrument, a multi-frequency segmental body composition analyser, or a digital floor scale Seca-876 for manual weight measurement when Tanita was contraindicated. Participants were then instructed to stand still on the scales' platform for Tanita to measure their weight. After the weight was measured, the participant held on the grips with both hands down by their sides to measure body composition. The results were then automatically sent to the database.

BMI was calculated as weight in kilograms divided by height in metres squared (kg/m^2). These measurements were taken from Tanita for weight and Seca for height.

Waist and hip measurements were measured using a non-stretchable sprung measuring tape by Seca. The waist was identified as the smallest part of the trunk after folding their arms across their chest. If it was not possible to find a natural indent of the trunk, the circumference around the umbilicus was measured in centimetres. This data was then entered manually into the computer. For the hip measurement, the Seca measuring tape was placed at the widest part of the hips and measured in centimetres. Waist to hip ratio was calculated as waist in centimetres divided by hip circumference also in centimetres.

Metabolic syndrome components

MetS was defined according to the new IDF definition as being centrally obese (defined as waist circumference ≥94 cm for males and ≥80 cm for females) as well as two of the following four components: (1) raised triglycerides (≥1.7 mmol/L), (2) reduced HDL cholesterol (<1.03 mmol/L in males and <1.29 mmol/L for females), (3) raised blood pressure (SBP ≥130 or DBP ≥85 mmHg), (4) raised fasting plasma

glucose (defined as HbA1c levels ≥ 5.7 mmol/L) (198). HbA1c was measured with turbidimetric inhibition immunoassay (TINIA). Cholesterol was measured using CHOD-PAP GEN2 STAND ID/MS, HDL cholesterol was measured using the HDL–C plus 3rd generation.

Blood pressure was measured by a trained nurse using an Omron M10-IT automated upper arm blood pressure monitor with arm cuffs of different sizes. Participants were seated in a chair with feet flat on the floor with an arm outstretched on a table, palm facing upwards. The reading was collected twice if the results were similar, but if different (differed by more than 5 mmHg) a third measurement was performed. The participant rested for 30 seconds between the measurements. If the measurements did not differ, the average of the first two measurements was taken. However, if the measurements did differ, the average of the second and third measurement was used. All readings were automatically imported and stored on the database.

Diabetes was reported in two ways: (1) self-reported; where 15.4% reported having diabetes and (2) laboratory measurements of HbA1c levels (≥6.5% was considered evidence for diabetes) (213), where 17.4% of participants were considered diabetic.

1.6.2 European Prospective into Cancer and Nutrition (EPIC)

Data cohort

EPIC is a multi-centre prospective cohort of approximately half a million men and women (approximately 70% women), which aims to investigate the nature of diet-cancer associations. EPIC was initiated in 1990 and data collection was completed in 1999. Information on diet, lifestyle, anthropometric measurements, and medical

history was collected from participants via 23 study centres in ten European countries: France, Germany, Greece, Italy, the Netherlands, Spain, Sweden, UK, Denmark, and Norway. Potential study subjects were contacted using several methods including mail, phone, and personal contact. Questionnaires on diet and lifestyle were either mailed or given by hand. Blood samples were collected at baseline according to standardised procedures, and stored at the International Agency for Research on Cancer (IARC) at -196°C liquid nitrogen for all countries except Sweden, in which it was stored in -80°C freezers (**Table 1.7**. All study participants provided written informed consent. Ethical approval was obtained from the review boards at IARC and local participating centres.

Anthropometric measurement assessments

The standard common protocol for anthropometric measurement was shared amongst the centres. Body weight were measured with subjects wearing light underwear using an electronic digital scale (Soehnle, type 7720/23). Height was measured using a flexible anthropometer. Waist circumference was measured midway between the lower rib margin and the superior anterior iliac spine (214). Hip circumference was measured as the widest point over the greater trochanters (214). BMI was calculated as body weight divided by height squared and waist to hip ratio (WHR) was calculated as waist circumference divided by hip circumference.

Serum vitamin D measurements

Biological samples, which included plasma and serum were collected from approximately 388,000 individuals. Serum 25(OH)D concentrations was measured using the enzyme immunoassay OCTEIA 25-Hydroxy Vitamin D kit in the Medical Research Council Human Nutrition Research laboratories in Cambridge. The intra-

assay and inter-assay CV were less than 8% at any time point. The lowest limit of detection was 5 nmol/L (215,216). Vitamin D deficiency was defined by the USES guidelines as less than 50 nmol/L in circulation, while individuals with >50 - ≤75 nmol/L were considered insufficient, and individuals with greater than 75 nmol/L were considered sufficient (93).

Metabolic syndrome assessment

MetS was defined in two ways; one according to the new IDF as being centrally obese (defined as waist circumference ≥94 cm for males and ≥80 cm for females) as well as two of the following four components: (1) raised triglycerides (≥1.7 mmol/L), (2) reduced HDL cholesterol (<1.03 mmol/L in males and <1.29 mmol/L for females), (3) raised blood pressure (SBP ≥130 or DBP ≥85 mmHg), (4) raised fasting plasma glucose (defined as HbA1c levels ≥ 5.7 mmol/L). The other definition was by the ATPIII which is defined as having any three of the following components: (1) waist circumference >102 cm in males and >88 cm in females, (2) raised triglycerides 1.7 mmol/L, (3) reduced HDL cholesterol (<1.03 mmol/L in males and <1.29 mmol/L in females), (4) raised blood pressure (SBP ≥130 or DBP ≥85 mmHg), (5) raised fasting plasma glucose (defined as HbA1c levels ≥ 110 mg/dL).

Serum HDL cholesterol and triglyceride concentrations were determined using a Synchron LX-20 Pro autoanalyzer (Beckman-Coulter; ref. 6). Measurements of HbA1c was done using high-performance liquid chromatography with a Bio-Rad Variant IITM instrument (217).

SBP and DBPs were measured in millimetres of mercury at baseline. Most centres applied techniques in a similar manner, where they performed two measurements on

the right arm while the participant was seated. However, the devices used to measure blood pressure, as well as the time between measurements varied between the centres (218). The average of the two measurements was used for both SBP and DBP.

Colorectal cancer assessment

Colon and rectal cancers were defined according to the International Classification of Diseases (ICD) for Oncology (ICD-O-3) site codes. Colon cancers were defined as cancers of the cecum, appendix, ascending colon, hepatic flexure, transverse colon, splenic flexure, descending colon, sigmoid colon, overlapping lesions of the colon and unspecified origin tumours (C18.0 – C18.9). Incident cancer cases were identified using population cancer registries in Italy, the Netherlands, Spain, and the UK, while in France, Germany and Greece, cases were identified during follow-up by a combination of methods including health insurance records, cancer and pathology registries, and by active follow-up directly through study participants or through next-of-kin.

Confounding factors assessment

BMI was measured in kg/m^2. Most centres measured weight and height, however the French centres had self-reports for weight and height in all participants, with only a subgroup that had direct measurements. In Norway and the Oxford Health conscious study, participants self-reported weight and height. Waist circumference was measured at either the narrowest torso circumference (France, Italy, Utrecht, Heidelberg, Denmark) or at the midway between the lower ribs and iliac crest (Bilthoven, Potsdam, Malmö, Oxford general

population). No instructions on self-waist measurement were provided to the Oxford health conscious group. In Spain, Greece, Heidelberg and Cambridge, a combination of methods was used, whereby the majority of participants were measured at the narrowest circumference. If the narrowest circumference could not be recognised, waist circumference was measured at the midway between the lower ribs and iliac crest.

The EPIC questionnaire consists of three different types of physical activity assessment: occupational, recreational, and household activity. A combined total physical activity index was generated to combine data for occupational, recreational, and household physical activity in sex-specific quartiles (inactive, moderately inactive, moderately active, and active). Questionnaire on physical activity can be found in Appendix D.

Information on current smoking habits was reported in the EPIC non-dietary questionnaire. The combined status of smoking and intensity of smoking was used in this study (never, former smoker and smoked less than 10 years, former smoker and smoked greater than or equal 10 years, former smoker with unknown duration of smoking, current smoker and low smoking intensity (1-15 cigarettes/day), current smoker and moderate smoking intensity (16-25 cigarettes/day), current smoker and high smoking intensity (≥26 cigarettes/day), and unknown). Questionnaire on smoking can be found in Appendix D.

Education was classified in a unique way where it synthesised information from different education systems across Europe (none, primary school, technical/professional school, secondary school, university, and unspecified),

Alcohol intake was measured at recruitment with current intake measured in the dietary questionnaire and past consumption in the non-dietary questionnaire. Information on past alcohol consumption was assessed as glasses of beverage consumed per week at different ages. The 24-hour dietary recalls (HDR) data were used to estimate average alcohol content in grams/day. The alcohol measurement used in this study was alcohol intake (beer, wine, fortified wine, and liquor) at different ages (20, 30, 40, and 50 years) at recruitment in grams/day. Questionnaire on alcohol consumption can be found in Appendix D.

Food items available in the dietary questionnaire are listed in two separate reports: foods/ingredients and mixed recipe, which is broken down into several ingredients. The dietary methods used in most EPIC countries are detailed dietary history questionnaires, structured by meals or occasions or not (Spain, France, Germany, Netherlands, Greece), non/semi- quantitative Food Frequency Questionnaires (FFQ) (United Kingdom, Denmark, Umeä and Norway), and a complex method combining a FFQ and a menu book in Malmö. In United Kingdom, two dietary methods were used at baseline (i.e a FFQ and a 7 day diary) and subsequently repeated. All dietary questionnaires are self-reported except in Spain and Ragusa, where the dietary information was collected by a face-to-face interview. The same definition and riles to break down recipes into ingredients used for the 24-hour diet recalls were applied to the EPIC dietary questionnaires, which increased the comparability of the data across countries. Fruits measured in grams/day included fruits, nuts and seeds, mixed fruits and olives, vegetables measured in grams/day included leafy vegetables, fruiting vegetables, root vegetables, cabbages, mushrooms, grain and pod vegetables, onion, garlic, stalk vegetables, sprouts, mixed salad, and mixed vegetables. Meat consumption measured in grams/day included fresh meat, beef, veal, pork,

mutton/lamb, horse, goat, poultry, chicken, hen, turkey, duck, goose, rabbit, game, processed meat, and offal. Total dietary energy consumption was generated as the sum of factored contribution from fat, protein, carbohydrates, and alcohol in kcal.

Genetic data

The current study includes participants from all EPIC centres except Norway and Denmark, where 2,400 cases of CRC and their matched controls were genotyped at the Centre for Inherited Disease Research (CIDR) at John Hopkins University. Samples were genotyped in batches with 96-well plates with one batch per plate. Each batch contained two HapMap controls.

Genotyping was performed using the Illumina HumanOmniExpress + Exome array (HumanOmniExpressExome-8v1-2, BPM annotation version A, genome build GRCh37/hg19) and using the calling algorithm GenomeStudio version 2011.1, Genotyping Module version 1.9.4, GenTrain Version 1.0. The array consists of a total of 964,193 SNPs. Genotypic data that passed initial quality control at CIDR were released to the Quality Assurance/Quality Control (QA/QC) analysis team at the University of Washington Genetic Analysis Centre (UW GAC), and to dbGaP. Approximately 98% of the SNPs with a missing call rate reported that less than 2% of these SNP showed no significant difference between missing call rates of cases and control status, since a missing call rate difference between cases and control can lead to false association (219).

Table 1.7. Number of participants included for questionnaire and blood sample in each country in EPIC

Participants included from each country		
Country	**Questionnaire**	**Questionnaire and Blood sample**
Denmark	57,054	56,131
France	74,524	21,053
Greece	28,555	28,483
Italy	47,749	47,725
Netherlands	40,072	36,318
Norway	37,215	11,000
Spain	41,440	39,579
Sweden	53,826	53,781
United Kingdom	87,942	43,141
Germany	53,091	50,678

1.6.3 UK Biobank

Data cohort

The UK Biobank is a national and international health resource aimed to improve the prevention, diagnosis and treatment of diseases, including cancer, heart diseases, stroke, diabetes, arthritis, depression, and dementia (220). The UK Biobank has recruited approximately 500,000 participants aged between 40-69 years between the years 2006 to 2010 from across the country. Potential participants received, by mail, information about the UK Biobank with an invitation to attend a local study Assessment Centre and after staff answered questions the potential participant had, the individual decided to either take part in the study and signs the consent form, or not (221). Questionnaires about lifestyle and other factors (medical history, mood, cognitive function, etc), as well as baseline physical measurement were performed at enrolment.

Biological samples were also taken, including urine, saliva and blood samples for future analysis (221).

Serum vitamin D measurements

To date, the biomarker for circulating 25(OH)D in the UK Biobank has yet to be measured.

Colorectal cancer assessment in the UK Biobank

Cancer was classified according to the ICD that was obtained through linkage to the national cancer registries with data available for prevalent and incident cases of cancers. CRC was classified according to the ICD 10 classification as C18.0 – C18.9, C19, and C20 and ICD 9 classification which include 153.0 – 153.9 for cancer of the colon and 154.0 and 154.2 for cancer of the rectum.

Confounding factors assessment

BMI was measured in kg/m^2 and was constructed from height and weight measured during the initial assessment centre visit. Weight (kg) was measured following the removal of shoes and heavy outer clothing using Tanita BC-148MA body composition analysers. Height (m) was measured using a Seca 202 device.

Smoking behaviour was collected by questionnaire or interview during the initial assessment visit. Smoking status was categorised into never smoker, previous smoker, and current smoker.

Education was categorised into seven categories (none, college or university degree, A levels/AS levels or equivalent, O levels/GCSE or equivalent, CSEs or equivalent, NVQ or HND or HNC or equivalent, and other professional qualification).

Ethnicity was asked during the initial assessment centre visit and was categorised into white, mixed, Asian or Asian British, Black or Black British, Chinese, or other ethnic group.

Genetic data in the UK Biobank

The UK Biobank has genotyped approximately 488,000 participants. Genotype calling was performed by Affymetrix on two arrays: approximately 50,000 participants samples were run on the UK BiLEVE array and the remaining were run on the UK Biobank Axiom array where both datasets have been merged to produce a single format. There are 805,426 markers in the genotype data and approximately 96 million genotypes imputed. The process of imputation involved pre-phasing the directly genotyped markers on both the UK BiLEVE and UK Biobank Axiom arrays with the 1000 Genome Phase 3 dataset used as a reference panel followed by a haploid imputation step using the program IMPUTE4. Further information about the genetic data and quality control is reported elsewhere (222). The positions of the markers were on the GRCh37 coordinates. The data was obtained from the UK Biobank portal, where it was downloaded, decrypted and converted into a Stata format. The 25(OH)D-associated SNPs and the bridging file, used to combine the 25(OH)D-associated SNPs with the general dataset, were also obtained through the UK Biobank portal.

Study subjects with no data available on the four 25(OH)D-associated SNPs were excluded from the analysis (N=15,252). Participants with prevalent and incident

cancers other than CRC or diagnosed prior to age 20 years were also excluded from the study (N=76,637). Participants consuming vitamin D supplements were also removed (N=15,993). Moreover, two participants were excluded due to withdrawal from the study, in which 394,774 participants remained. Individuals that were related (to the 3rd degree or closer) were excluded (N=86,487) in a sensitivity analysis

1.7 Aims

The overall aim of this thesis was to investigate the relationship between vitamin D, MetS and its components, and the risk of CRC using a molecular epidemiology approach to highlight the potential causal pathways linking these phenotypes.

The specific aims were:

(1) To assess the relationship between vitamin D and MetS and its components using data from Qatar Biobank;

(2) To assess whether the association between vitamin D and CRC is mediated through MetS or its individual components in the EPIC study;

(3) To assess the potential causal relationship between vitamin D and CRC using a MR approach;

(4) To assess the potential causal relationship as well as the direction of the associations between vitamin D and MetS components (BMI, HDL cholesterol, triglycerides, blood pressure and HbA1c levels).

CHAPTER 2 VITAMIN D AND METABOLIC SYNDROME AND ITS COMPONENTS IN A QATARI POPULATION

** This work has been published in the Journal of Nutrition and Diabetes in 2017 with the title of "Prevalence of vitamin D deficiency and association with metabolic syndrome in a Qatari population". I wrote the article as well as performed all the analyses.*

In this Chapter, I will estimate the prevalence of vitamin D deficiency and MetS within the Qatari population and investigate the association between MetS and its components with vitamin D, in the well-characterised Qatar Biobank (QBB).

Introduction

Vitamin D deficiency is highly prevalent worldwide and is associated with many adverse health outcomes (76,104). Vitamin D is acquired in three ways; from sun exposure, diet, and supplements; however, the greatest proportion is obtained from sun exposure. For example, exposure to 0.5 minimal erythemal dose is equivalent to supplementing approximately 3,000 IU of vitamin D_3 (76). One of the main physiological functions of vitamin D is to maintain calcium and phosphorous levels in the body to sustain various metabolic functions including bone metabolism (74). The most abundant circulating biomarker of vitamin D status is 25(OH)D which also has a longer half-life (25 days) compared to the active metabolite; $1,25(OH)_2D$ (7 hours) (223). However, the threshold used to define vitamin D deficiency often varies according to the population and outcome of interest (96,98,224). The most common definition for optimal 25(OH)D levels is the concentration at which it supresses the parathyroid hormone to its minimum, however, using this definition results in a wide range of minimal optimal values from 30 nmol/L to 100 nmol/L (96).

A number of studies have shown that vitamin D levels are inversely associated with the risk of a diverse set of diseases including several cancers, diabetes, and cardiovascular diseases (100,225–228). The association between vitamin D levels and a range of disease outcomes could be explained by intermediate disease risk factors, such as MetS; a constellation of risk factors including increased obesity, hypertriglyceridemia, hypertension, and diabetes. Some studies have suggested an inverse associations between serum vitamin D and MetS (104,174,228–231), while others have not confirmed this observation (232,233). It is also unclear which

components of the MetS might drive this association with vitamin D with some studies suggesting obesity and others glucose haemostasis (228,231,234,235).

Despite the long hours of sunlight in Qatar and surrounding regions, vitamin D deficiency has been shown to be highly prevalent in this region (113–118). For example, a Kuwaiti and an Emirati study reported that approximately 98% and 83% (respectively) of the participants had serum 25(OH)D less than 50 nmol/L (116,117) . Studies thus far conducted in Qatar on vitamin D were specific to a certain population (elderly and health professionals) (114,118). At the same time, the prevalence of MetS was measured to be approximately 26.5%, according to the IDF criteria, amongst Qataris aged 20 and over (236). Similarly, there was a high prevalence of MetS, according to both the IDF and Adult Treatment Panel III (ATPIII) criteria, and its components in neighbouring regions (237–240). Given the high prevalence of these two conditions in the region and the limited epidemiological evidence in this population so far, we aimed to investigate the prevalence of vitamin D deficiency, as well as the association between MetS and its components with vitamin D, in the well-characterised QBB.

Methods

Data for this analysis was obtained from the QBB. Information on study design, anthropometric measurement, exposure, and outcome measurements are provided in Chapter 1 section 6.

Statistical analysis

Descriptive analysis

To test the difference in the distributions of categorical variables, a chi-square test was performed, while a t-test was performed to test the differences between a categorical and a continuous variable after stratifying by sex. A geometric mean for serum 25(OH)D was calculated by performing a one-way Analysis of Covariance (ANCOVA) adjusting for sex, age, and season of blood collection. Partial Spearman's correlation coefficient was used to measure the cross-sectional association between anthropometric measurements, circulating lipids, HbA1c, glucose, insulin, C-peptide, folate, vitamin D, calcium, and sex hormone binding globulin.

Vitamin D and metabolic syndrome

Vitamin D was modelled as natural log transformed continuous measurements. MetS and its components were categorised into having abnormal levels or normal levels based on the IDF definition (198). Two models were generated, Model 1 was adjusted for age (continuous) and sex and a multivariable model (Model 2) was adjusted for age, sex, ethnicity (Qatari and non-Qatari), education (less than primary school, primary school, secondary school, technical/professional school, and university/postgraduate), physical activity (hours/week) and season (winter, spring,

summer and fall). Participants with missing ethnicity values (n=242) were excluded from all analyses.

Results

Descriptive Characteristics

Approximately 58% of the study population in the QBB were females; mean age was similar between males and females (**Table 2.1**). There were several differences between males and females with regards to anthropometric factors: men had a mean visceral fat of 97 cm compared to 85 cm for women and reported to exercise more frequently (19 hrs/week) than women 11 hrs/week). BMI as well as the prevalence of diabetes (~15%) and MetS (~28%) were similar between males and females (**Table 2.1**).

Prevalence of vitamin D deficiency

Approximately 64% of the participants in this study were vitamin D deficient (<50 nmol/L) with slightly more men (69%) being vitamin D deficient compared to women (61%). Another 25% of the population had insufficient vitamin D levels. **Table 2.1** reports the prevalence of stages of vitamin D deficiency according to USES definitions (99). The percentage of individuals reported taking vitamin D supplements was high with 49% of females and 25% of males (**Table 2.1**). Women were more likely to undergo vitamin D screening, and take vitamin D and calcium supplements compared to men.

Table 2.2 shows crude and adjusted mean vitamin D levels across categories of participant characteristics. Females had a higher concentration of serum 25(OH)D compared to males in both the adjusted and unadjusted mean 25(OH)D, although this difference was not statistically significant. Also, higher 25(OH)D levels were observed

in older compared to younger participants. The remaining characteristics did not show differences according to 25(OH)D levels (**Table 2.2**).

Associations between serum 25(OH)D levels metabolic syndrome

Table 2.3 reports the association between vitamin D and MetS and its components. A 10% increase in 25(OH)D levels was associated with less presence of MetS (OR: 0.94, 95%CI: 0.91 – 0.98). Serum 25(OH)D levels were inversely associated with some components of the MetS including waist circumference (OR per 10% increase in 25(OH)D: 0.96, 95%CI: 0.93 – 0.996) and triglyceride levels (OR per 10% increase in 25(OH)D: 0.93, 95%CI: 0.90 – 0.97). No association was found with HDL cholesterol, blood pressure, or HbA1c (per 10% increase in serum 25(OH)D OR: 0.99, 95%CI: 0.96 – 1.03, OR: 0.96, 95%CI: 0.92 – 1.01, and OR: 0.99, 95%CI: 0.95 – 1.04 respectively).

When vitamin D was categorised into deficient and non-deficient, individuals that were deficient were more likely to have MetS (OR: 1.54, 95%CI: 1.09 – 2.18, p-value: 0.01), high blood pressure (OR: 1.60, 95%CI: 1.06 – 2.42, p-value: 0.03), high waist circumference (OR: 1.39, 95%CI: 1.01 – 1.92, p-value: 0.04), and high triglyceride levels (OR: 2.30, 95%CI: 1.58 – 3.34, p-value <0.01) (**Table 2.4**). Moreover, age was the main confounder responsible for the statistically significant association between 25(OH)D concentrations and MetS and its components.

Discussion

In this unique study in Qatar, I showed that despite being a country with high levels of sun exposure, the prevalence of vitamin D deficiency is very high (64%) with the majority of the examined population showing signs of deficiency. This is consistent with previous observations which estimated the prevalence of vitamin D deficiency (<50 nmol/L) to range between 72% - 87% (114,118). At the same time, the percentage of participants reporting supplementation for vitamin D, especially in women was very high (49%), higher than that observed in white populations (156,241,242), potentially showing a greater awareness of vitamin D deficiency in this population.

Approximately 64% of the participants in the QBB were vitamin D deficient (<50 nmol/L), with slightly more men being deficient compared to women. Previous evidence from studies in the Middle Eastern region supports higher levels of vitamin D in males compared to females. A Saudi Arabian study performed on 10,709 participants reported more females being severely vitamin D deficient (<25 nmol/L) compared to males (115). A Bahraini and Lebanese study also reported lower mean serum vitamin D in females compared to males (113,243). One hypothesis for this difference could be explained by vitamin D supplementation, as women in the QBB were more likely to take vitamin D supplements compared to men. This was not reported in both the Saudi and Bahraini studies due to exclusion of participants that consumed vitamin D supplements (113,115).

Vitamin D deficiency was positively associated with the risk of MetS in the QBB. Waist circumference and triglycerides were the only two components of MetS that reported

a positive association with low levels of vitamin D. Several cross-sectional and few prospective studies have reported similar associations between vitamin D deficiency and the risk of MetS (228,230,244–246). In addition, prospective studies from the PROMISE cohort and the Australian Diabetes, Obesity and Lifestyle (Aus-Diab) study, reported significant inverse associations between continuous serum vitamin D and overall MetS, based on the IDF criteria (228,244). Studies examining the direction of the effect between vitamin D, cholesterol and lipids suggest that lipids levels and BMI may be a cause for decreased vitamin D and not vice versa (176,247). Other cross-sectional studies did not support these associations (248,249). However, this might have been due to the small sample sizes (172,248) and small variation in the exposure (e.g. high levels of vitamin D in the study by Reis *et al*. in both men and women) (249). Studies on Middle Eastern population where the levels of vitamin D could be different due to different lifestyles and darker skin pigmentation, are very limited and our results support an inverse association between the two examined phenotypes.

Many mechanisms have been proposed to explain the association between vitamin D and future risk of MetS components. Since vitamin D is fat soluble and could be stored in adipose tissue, it can be sequestered in the subcutaneous fat in obese individuals, reducing the levels of circulating vitamin D in the blood leading to less release of vitamin D into the blood (250,251). Vitamin D has also been shown to inhibit the release of cytokines from the immune cell which is harmful to β cells (252).

Vitamin D deficiency was positively associated with obesity, measured as BMI, waist circumference, and waist-to-hip ratio, with stronger associations observed with waist circumference. Similar results were found in both prospective and cross-sectional studies, including the Aus-Diab study (244), the British Birth Cohort (172), Middle

Eastern populations, (117,253) and in other populations around the world (230,254,255). Conversely, several cross-sectional studies did not report a correlation between obesity and vitamin D deficiency (248,249,256). However, these are mainly small studies with limited power to observe associations (248,249).

Lower levels of vitamin D was also associated with high triglyceride levels. Studies reported significant inverse associations between vitamin D with high triglyceride levels (173,254). However, when stratified by sex, only the association in men remained significant in the Ling *et al.* study (254). On the contrary, Reis *et al.* reported a positive trend between vitamin D deficiency and high triglyceride levels only in women (249).

This is the first study of vitamin D and MetS performed in the Qatari population and long-term residents (≥15 years living in Qatar) with high quality data which involves collection of extensive questionnaire information, clinical phenotyping and biological samples (212). QBB, as of yet is not considered to have a fully representative sample, since the most common source of participant recruitment is through word of mouth to friends and families (212). However, the population of our study shows small differences with the Qatar Stepwise Survey (QSS) which is based on the WHO STEPwise approach to noncommunicable sureivellance (STEPS) survey and consists of 2,496 participants. In the QSS, measurements of height, weight, and BMI were similar to this study. However, in QBB, over 60% of the participants had a university degree compared to 35% of the QSS (257) suggesting a higher representation of more educated and higher socioeconomic status individuals in QBB. Qataris in Qatar are considered a minority, in which they represent approximately 12% of the population in Qatar (258).

Since the QBB study is cross-sectional, the directionality of the relationship between vitamin D and MetS could not be elucidated. The non-significant results for the association between vitamin D deficiency and HbA1c levels, may have been due to the relatively small sample size and low response rate for some questions regarding physical activity, smoking, supplements, etc. It could also have been due to other confounding factors such as specific medication use and dosage of supplements, which were not captured. Approximately 84% of participants were missing data on medication use for hypertension and 68% of participants were missing information for cholesterol medication, which may have caused some misclassification. Additionally, the low levels of circulating vitamin D found in this population may have been insufficient to observe any inverse relationships with MetS components and diabetes.

To conclude, the findings from this study support a positive association between vitamin D deficiency and the presence of MetS. We also observed that MetS components, such as obesity and high triglycerides, were inversely associated with circulating serum vitamin D levels. Future prospective studies should elucidate the potential causal association between vitamin D and MetS using MR approaches or through supplementation with vitamin D. Moreover, mechanistic studies should concentrate on identifying pathophysiological pathways and molecular mechanisms linking vitamin D deficiency and MetS.

Table 2.1. Main characteristics of Qatar Biobank study participants stratified by sex

Characteristic	Females (N=702)	Males (N=503)	Total (N=1 205)
	Mean (SD)	**Mean (SD)**	**Mean (SD)**
Age (IQR) (N=1,205)	39.3 (28 – 50)	40.8 (30 – 51)	39.9 (29 – 50)
Sex‡ (%) (N=1,205)	58.2	41.8	
BMI (kg/m^2) (N=1,199)	29.0 (6.8)	28.7 (5.5)	28.9 (6.3)
Height (cm) (N=1,203)	158.4 (6.0)	172.5 (6.6)	164.3 (9.3)
Weight (kg) (N=1,199)	72.6 (17.2)	85.5 (18.1)	78.0 (18.7)
Waist circumference (cm) (N=1,994)	85.0 (14.7)	97.2 (13.6)	90.1 (15.5)
WHR∫ (N=1,194)	0.8 (0.1)	0.9 (0.1)	
Education‡ (%) (N=1,202)			
Less than primary school	4.0	0.2	2.4
Primary school	4.3	2.0	3.3
Secondary School	5.3	7.2	6.1
Technical/Professional school	22.1	22.9	22.5
University/Postgraduate	74.2	67.7	65.7
Ethnicity‡ (%) (N=954)			
Non-Qatari	18.6	39	27.4
Qatari	81.4	61.0	72.6
Season of blood draw‡ (%) (N=1,205)			
Winter Ŧ	33.0	33.8	33.4
SpringŦ	22.9	33.0	27.1
Summer Ŧ	14.2	13.3	13.9
Fall Ŧ	29.8	19.9	25.6
MET∫ score (hr/wk) (N=1,205)	10.8 (24.8)	19.5 (43.8)	14.5 (34.3)
Self-reported diabetic‡ (%) (N=1,198)	14.6	16.4	15.3
Metabolic syndrome‡ (%) (N=1,205)	29.2	27.8	28.6
Vitamin D supplementation‡ (%) (N=1,205)	49.0	24.8	38.9
25-hydroxyvitamin D nmol/L‡ (%) (N=1,176)			
Severely deficient (<25 nmol/L)	10.6	5.8	8.6
Deficient (25 - <50 nmol/L)	50.7	62.8	55.8
Insufficient (50 - <75 nmol/L)	26.1	24.7	25.5
Sufficient (≥75 nmol/L)	12.7	6.6	10.1

‡ Categorical variables with (%) indicate percentages rather than means and SD. The percentages were taken after removing missing, prefer not to answer, and I do not know categories

∫ MET: Metabolic equivalent WHR: waist to hip ratio

Ŧ winter: December-February, spring: March-May, summer: June-August, and fall: September-November

Source reproduced from Al-Dabhani *et al.* (259) with permission under the terms of the Creative Commons Attribution License.

Table 2.2. Unadjusted and adjusted mean 25-hydroxyvitamin D levels by participant characteristics in nmol/L in the Qatar Biobank*

Characteristics	Unadjusted mean	p-value	Adjusted mean 25(OH)D (SD) ₸	p-value	P-Trend
Sex (N=1,176)		0.06			0.08
Female (N=679)	47.5 (25.0)		42.5 (0.2)		
Male (N=497)	45.0 (22.5)		41.2 (0.2)	0.08	
Age (N=1,176)		**<0.001**			**<0.001**
<25 (N=140)	37.7 (22.7)		33.7 (0.0)		
25-34 (N=352)	41.7 (23.0)		37.5 (0.0)	0.01	
35-44 (N=224)	46.0 (23.7)		41.5 (0.0)	<0.001	
45-54 (N=283)	50.7 (25.0)		46.7 (0.0)	<0.001	
≥55 (N=177)	56.2 (20.7)		52.7 (0.0)	<0.001	
Education (N=1,173)		**<0.001**			0.46
Less than primary school (N=29)	57.7 (21.7)		53.5 (0.2)		
Primary school (N=39)	59.7 (26.7)		55.2 (0.2)	0.53	
Secondary School (N=69)	48.2 (17.0)		45.2 (0.2)	0.72	
Technical/Professional school (N=263)	43.7 (28.5)		38.7 (0.2)	0.17	
University/Postgraduate (N=773)	45.7 (22.5)		42.0 (0.2)	0.64	
Ethnicity (N=934)		**<0.001**			0.29
Non-Qatari (N=261)	44.0 (16.2)		43.0 (0.2)		
Qatari (N=673)	46.2 (23.0)		41.7 (0.2)	0.29	
Season blood draw (N=1,176)		0.99			0.43
Winter ₣ (N=400)	46.2 (25.5)		42.7 (0.2)		
Spring ₣ (N=318)	46.5 (22.0)		42.0 (0.2)	0.46	
Summer ₣ (N=164)	45.5 (22.7)		42.0 (0.2)	0.62	
Autumn ₣ (N=294)	47.2 (24.2)		41.0 (0.2)	0.41	
MET ʃ score (hr/wk) (N=1,176)		0.24			0.15
0 (N=415)	47.0 (25.7)		42.0 (0.2)		
>0 - 9.99 (N=401)	46.2 (20.5)		42.2 (0.2)	0.77	
10 - 49.99 (N=281)	45.0 (18.7)		41.2 (0.2)	0.73	
≥50 (N=79)	51.0 (36.5)		44.7 (0.2)	0.02	
Vitamin D supplementation		**<0.001**			**<0.001**
None	38.5 (16.0)		35.7 (0.2)		
Yes	59.0 (28.5)		54.2 (0.2)		

* An ANCOVA test was used to measure significance
ʃ MET: Metabolic equivalent,
₣ Seasons defined as Winter: December-February, Spring: March-May, Summer: June-August, and Autumn: September-November
₸ Mean vitamin D was adjusted for age, sex and season of blood collection
Source reproduced from Al-Dabhani et al. (259) with permission under the terms of the Creative Commons Attribution License

Table 2.3. Linear regression analyses between vitamin D and metabolic syndrome and its components in the Qatar Biobank

	Model 1: adjusted for age and sex		Model 2*	
	OR per 10% increase in VD‡ (95%CI)	**P-value**	**OR per 10% increase in VD‡ (95%CI)**	**P-value**
MetS‡				
Normal	Ref			
MetS	0.97 (0.94 - 1.01)	0.10	0.94 (0.91-0.98)	**0.01**
WC‡				
Normal	Ref			
High	0.96 (0.94 – 0.99)	**0.01**	0.96 (0.93 – 0.99)	**0.03**
TG‡				
Normal	Ref			
High	0.95 (0.92 – 0.99)	**0.01**	0.93 (0.90 – 0.97)	**0.001**
BP‡				
Normal	Ref			
High	0.99 (0.96 - 1.03)	0.73	0.96 (0.92 – 1.01)	0.11
HDL-C‡				
Normal	Ref			
Low	0.99 (0.97 - 1.02)	0.83	0.99 (0.96 – 1.03)	0.79
HbA1c				
No	Ref			
Yes	0.97 (0.94 – 0.99)	**0.03**	0.99 (0.95 – 1.04)	0.76

*Model 2: logistic regression adjusted for age, sex, ethnicity, MET score, education, and season of blood draw
‡VD: vitamin D, WC: waist circumference, WHR: waist-to-hip ratio, BMI: body mass index, MetS: Metabolic Syndrome, according to the IDF criteria, TG: triglycerides, BP: blood pressure, HDL-C: high density lipoprotein-cholesterol
Source reproduced from Al-Dabhani et al. (259) with permission under the terms of the Creative Commons Attribution License

Table 2.4. Logistic regression analyses between vitamin D deficiency and metabolic syndrome and its components

	Vitamin D deficiency (<50 nmol/L)			
	Model 1: adjusted for age and sex		Model 2*	
	OR (95%CI)	P-value	OR (95%CI)	P-value
MetS‡				
Normal	Ref		Ref	
MetS	1.07 (0.79 – 1.45)	0.65	1.54 (1.09 - 2.18)	0.01
WC‡				
Normal	Ref		Ref	
High	1.29 (0.98 – 1.71)	0.07	1.39 (1.01 - 1.92)	0.04
TG‡				
Normal	Ref		Ref	
High	1.57 (1.15 – 2.14)	<0.01	2.30 (1.58 - 3.34)	<0.01
BP‡				
Normal	Ref		Ref	
High	1.12 (0.81 - 1.4)	0.49	1.60 (1.06 - 2.42)	0.03
HDL-C‡				
Normal	Ref		Ref	
Low	1.00 (0.77 – 1.30)	0.98	1.01 (0.75 - 1.37)	0.93
HbA1c				
No	Ref		Ref	
Yes	1.36 (1.02 – 1.82)	0.04	1.24 (0.83 - 1.84)	0.30

*Model 2: Logistic regression adjusted for age, sex, ethnicity, MET score, education, and season of blood draw
‡WC: waist circumference, WHR: waist-to-hip ratio, BMI: body mass index, MetS: Metabolic Syndrome, according to the IDF criteria, TG: triglycerides, BP: blood pressure, HDL: high density lipoprotein-cholesterol
‡Cutoffs were defined according to the International Diabetes Federation

CHAPTER 3 ASSOCIATION BETWEEN VITAMIN D AND COLORECTAL CANCER: IS THIS ASSOCIATION MEDIATED BY METABOLIC SYNDROME?

In this Chapter, I will investigate whether the association between vitamin D and CRC is mediated by MetS, or its components.

Introduction

According to the WHO, CRC is the third most commonly diagnosed cancer in men and the second in women with an estimated 1.85 million new cases and approximately 880 thousand deaths worldwide in 2018 (1). In the UK, CRC is also the third most common cancer and the second most common cause of cancer mortality (6). A substantial number of epidemiological studies have identified risk factors associated with CRC. Among the various environmental factors, higher levels of circulating vitamin D has been consistently associated with lower CRC risk in observational studies (19). Although the main physiological role of vitamin D is to control calcium homeostasis, it has been hypothesized that the active metabolite of vitamin D, $1,25(OH)_2D$, controls cell growth and improves expression of various genes regulating the normal structure and function of the colon crypt, as well as its function in apoptosis (22,49).

A systematic review and meta-analysis of 20 observational studies on serum 25(OH)D and the risk of CRC, using study specific cut off points, reported a statistically significant inverse association when comparing the highest 25(OH)D category with the lowest (pooled RR: 0.63, 95%CI: 0.55 – 0.71, I^2=14.7%) (54). The association between vitamin D levels and CRC could be explained by intermediate disease risk factors, including MetS. Some studies have suggested an inverse association between serum vitamin D and MetS (104,174,228–231), while others have not confirmed this observation (232,233). It is also unclear which components of MetS that might drive this association with some studies suggesting obesity and others glucose haemostasis (228,231,234,235).

There have been inconsistent findings regarding obesity-related metabolic abnormalities, such as hypertension, hypertriglyceridemia, high cholesterol, and hyperglycaemia, in relation to the risk of CRC (260–263). A meta-analysis of 12 studies on men and 10 studies on women reported that MetS significantly increased the risk of CRC in men (RR: 1.25, 95%CI: 1.19 – 1.32, p-value< 0.01, I^2=35%) and women (RR: 1.34, 95%CI: 1.09 – 1.64, p-value: 0.006, I^2=60%) (45).

Another systematic review and meta-analysis of 8 cohort and nested case-control studies assessing the association between HbA1c levels and CRC suggested a positive association (RR: 1.22, 95%CI: 1.02 – 1.47, p-value: 0.03, I^2=25%) (46). However, when stratified by subgroups: gender, study design, or cancer subtype, no statistically significant association was found between HbA1c levels and CRC. A meta-analysis of 9 prospective studies reported a significant association between high versus low concentration of serum triglycerides and CRC risk (RR: 1.18, 95%CI: 1.04 – 1.34, I^2=48%) (48). However, no association was found for high versus low serum HDL cholesterol levels and CRC risk (RR: 0.84, 95%CI: 0.69 – 1.02, I^2=42%) (48).

In Chapter 1, I've discussed the well-established relationship between vitamin D and MetS and its components. For example, individuals with low levels of vitamin D are more likely to be overweight and obese. In turn, obese individuals had a higher risk of developing CRC. Studies have also reported a significant association between low levels of vitamin D concentration with CRC. Similar pattern of associations has been shown for other metabolic factors such as insulin resistance. Therefore, I hypothesised that the association between vitamin D and CRC could be mediated by MetS components.

In the next section I will assess whether MetS and its components mediates the association between vitamin D and CRC using the EPIC nested case cohort.

Methods

Data for this analysis was obtained from the EPIC cohort. Information on study design, anthropometric measurement, vitamin D, MetS and its components, and CRC were provided in Chapter 1 section 6.

A nested case-control study in EPIC was used for this analysis. Controls were selected (1:1) from the full cohort of individuals who were alive and free of cancer (except non-melanoma skin cancer) at the time of diagnosis of the cases, using incidence density sampling and matched by age (+/- 5 years at recruitment), sex, study centre, follow-up time since blood collection, time of day at blood collection (+/- 4 hours), fasting status (<3 hours, 3-6 hours, and >6 hours), menopausal status (pre-menopausal, post-menopausal, peri-menopausal/unknown), phase of menstrual cycle at blood collection, and usage of hormone replacement therapy at time of blood collection (yes/no). The current study included 1,150 incident CRC cases and 1,150 matched controls.

Statistical analysis

Descriptive analysis

To test the difference in the distributions of categorical variables, a chi-square test was performed, while a t-test was performed to test the differences between a categorical and a continuous variable after stratifying by colon and rectal cancer for the vitamin D and CRC association, and by MetS, for the vitamin D and MetS association.

Association between Vitamin D and colorectal cancer risk

A conditional logistic regression was performed on the association between serum vitamin D and CRC. Model 1 was adjusted for BMI (kg/m^2), physical activity (sex-specific quintiles), smoking status (never, former smoker and smoked less than 10 years, former smoker and smoked greater than or equal 10 years, former smoker with unknown duration of smoking, current smoker and low smoking intensity (1-15 cigarettes/day), current smoker and moderate smoking intensity (16-25 cigarettes/day), current smoker and high smoking intensity (≥26 cigarettes/day), and unknown), education (none, primary school, technical/professional school, secondary school, university, and unspecified), alcohol (sex-specific quartiles), fruits (g/day), vegetables (g/day), meat consumption (g/day), and total dietary energy consumption (quartiles). Model 2 was further adjusted for month of blood collection due to variation of serum 25(OH)D levels in each season. A sensitivity analysis was performed excluding CRC cases diagnosed within two years of follow up.

Association between vitamin D and metabolic syndrome

For this analysis, all cases of CRC were removed leaving 1,150 participants. Since Sweden had no cases of MetS, they were excluded from the study (N=51).

A logistic regression analysis was used to investigate the association between vitamin D (divided into quintiles) and MetS and its components (binary). Two models were examined for this analysis; Model 1 was adjusted for age (continuous), sex and centre, Model 2 was further adjusted for smoking status (never, former, smoker, and unknown), CRP (mg/L), education (none/primary school. technical/professional, secondary school, university, and unspecified), physical activity (inactive, moderately inactive, moderately inactive, and active), season of blood collection (winter, spring,

summer and fall), meat products (g/day), fish (g/day), fruit (g/day), vegetables (g/day), and alcohol consumption (g/day).

Association between metabolic syndrome and colorectal cancer

A conditional logistic regression was performed on the association between MetS and CRC. Two models were investigated; the first model did not adjust for anything other than the matching factors. Model 2 was further adjusted for smoking status (never, former, current, and unknown), education (none, primary school, technical/professional school, secondary school, university, and unknown), physical activity (inactive, moderately inactive, moderately active, active, and unknown), fibre (g/day), fruit (g/day), vegetables (g/day), meat (g/day), fish (g/day), and alcohol consumption (g/day). A sensitivity analysis removed participants that were diagnosed with CRC within two years of follow up.

Mediation analysis

Three mediation methods were used in this analysis on EPIC: I) the difference of coefficient methods (265,266), II) the product of coefficient method (265,266) and III) the causal inference method (267). Vitamin D was natural log transformed in these models. All models used were adjusted for age (continuous), sex, study centre, smoking status (never, former, current, and unknown), and alcohol consumption (g/day).

Results

Association between vitamin D and colorectal cancer

Measurements of obesity (BMI, waist circumference, waist-to-hip ratio, and weight) and circulating vitamin D levels were different between colon cancer cases and controls with cases having slightly higher levels compared to controls, but no difference was found for rectal cancer. Moreover, there was a difference in alcohol consumption between cases and controls in rectal cancer (p-value: 0.02), where cases tended to consume more alcohol than controls, however no difference was found for colon cancer cases and control (p-value: 0.20). No other differences were observed between cases and controls (**Table 3.1**).

A significant trend was found for circulating vitamin D and colorectal and colon cancer in both models (P-trend <0.05) (**Table 3.2**). A 10% increase in circulating vitamin D levels was associated with a 2% decrease in the risk of CRC (OR: 0.98, 95%CI: 0.96 – 1.00). An association was also found for colon cancer, but not with rectal cancer (**Table 3.2**). When comparing each category to the reference category (≥50 - <75 nmol/L serum 25(OH)D levels), no association in either CRC or colon cancer was found. No trend was found for rectal cancer in the multivariable model (p-trend: 0.48). When stratified by the latitude (>50°C and <50°C), there was no significant association between vitamin D and CRC (**Table 3.3)**. Participants that were diagnosed with CRC (n=24) two years after follow up were excluded in a sensitivity analysis and the results remained qualitatively similar. These results have been published previously in 2010 by Jenab *et al.* (268) and were replicated here for completeness of this study.

Associations between vitamin D and metabolic syndrome

Approximately 56% of cases with MetS were men. Moreover, participants with MetS tended to be older than participants without MetS. There was also a significant difference in measurements of obesity between cases of MetS and controls. In regards to dietary consumption, only meats and meat products showed significant differences among cases and controls. Higher levels of CRP were found in cases compared to controls. Approximately 37% of the participants in this study were vitamin D deficient (<50 nmol/L) with more participants with MetS (45%) being vitamin D deficient compared to participants without MetS (34%) (**Table 3.4**).

Low levels of serum vitamin D (<50 nmol/L) were statistically significantly associated with MetS compared to vitamin D levels between 50 nmol/L to 75 nmol/L (OR: 1.49, 95%CI: 1.07 – 2.08). Moreover, statistically significant trends were found for the association between low levels of serum vitamin D with elevated triglyceride levels and high waist circumference in all three models (**Table 3.5**). No association was found with blood pressure, HDL cholesterol, and HbA1c levels (**Table 3.5**).

Association between metabolic syndrome and the risk of colorectal cancer

A higher prevalence of MetS defined as IDF and ATPIII was found for cases of colon cancer compared to matched controls (p-value <0.001), but not for rectal cancer (p-value: 0.13 and 0.49 for IDF and ATPIII respectively). Measurements of obesity (BMI, waist circumference, waist-to-hip ratio and weight) and circulating vitamin D levels were significantly higher in cases compared to controls in colon cancer, but no significant difference was found in rectal cancer (**Table 3.1**). The other components of MetS including HbA1c, HDL-cholesterol, and triglyceride levels were associated in

colon cancer only. Higher levels of triglyceride and HbA1c and lower levels of HDL cholesterol were found in colon cancer cases compared to controls. However, both SBP and DBP did not show evidence of an association in either colon or rectal cancer. (**Table 3.1**).

In the multivariable conditional logistic regression, MetS defined according to both IDF and ATPIII was associated with 63% and 46% increase in odds of CRC (OR: 1.63, 95%CI: 1.35 – 1.98 and OR: 1.46, 95%CI: 1.17 – 1.83, respectively) (**Table 3.6**). When stratified by subsite (colon and rectum) the association between MetS and CRC subsite was strengthened for colon cancer (OR: 1.92, 95%CI: 1.50 – 2.44 and OR: 1.69, 95%CI: 1.27 – 2.25, for IDF and ATPIII definitions respectively), but was attenuated for rectal cancer. Abdominal obesity, reduced HDL cholesterol, and high HbA1c were associated with colon cancer in the multivariable adjusted model (OR: 1.73, 95%CI: 1.35 – 2.22, OR: 1.33, 95%CI: 1.01 – 1.75, and RR: 1.78, 95%CI: 137 – 2.33 respectively). Only elevated levels of HbA1c showed evidence of an association with rectal cancer (OR: 2.01, 95%CI: 1.41 – 2.95) (**Table 3.6**). High blood pressure reported no evidence of an association with either colon or rectal cancer (**Table 3.6**). Exclusion of participants that were diagnosed with CRC (N=22) two years after follow up, resulted in qualitatively similar estimates.

Mediation analysis

From the analyses above, I showed that the three assumptions that were needed to conduct a mediation analysis held true for MetS and only two of its components; waist circumference and triglyceride levels; vitamin D was associated with MetS and CRC and MetS was associated with CRC. Therefore, I continued with the mediation analysis for only those components with both CRC and colon cancer, and not for rectal

cancer, since no strong evidence of an association between vitamin D and rectal cancer was observed. Since little difference in the association estimates was found after removing individuals that developed CRC after two years of follow-up, these individuals were not excluded from the mediation analysis.

Results from the difference of coefficient method (**Table 3.7**) did not support mediation between vitamin D and CRC through MetS. However, there was evidence of mediation between vitamin D and CRC occurring through waist circumference.

The results from the product of coefficient methods (**Table 3.8**) reported that MetS and waist circumference were both significant mediators in the vitamin D and colorectal/colon cancer pathway.

The counterfactual approach allowed for the decomposition of the total effect into a direct effect and an indirect (mediated) effect. Results from this mediation analysis reflected some evidence of mediation by MetS (18%) between vitamin D and colorectal (**Table 3.9**). Waist circumference also partially mediated (18%) the association between vitamin D and CRC. There was also very small amount of mediation occurring through triglycerides (4%) (**Table 3.9**). Partial mediation was also found between these components with colon cancer (**Table 3.9**).

Discussion

Results from this study suggests that MetS, waist circumference, and triglyceride levels partially mediate the association between vitamin D with CRC and colon cancer, with the most contributing factor of MetS for this mediation analysis being waist circumference, which leads to believe that waist circumference and not MetS mediates the association between vitamin D and CRC.

A positive association was found between individuals with MetS with the risk of CRC and colon cancer. High waist circumference and high triglyceride levels all increased the risk of CRC and colon cancer. Previous studies also supported an association between MetS, waist circumference and triglyceride levels with CRC. A systematic reviews and meta-analysis on MetS reported a 33% increased risk of CRC in men and a 41% increased risk of CRC in women (44). Common confounders such as diet and physical activity were accounted for in the present analyses and were unlikely to explain the association between MetS, waist circumference, and triglyceride levels with CRC risk. One potential mechanism for the relationship between waist circumference and CRC is that increased body size is associated with increased insulin secretion, which can promote cell growth and inhibit apoptosis, increasing the risk of CRC (19). Alternatively, the association between triglyceride levels and CRC can be explained by the association of high levels of triglycerides with increased insulin levels, which in turn increases the risk of CRC (269).

Evidence from a meta-analysis of 6 prospective studies on circulating serum 25(OH)D reported a significant inverse association with CRC risk per 100 IU/L increase of 25(OH)D (RR: 0.96, 95%CI: 0.94 – 0.97, I^2=0%) (270). On the contrary, MR and RCTs

studies to date have failed to confirm the protective effect of vitamin D and CRC risk seen in observational studies (53,271,272). Results from two MR studies also reported no causal relationship between vitamin D and CRC (271,272). A recent study of 2,303 postmenopausal women provided participants in the treatment group with 2,000 IU of cholecalciferol and 1,500 mg of calcium carbonate reported no difference in all-type cancer incidence between the treatment group and the placebo group (p-value: 0.06) after a four year follow up period (53).

This discrepancy between observational studies, RCTs and MR could be due to potential limitations of observational studies including reverse causation, short follow up time, and residual confounding. It could also be due to the low power in the MR study.

This present study verified results from Aleksandrova *et al.* (217) on the association between MetS and its components with the risk of colon and rectal cancer in EPIC, Rinaldi *et al.* (47) on the association between HbA1c levels and the risk of CRC, and Jenab *et al.* (268) on the association between vitamin D and CRC. These analyses were essential to perform the mediation analysis. Overall, the present results were similar to Aleksandrova *et al.* (217), Rinaldi *et al.* (47)*, and Jenab et al.* (268).

The difference of coefficient method reported significant mediation between vitamin D and CRC occurring only through waist circumference. The product of coefficient method and the causal method confirmed this relationship. Since waist circumference was the only common component that mediated the association between vitamin D and CRC in all three methods used, the association between vitamin D and CRC could possibly be driven by obesity, where some studies viewed obesity as the main initiator of MetS (264).

This is the first study to have investigated whether MetS or its components mediate the association between vitamin D and CRC. Three different methods for mediation analysis were used to support whether MetS, or its components were in the causal pathway between vitamin D and CRC. However, since this is an observational study, it suffered from several limitations including reverse causation. Individuals that developed CRC within two years of the study were excluded from the vitamin D to CRC association (n=24) as well as the MetS to CRC association (n=22) and the results remained the qualitatively similar. Residual confounding due to family history of CRC or the use of nonsteroidal anti-inflammatory drugs could also have affected the results for some of the MetS components (273,274). Furthermore, I assumed that MetS/components were on the causal pathway between vitamin D and CRC, since observational studies supported this causal pathway. However, vitamin D could possibly mediate the association between MetS and CRC. A bi-directional MR is presented in Chapter 5 of this thesis and aims to investigate the bi-directional association between vitamin D and MetS.

There were also limitations to mediation analysis, including inconsistent mediation. If the direction of association between exposure, mediator, and outcome differ, the effect size will be small and could cancel out (266). In this study, low levels of vitamin D increased CRC risk and MetS. However, individuals with MetS had a higher risk of CRC. Therefore, an explanation for the non-significant result could have been due to inconsistent mediation. Overall, the associations may not be statistically significant, yet mediation may still exist in a study. Both the product and the difference method do not provide a clear framework for generalising the tests for more than one mediator (265). In this analysis I only investigated one mediator at a time, however if more mediators were to be included in the analysis, for example, investigating whether waist

circumference and insulin mediates the association between vitamin D and CRC, using the product and difference of coefficient method would be complicated to interpret. The causal method, provides more information on mediation by estimating the direct and indirect effect of an association between exposure and outcome from the counterfactuals. This method also becomes more difficult to disentangle when models incorporate multiple mediators.

To conclude, the findings from this study suggest that waist circumference partially mediates the association between vitamin D and CRC, which could provide explanations for potential mechanisms on the relationship between vitamin D and CRC and the discrepancies found between observational and MR studies.

In the coming chapters I will discuss the potential causal effect between vitamin D and CRC and between vitamin D with MetS components.

Table 3.1. Descriptive analysis of nested case control dataset stratified by colon and rectal cancer in EPIC

	Colon cancer			Rectal cancer		
	Cases	**Matched controls**	**p-value**	**Cases**	**Matched controls**	**p-value**
Sex						
Men, n	351	351		246	246	
Women, n	366	366		187	187	
Mean age (SD)						
at blood collection	58.9 (7.3)	58.9 (7.3)	0.52	58.2 (6.8)	58.2 (6.8)	0.71
Anthropometric measures						
Mean weight (SD)	76.3 (14.7)	73.4 (12.4)	<0.001	75.8 (13.9)	75.4 (14.2)	0.59
Mean waist circumference (SD)	91.1 (13.3)	88.4 (12.1)	<0.001	91.0 (13.0)	90.1 (13.2)	0.25
Mean waist-to-hip ratio (SD)	0.9 (0.1)	0.9 (0.1)	<0.001	0.9 (0.1)	0.9 (0.1)	0.14
Mean BMI (SD)	26.9 (4.5)	26.3 (3.8)	0.003	26.7 (4.1)	26.5 (3.9)	0.45
Mean years Follow up time (SD)	3.8 (2.2)			3.9 (2.2)		
Metabolic syndrome/components						
MetS (IDF)	298 (43.6)	204 (29.8)	<0.001	163 (39.3)	142 (34.2)	0.13
MetS (NCEP/ATPIII)	166 (24.3)	109 (15.9)	<0.001	87 (21.0)	79 (19.0)	0.49
Triglyceride levels (SD)	1.81 (1.3)	1.67 (1.0)	0.02	1.86 (1.4)	1.82 (1.1)	0.61
HDL-C (SD)	1.4 (0.4)	1.5 (0.5)	<0.001	1.4 (0.4)	1.5 (0.4)	0.85
Systolic blood pressure (SD)	139.5 (22.4)	137.5 (19.4)	0.09	139.2 (21.3)	138.7 (21.4)	0.76
Diastolic blood pressure (SD)	84.0 (12.3)	82.8 (10.4)	0.07	83.9 (11.3)	83.3 (10.9)	0.49
HbA1c (SD)	5.9 (0.8)	5.8 (0.7)	0.02	5.9 (0.9)	5.8 (0.8)	0.34
Smoking status/duration/intensity, n (%)			0.90			0.35
Never	297 (44.0)	325 (47.4)		162 (41.7)	163 (40.0)	
Ex-smokers, duration of smoking <10 years	40 (5.9)	36 (5.2)		18 (4.6)	22 (5.4)	
Ex-smokers, duration of smoking ≥10 years	188 (27.8)	185 (27.0)		123 (31.7)	107 (26.3)	
Ex-smokers, missing duration of smoking	18 (2.7)	12 (1.7)		4 (1.0)	8 (2.0)	
Smokers, <15 cigarettes a day	68 (10.1)	69 (10.1)		44 (11.3)	55 (13.5)	

	Colon cancer			Rectal cancer		
	Cases	**Matched controls**	**p-value**	**Cases**	**Matched controls**	**p-value**
Smokers, ≥15-<25 cigarettes a day	48 (7.1)	44 (6.4)		25 (6.4)	40 (9.8)	
Smokers, ≥15 cigarettes a day	10 (1.5)	9 (1.3)		9 (2.3)	7 (1.7)	
Physical activity, n (%)			0.34			0.13
Inactive	182 (25.4)	155 (21.6)		97 (22.4)	104 (24.0)	
Moderately inactive	162 (22.6)	158 (22.0)		111 (25.6)	82 (18.9)	
Moderately active	149 (20.8)	176 (24.5)		94 (21.7)	106 (24.5)	
Active	186 (25.9)	188 (26.2)		112 (25.9)	113 (26.1)	
Education level, n (%)			0.24			0.37
None	39 (5.5)	33 (4.3)		20 (4.6)	18 (4.2)	
Primary School	252 (35.4)	286 (40.1)		140 (32.6)	163 (37.9)	
Technical/Professional	167 (23.5)	169 (23.7)		117 (27.2)	121 (28.1)	
Secondary School	115 (16.1)	90 (12.6)		56 (13.0)	42 (9.8)	
University	118 (16.6)	120 (16.8)		88 (20.5)	81 (18.8)	
Unspecified	21 (2.9)	15 (2.1)		9 (2.1)	5 (1.2)	
Dietary variables (g/day)						
Total energy kcal	2154.6 (756.9)	2123.5 (655.6)	0.34	2220.3 (689.6)	2173.5 (656.6)	0.28
Total vegetables (SD)	183.8 (123.7)	192.2 (127.3)	0.12	183.2 (152.0)	183.6 (125.1)	0.96
Total fruits (SD)	225.8 (185.0)	234.7 (180.0)	0.31	209.5 (155.5)	213.1 (158.8)	0.71
Meats and meat products (SD)	113.1 (79.6)	110.6 (59.4)	0.45	124.6 (65.8)	119.5 (64.9)	0.18
Fish (SD)	26.9 (27.6)	29.0 (27.9)	0.08	27.2 (23.3)	29.4 (29.5)	0.16
Alcohol (SD)	238.9 (364.8)	217.7 (346.6)	0.20	325.8 (448.5)	262.3 (395.3)	0.02
Vitamin D (µg/day) (SD)	4.0 (2.6)	3.9 (2.4)	0.83	4.1 (2.4)	4.3 (2.6)	0.34
Circulating vitamin D (nmol/L) (SD)	57.7 (27.9)	62.8 (28.6)	<0.001	59.8 (27.3)	61.7 (30.5)	0.30

MetS: metabolic syndrome, HDL-C: high density lipoprotein-cholesterol, IDF: International Diabetes Federation, ATPIII: Adult Treatment Panel III, SD: standard deviation,

Table 3.2. Conditional logistic regression between vitamin D and colorectal cancer and subtypes matched on age, sex, centre, blood collection, fasting status, menopausal status, phase of menstrual cycle, and HRT

	Vitamin D cut off points (nmol/L)						
	<25	**≥25.0 - <50**	**≥50 - <75**	**≥75 - <100**	**≥100**	**P-trend**	**10% increase in vitami**
Colorectum							
Mean (SD), median (nmol/l)	19.9 (4.2) 21.0	38.6 (7.0) 38.9	61.3 (7.0) 60.6	84.5 (6.6) 83.0	125.2 (29.0) 114.3		
No of cases/controls	65/50	432/371	406/418	161/199	86/112		
Matching components	1.40 (0.93 - 2.09)	**1.24 (1.01 - 1.51)**	1.00	0.84 (0.65 - 1.08)	0.77 (0.56 - 1.06)	<0.001	
Model 1	1.25 (0.79 - 1.98)	1.14 (0.91 - 1.43)	1.00	0.80 (0.60 - 1.05)	0.89 (0.62 - 1.29)	0.01	
Model 2	1.16 (0.72 - 1.86)	1.13 (0.89 - 1.43)	1.00	0.79 (0.60 - 1.05)	0.87 (0.60 - 1.26)	0.04	
Colorectal cancer (yes vs no)							**0.98 (0.96 - 1.00)**
Colon							
Mean (SD), median (nmol/l)	19.7 (4.2) 20.3	38.4 (7.1) 38.9	60.8 (7.1) 59.8	84.5 (6.4) 83.5	127.6 (32.2) 118.7		
No of cases/controls	46/26	266/229	258/265	98/130	49/67		
Matching factors	**1.93 (1.15 - 3.26)**	1.26 (0.98 - 1.64)	1.00	0.80 (0.59 - 1.09)	0.74 (0.49 - 1.11)	<0.001	
Model 1	1.67 (0.92 - 3.04)	1.18 (0.89 - 1.58)	1.00	0.79 (0.56 - 1.12)	0.85 (0.53 - 1.34)	0.01	
Model 2	1.55 (0.82 - 2.91)	1.18 (0.86 - 1.62)	1.00	0.77 (0.54 - 1.11)	0.85 (0.52 - 1.38)	0.03	
Colon cancer (yes vs no)							**0.96 (0.94 - 0.99)**
Rectum							
Mean (SD), median (nmol/l)	20.6 (4.4) 22.1	38.9 (6.8) 39.4	62.1 (6.7) 62.2	84.4 (6.9) 82.7	122.1 (24.2) 112.9		
No of cases/controls	19/24	166/142	148/153	63/69	37/45		
Matching factors	0.79 (0.40 - 1.56)	1.23 (0.89 - 1.69)	1.00	0.93 (0.61 - 1.42)	0.83 (0.50 - 1.36)	0.22	

	Vitamin D cut off points (nmol/L)						
	<25	**≥25.0 - <50**	**≥50 - <75**	**≥75 - <100**	**≥100**	**P-trend**	**10% increase in vitami**
Model 1	0.88 (0.40 - 1.95)	1.13 (0.77 - 1.67)	1.00	0.83 (0.50 - 1.36)	0.97 (0.52 - 1.83)	0.48	
Model 2	0.97 (0.41 - 2.25)	1.24 (0.83 - 1.87)	1.00	0.84 (0.50 - 1.41)	0.88 (0.46 - 1.70)	0.24	
Rectal cancer (yes vs no)							1.00 (0.96 - 1.03)

Model 1: further adjusted for BMI, education, alcohol, smoking status, PA, meat, fish, fruit, vegetables, and energy
Model 2: Model 1 further adjusted to month of blood collection

Table 3.3. Conditional logistic regression of vitamin D and the risk of colorectal cancer matched on age, sex, centre, blood collection, fasting status, menopausal status, phase of menstrual cycle, and HRT stratified by latitude

Country	Cases	mean Vitamin D	OR (95%CI) per 10% increase
North latitude (>50°N)	852	61.7	0.97 (0.95 – 0.99)
South Latitude (<50°N)	298	56.8	0.97 (0.94 – 1.01)

Table 3.4. Description of study population stratified by metabolic syndrome cases and control in EPIC

	Metabolic Syndrome		
	Cases	**Controls**	**p-value**
Mean age at blood collection (SD)	60.6 (6.4)	57.9 (7.1)	**<0.001**
Sex, n (%)			**0.02**
Men	194 (56.1)	366 (48.6)	
Women	152 (43.9)	387 (51.4)	
Obesity measures			
Mean waist circumference (SD)	98.7 (10.4)	84.6 (10.8)	**<0.001**
Mean waist-to-hip ratio (SD)	0.9 (0.1)	0.8 (0.1)	**<0.001**
Mean BMI (SD)	29.3 (3.5)	25.1 (3.2)	**<0.001**
Smoking status, n (%)			0.22
Never	131 (37.9)	330 (43.8)	
Former	124 (35.8)	239 (31.7)	
Smoker	89 (25.7)	176 (23.4)	
Unknown	2 (0.6)	8 (1.1)	
Physical activity, n (%)			0.88
Inactive	67 (19.6)	135 (18.2)	
Moderately inactive	82 (24.0)	173 (23.4)	
Moderately active	85 (24.8)	200 (27.0)	
Active	108 (31.6)	232 (31.3)	
Season, n (%)			0.98
Winter	87 (25.1)	183 (24.3)	
Spring	96 (27.7)	215 (28.5)	
Summer	73 (21.1)	156 (20.7)	
Fall	90 (26.0)	199 (26.4)	
Education level, n (%)			0.15
None/Primary school	166 (48.4)	311 (41.5)	
Technical/Professional	86 (25.1)	189 (25.2)	
Secondary School	38 (11.1)	85 (11.3)	
University	49 (14.3)	148 (19.8)	
Unspecified	4 (1.2)	16 (2.1)	
Dietary variables (g/day)			
Total energy kcal	2147.0 (685.8)	2162.3 (639.2)	0.72
Total vegetables	197.0 (136.3)	193.3 (121.5)	0.65
Total fruits	220.0 (169.3)	236.3 (175.9)	0.15
Meats and meat products	125.8 (66.2)	111.5 (58.8)	**<0.001**
Fish	29.7 (28.0)	30.2 (28.9)	0.78

	Metabolic Syndrome		
	Cases	**Controls**	**p-value**
Eggs and egg products	20.1 (18.5)	18.5 (17.2)	0.17
Alcohol	253.3 (403.7)	236.3 (356.9)	0.48
Vitamin D (µg/day)	4.2 (2.7)	3.9 (2.4)	0.08
Calcium (µg/day)	1018.7 (415.7)	1040.7 (416.7)	0.42
CRP (mg/L)	4.6 (5.9)	3.2 (5.5)	**<0.001**
25-hydroxyvitamin D nmol/L (%) (N=1,150)			**0.001**
Severely deficient (<25 nmol/L)	22 (6.4)	28 (3.7)	
Deficient (25 - <50 nmol/L)	133 (38.4)	227 (30.1)	
Insufficient (50 - <75 nmol/L)	119 (34.4)	273 (36.2)	
Sufficient (≥75 nmol/L)	72 (20.8)	225 (29.9)	

SD: standard deviation, CRP: C - reactive protein

Table 3.5. Crude and multivariate logistic regression for the association between vitamin D (categorical) and metabolic syndrome/components

	Vitamin D cut off points (nmol/L)					
	<50.0	**≥50.0 - <75.0**	**≥75.0**	**P-trend**	**OR (95%CI) per 10% increase**	**p-value**
Metabolic syndrome						
Model1:	1.47 (1.08 - 1.99)	1.00	0.75 (0.52 - 1.07)	**<0.001**	0.94 (0.92 – 0.97)	**<0.001**
Model 2:	1.49 (1.07 - 2.08)	1.00	0.72 (0.50 - 1.07)	**<0.001**	0.93 (0.90 – 0.97)	**<0.001**
Triglyceride levels						
Model1:	1.24 (0.92 - 1.68)	1.00	0.80 (0.57 - 1.11)	**0.01**	0.96 (0.93 – 0.98)	**0.002**
Model 2:	1.24 (0.89 - 1.71)	1.00	0.78 (0.55 - 1.11)	**0.02**	0.95 (0.93 – 0.98)	**0.004**
Model 3:	1.94 (1.20 – 3.13)	1.00	1.07 (0.64 – 1.80)	**0.01**	0.93 (0.90 – 0.98)	**0.004**
HDL-C						
Model1:	0.98 (0.69 - 1.38)	1.00	0.73 (0.49 - 1.09)	0.29	0.98 (0.95 – 1.01)	0.16
Model 2:	1.02 (0.70 - 1.48)	1.00	0.76 (0.50 - 1.16)	0.30	0.97 (0.94 – 1.01)	0.18
Model 3:	0.72 (0.39 – 1.32)	1.00	0.91 (0.47 – 1.78)	0.37	0.97 (0.97 – 1.09)	0.32
Blood Pressure						
Model1:	1.43 (1.01 – 2.02)	1.00	1.14 (0.78 - 1.64)	0.11	0.98 (0.95 – 1.01)	0.11
Model 2:	1.32 (0.90 - 1.94)	1.00	1.24 (0.83 - 1.88)	0.51	0.99 (0.96 – 1.03)	0.62
Model 3:	1.20 (0.72 – 2.00)	1.00	1.40 (0.81 – 2.42)	0.83	1.00 (0.95 – 1.05)	0.94
Waist circumference						
Model1:	1.50 (1.12 - 2.02)	1.00	0.64 (0.46 - 0.89)	**<0.001**	0.93 (0.90 – 0.95)	**<0.001**
Model 2:	1.55 (1.11 - 2.15)	1.00	0.73 (0.51 - 1.04)	**<0.001**	0.93 (0.90 – 0.96)	**<0.001**
Model 3:	1.32 (0.82 – 2.13)	1.00	0.58 (0.35 – 0.96)	**0.004**	0.94 (0.90 – 0.98)	**0.005**

	Vitamin D cut off points (nmol/L)					
	<50.0	**≥50.0 - <75.0**	**≥75.0**	**P-trend**	**OR (95%CI) per 10% increase**	**p-value**
HbA1c						
Model1:	1.24 (0.86 - 1.78)	1.00	1.37 (0.92 - 2.03)	0.97	0.99 (0.96 – 1.02)	0.71
Model 2:	1.31 (0.88 - 1.97)	1.00	1.10 (0.70 - 1.71)	0.29	0.97 (0.93 – 1.01)	0.14
Model 3:	1.38 (0.87 – 2.18)	1.00	1.02 (0.62 – 1.66)	0.17	0.97 (0.93 – 1.01)	0.17

*Model 1 was adjusted for age, sex and centre
*Model 2 was further adjusted for smoking status, CRP, education, alcohol, physical activity, season, meat, fish, fruit, and vegetables
*Model 3 was further adjusted for metabolic syndrome components
HDL-C: high density lipoprotein-cholesterol, HbA1c: glycated haemoglobin A1c

Table 3.6. Multivariate logistic regression of the association between metabolic syndrome and components with colorectal, colon and rectal cancer

	Colorectal cancer		Colon cancer		Rectal cancer	
	OR (95%CI)	**p-value**	**OR (95%CI)**	**p-value**	**OR (95%CI)**	**p-value**
Metabolic Syndrome (IDF)						
Model 1	**1.62 (1.35 - 1.95)**	**<0.001**	**1.88 (1.49 - 2.37)**	**<0.001**	1.27 (0.94 - 1.71)	0.11
Model 2	**1.63 (1.35 - 1.98)**	**<0.001**	**1.92 (1.50 - 2.44)**	**<0.001**	1.26 (0.91 - 1.73)	0.16
Metabolic Syndrome (NCEP/ATPIII)						
Model 1	**1.47 (1.18 – 1.83)**	**<0.001**	**1.70 (1.29 – 2.24)**	**<0.001**	1.14 (0.80 – 1.63)	0.47
Model 2	**1.46 (1.17 – 1.83)**	**0.001**	**1.69 (1.27 – 2.25)**	**<0.001**	1.17 (0.79 – 1.71)	0.73
High Triglyceride levels						
Model 1	1.14 (0.95 - 1.38)	0.16	**1.30 (1.03 - 1.64)**	**0.03**	0.90 (0.66 - 1.23)	0.52
Model 2	1.11 (0.91 - 1.35)	0.32	1.27 (0.99 - 1.63)	0.06	0.91 (0.64 - 1.29)	0.59
High Blood Pressure						
Model 1	0.97 (0.80 - 1.19)	0.80	0.94 (0.73 - 1.22)	0.65	1.03 (0.75 - 1.41)	0.87
Model 2	0.96 (0.78 - 1.19)	0.74	0.98 (0.75 - 1.30)	0.92	0.89 (0.63 - 1.27)	0.53
Low HDL-C						
Model 1	**1.26 (1.03 - 1.54)**	**0.02**	**1.41 (1.09 - 1.82)**	**0.01**	1.03 (0.73 - 1.44)	0.86
Model 2	**1.25 (1.00 - 1.54)**	**0.04**	**1.33 (1.01 - 1.75)**	**0.04**	1.08 (0.75 - 1.56)	0.66
High waist circumference						
Model 1	**1.50 (1.25 - 1.80)**	**<0.001**	**1.74 (1.37 - 2.21)**	**<0.001**	1.03 (0.73 - 1.44)	0.86
Model 2	**1.48 (1.23 - 1.79)**	**<0.001**	**1.73 (1.35 - 2.22)**	**<0.001**	1.16 (0.86 - 1.57)	0.33
High HbA1c						
Model 1	**1.71 (1.39 - 2.09)**	**<0.001**	**1.65 (1.29 - 2.12)**	**<0.001**	**1.82 (1.29 - 2.57)**	**0.001**
Model 2	**1.86 (1.50 - 2.30)**	**<0.001**	**1.78 (1.37 - 2.33)**	**<0.001**	**2.04 (1.41 - 2.95)**	**<0.001**

Model 1: adjusted for matching factors

Model 2 further adjusted for smoking status, education, physical activity, fibre (g/d), fruit (g/d), vegetables (g/d), meat (g/d), fish (g/d), and alcohol consumption (g/d).
HDL-C: high density lipoprotein-cholesterol, HbA1c: glycated haemoglobin A1c

Table 3.7. Mediation analysis, association between vitamin D and colorectal cancer and colon cancer mediated by metabolic syndrome/components (**difference method**) in EPIC

Mediation analysis: VD-->MetS-->CRC			**Mediation analysis: VD-->MetS-->CC**		
	Without Mediator	With mediator		Without Mediator	With Mediator
β	-0.32	-0.25	β	-0.49	-0.39
SE	0.11	0.11	SE	0.14	0.14
SE_{both}	0.04		SE_{both}	0.09	
P-value	0.25		P-value	0.25	
Mediation analysis: VD-->TG-->CRC			**Mediation analysis: VD-->TG-->CC**		
	Without Mediator	With mediator		Without Mediator	With Mediator
β	-0.32	-0.28	β	-0.49	-0.44
SE	0.11	0.11	SE	0.14	0.14
SE_{both}	0.05		SE_{both}	0.09	
P-value	0.39		P-value	0.58	
Mediation analysis: VD-->WC-->CRC			**Mediation analysis: VD-->WC-->CC**		
	Without Mediator	With Mediator		Without Mediator	With Mediator
β	-0.32	-0.22	β	-0.49	-0.35
SE	0.11	0.11	SE	0.14	0.14
SE_{both}	0.04		SE_{both}	0.08	
P-value	**0.02**		P-value	0.10	

Model adjusted for age, sex, study centre, smoking status, and alcohol consumption
VD: vitamin D, MetS: metabolic syndrome, TG: triglycerides, WC waist circumference, CRC: colorectal cancer, CC: colon cancer

Table 3.8. Mediation analysis, association between vitamin D and colorectal cancer and colon cancer mediated by metabolic syndrome/components (**product method**) in EPIC

Mediation analysis: VD-->MetS-->CRC			**Mediation analysis: VD-->MetS-->CC**		
	Coefficient	SE		Coefficient	SE
β (M-->X) (X)	-0.58	0.16	β (M-->X) (X)	-0.58	0.16
Θ_2 (Y-->X|M) (M)	0.57	0.1	Θ_2 (Y-->X|M) (M)	0.71	0.13
βΘ_2 (indirect effect)	-0.33		βΘ_2 (indirect effect)	-0.41	
SE_{both}	0.11		SE_{both}	0.14	
P-value	**0.002**		P-value	**0.002**	
Mediation analysis: VD-->TG-->CRC			**Mediation analysis: VD-->TG-->CC**		
	Coefficient	SE		Coefficient	SE
β (M-->X) (X)	-0.39	0.15	β (M-->X) (X)	-0.39	0.15
Θ_2 (Y-->X|M) (M)	0.12	0.1	Θ_2 (Y-->X|M) (M)	0.28	0.13
βΘ_2 (indirect effect)	-0.045		βΘ_2 (indirect effect)	-0.11	
SE_{both}	0.04		SE_{both}	0.07	
P-value	0.29		P-value	0.09	
Mediation analysis: VD-->WC-->CRC			**Mediation analysis: VD-->WC-->CC**		
	Coefficient	SE		Coefficient	SE
β (M-->X) (X)	-0.85	0.15	β (M-->X) (X)	-0.85	0.15
Θ_2 (Y-->X|M) (M)	0.42	0.1	Θ_2 (Y-->X|M) (M)	0.49	0.13
βΘ_2 (indirect effect)	-0.36		βΘ_2 (indirect effect)	-0.42	
SE_{both}	0.11		SE_{both}	0.13	
P-value	**0.001**		P-value	**0.002**	

Model adjusted for age, sex, study centre, smoking status, and alcohol consumption
VD: vitamin D, MetS: metabolic syndrome, TG: triglycerides, WC waist circumference, CRC: colorectal cancer, CC: colon cancer

Table 3.9. Mediation analysis, association between vitamin D and colorectal cancer and colon cancer mediated by metabolic syndrome/components (**Causal method**)

Mediation analysis: VD-->MetS-->CRC			**Mediation analysis: VD-->MetS-->CC**	
	OR	95%CI	OR	95%CI
Natural direct effect	1.25	1.04 - 1.49	1.27	1.01 - 1.60
Natural indirect effect	1.05	1.00 - 1.10	1.04	0.99 - 1.10
total effect	1.31	1.09 - 1.57	1.33	1.05 - 1.68
% mediation		18%		14%
Mediation analysis: VD-->TG-->CRC			**Mediation analysis: VD-->TG-->CC**	
	OR	95%CI	OR	95%CI
Natural direct effect	1.30	1.09 - 1.56	1.32	1.05 - 1.66
Natural indirect effect	1.01	0.92 - 1.11	1.01	0.97 - 1.04
total effect	1.31	1.07 - 1.60	1.33	1.06 - 1.68
% mediation		4%		3%
Mediation analysis: VD-->WC-->CRC			**Mediation analysis: VD-->WC-->CC**	
	OR	95%CI	OR	95%CI
Natural direct effect	1.24	1.04 - 1.50	1.25	0.99 - 1.58
Natural indirect effect	1.05	0.93 - 1.19	1.05	0.99 - 1.12
total effect	1.31	1.06 - 1.63	1.32	1.04 - 1.67
% mediation		18%		17%

Model adjusted for age, sex, study centre, smoking status, and alcohol consumption
VD: vitamin D, MetS: metabolic syndrome, TG: triglycerides, WC waist circumference, CRC: colorectal cancer, CC: colon cancer

CHAPTER 4 ASSOCIATION BETWEEN VITAMIN D AND COLORECTAL CANCER USING A MENDELIAN RANDOMISATION APPROACH

**Part of this work has been published in the BMJ in 2017 with the title of "Circulating vitamin D concentration and risk of seven cancers: Mendelian randomisation study". I performed and interpreted the Mendelian randomisation analysis and sensitivity analyses of vitamin D and the seven cancers.*

In this Chapter I will use a Mendelian randomisation approach to infer whether low levels of serum 25(OH)D causally increase the risk of CRC.

Introduction

In the previous Chapter, I discussed the relationship between vitamin D and CRC using observational data, where I found that lower levels of vitamin D was associated with an increased risk of CRC. This observation was also supported by previous studies. Although observational epidemiology is important for understanding disease aetiology, causality is difficult to be inferred due to limitations including reverse causation and residual confounding. Causal inference is therefore mainly studied through RCTs, which by randomisation of participants to different exposures, address the aforementioned limitations.

In relation to vitamin D, RCTs to date have failed to confirm the protective effect of vitamin D supplementation and CRC risk seen in observational studies. For instance, one large RCT from the 40 WHI centres involving 36,282 postmenopausal women, in which 322 developed CRC, reported that daily supplementation of vitamin D and calcium for seven years had no effect on the incidence of CRC (HR: 1.08, 95%CI: 0.86 – 1.34). It could be argued that the dosage of vitamin D was low (400 IU per day), which is below the IOM recommendation of 600 – 800 IU per day (51,52) and therefore the efficacy of the dose may have been too weak. However, a more recent study of 2,303 postmenopausal women that provided participants in the treatment group with 2,000 IU of cholecalciferol and 1,500 mg of calcium carbonate also reported no statistically significant difference in all-type cancer incidence between the treatment group and the placebo group (p-value: 0.06) after a four year follow up period (53). Nevertheless, RCTs also suffer from limitations including poor compliance with the study protocol and short follow up periods.

MR is a relatively new approach which may overcome limitations of observational and RCT studies. MR is an IV analysis that uses IVs as proxies for environmental exposures. Due to the random allocation of alleles from parents to offspring, at the population level, alleles are generally independent of confounding factors. Moreover, since diseases cannot change our genotype, reverse causation does not affect MR analysis. Because we are exposed to the genotype over the long term, this analysis will show long term exposure. To perform a MR analysis, three assumptions must be met: 1) the IV must be associated with the exposure of interest, 2) the IV is independent of the confounding factors that confound the association between the exposure and the outcome, 3) the IV is associated with the outcome only through the exposure (207).

The SUNLIGHT consortium, a GWAS of approximately 42,000 individuals of European descent from 15 cohorts, has found four IVs that were associated with serum 25(OH)D levels (125): rs2282679, rs12785878, rs10741657, and rs6013897. This analysis has recently been updated and the new GWAS which includes 79,366 individuals, highlighted two additional IVs associated with 25(OH)D, rs8018720 and rs10745742. These variants can be used to construct a genetic instrument for vitamin D levels and subsequently study the causal role of vitamin D with diseases using a MR approach.

There have been a few MR studies on vitamin D with multiple outcomes. Some MR studies with 25(OH)D reported no evidence for causality. These include studies on ischaemic heart disease (275), T2D (276), coronary artery disease (CAD) (277), and schizophrenia (278) amongst other diseases. While other MR studies suggested a causal inverse association between low levels of vitamin D and Alzheimer's disease (279), multiple sclerosis (280), and ovarian cancer (281). **Table 4.1** reports the MR

studies on vitamin D with multiple outcomes using 1-sample and 2 sample MR methodologies after performing a narrative review of the literature, as well as different genetic variants associated with serum vitamin D.

Two MR studies have investigated the association between vitamin D and CRC which provided no evidence for a causal association (271,272). Here, I aim to repeat this analysis in a large sample and to assess the relationship between vitamin D and CRC by subsite. This analysis will be done on three studies 1) EPIC, where I will be using a 1-sample MR approach, 2) the UK Biobank with the SUNLIGHT consortium, using a 2-sample MR approach, and 3) the GECCO consortium also with the SUNLIGHT consortium, where I'll also be using a 2-sample MR approach.

Methods

SNP selection

To run a MR analysis, IVs associated with the exposure of interest, serum 25(OH)D, were needed. These were chosen based on evidence of their genome-wide significance in previous studies from the SUNLIGHT consortium; a large GWAS on vitamin D (125). The SNPs were rs12785878; located near *DHCR7/NADSYN1* on chromosome 11, rs2282679; located on chromosome 4 and encodes for the *GC* gene, rs6013897; located near *CYP24A1* on chromosome 20, and rs10741657 located on chromosome 11 in *CYP2R1*. SNPs used in the analyses in this Chapter were based on the GWAS study of 25(OH)D by Wang *et al*. published in 2010 (125). This study by Wang *et al.* was performed among 30,000 Europeans from 15 cohorts. In 2018, a GWAS by Jiang *et al.* found two additional IVs that were associated with 25(OH)D (p-value < $5x10^{-8}$), rs8018720 and rs10745742. This GWAS by Jiang *et al.* was performed on approximately 79,000 European individuals from 31 cohorts (128). Only the first four SNPs, which explained approximately 2.6% of the variance of 25(OH)D, were used in the analyses in this Chapter. . The analyses was not repeated including the two newly discovered SNPs, since the analyses were completed approximately a year before the second GWAS was published. Furthermore, since the two new SNPs explained little of the variance of serum 25(OH)D (~0.24%), I do not expect to observe a difference in results if the two additional SNPs were included in the analyses.

4.1 1-sample Mendelian randomisation of vitamin D and colorectal cancer risk in EPIC

Data for the potential causal association between vitamin D and CRC in a 1-sample study design, where the association between SNP with exposure and the association between SNP with outcome come from the same dataset, was obtained from the EPIC cohort. Information on study design, anthropometric measurement, vitamin D, MetS and its components, CRC, and genetic data, are provided in Chapter 1 section 6.

Since information at the individual level was available for serum 25(OH)D measurements, CRC status and the genotype data in the same data cohort, a 1-sample MR approach was used to estimate the causal effect of serum 25(OH)D on CRC risk. Genotype data for the four IVs associated with vitamin D is shown in **Table 4.2**. Since the genotype data did not completely match up with the nested case-control study for CRC, only 1,567 observations were included in the analysis. Observations that were greater than four SDs for serum 25(OH)D were considered outliers and were removed (N=2). Of those remaining, 831 were CRC cases and 734 were controls.

Before applying the 2-stage least square (2SLS) method, an unweighted risk score for the vitamin D-associated SNPs was generated (282). The unweighted risk score was calculated as the sum of all the 25(OH)D lowering alleles, with the homozygote risk alleles coded as 2, heterozygote risk allele as 1 and homozygote non-risk allele as 0. This was used in the first stage, where the unweighted risk score was regressed on the exposure, serum 25(OH)D, and in the second stage the predicted values from the first stage were regressed on the outcome, CRC.

Assumptions of HWE were tested using a chi-square test in Stata. Sensitivity analyses were performed by dividing the 4 SNPs into a synthesis allele score, alleles that were involved in the synthesis pathway of vitamin D (*DHCR7* and *CYP2R1*), and a

metabolism allele score, alleles that were in the metabolism pathway of vitamin D (*CYP24A1* and *GC*).

To infer a causal relationship between vitamin D and CRC, the assumptions of MR that were mentioned previously must be observed. Only SNPs that had biological plausibility and a genome-wide significant association (p-value< $5x10^{-8}$) with 25(OH)D were considered in the analysis. To assess potential violation to the second and third assumptions, associations between the IVs with potential confounders were assessed. Moreover, a power calculation was performed using a web-based application (https://cnsgenomics.shinyapps.io/mRnd/) which assessed power based on sample size, type 1 error rate which was set to 0.05, proportion of cases in the study, the true odds ratio of the outcome per standard deviation of the exposure, and the proportion of variance explained of the 25(OH)D-associated SNPs.

4.2 2-sample Mendelian randomisation of vitamin D and colorectal cancer risk in the UK Biobank and GECCO

Data for the potential causal association between vitamin D and CRC in a 2-sample (summary) study design, where the association between SNP to exposure and the association between SNP to outcome come from different datasets, was obtained from both the UK Biobank and the GECCO consortium. Information on study design, CRC, and genetic data, for the UK Biobank are provided in Chapter 1 section 6.

Data for genetic epidemiology of 25(OH)D concentration using the Study of Underlying Genetic Determinants of Vitamin D and Highly Related Traits (SUNLIGHT) consortium

Since the biomarker for circulating 25(OH)D in the UK Biobank has yet to be measured, the summary estimates of 25(OH)D-associated SNPs were obtained from the Study of Underlying Genetic Determinants of Vitamin D and Highly Related Traits (SUNLIGHT) consortium. This GWAS consisted of 42,000 individuals of European descent from 15 cohorts from Europe, Canada and the USA. Several different arrays were used for genotyping the 15 cohorts. Serum 25(OH)D was measured using either radioimmunoassay, chemiluminescent assay, ImmunoDiagnostic Systems OCTEIA ELISA analyser, or high performance liquid chromatography-tandem mass spectrometry. Serum 25(OH)D was naturally log transformed and adjusted for age, sex, BMI and season.

Data for genetic epidemiology of colorectal cancer in GECCO

Summary data for CRC was obtained from the Genetics and Epidemiology of Colorectal Cancer Consortium (GECCO), a collaboration of 23 studies which include data for over 40,000 participants (219). Results from individual GWAS for CRC were

combined using standard fixed-effects meta-analysis methods. These studies were genotyped using either Illumina, Affymetrix, or the OmniExpress platforms. SNPs were imputed to the Haplotype Reference Consortium using the program MACH. SNP were identified using the Human Genome Browser version hg19. Each SNP was adjusted for age, sex, genotyping phase and principal components to account for population substructure. Further information regarding the statistical analysis, imputation and quality control steps in the aforementioned GWAS can be found elsewhere (283,284).

Statistical analysis

The association of baseline characteristics of participants in the UK Biobank in relation to CRC case-control status was assessed using chi-square or t-test approaches after adjustment for age and sex. The associations between each of the SNPs of 25(OH)D with the risk of CRC using the additive model for the SNPs were done first in the UK Biobank. Moreover, an unweighted and weighted risk score for all four 25(OH)D-associated SNPs was generated in the UK Biobank. For the weighted risk score, the effect estimate of the association between the SNP with 25(OH)D from a previous study (176) was multiplied by each SNP and summed. Crude and adjusted logistic regressions were used to analyse the association between each of the four 25(OH)D-associated SNPs and genetic risk scores, with CRC risk in the UK Biobank adjusting for age, sex, BMI, ethnicity, smoking status, and education. Assumptions of HWE were tested using a chi-square test in Stata.

Common confounders including age, sex, smoking status, alcohol consumption, and BMI were tested for associations with 25(OH)D-associated SNPs in the UK Biobank to test for violation of the second and third MR assumptions.

2-sample MR analyses were then conducted to test the potential causal association between circulating 25(OH)D concentration and the risk of CRC using summary-level data for the effect estimates of 25(OH)D-associated SNPs with 25(OH)D concentration from the SUNLIGHT consortium and effect estimates of 25(OH)D-associated SNPs with the risk of CRC from the UK Biobank, as well as the GECCO consortium. MR was performed using two methods for the estimation of a causal effect using summary-level data: the fixed-effects IVW method and the likelihood-based method (211). The IVW method combines the ratio estimates from the individual variants using inverse-variance weights. The MR estimate and SE were calculated using the following equations:

- $\beta_{IVW} = \frac{\sum_k X_k Y_k \sigma^{-2} Y_k}{\sum_k X^2{}_k \sigma^{-2} Y_k}$ $\qquad$ $se(\beta_{IVW}) \sqrt{\frac{1}{\sum_k X^2{}_k \sigma^{-2} Y_k}}$

Where X_k is the estimate of the association between SNP_k and 25(OH)D concentration, Y_k is the association between SNP_k and CRC risk, and σ_{Yk} is the SE of the association between SNP_k and CRC risk. The second method was the likelihood-based method that assumes a linear relationship between the risk factor and the outcome. Serum 25(OH)D and CRC risk were jointly modelled using a bivariate normal distribution for each of the SNPs. The likelihood-based analysis was performed using a web-based platform (https://sb452.shinyapps.io/summarized/) with zero correlation. The IVW method was coded in Stata. All statistical analysis was performed on Stata version 13.

Analyses for CRC subtypes in the GECCO consortium were also performed: CRC in men and women, colon cancer, rectal cancer, proximal colon cancer, and distal colon cancer.

Sensitivity analyses were performed by dividing the four SNPs into synthesis allele score, alleles that were involved in the synthesis pathway of vitamin D (*DHCR7* and *CYP2R1*) and metabolism allele score, alleles that were in the metabolism pathway of vitamin D (*CYP24A1* and *GC*). Furthermore, related participants were also removed in the UK Biobank.

To infer a causal relationship between vitamin D and CRC, the assumptions of MR that were mentioned previously must be observed. To assess potential violation to the second and third assumptions, a goodness-of-fit (Cochran Q) test was assessed for each of the SNPs under the null hypothesis stating that each of the SNPs has an association with CRC risk that is proportional to its association with 25(OH)D. Rejection of the null hypothesis (p-value< 0.05), indicated heterogeneity of the association between the SNPs and CRC risk relative to the association between SNPs and 25(OH)D. MR-Egger regression, which is an adaptation of the Egger regression that tests for small study biases in meta-analysis, provides information on pleiotropy. The intercept in the MR-Egger regression provides a test for directional pleiotropy and the slope provides a potential causal estimate after adjusting for pleiotropic SNPs. A weighted median analysis was also conducted to protect against invalid instruments. The weighted median approach provides a potential causal effect estimate if at least 50% of the weight comes from valid SNPs (285). These sensitivity analyses measure the consistency of the effect estimates to the IVW method rather than the significance (286). If all sensitivity methods provide similar estimates, then a causal effect is more plausible (286). Moreover, a power analysis was performed to estimate the minimum detectable magnitude of association for CRC per 1kb increase in 25(OH)D using a web-based application (https://cnsgenomics.shinyapps.io/mRnd/).

Additionally, a meta-analysis of EPIC, the UK Biobank and GECCO was performed to obtain greater statistical power.

Results

4.1 1-sample Mendelian randomisation of vitamin D and colorectal cancer risk in EPIC

Based on previous literature, four independent SNPs, which have shown genome-wide significance (p-value< $5x10^{-8}$) with 25(OH)D levels, were selected (11,22). All four 25(OH)D-associated SNPs were in HWE (rs2282679 p-value: 0.52, rs10741657 p-value: 0.99, rs12785878 p-value: 0.29, and rs6013897 p-value: 0.21).

The four SNPs were associated with a decrease of log transformed 25(OH)D (based on the serum 25(OH)D lowering allele) (**Table 4.3**). However, rs6013897 was not statistically significantly associated with serum 25(OH)D (p-value: 0.55) in EPIC. The variance explained (R^2) and F-statistics were 0.02 and 8.11 respectively for all four SNPs combined denoting weak instruments due to the small sample size of the EPIC nested case-control study. When rs6013897 was removed, due to it not being statistically significantly associated with 25(OH)D in EPIC, the R^2 did not change, but the F-statistics increased to 10.71.

The association between the four 25(OH)D-associated SNPs with 831 CRC cases and 734 controls were assessed and found a statistically significant association between rs10741657 and the risk of CRC (β: -0.21, 95%CI: -0.36 – -0.07) (**Table 4.4**), while the other three SNPs reported no significant association with CRC risk (**Table 4.4**).

Only rs10741657 was found to be causally associated with CRC risk using the 2SLS approach (**Table 4.5**). An unweighted allele score was generated for the four SNPs and the results reported a borderline positive association between 25(OH)D lowering

alleles and the risk of CRC (OR: 1.03, 95%CI: 0.99 – 1.06, p-value: 0.09) (**Table 4.5**). An allele score was also generated for the alleles involved in 25(OH)D synthesis (rs12785878 and rs10741657) and metabolism (rs2282679 and rs6013897) with the synthesis allele score reporting a positive borderline causal association per 25(OH)D lowering allele (OR: 1.04, 95%CI: 1.00 – 1.08, p-value: 0.07) (**Table 4.5**).

To test whether the 25(OH)D-associated SNPs were associated with confounders, chi-square tests were performed for each of the SNPs with sex, age at blood recruitment, smoking status, alcohol, BMI, energy intake, and physical activity. The analysis reported significant associations between rs6013897 with physical activity and age (p-value< 0.05) and between rs12785878 with smoking (p-value: 0.02). However, after adjusting for multiple testing (p-value< 0.007), none of these associations remained significant (**Table 4.6**). To obtain at least 80% power with a sample size of ~1,500 and 2.6% variance explained by the IVs, a minimum odds ratio of 2.30 or greater or 0.46 or less was needed.

4.2 2-sample Mendelian randomisation of vitamin D and colorectal cancer risk in the UK Biobank and GECCO

Descriptive statistics in the UK Biobank

Among the 394,774 participants, 5,009 incident and prevalent CRC cases were identified until the year 2013, through linkage with cancer registry database. The majority (~95%) of the study population were of white ethnicity. There were more cases of CRC in males than females (58% and 42% respectively) (**Table 4.7**). Moreover, there was a significant difference in age, mean age for CRC cases was 61.5, while the mean age for controls was 55.8, (p-value <0.01). Cases had higher BMI, waist circumference and waist-to-hip ratio (p-values <0.01) and were more likely to have smoked (previously and current) (p-value <0.01). There was also a significant difference in alcohol consumption between CRC cases and controls with slightly more cases consuming/previously consumed alcohol compared to controls. There was a significant difference between cases and control in regards to education (p-value <0.01) and physical activity (p-value <0.01) (**Table 4.7**). Only baseline information was used for these variables.

Association estimates for individual SNPs with 25-hydroxyvitamin D

From the paper by Vimaleswaran *et al.*, the risk allele for each of the SNPs that were identified to be associated with 25(OH)D were obtained, as well as the estimated effect size between the SNPs and circulating 25(OH)D (**Table 4.8**). The SE was calculated by subtracting the upper limit 95% confidence interval by the lower limit and then dividing by 3.92 (287).

To obtain an 80% power with a 5% alpha level and 2.6% of the 25(OH)D variance explained by the four SNPs, the minimum detectable odds ratio should be less than 0.77 or greater than or equal to 1.23 (**Table 4.9**) for the UK Biobank. While in GECCO, with a sample size of 11,488 cases and 11,679 controls, to detect effect sizes of moderate magnitude, ranging from odds ratios of 0.68 per SD increase in serum 25(OH)D concentrations for rectal cancer to 0.81 for CRC (**Table 4.9**). Similar minimum detectable effect (MDE) sizes were estimated for CRC subtypes in GECCO (**Table 4.9**).

Association between SNPs and colorectal cancer in the UK Biobank

There was evidence of an association between only one of the four SNPs (rs2282679) with CRC risk per G allele (β: 0.06, 95%CI: 0.01 – 0.11, p-value: 0.02) after adjusting for age, sex, BMI, ethnicity, smoking status, and education (**Table 4.10**). The weighted and unweighted risk score had a normal distribution and reported no association with CRC risk (**Table 4.10**). Moreover, when the allele score was divided into synthesis allele score and metabolism allele score, an association was observed for the metabolism score (β: -0.01, p-value: 0.02), but no association was found for the synthesis score (β: 0.01, p-value: 0.11).

Association between SNPs and colorectal cancer in GECCO

None of the four SNPs were associated with CRC risk in GECCO. However, when divided into subtypes, rs6013897 showed an association with colon cancer (OR: 0.94, 95% CI: 0.89 - 0.99) (**Table 4.11**).

Mendelian randomisation estimates based on multi-SNP scores in the UK Biobank

Table 4.12 shows the MR estimates from the fixed-effects IVW method and the likelihood-based method. Both methods showed similar results of no causal association between 25(OH)D and CRC risk when all four SNPs were used (OR: 0.86, 95%CI: 0.68 – 1.08, p-value: 0.20). MR estimates between the allele scores is also displayed in **Figure 4.1**. A positive slope indicated that decreasing 25(OH)D concentration was associated with an increased risk of CRC, while a negative slope indicated that decreasing 25(OH)D concentration was associated with a decreased risk of CRC. However, since the 95%CI lines pass through zero, the association was not statistically significant for the allele score.

Mendelian randomisation estimates based on multi-SNP scores in GECCO

Based on MR analyses using either the IVW method or the likelihood-based method, there was little evidence that 25(OH)D concentrations was causally associated with the risk of CRC or their subtypes (OR: 0.92, 95%CI: 0.76 – 1.10 for CRC, OR: 0.92, 95%CI: 0.71 – 1.18 for CRC in women, OR: 0.91, 95%CI: 0.70 – 1.20 for CRC in men, OR: 0.90, 95%CI: 0.73 – 1.11 for colon cancer, OR: 0.93, 95%CI: 0.68 – 1.26 for rectal cancer, OR: 0.97, 95%CI: 0.73 – 1.28 for distal colon cancer, and OR: 0.83, 95%CI: 0.64 – 1.07 for proximal colon cancer) (**Table 4.12**). These associations were also displayed graphically in **Figure 4.1**.

Assessment of MR assumptions

The MR estimates have a causal interpretation only if the IV assumptions are valid. Statistical tests and sensitivity analyses were conducted to evaluate the potential violation of the second and third assumptions.

As a sensitivity analysis, related participants were removed from the study in the UK Biobank, which might introduce some bias. This may cause statistical tests of association to become invalid due to lack of true independence between individuals (288). However, in this study the results remained qualitatively similar (**Tables 4.13 – 4.14**), after the removal of related participants.

Further sensitivity analyses were performed to assess potential violations to the MR assumptions in both the UK Biobank and GECCO. The goodness of fit test (Cochran's Q test) reported no heterogeneity of the association between the SNPs and CRC risk relative to the association between SNPs and 25(OH)D (p-value >0.05) (**Table 4.15**). Moreover, the MR-Egger regression method, which tests for bias from instruments due to pleiotropy, reported no evidence of pleiotropy, and after adjusting for potential bias in the UK Biobank and GECCO (intercept p-value: 0.27, OR: 0.62, 95%CI: -1.16 – 1.21 and intercept p-value: 0.25, OR: 0.70, 95%CI: 0.49 – 1.02, respectively), the estimates remained similar to the IVW and likelihood based method (**Table 4.16**). The weighted median method also reported similar estimates to the IVW and the likelihood based method in the UK Biobank and GECCO (OR: 0.98, 95%CI: 0.64 – 1.49 and OR: 0.73, 95%CI: 0.73 – 1.08, respectively) (**Table 4.16**). MR analyses using two separate allelic scores (vitamin D synthesis and metabolism) were evaluated, and the results were consistent with the IVW effect estimates (**Tables 4.17 – 4.18**).

Common confounders including age, sex, smoking status, alcohol consumption, and BMI were tested for associations with 25(OH)D-associated SNPS in the UK Biobank to test for pleiotropy and found significant association with age, smoking status, alcohol consumption, education, and physical activity for the four 25(OH)D-associated SNPs (**Table 4.19**).

Meta-analysis of EPIC, the UK Biobank, and GECCO

EPIC, the UK Biobank, and GECCO were meta-analysed with a sample size of 419,506, of which 17,328 were CRC cases. The gene to exposure and the gene to outcome effect estimates for each SNP were meta- analysed across the three studies. The random effect estimates obtained from the meta-analyses of the gene to exposure and the gene to outcome were then used to rerun the Mendelian randomisation analysis using Fieller's Theorem. Random effect meta-analysis of these three studies also showed no evidence of causal association between vitamin D lowering alleles and the risk of CRC (OR: 1.01, 95%CI: 0.99 – 1.03, p-value: 00.33.

Discussion

In this Chapter, I have assessed whether low levels of vitamin D was causally associated with an increased risk of CRC and CRC subtypes using several MR approaches, including 2SLS analysis, IVW, and likelihood based methods. I used four SNPs shown to be associated with circulating 25(OH)D that explained 2.6% of the variance. The results from the MR analyses, including the meta-analysis of the three studies, did not support a causal effect between low levels of vitamin D and CRC risk.

In Chapter 3 a significant association between low levels of vitamin D and an increased risk of CRC in the EPIC cohort was observed, which was supported by previous observational studies. Due to the limitations of observational studies, one method that can assess causality is MR, in which it uses IVs as proxies for environmental exposures, eliminating residual confounding, reverse causation, and the need for a long follow up time.

Results from the 2SLS MR analysis in EPIC found no causal relationship between vitamin D and CRC. A similar study by Theodoratou *et al.* (271) with 2,001 cases and 2,237 controls reported similar results. Both studies did not provide evidence of a causal relationship between vitamin D and CRC risk. This analysis was severely underpowered and results may suffer from weak instrument bias, due to the small sample size (831 cases and 734 control) in the present study, with only 5% power to detect a true association. Two of the IVs (rs6013897 and rs10741657) were weak instruments (F-statistics <10) which could have biased the MR estimate towards the null. Moreover, the small confidence intervals in my study could also be a sign of weak

instrument bias (207). This analysis was only performed for descriptive and educational reasons.

The study by Theodoratou *et al.* also lacked power to detect a true effect of vitamin D on CRC. Another more recent 1-sample MR study by He *et al.*, included 10,725 CRC cases and 30,704 control from Scotland, Croatia, and the UK Biobank, in which they used 6 IVs for circulating vitamin D. Even with this larger study, He *et al.* also reported no causal relationship between vitamin D and CRC risk. The proportion of variance explained by the IVs (2.84%) was small and therefore did not achieve 80% power (272).

The validity of MR estimates requires several assumptions to be held. The first is that the SNPs are associated with the exposure of interest. This was done by selecting IVs that were significantly associated with serum 25(OH)D at the genome-side threshold of $< 5x10^{-8}$. These IVs also supported a biological significance in which they are located nearby genes that are involved in the synthesis and metabolism of vitamin D. However, when assessing the association between 25(OH)D-associated SNP and 25(OH)D, the SNP rs6013897, which was found to be strongly associated with 25(OH)D in the SUNLIGHT consortium was not statistically associated with 25(OH)D in this present study. Due to the potential violation of the first assumption of MR, I performed a sensitivity analysis by removing rs6013897 from the analysis and the results remained consistent. When further stratified into synthesis and metabolism allele scores, no evidence of causal association was found with CRC.

The EPIC study was underpowered to detect a potential causal association, due to the small sample size as well as weak instruments. Twin based studies suggested that genetics contribute to the variability of vitamin D, but to different extents. One study

reported no genetic contribution during the summer (121), while another study reported no genetic contribution during the winter (124). Due to the inconsistency to the genetic contribution for the variation of serum 25(OH)D, further analysis needs to be performed to determine the best season for blood withdrawal for GWAS testing.

To obtain more power, a 2-sample MR using a larger dataset to assess the causal association between vitamin D and CRC was performed, in which the SNP to CRC effect estimates were obtained from the UK Biobank, and the SNP to 25(OH)D levels effect estimates from the SUNLIGHT consortium. No causal association was found between vitamin D and CRC in the UK Biobank. This analysis consisted of a much larger sample size (5,009 cases and 389,765 controls) compared to the 1-sample MR in EPIC (831 cases and 734 control). Although this study was quite large, no association was found between serum vitamin D and CRC, where it was powered to find minimum detectable odds ratios ranging from 0.77 to 1.23 per 25 nmol/L in 25(OH)D. The study by He *et al.* also performed a 2-sample MR analysis on 18,967 CRC cases and 48,168 controls from seven studies (CCFR1, CCFR2, COIN, FINLAND, UK1, VQ58, SOCCS, Croatia, and UK Biobank) and reported no causal relationship between vitamin D and CRC (272). This study was also underpowered to detect small causal associations.

All three assumptions need to be met to have a valid MR result. The SNPs that were selected were significantly associated with serum 25(OH)D at the genome-wide threshold of $< 5x10^{-8}$ in the SUNLIGHT consortium. The second and third assumptions of MR cannot be proven. However, multiple tests were done to check for violations of these assumptions. Pleiotropy was tested using the Cochran's Q test, MR-Egger regression, and the weighted median analysis. These three methods reported no

evidence of pleiotropy for the association between 25(OH)D and CRC in the UK Biobank. However, when testing for associations between the SNPs and potential confounders in the UK Biobank, all four SNPs were associated with at least one of the measured confounders at the genome-wide significance threshold of $5x10^{-8}$, which could be a sign of pleiotropy. The study by He *et al*. reported no evidence of pleiotropy (272).

The advantages of summary MR is that summary level statistics of the IV to exposure and IV to outcome could be used from different sources to obtain causal estimates. And since the UK Biobank had no data on serum 25(OH)D levels at the time of analysis, this method was useful to estimate the causal effect. Moreover, the main advantage of using a summary MR approach is to increase the statistical power of the study, particularly with a binary outcome (289). Although I used a larger sample size compared to EPIC, this study remained underpowered.

To try to obtain sufficient power to detect a potential causal association, I used the GECCO consortium which had approximately double the number of cases of CRC compared to the UK Biobank (11,488 cases in GECCO compared to 5,009 cases in the UK Biobank). As with the UK Biobank, the effect estimates from the associations between SNPs and 25(OH)D were obtained from the SUNLIGHT consortium and the SNPs to CRC effect estimates from the GECCO. Although there was a larger number of cases in this dataset, no causal association was found between vitamin D and CRC in GECCO. Additionally, this study was powered to find minimum detectable odds ratios ranging from 0.81 to 1.23 per 25 nmol/L in 25(OH)D.

When further stratified by subtype, the results were not statistically significant in any of the colorectal subtypes. However, these analyses lacked even more power since there were fewer cases included in these sub-analyses.

The Cochran's Q test, MR-Egger regression, and weighted median analyses were used to assess for pleiotropy. These methods reported no evidence of pleiotropy for the association between 25(OH)D and CRC in the GECCO.

To conclude, no causal association was found between vitamin D and CRC in any of the study datasets. Due to the lack of power and limitations of the MR method, it cannot be said with certainty that there is no causal relationship between vitamin D and CRC. Instead, larger sample size and stronger IVs are needed to increase the power of the study which can then determine whether a potential causal association truly exists between vitamin D and CRC. GECCO has now released data on approximately 60,000 CRC case-control pairs, in which this analysis could be done in the future. Although no association was found between vitamin D and CRC subtypes in this current study, with the newly released GECCO dataset, which is three times the size of the current dataset, a potential causal association could be found between vitamin D and CRC subtypes.

Summary level data that is stratified by vitamin D deficiency status can also establish a better understanding of a causal relationship between vitamin D and CRC. A large randomised, double-blind, placebo-controlled, 2x2 factorial VITamin D and Omega-3 TriaL (VITAL) clinical trial on approximately 25,000 participants, recently published results after a mean treatment period of 5 years with 2,000 IU/day of vitamin D_3 (290). Results from this reported 51 cases of CRC in the vitamin D supplementation group and 47 cases of CRC in the placebo group. The study reported that daily

supplementation of high dose of vitamin D did not reduce the incidence of CRC (HR: 1.09, 95%CI: 0.73 – 1.62). Further trials are ongoing to add information on other vitamin D doses with cancer and cardiovascular outcomes. A 2-year post intervention follow-up of the VITAL trial is currently ongoing to understand latency effects of vitamin D supplementation, as well as to increase statistical power (291).

Table 4.1. Previous Mendelian randomisation studies on vitamin D and multiple outcomes

Study Author	Outcome	Year	MR approach	Cohort	Population	25(OH)D associated SNPs	cases/ controls	MR Findings
Theodoratou, Evropi *et al.*	Colorectal cancer	2012	1-sample	Study of CRC in Scotland (SOCCS)	Scotland Males and Females	rs2282679 rs12785878 rs10741657 rs6013897	2001/2237	There was NO association between any of the 4 SNF and CRC risk The estimated causal effect was for genetically lower 25(OH)D and the risk of CRC was: OR 1.16 (95% CI 0.60, 2.23), whilst it was 0.94 (95% CI 0.46 - 1.91) ar 0.93 (0.53 - 1.63) when using an upstream and a downstream allele score, respectively
Skaaby, Tea *et al.*	Cardiovascular risk factors	2013	1-sample	Inter99 Monica 10 Health2006	Denmark males and females	Filaggrin gene: R510X, 2282del4 and R2447X	11,983	IV analyses showed a 23.8% (95%CI: 3.0 - 48.6, pva 0.02) higher HDL level and a 30.5% (95%CI: 0.8 - 51 pval:0.04) lower serum level of triglyceride per doubli of vitamin D. No causal association was found for LDL, Total cholesterol, SBP, DBP, BMI, WC, and MetS
Vimaleswaran, Karani, *et al.*	Obesity	2013	2-sample	meta-analysis of 21 cohorts	UK, US, Canada, Finland, Germany, and Sweden Males and Females	rs2282679 rs10741657 rs6013897 rs12785878	42,024	The IV ratio reported that a 10% higher genetically instrumented BMI was associated with 4.2% lower 25(OH)D concentration (95%CI: -7.1 - -1.3, pval:0.00 however little evidence was found for the reverse association.
Kunutsor, Setor *et al.*	High Blood pressure	2013	2-sample	16 RCT studies	Europe, North America, Asia Males and females	rs2282679 rs12785878 rs10741657 rs6013897	1,879	MR results showed no causal association between vitamin D and SBP (B: -0.11, 95%CI: -0.31 - 0.09, pv 0.27) and (B: -0.10 mmHg, 95%CI: -0.22 - 0.03, pval: 0.13) for DBP
Afzal, Shoaib *et al.*	Mortality	2014	1-sample	The Copenhagen General Population Study The Copenhagen City Heart Study The Copenhagen Ischemic Heart Disease Study	Denmark males and females	rs7944926 rs11234027 rs10741657 rs12794714	35,334	Genetically low 25(OH)D were associated with increa risk for cancer mortality (OR per 20nmol/L decreases 25(OH)D OR:1:10, 1.02 - 1.19) Genetically low 25(OH)D were associated with increa risk of all-cause mortality (OR per 20nmol/L decrease 25(OH)D OR:1.30, 1.05 - 1.61) No causal association was found between genetically low 25(OH)D and cardiovascular mortality (OR per 20nmol/L decreases in 25(OH)D OR:0.77, 95%CI: 0. 1.08)

Study Author	Outcome	Year	MR approach	Cohort	Population	25(OH)D associated SNPs	cases/ controls	MR Findings
Vimaleswaran, Karani, *et al.*	Arterial BP Hypertension	2014	2-sample	D-CarDia (meta-analysis)	Europe and North America	rs2282679 rs12794714 rs6013897 rs12785878	51,122	Per 10% increase in genetically low 25(OH)D was associated with a change of -0.29 mmHg on DBP (95%CI: -0.52 - -0.07, pval:0.01) Per 10% increase in genetically low 25(OH)D was associated with a change of -0.37 mmHg on SBP (95%CI: -0.73 - 0.003, pval:0.052) Per 10% increase in genetically low 25(OH)D was associated with an hypertension OR: 0.92 (95%CI: 0 - 0.97, pval:0.002)
Husemoen, *et al.*	Adiponectin	2014	1-sample	Inter99 Study MONICA 10 study	Denmark	rs2282679 FLG loss of function	9,061	Genetically lowered 25(OH)D concentration supporte positive causal association with adiponectin (effect estimate per doubling of 25(OH)D was 37.1%, 95%C 3.7% - 95.2%, pval:0.08)
Ooi, Esther *et al.*	Cholesterol	2014	1-sample	Copenhagen General Population Study Copenhagen City Heart Study	Danish Males and females	rs11234027 rs7944926 rs10741657 rs12794714	85,868	Genetically elevated cholesterol was associated with 25(OH)D (-8.9%, 95CI: -15% - -2.3%)
Mokry, Lauren *et al.*	Multiple Sclerosis	2015	2-sample	SUNLIGHT (15 cohorts) CaMos (SNP analysis) IMSGC Immunochip study	European descent Male and female	rs10741657 rs12785878 rs2282679 rs6013897	14,498/24,0 91	1 SD decrease in ln25(OH)D level was associated wi an increased risk of MS (OR:2.02, 95%CI:1.65 - 2.46
Brondum-Jacobsen, Peter *et al.*	Ischaemic heart disease & Myocardial infarction	2015	1-sample	The Copenhagen General Population Study The Copenhagen City Heart Study The Copenhagen Ischemic Heart Disease Study	Denmark Male and female	rs7944926 rs11234027 rs10741657 rs12794714	10,170	No evidence was found to suggest that genetically reduced 25(OH)D was associated with increased risk IHD (OR per 25nmol/L of 25(OH)D: 0.98, 95%CI: 0.7 1.26, pval:0.86) or MI (OR per 25nmol/L of 25(OH)D:1.15, 95%CI: 0.83 - 1.59, pval:0.49)
Ye, Zhang *et al.*	Type 2 diabetes	2015	2-sample	EPIC-Norfolk EPIC-InterAct DIAGRAM consortium, ADDITION-Ely, Norfolk Diabetes, and Cambridgeshire	European Male and female	rs12785878 rs10741657 rs4588 rs17217119	28,144/76,3 44	MR showed no significant association between VD a T2D (OR per 25 nmol/L lower 25(OH)D concentration 1.01, 95%CI: 0.75 - 1.36, pval:0.94)

Study Author	Outcome	Year	MR approach	Cohort	Population	25(OH)D associated SNPs	cases/ controls	MR Findings
Liefaard, Marte *et al.*	C-Reactive protein	2015	1-sample	Rotterdam Study	Netherlands males and females	rs2282679 rs12785878 rs10741657 rs6013897 18SNPs for CRP	9,649	A Bi-directional MR analyses showed no association between 25(OH)D genetic risk score and lnCRP (β p SD:-0.018, pval:0.08). No association was found between the CRP genetic score and 25(OH)D (B per SD: 0.001; pval: 0.998)
Dudding, Tom	Dental Childhood Caries	2015	1-sample	Avon Longitudinal Study of Parents and Children (ALSPAC)	Southwest England Children both gender	rs2282679 rs10741657 rs7944926	5,545	MR analyses reported no evidence of a causal association between increased 25(OH)D and the odc caries experience (OR per 10 nmol/L increase in 25(OH)D 0.93 (95% CI: 0.83, 1.05; P = 0.26))
Trummer, Olivia *et al.*	Prostate cancer prognosis	2015	1-sample	Austrian Prostate cancer genetics (PROCAGENE)	Austria	rs2282679	703 prostate cases	NO association between single SNP with prostate cancer outcomes
Mokry, Lauren *et al.*	Alzheimer disease	2016	2-sample	SUNLIGHT consortia International Genomics of Alzheimer's Project	European Male and female	rs2282679 rs12785878 rs10741657 rs6013897	17,008/37,154	MR analyses demonstrated that a 1-SD decrease in natural log–transformed 25OHD increased AD risk by 25% (OR: 1.25, 95% CI: 1.03–1.51, pval: 0.02).
Rhead, Brooke *et al.*	Multiple sclerosis	2016	1-sample	Kaiser Permanente Medical Care Plan, Northern California Region (KPNC) The Epidemiological Investigation of Multiple Sclerosis the Genes and Environment in Multiple Sclerosis	USA males and females	rs2282679 rs2060793 rs3829251 rs10741657	7,391/14,777	MR analyses reported that increasing levels of 25(OH are associated with a decreased risk of MS in both populations: KPNC: OR:0.79, 95%CI: 0.64 - 0.99, pval:0.04 EIMS/GEMS: OR: 0.86, 95%CI: 0.76 - 0.98, pval:0.0 Meta-analysis: OR: 0.85 95%CI: 0.76 -0.94, pval: 0.0
Manousaki, Despoina *et al.*	Coronary artery disease	2016	2-sample	SUNLIGHT consortia CARDIoGRAM study CaMos	males and females	rs2282679 rs12785878 rs10741657 rs6013897	22,233/64,762	Genetically lowered 25(OH)D was not associated wit increased risk of CAD The MR for CAD was OR: 0.99 (95% CI: 0.84–1.17; P=0.93) per SD decrease in log-transformed 25OHD levels for all four SNPs
Ong, Jue-Sheng *et al.*	Ovarian cancer	2016	2-sample	Ovarian Cancer Association Consortium	European women	rs2282679 rs12794714 rs7944926	10,065/21,654	Genetically lowered 25(OH)D were associated with higher ovarian cancer susceptibility (OR per 20 nmol decrease in 25(OH)D: 1.27, 95%CI: 1.06 - 1.51)
Li, Shan-Shan *et al.*	Bone mineral density	2016	1-sample	Department of Osteoporosis and Bone Diseases Outpatient Clinic of Shanghai Jiao Tong University Affiliated	Chinese postmenopausal women	rs2282679 rs12785878 rs10741657 rs6013897	1,824	No causal association between genetically low serum 25(OH)D and BMD

Study Author	Outcome	Year	MR approach	Cohort	Population	25(OH)D associated SNPs	cases/ controls	MR Findings
				Sixth People's Hospital				
Taylor, Amy *et al.*	Schizophrenia	2016	2-sample	SUNLIGHT consortia Vimaleswaran *et al.* Psychiatric Genetics Consortium (PGC)	European male and female	rs2282679 rs12785878 rs10741657 rs6013897	34,241/45,604	No evidence for ca causal effect of 25(OH)D on schizophrenia (OR per 10% increase in 25(OH)D: 0.9 95%CI: 0.97 - 1.01) Positive suggestive evidence for a causal effect of schizophrenia on vitamin D levels (OR: 1.05, 95%CI: 0.99-1.12)
Hysinger, Erik *et al.*	Paediatric asthma	2016	1-sample	The children's hospital of Philadelphia Center for Applied Genomics-The Asthma cohort	US	rs2282679 rs10741657	1388	No association between vitamin D genetic risk score severe asthma exacerbations
Gianfrancesco, Milena *et al.*	paediatric-onset multiple sclerosis	2017	1-sample	Multiple paediatric MS centres in the US Epidemiologic Investigation of risk factors for MS Genes and Environment in MS	US and Sweden	rs2282679 rs2060793 rs3829251	415 cases US 262 cases Sweden	Strong evidence for causal association between low serum vitamin D and risk of paediatric-onset of MS
Olsson, Erika *et al.*	Dementia Cognitive impairment	2017	1-sample	Uppsala Longitudinal Study of Adult men	Uppsala, Sweden men only	rs12785878 rs12794714	1182	A genetic risk score was generated and NO associati was found between the risk score with Alzheimer's disease, Vascular dementia, All-cause dementia, and cognitive impairment
Noordam, Raymond *et al.*	Features of skin aging	2017	bi-directional	Rotterdam Study Leiden Longevity Study	Netherland	rs2282679 rs3829251 rs2060793	3,831 661	Higher genetically determined 25(OH)D concentratio was not associated with aging Genotype and GRS for pigment spots or perceived a were not associated with higher 25(OH)D concentrati
Maddock, Jane *et al.*	cognitive function	2017	1-sample	17 cohorts	European	rs12785878 rs12794714	172,349	No evidence for a causal association between 25(OH concentrations and cognitive performance in mid to later life

Table 4.2. Characteristics of genetic variants associated with 25-hydroxyvitamin D concentration in EPIC

Chromosome	SNP	Position	Gene	Risk Allele	Percent missing
11	rs12785878	71456403	DHCR7/NADSYN1	G	0.00%
11	rs10741657	14893332	CYP2R1	G	0.06%
20	rs6013897	54125940	CYP24A1	A	11.00%
4	rs2282679	71752606	GC	G	0.10%

Table 4.3 Association between genetic variants associated with 25-hydroxyvitamin D and serum 25(OH)D in EPIC

SNP	Effect Allele	Coefficient	SE	95%CI	p-value	F-statistic
rs12785878	G	-0.07	0.02	-0.10 – -0.03	<0.001	16.58
rs10741657	G	-0.03	0.02	-0.06 – -0.00	0.05	3.86
rs6013897	A	-0.01	0.02	-0.05 – 0.02	0.55	0.36
rs2282679	G	-0.05	0.02	-0.09 – -0.03	0.001	11.16

Table 4.4. Association between genetic variants associated with 25-hydroxyvitamin D and colorectal cancer risk in EPIC

SNP	Effect Allele	Coefficient	SE	95%CI	p-value
rs12785878	G	0.02	0.08	-0.13 – 0.17	0.81
rs10741657	G	-0.21	0.07	-0.36 – -0.07	0.004
rs6013897	A	-0.02	0.09	-0.19 – 0.15	0.79
rs2282679	G	-0.03	0.08	-0.18 – 0.12	0.69
Synthesis allele score*		-0.10	0.05	-0.21 – 0.004	0.06
Metabolism allele score*		-0.03	0.06	-0.14 – 0.09	0.64

*Synthesis SNPs: rs12785878, rs10741657, Metabolism SNPs: rs2282679 and rs6013897

Table 4.5. Mendelian randomisation estimates between multi-SNP risk scores of continuous 25(OH)D and colorectal cancer risk in EPIC using the Two Stage Least Square approach

Vitamin D associated SNP	OR	95%CI	P-value
rs12785878	0.99	0.95 – 1.04	0.79
rs10741657	1.13	1.04 – 1.23	0.005
rs6013897	1.04	0.79 – 1.37	0.79
rs2282679	1.01	0.96 – 1.06	0.68
unweighted allele score 2SLS*	1.03	1.00 – 1.06	0.08
Synthesis score*	1.04	1.00 – 1.08	0.07
Metabolism score*	1.01	0.96 – 1.07	0.63

*Synthesis SNPs: rs12785878, rs10741657, Metabolism SNPs: rs2282679 and rs601389

Table 4.6. Testing for association between vitamin D-associated SNPs with potential confounders in EPIC

rs10741657	AA	GA	GG	p-value from chi-sq test
	%	%	%	
Sex (male)	15.3	45.1	39.6	0.21
MET (inactive)*	13.4	46.8	39.8	0.71
smoking (never)	13.8	46.7	39.5	0.92
rs10741657	**AA**	**GA**	**GG**	**p-value from ANOVA**
	Mean (SD)	**Mean (SD)**	**Mean (SD)**	
age	59.0 (8.1)	58.4 (8.0)	58.4 (7.9)	0.64
BMI	26.4 (4.0)	26.6 (3.9)	26.6 (4.3)	0.85
Energy (kcal)	2084 (682)	2127 (746)	2057 (682)	0.19
Alcohol (g/d)	14.2 (20.7)	13.3 (19.7)	13.9 (20.8)	0.77

rs6013897	TT	TA	AA	p-value from chi-sq test
	N	N	N	
Sex (male)	454	238	31	0.17
MET (inactive)*	179	96	9	**0.03**
smoking (never)	429	264	33	0.29
rs6013897	**TT**	**TA**	**AA**	**p-value from ANOVA**
	Mean (SD)	**Mean (SD)**	**Mean (SD)**	
age	58.7 (8.1)	58.5 (7.5)	56.0 (8.6)	**0.02**

BMI	26.6 (4.1)	26.6 (4.0)	26.2 (4.6)	0.76
Energy (kcal)	2094 (752)	2080 (654)	2193 (612)	0.44
Alcohol (g/d)	13.3 (19.2)	14.1 (22.4)	13.9 (16.6)	0.75

rs12785878	**TT**	**GT**	**GG**	**p-value from chi-sq test**
	N	**N**	**N**	
Sex (male)	388	269	66	0.14
MET (inactive)*	154	104	26	0.53
smoking (never)	356	299	71	**0.02**
rs12785878	**TT**	**GT**	**GG**	**p-value from ANOVA**
	Mean (SD)	**Mean (SD)**	**Mean (SD)**	
Age	58.9 (7.6)	58.0 (8.2)	58.4 (8.5)	0.12
BMI	26.5 (4.2)	26.5 (3.9)	27.0 (4.0)	0.40
Energy (kcal)	2090 (687)	2094 (771)	2108 (579)	0.96
Alcohol (g/d)	14.1 (21.6)	13.2 (18.9)	12.3 (18.3)	0.51

rs2282679	**TT**	**GT**	**GG**	**p-value from chi-sq test**
	N	**N**	**N**	
Sex (male)	345	314	64	0.27
MET (inactive)*	143	114	27	0.60
smoking (Never)	358	299	69	0.91
rs2282679	**TT**	**GT**	**GG**	**p-value from ANOVA**
	Mean (SD)	**Mean (SD)**	**Mean (SD)**	
Age	58.3 (8.0)	58.7 (7.9)	58.7 (8.0)	0.63
BMI	26.6 (3.9)	26.5 (4.2)	26.8 (4.5)	0.67
Energy (kcals)	2080 (743)	2106 (688)	2111 (661)	0.76
Alcohol	13.0 (20.9)	14.7 (20.4)	12.2 (15.7)	0.19

*MET: metabolic equivalent

Table 4.7. Baseline characteristics of participants in the UK Biobank stratified by colorectal cancer cases and adjusted for age and sex

Baseline characteristics	CRC* cases (N=5,009)	Controls (N=389,765)
	Mean(SD)	Mean (SD)
Sex (%)		
Female	42.2	52.5
Male Δ	57.8	47.5
Age Δ	61.5 (6.2)	55.8 (8.1)
BMI* (kg/m^2) (N=393,190) Δ	27.9 (4.5)	27.5 (4.7)
Waist circumference (cm) (393,919) Δ	94.8 (13.2)	91.4 (13.3)
WHR* Δ	0.9 (0.1)	0.9 (0.1)
Smoking (%)		
Never	45.2	55.4
Previous	45.0	33.6
Current	9.0	10.5
Alcohol (%)		
Never	3.9	4.4
Previous	4.1	3.4
Current	92.0	92.1
MET score (%)		
Low	35.0	32.6
Moderate	45.4	47.7
High	19.6	19.7
Education (%)		
College or university	36.6	39.7
A levels	12.9	13.6
O levels	27.8	25.6
CSEs	4.8	6.9
NVQ/HND/HNC	9.7	8.0
Other professions	8.2	6.1
Ethnicity (%)		
White	96.7	94.3
Non-White	2.8	5.7
Family history of CRC* at baseline (%)		
No	82.3	89.4
Yes	17.7	10.6

* CRC: colorectal cancer, BMI: body mass index, WHR: waist-to-hip ratio, MET: metabolic equivalent
Ŧ A chi2 test was used for categorical variables and a t-test was used for continuous variables
Δ Age was adjusted for sex and sex was adjusted for age. All other variables were adjusted for age and sex.

Table 4.8. Characteristics of genetic variants associated with 25-hydroxyvitamin D concentration reported in prior Genome-Wide Association Study (Vimaleswaran *et al.*(176))

Chromosome	SNP	Position	Gene	Risk Allele	β* estimates	SE
11	rs12785878	71456403	DHCR7/NADSYN1	G	-2.114	0.17
11	rs10741657	14893332	CYP2R1	G	-1.724	0.166
20	rs6013897	54125940	CYP24A1	A	-0.978	0.185
4	rs2282679	71752606	GC	G	-4.671	0.175

* Reported in per unit increase in log-transformed continuous 25-hydroxyvitamin D concentrations

Table 4.9. Number of colorectal cancer cases and controls and statistical power in EPIC, the UK Biobank and GECCO

Colorectal Cancer Type	Study	Cases	Controls	Minimum detectable OR* ($R^2 = 0.03$)	Minimum detectable OR* ($R^2 = 0.05$)
All	EPIC	831	734	0.46/2.30	0.54/1.90
All	UK Biobank	5,009	389,765	0.77/1.23	0.82/1.18
All	GECCO	11,488	11,679	0.81/1.23	0.85/1.18
All (women)	GECCO	6,132	6,380	0.75/1.33	0.80/1.25
All (men)	GECCO	5,356	5,297	0.73/1.37	0.78/1.28
Colon	GECCO	7,678	11,679	0.78/1.28	0.83/1.20
Rectal	GECCO	2,783	11,679	0.68/1.47	0.75/1.33
Distal Colon	GECCO	3,354	11,679	0.70/1.43	0.77/1.30
Proximal Colon	GECCO	4,185	11,679	0.73/1.37	0.79/1.27

*Minimum detectable odds ratio per 1 standard deviation increase/decrease in 25(OH)D concentrations; assume 80% power, 5% alpha level, and that 3% or 5% of the 25(OH)D variance is explained by the four SNPs (rs2282679, rs10741657, rs12785878, rs6013897). One standard deviation in 25(OH)D corresponds approximately to 25 nmol/L.

Table 4.10. Association between rs12785878, rs10741657, rs6013897, and rs2282679 with colorectal cancer risk in the UK Biobank participants

			Crude model		Adjusted model*		
SNP ID	**Alleles (N)**	**Cases**	**Coefficient (95%CI)**	**p-value**	**Coefficient (95%CI)**	**p-value**	**Standa Error**
rs2282679	TT	2,468	REF		REF		0.02
	GT	2,101	0.07 (0.01– 0.13)	0.02	0.06 (-0.01 – 0.13)	0.08	
	GG	446	0.12 (0.02 – 0.22)	0.02	0.11 (-0.01 – 0.23)	0.06	
	Per G allele		0.06 (0.02 – 0.11)	<0.01	0.06 (0.01 – 0.11)	0.02	
rs10741657	AA	847	REF		REF		0.02
	AG	2,406	-0.06 (-0.13 – 0.02)	0.17	-0.05 (-0.14 – 0.04)	0.32	
	GG	1,762	-0.10 (-0.18 – 0.02)	0.02	-0.07 (-0.17 – 0.02)	0.13	
	Per G allele		-0.05 (-0.09 – -0.01)	0.02	-0.04 (-0.08 – 0.01)	0.13	
rs12785878	TT	3,039	REF		REF		0.03
	GT	1,677	-0.04 (-0.10 – 0.02)	0.16	-0.02 (-0.09 – 0.05)	0.50	
	GG	299	-0.26 (-0.38 – -0.14)	<0.01	-0.05 (-0.19 – 0.10)	0.53	
	Per G allele		-0.09 (-0.13 – -0.04)	<0.01	-0.02 (-0.08 – 0.03)	0.45	
rs6013897	TT	3,225	REF		REF		0.03
	AT	1,593	0.00 (-0.06 – 0.06)	0.97	0.04 (-0.03 – 0.11)	0.27	
	AA	185	-0.09 (-0.24 – 0.06)	0.24	-0.01 (-0.18 – 0.16)	0.88	
	Per A allele		-0.02 (-0.07 – 0.03)	0.52	0.02 (-0.03 – 0.08)	0.43	
Unweighted risk score ρ			-0.02 (-0.04 – 0.00)	0.05	0.004 (-0.02 – 0.03)	0.76	
Weighted risk score ρ			0.00 (-0.01 – 0.01)	0.91	-0.01 (-0.01 – 0.00)	0.20	
Synthesis score ¥ρ			0.03 (0.02 – 0.05)	<0.01	0.01 (0.00 – 0.03)	0.11	
Metabolism score τ ρ			-0.01 (-0.02 – 0.00)	0.01	-0.01 (-0.02 – 0.00)	0.02	

* The model was adjusted for age, sex, BMI, ethnicity, education, and smoking status
¥ Synthesis allele score is made up from rs12785878 and rs10741657
τ Metabolism allele score is made up from rs6013897, and rs2282679

Table 4.11. Association between rs12785878, rs10741657, rs6013897, and rs2282679 and colorectal cancer risk in the GECCO consortium

SNP	Effect allele	β (95%CI)	Standard error
Colorectal cancer			
rs2282679	G	0.03 (-0.01 ,0.07)	0.019439
rs10741657	G	-0.01 (-0.04 ,0.03)	0.020209
rs12785878	G	0.00 (-0.04 ,0.05)	0.024893
rs6013897	A	-0.03 (-0.08 ,0.01)	0.020617
Colon cancer			
rs2282679	G	0.03 (-0.02 ,0.08)	0.024185
rs10741657	G	-0.01 (-0.04 ,0.04)	0.025138
rs12785878	G	0.02 (-0.03 ,0.07)	0.024416
rs6013897	A	-0.06 (-0.12 ,-0.01)	0.026441
Rectal cancer			
rs2282679	G	0.05 (-0.02 ,0.11)	0.032928
rs10741657	G	-0.02 (-0.08 ,0.04)	0.030318
rs12785878	G	-0.04 (-0.12 ,0.03)	0.035909
rs6013897	A	-0.01 (-0.08 ,0.07)	0.039647
Women			
rs2282679	G	0.02 (-0.04 ,0.08)	0.029162
rs10741657	G	-0.01 (-0.05 ,0.05)	0.030021
rs12785878	G	0.02 (-0.04 ,0.08)	0.029162
rs6013897	A	-0.03 (-0.09 ,0.03)	0.030621
Men			
rs2282679	G	0.04 (-0.02 ,0.10)	0.028617
rs10741657	G	-0.01 (-0.06 ,0.05)	0.030021
rs12785878	G	-0.02 (-0.08 ,0.04)	0.030318
rs6013897	A	-0.04 (-0.11 ,0.03)	0.035909
Distal			
rs2282679	G	0.01 (-0.05 ,0.07)	0.029443
rs10741657	G	0.01 (-0.04 ,0.07)	0.029443
rs12785878	G	0.02 (-0.05 ,0.08)	0.029162
rs6013897	A	-0.05 (-0.13 ,0.02)	0.036273
Proximal			
rs2282679	G	0.05 (-0.01 ,0.10)	0.028352
rs10741657	G	-0.01 (-0.06 ,0.05)	0.030021
rs12785878	G	0.03 (-0.03 ,0.09)	0.028887
rs6013897	A	-0.06 (-0.13 ,0.00)	0.031569

Table 4.12. Mendelian randomisation estimates between multi-SNP risk scores of continuous 25(OH)D and colorectal cancer risk calculated using the inverse-variance weighted method and the likelihood method from the UK Biobank and GECCO

		Inverse-Variance Weighted			Likelihood Method		
Colorectal Cancer Type	**Study**	**OR[a]**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**
All	UK Biobank	0.86	0.68 – 1.08	0.20	0.86	0.68 – 1.08	0.20
All	GECCO	0.92	0.76 – 1.10	0.36	0.92	0.76 – 1.10	0.36
All (women)	GECCO	0.92	0.71 – 1.18	0.52	0.92	0.71 – 1.18	0.52
All (men)	GECCO	0.91	0.70 – 1.20	0.52	0.91	0.70 – 1.20	0.52
Colon	GECCO	0.90	0.73 – 1.11	0.33	0.90	0.73 – 1.11	0.33
Rectal	GECCO	0.93	0.68 – 1.26	0.64	0.93	0.68 – 1.26	0.64
Distal Colon	GECCO	0.97	0.73 – 1.28	0.83	0.97	0.73 – 1.28	0.83
Proximal Colon	GECCO	0.83	0.64 – 1.07	0.14	0.82	0.64 – 1.07	0.14

[a]The odds ratios (ORs) represent increase/decrease of risk per standard deviation decrease in nmol/L in the natural scale of 25(OH)D.

Table 4.13. Association between rs12785878, rs10741657, rs6013897, and rs2282679 and colorectal cancer risk in 407,295 UK Biobank participants after removing related participants

			Crude model		Adjusted model*		
SNP ID	**Alleles (N)**	**Cases**	**Coefficient (95%CI)**	**p-value**	**Coefficient (95%CI)**	**p-value**	**Standard Error**
rs2282679	TT	2,029	REF		REF		
	GT	1,713	0.07 (0.00 – 0.13)	0.04	0.07 (-0.01 – 0.14)	0.07	
	GG	374	0.15 (0.03 – 0.26)	0.01	0.13 (0.01 – 0.26)	0.04	0.03
	Per G allele		0.07 (0.02 – 0.12)	<0.01	0.07 (0.01 – 0.12)	0.02	
rs10741657	AA	706	REF		REF		
	AG	1,952	-0.08 (-0.17 – 0.00)	0.06	-0.07 (-0.17 – 0.03)	0.19	
	GG	1,458	-0.11 (-0.20 – -0.02)	0.02	-0.09 (-0.19 – 0.02)	0.10	0.03
	Per G allele		-0.05 (-0.09 – -0.01)	0.02	-0.04 (-0.09 – 0.01)	0.12	
rs12785878	TT	2,468	REF		REF		
	GT	1,397	-0.03 (-0.09 – 0.04)	0.41	-0.02 (-0.10 – 0.06)	0.62	
	GG	251	-0.28 (-0.41 – -0.15)	<0.01	-0.06 (-0.22 – 0.10)	0.47	0.03
	Per G allele		-0.09 (-0.14 – -0.04)	<0.01	-0.02 (-0.08 – 0.04)	0.43	
rs6013897	TT	2,659	REF		REF		
	AT	1,293	-0.02 (-0.09 – 0.04)	0.52	0.03 (-0.04 – 0.11)	0.36	
	AA	154	-0.09 (-0.25 – 0.07)	0.29	-0.02 (-0.20 – 0.16)	0.83	0.03
	Per A allele		-0.03 (-0.08 – 0.02)	0.23	0.02 (-0.04 – 0.08)	0.59	
Unweighted risk score ρ			-0.02 (-0.05 – 0.00)	0.06	0.004 (-0.02 – 0.03)	0.78	
Weighted risk score ρ			0.00 (-0.01 – 0.01)	0.80	-0.01 (-0.02 – 0.00)	0.17	
Synthesis score ¥ρ			0.03 (0.02 – 0.05)	<0.01	0.02 (0.00 – 0.04)	0.10	
Metabolism score τ ρ			-0.01 (-0.02 – 0.00)	0.01	-0.01 (-0.03 – 0.00)	0.01	

* The model was adjusted for age, sex, BMI, ethnicity (other, vs white), education, and smoking status

¥ Synthesis allele score is made up from rs12785878 and rs10741657

τ Metabolism allele score is made up from rs6013897, and rs2282679

Table 4.14. Mendelian randomisation estimates between multi-SNP risk scores, synthesis score, and metabolism score of continuous 25-hydroxyvitamin D and colorectal cancer risk using the inverse-variance weighted method and the likelihood-based method for the estimation of a causal effect using summarised data from the UK Biobank and SUNLIGHT consortium after the removal of related participants

	Inverse-Variance Weighted estimate			Likelihood estimate		
	OR	**95%CI**	**p-value**	**OR**	**95%CI**	**p-value**
Allele score	0.84	0.65 – 1.08	0.17	0.84	0.65 – 1.08	0.17
Synthesis allele score	1.54	0.92 – 2.57	0.10	1.54	0.92 – 2.58	0.10
Metabolism allele score	0.70	0.52 – 0.93	0.01	0.70	0.52 – 0.93	0.01

Table 4.15. P-values for the goodness-of-fit test for continuous 25(OH)D and colorectal cancer risk and subtypes

Colorectal Cancer Type	P-value of goodness-of-fit test including all SNPs
UK Biobank	
All	0.14
GECCO	
All	0.36
All (women)	0.64
All (men)	0.45
Colon	0.11
Rectal	0.35
Distal Colon	0.43
Proximal Colon	0.13

Table 4.16. Mendelian randomisation estimates between multi-SNP risk scores of continuous 25(OH)D and colorectal cancer risk and subtypes calculated using Egger's regression and weighted median approach[a] in the UK Biobank and GECCO

	MR Egger		Weighted median
Colorectal Cancer Type	**Intercept p-value**	**Slope OR (95% CI)**	**OR (95% CI)**
UK Biobank			
All	0.27	0.62 (0.31 – 1.21)	0.98 (0.64 – 1.49)
GECCO			
All	0.25	0.70 (0.49 – 1.02)	0.89 (0.73 – 1.08)
All (women)	0.51	0.77 (0.46 – 1.28)	0.90 (0.68 – 1.18)
All (men	0.27	0.63 (0.36 – 1.10)	0.88 (0.66 – 1.18)
Colon	0.33	0.66 (0.42 – 1.02)	0.86 (0.69 – 1.08)
Rectal	0.28	0.62 (0.33 – 1.14)	0.91 (0.65 – 1.28)
Distal Colon	0.59	0.82 (0.46 – 1.44)	0.93 (0.69 – 1.26)
Proximal Colon	0.25	0.55 (0.32 – 0.92)	0.77 (0.58 – 1.02)

[a]To further assess potential violation of the second assumption of Mendelian randomisation (MR) due to pleiotropic SNP effects, I employed the MR-Egger regression method, which is an adaptation of the Egger regression in a meta-analysis. The p-value of the intercept is as a valid test of directional pleiotropy, whereas the slope of the MR-Egger regression is the pleiotropy-adjusted causal effect estimate. I further used the weighted median method to diagnose and protect against invalid genetic instruments.

Table 4.17. Mendelian randomisation estimates between multi-SNP risk scores of continuous 25(OH)D **synthesis** (rs10741657, rs12785878) and colorectal cancer risk and subtypes calculated using the inverse-variance weighted method (left) and the likelihood method (right).

	Inverse-Variance Weighted Method			Likelihood Method		
Colorectal Cancer Type	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**
UK Biobank						
All	1.47	0.92 – 2.36	0.10	1.47	0.92 – 2.36	0.11
GECCO						
All	1.00	0.69 – 1.45	0.99	1.00	0.69 – 1.45	0.99
All (women)	0.89	0.54 – 1.48	0.67	0.89	0.54 – 1.48	0.67
All (men)	1.16	0.67 – 1.99	0.61	1.16	0.67 – 2.00	0.61
Colon	0.88	0.58 – 1.34	0.55	0.88	0.57 – 1.34	0.55
Rectal	1.50	0.80 – 2.80	0.20	1.50	0.80 – 2.81	0.21
Distal Colon	0.84	0.48 – 1.46	0.54	0.84	0.48 – 1.46	0.54
Proximal Colon	0.86	0.51 – 1.43	0.56	0.86	0.51 – 1.43	0.56

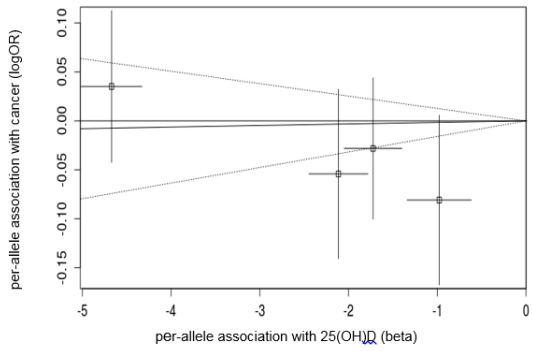
Table 4.18. Mendelian randomisation estimates between multi-SNP risk scores of continuous 25(OH)D **metabolism** (rs2282679, rs6013897) and colorectal cancer risk and subtypes calculated using the inverse-variance weighted method (left) and the likelihood method (right).

	Inverse-Variance Weighted Method			Likelihood Method		
Colorectal Cancer Type	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**
UK Biobank						
All	0.73	0.56 – 0.95	0.02	0.73	0.56 – 0.95	0.02
GECCO						
All	0.89	0.72 – 1.10	0.28	0.89	0.72 – 1.10	0.28
All (women)	0.93	0.69 – 1.24	0.63	0.93	0.69 – 1.24	0.63
All (men)	0.84	0.62 – 1.16	0.30	0.84	0.62 – 1.16	0.30
Colon	0.91	0.71 – 1.16	0.45	0.91	0.71 – 1.16	0.45
Rectal	0.79	0.56 – 1.13	0.20	0.79	0.56 – 1.13	0.20
Distal Colon	1.02	0.74 – 1.40	0.93	1.02	0.74 – 1.40	0.93
Proximal Colon	0.82	0.61 – 1.10	0.18	0.81	0.61 – 1.10	0.18

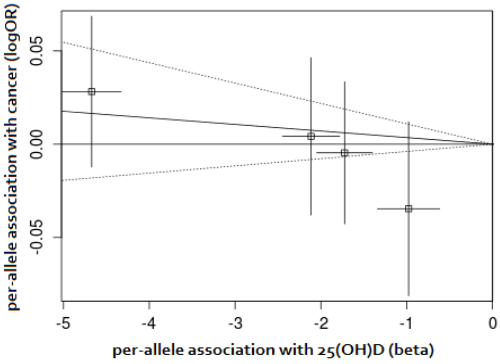
Table 4.19. Correlation between each of the vitamin D associated SNPs with confounders in the UK Biobank

rs2282679					
variable	**Coefficient**	**SE**	**P-value**	**LCI**	**UCI**
age	0.00	0.00	**2.65E-10**	0.00	0.00
sex	0.00	0.01	7.66E-01	-0.01	0.01
smoking status	0.01	0.00	**6.75E-03**	0.00	0.02
alcohol consumption	0.06	0.01	**1.04E-17**	0.05	0.08
MET	0.01	0.00	5.97E-02	0.00	0.02
BMI	0.00	0.00	**2.69E-03**	0.00	0.00
education	0.00	0.00	3.14E-01	0.00	0.01
rs10741657					
variable	**Coefficient**	**SE**	**P-value**	**LCI**	**UCI**
age	0.00	0.00	**4.17E-05**	0.00	0.00
sex	0.00	0.01	7.74E-01	-0.01	0.02
smoking status	0.00	0.01	7.58E-01	-0.01	0.01
alcohol consumption	-0.07	0.01	**4.74E-11**	-0.09	-0.05
MET	-0.01	0.01	1.32E-01	-0.02	0.00
BMI	0.00	0.00	8.11E-01	0.00	0.00
education	0.00	0.00	1.68E-01	-0.01	0.00
rs12785878					
variable	**Coefficient**	**SE**	**P-value**	**LCI**	**UCI**
age	-0.01	0.00	**9.11E-91**	-0.01	-0.01
sex	0.01	0.01	2.14E-01	0.00	0.02
smoking status	-0.04	0.00	**8.56E-19**	-0.05	-0.03
alcohol consumption	-0.28	0.01	0.00E+00	-0.30	-0.27
MET	-0.02	0.00	**1.36E-06**	-0.03	-0.01
BMI	0.00	0.00	6.54E-01	0.00	0.00
education	-0.03	0.00	**2.36E-37**	-0.03	-0.02
rs6013897					
variable	**Coefficient**	**SE**	**P-value**	**LCI**	**UCI**
age	0.00	0.00	1.46E-02	0.00	0.00
sex	0.00	0.01	9.48E-01	-0.01	0.01
smoking status	-0.01	0.00	2.00E-02	-0.02	0.00
alcohol consumption	-0.07	0.01	**1.39E-18**	-0.08	-0.05
MET	-0.01	0.00	4**.95E-03**	-0.02	0.00
BMI	0.00	0.00	2.61E-01	0.00	0.00
education	-0.01	0.00	**6.75E-04**	-0.01	0.00

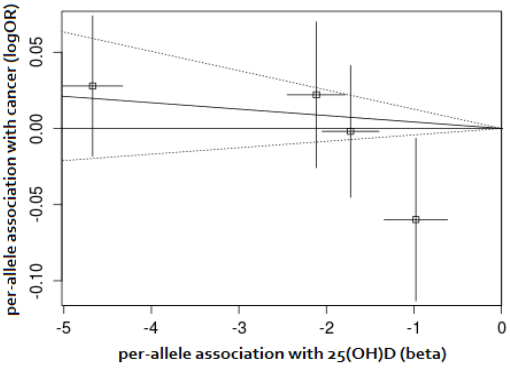
Colorectal cancer (UK Biobank)



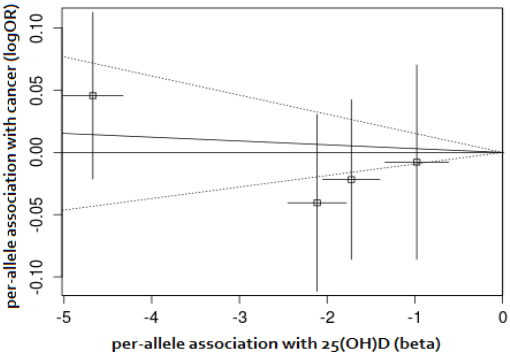
Colorectal cancer (GECCO)



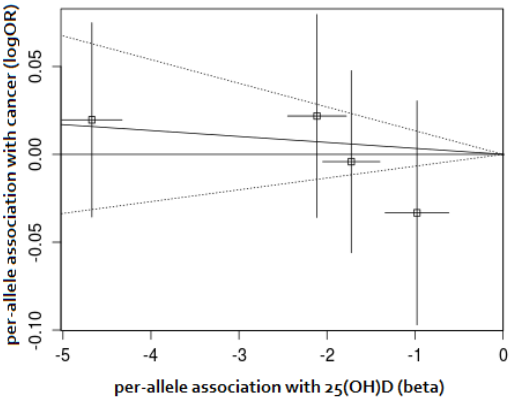
Colorectal cancer (colon; GECCO)



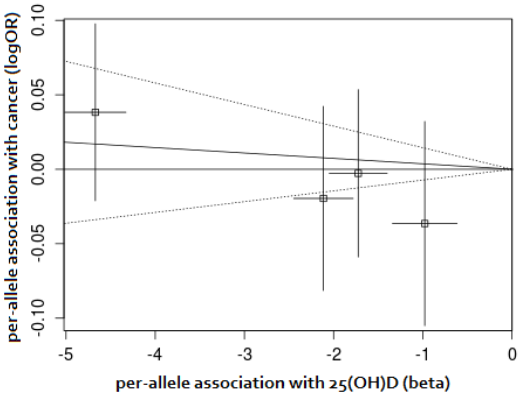
Colorectal cancer (rectal; GECCO)



Colorectal cancer (women; GECCO)



Colorectal cancer (men; GECCO)



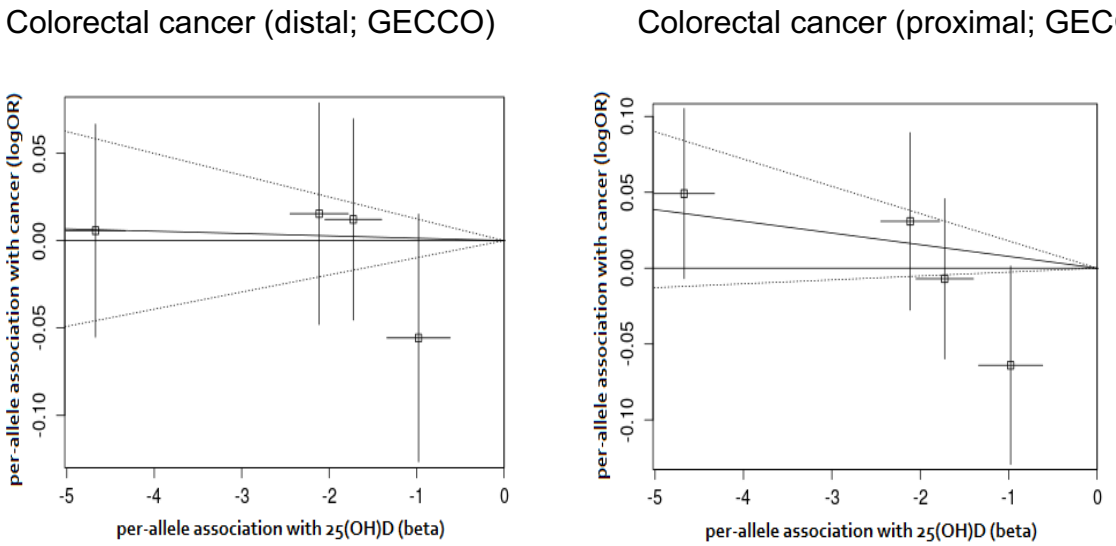


Figure 4.1. Scatter plots of associations between vitamin D associated SNPs with cancer risk and serum 25(OH)D concentrations in the UK Biobank and GECCO consortia. Per-allele associations with cancer risk are plotted against per-allele associations with continuous serum 25(OH)D concentrations (vertical and horizontal black lines show the 95% confidence interval (CI) for each SNP). The plots are overlaid by the Mendelian randomisation estimate (slope of solid line) and its 95% CI (dotted lines) of the multi-SNP score of continuous serum 25(OH)D on risk of seven colorectal cancer and subtypes

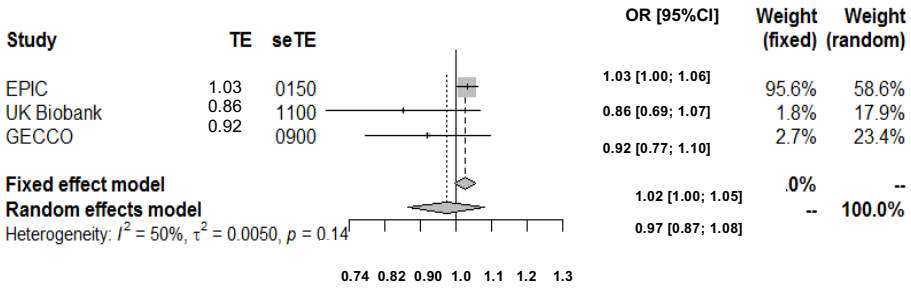


Figure 4.2. Meta-analysis of EPIC, the UK Biobank, and GECCO for the potential causal association between vitamin D and colorectal cancer

CHAPTER 5 BI-DIRECTIONAL MENDELIAN RANDOMISATION OF VITAMIN D AND COMPONENTS OF METABOLIC SYNDROME

In this Chapter I will investigate the potential causal associations between vitamin D and MetS components using a MR approach. I will investigate the directionality of these relationships using a bi-directional MR methodology.

Introduction

It has been estimated that around 15 – 30% of the world's adult population has MetS and are mostly residing in developed countries (292). Results from Chapter 2 and 3 reported an inverse relationship between vitamin D with MetS and its components, which was supported by previous studies (171–173). Due to the associations found in Chapter 2 and 3, this Chapter investigated the potential causal relationship between vitamin D and MetS components, as well as the direction of the association.

In the section below I discuss evidence from observational studies, RCTs and MR studies investigating the associations between vitamin D and MetS components.

Blood pressure and vitamin D

Conflicting results have been found between vitamin D and blood pressure across study types. A systematic review and meta-analysis on 11 observational studies reported a significant inverse association between circulating 25(OH)D levels and the risk of hypertension (RR: 0.88, 95%CI: 0.81 – 0.97) (186). While a systematic review and meta-analysis of 46 RCTs (4,541 individuals) showed no statistically significant effect on SBP or DBP in participants taking vitamin D supplements (β: 0.0, 95%CI: -0.8 – 0.8, I^2=21% and β: -0.1, 95%CI: -0.6 – 0.5, I^2=20% respectively) (187). However, 36 of the 46 RCTs had a follow up duration < 1 year which might have been too short a duration to observe significant results (187). A MR assessing the potential causal association between vitamin D and blood pressure using data from the D-CarDia study for 25(OH)D, consisting of approximately 99,500 participants of European ancestry from the UK, US, Canada, Finland, Germany and Sweden. Data for blood pressure

was obtained from ICBP (N=146,581), CHARGE (N=29,136) and Global BPGen (N=34,433). A 10% increase in 25(OH)D concentration was associated with lower DBP (β: -0.29, 95%CI:-0.52 – -0.07, p-value: 0.01) as well as reduced odds of hypertension (OR: 0.92, 95%CI: 0.87 – 0.97, p-value: 0.002). No association was found between 25(OH)D concentrations and SBP (per 10% increase in 25(OH)D; β: -0.37, 95%CI:-0.73 – 0.003, p-value: 0.052) (180).

HDL cholesterol, triglycerides, and vitamin D

A NHANES III study on 8,421 participants reported an inverse association comparing high levels of serum 25(OH)D (≥96.4 nmol/L) to low levels (≤48.4 nmol/L) with low HDL cholesterol (OR: 0.71, 95%CI: 0.56 – 0.90) after adjusting for multiple factors (173). However, after further adjustments for other MetS components, the association was not statistically significant (OR: 0.96, 95%CI: 0.76 – 1.22) (173). Furthermore, high levels of vitamin D was associated with lower risk of hypertriglyceridemia compared to low levels of vitamin D (OR: 0.59, 95%CI: 0.44 – 0.80) after adjusting for multiple factors (173). As with HDL cholesterol, the association between vitamin D and hypertriglyceridemia did not reach statistical significance after adjusting for the other MetS components (173). A meta-analysis of observational studies for the association between vitamin D and lipids in adults has not been performed.

A MR study on three population based studies from the Danish Central Personal Register (N=11,983) on vitamin D and cardiovascular diseases reported higher HDL cholesterol (relative difference in %: 23.8, 95%CI: 3.0 – 48.6) and lower triglyceride levels (relative difference in %: -30.5, 95%CI: -51.3 – -0.8) per doubling of vitamin D when using the filaggrin gene as the IV for vitamin D status. Previous studies have showed that the loss-of-function mutation in this gene increased the levels of serum

25(OH)D concentration (190). However, these associations became non-significant after Bonferroni corrections for the 10 outcomes they measured (190). A bi-directional MR on approximately 85,000 Danish individuals investigated the association between vitamin D and HDL cholesterol in which they reported a 50% decrease in plasma 25(OH)D levels was associated with lower HDL cholesterol levels (change in HDL cholesterol levels %: -6.0, 95%CI: -10.0 – -2.3) (191). Moreover, they reported that halving of HDL cholesterol levels increased plasma 25(OH)D levels (change in 25(OH)D %: 20.0, 95%CI: 7.4 –34) which contradicted their observational results of a low HDL cholesterol being associated with low 25(OH)D (change in 25(OH)D %: -1.5, 95%CI: -2.2 – -0.7 per halving of HDL cholesterol) after multivariable adjustment (191). This may have been due to the small variance explained for HDL cholesterol-associated IV, where the variance explained by the 3 IVs for HDL cholesterol was 0.4% and 1.9% for 25(OH)D, (191) which might lead to false positive results.

A meta-analysis of 8 RCTs investigating the association between vitamin D supplementation with HDL cholesterol and triglyceride levels reported no statistically significant relationship (Mean difference: -0.14 mg/dl, 95%CI: -0.99 – 0.71, I^2=16% and mean difference: -1.92 mg/dl, 95%CI: -7.72 – 3.88, I^2=46% respectively) (192). In a sensitivity analysis where the authors included studies that had an intervention duration greater than one year, they reported a significant association between vitamin D supplementation and HDL cholesterol (mean difference: -2.01, 95%CI: -3.83 – -0.18, p-value: 0.03), but not with triglycerides (mean difference: 0.96, 95%CI: -10.01 – 11.92, p-value: 0.86) (192).

HbA1c, T2D, and vitamin D

There have been consistent results for the association between vitamin D and T2D in observational studies and RCTs. A systematic review and meta-analysis of 14 prospective studies from Europe, US, Australia and Japan on 190,626 participants, investigated the association between vitamin D and T2D and reported a 19% lower risk of developing T2D among those with higher vitamin D levels (RR: 0.81, 95%CI: 0.71 – 0.92, I^2=67%) (174).

However, a MR study on 28,144 cases of T2D and 76,344 controls from the EPIC-InterAct and four case-control studies (DIAGRAM consortium, ADDITION-Ely, Norfolk Diabetes and Cambridgeshire) assessing the potential causal association between 25(OH)D and T2D, showed no causal association (OR: 1.01, 95%CI: 0.75 – 1.36, p-value: 0.94 per 1 SD decrease in 25(OH)D concentration). This study also assessed the association between vitamin D and HbA1c on 46,368 Europeans from the MAGIC consortium and also reported no statistically significant association (β: 0.01, 95%CI: -0.04 – 0.05) per 1 SD decrease in 25(OH)D. However, this study was underpowered (a minimum effect estimate of 0.08 or -0.09 is required with an R^2 of 0.05 to obtain sufficient power of 80%) (181).

A systematic review and meta-analysis of 24 RCTs reported reduced HbA1c levels with vitamin D supplementation, which supported results from observational studies (SMD: -0.25, 95%CI: -0.45 – -0.05, I^2=75.5%) (182).

Obesity and vitamin D

There have been inconsistent findings on the association between vitamin D and obesity. A meta-analysis of 13 observational studies on 48,882 adults and the elderly worldwide, investigated the association between obesity and the risk of vitamin D

deficiency, and reported that vitamin D deficiency was 33% higher in obese adults compared to the normal weight adults, irrespective of age and latitude (RR: 1.33, 95%CI: 1.15 – 1.54, I^2=91.5%) (175).

A bi-directional MR analysis on vitamin D and BMI was done on 42,000 participants from the D-CarDia study and the GIANT consortium. This study demonstrated that high BMI potentially caused lower 25(OH)D concentration (β: -0.42, 95%CI: -0.71 – 0.13, p-value: 0.005), while the effects of low 25(OH)D on BMI were small and non-significant (β: -0.00, 95%CI: -0.06 – 0.05 and β: -0.03, 95%CI: -0.06 – 0.01 for synthesis and metabolism scores respectively with BMI) (176). However, this study may suffer from weak instrument bias, in which the instruments explained a very small amount of variation, 12 SNPs for BMI explained only 0.97% of the variation in BMI while the 4 SNPs associated with vitamin D explained 1.9% variation of vitamin D, which could lead to low statistical power and biased results.

On the other hand, a systematic review and meta-analysis of 11 RCTs providing information on 2,114 participants reported no effect of vitamin D supplementation compared to placebo on BMI (weighted mean difference: -0.06, 95%CI: -0.14 – 0.03, I^2=0.0%)(177). When further stratified into vitamin D supplemental dosage (<1000, 1000 - <2000, 2000 - <4000, and ≥4000 IU) results were unchanged (177).

Only two of the five MetS components: BMI and HDL cholesterol, had been investigated in a bi-directional MR analysis with vitamin D and both these studies had fewer vitamin D-associated SNPs; four vitamin D-associated SNPs for both studies compared to 6 vitamin D associated SNPs in the present study, as well as a smaller sample size for vitamin D (42,024 for the study on BMI and 31,435 for the study on

HDL cholesterol) compared to ~80,000 in this present study. Summary of previous MR studies on MetS components can be found in **Table 5.1**.

The aim of this Chapter is to conduct a bi-directional MR study on large genetic consortia to understand the causality and directionality of the association between vitamin D and MetS components. This will be done using a 2-sample MR approach using data from the updated SUNLIGHT consortium for vitamin D and several consortia for MetS components including: the GIANT consortium for BMI, MAGIC consortium for HbA1c levels, GLGC for HDL cholesterol and triglyceride levels, and the UK Biobank for SBP and DBP.

Methods

Study Design

A 2-sample bi-directional MR analysis was performed to investigate the potential causal association between vitamin D and MetS components: BMI, HDL cholesterol, triglycerides, HbA1c, and SBP and DBP. SNPs associated with vitamin D from the SUNLIGHT consortium were used as proxies for serum 25(OH)D to investigate whether vitamin D caused an increased/decreased risk in MetS components (explained below). SNPs associated with MetS components were obtained from several GWAS consortia (180,235–237, see below) and were used as proxies to estimate the causal association between MetS components and vitamin D.

Genetic variants of serum 25-hydroxyvitamin D

To construct a genetic risk score for vitamin D, six IVs which were reported to reach genome-wide significance were used (128). This updated SUNLIGHT GWAS by Jiang *et al.* consisted of 79,366 individuals of European descent from 31 cohorts from Europe, Canada and the USA.

Several different arrays were used for genotyping the 31 cohorts. Details of genotyping methods and quality control can be found in the paper by Wang *et al.* (125). Serum 25(OH)D was measured using either radioimmunoassay, chemiluminescent assay, ImmunoDiagnostic Systems OCTEIA ELISA analyser, or high performance liquid chromatography-tandem mass spectrometry. SNPs used for imputation differed in each study, in regards to the stringent filters including a MAF ≥ 0.01, SNP call rate ≥ 0.97, and HWE (p-value ≥ $5x10^{-7}$). These SNPs were then used for imputation based

on the haplotypes of the HapMap CEU trios using the MACH or IMPUTE software. Over 2 million imputed SNPs passed the quality control for each of the studies and were analysed for association with 25(OH)D levels. Serum 25(OH)D was naturally log transformed and adjusted for age, sex, BMI and season. The dataset included the rsnumber, effect allele, reference allele, the beta coefficient and the SE. The beta estimates were reported as percent change in serum 25(OH)D per effect allele.

The six SNPs that were associated with serum 25(OH)D were rs12785878; located near *DHCR7/NADSYN1* on chromosome 11, rs2282679; located on chromosome 4 and encodes for the *GC* gene, rs6013897; located near the gene *CYP24A1* on chromosome 20, rs10741657 which is located on chromosome 11 in the gene *CYP2R1*, rs10745742 located near the gene *AMDHD1* on chromosome 12, and rs8018720 located near the gene *SEC23A* on chromosome 14. The combination of the six SNPs explained approximately 2.7% of the variance in 25(OH)D. None of these SNPs were in LD. The GWAS summary data for 25(OH)D levels were obtained from the Genome-Wide Repository of Associations Between SNPs and Phenotypes (GRASP) (https://grasp.nhlbi.nih.gov/FullResults.aspx) from the study by Manousaki *et al.* (296).

Genetic variants of metabolic syndrome components

To construct a genetic risk score for each of the MetS components, IVs which were reported to reach genome-wide significance and were not in LD (R^2< 0.1) in each of the relative GWAS consortium were used (except for BMI where I used a threshold of R^2< 0.001, due to the large number of genome-wide significant SNP that had an R^2<0.1 (822 SNPs)). The risk/effect allele was selected, such that the association with

each of the MetS components were directionally concordant. Details on the IVs for each of the MetS components and the consortia are discussed below.

Blood Pressure

The UK Biobank has genotyped approximately 488,000 participants. Genotype calling was performed by Affymetrix on two arrays: approximately 50,000 participants samples were run on the UK BiLEVE array and the remaining were run on the UK Biobank Axiom array where both datasets have been merged to produce a single format. There were 805,426 markers in the genotype data and approximately 96 million genotypes imputed. The process of imputation involved pre-phasing the directly genotyped markers on both the UK BiLEVE and UK Biobank Axiom arrays with the 1000 Genome Phase 3 dataset used as a reference panel followed by a haploid imputation step using the program IMPUTE4.

A phenome scan was performed in the UK Biobank using PHESANT (PHEnome Scan ANalysis Tool) for DBP. The GWAS for SBP was obtained from a paper by Carter *et al.*(297), which was also based on data from the UK Biobank. This data was further adjusted for anti-hypertensive treatment. The phenotype for DBP and SBP were measured in both males and females. Association analyses for inverse-rank normal transformed SBP and DBP were conducted using a linear regression. Models assumed an additive genetic model adjusting for age, age^2, sex, sex*age, sex*age^2 and the top twenty principal components as covariates (298). Because SBP and DBP were inverse normal transformed, the unit for blood pressure is one SD which is equivalent to 19.07 mmHg. The format of the dataset was in 'chr:position:ref_allele:alt_allele'. To obtain rsnumbers, SNPnexus was used. Variants that did not have an rsnumber were removed (~2,000 for DBP). SNPs associated with

SBP or DBP were selected based on genome wide significance (p-value<$5x10^{-8}$) and were in linkage equilibrium (R^2<0.1), which was done by clumping using the TwoSampleMR package in R. Clumping is used to keep only one representative SNP per region of LD.

A total of 284 SNPs were found for SBP from the UK Biobank, which explained 4.3% of the variance of SBP and 267 SNPs were found for DBP, which explained 3.8% of the variance of DBP (**Table 5.2**).

The percent variation explained for each of the IVs of MetS components were calculated from the following formula (299) (**Equation 3**):

$$\frac{2\beta^2 * MAF * (1 - MAF)}{2\beta^2 * MAF * (1 - MAF) + \left(se(\beta)\right)^2 * 2 * N * MAF(1 - MAF)} \quad (3)$$

β is the estimate of the SNP-exposure association, MAF is the minor allele frequency for the SNPs, se is the standard error for the SNP-exposure association and N is the sample size of the study population.

Triglycerides and HDL cholesterol

Summary statistics from 37 studies of European ancestry on triglycerides and HDL cholesterol were obtained from the Global Lipids Genetic Consortium (GLGC) (294). Approximately 197,000 SNPs were genotyped, and were selected based on previous GWAS for cardiovascular and metabolic phenotype using the Illumina iSelect Metabochip genotyping array on approximately 95,000 individuals. Blood lipid levels were measured after more than 8 hours of fasting. Individuals that were consuming lipid-lowering medication were excluded. Individual SNP associations were performed

using a linear regression with the lipid being inverse normal transformed. MACH was used on ~2.6 million SNPs to obtain imputed genotype. Lipids were adjusted for age, age^2, and sex. The data was downloaded from the following website (https://bit.ly/2OeRVVi).

Summarised effect sizes associated with the SNPs, the corresponding SE, the effect allele, and the reference allele were extracted for triglyceride levels and HDL cholesterol from the meta-analysis by Willer *et al.* (294). The files included both genotyped and imputed SNPs from the joint analysis of Metabochip and GWAS data. 211 SNPs were found to be associated with HDL cholesterol at genome-wide significance, in which it explained approximately 14% of the variance of HDL cholesterol, while 137 SNPs were found to reach genome-wide significance for triglyceride levels and explained approximately 9% of the variance of triglyceride levels (**Table 5.2**).

For the SNP-exposure (G-X) association, the beta estimates were inverse normal transformed and therefore HDL cholesterol and triglycerides were interpreted as per one SD, in which one SD for HDL cholesterol and triglycerides were equivalent to 17 mg/dL and 102 mg/dL, respectively. The SD was calculated using the formula from the following website “https://bit.ly/2SII4R1”.

HbA1C

Summary statistics on HbA1c was downloaded from the MAGIC (the Meta-Analyses of Glucose and Insulin-related traits Consortium) website (https://www.magicinvestigators.org/downloads/) which consisted of 56 cohorts of European ancestry (N=123,665). Participants were excluded if they were diabetic, on

diabetes medication, or had a fasting glucose level greater than 7mmol/L. HbA1c was measured as a National Glycohemoglobin Standardization Program (NGSP) percent (300).

Each cohort was genotyped on commercially available genome-wide arrays and quality control was conducted for each cohort. Approximately 2.5 million SNPs were available after imputation and quality control. HbA1c was untransformed and was measured in % haemoglobin. The effect estimates for HbA1c were adjusted for age, sex, and study-specific covariates. 64 SNPs were found to be associated with HbA1c levels at genome-wide significance and explained 3.5% of the variance of HbA1c levels (**Table 5.2**).

BMI

Summary data of meta-analysis of GWA data for BMI was obtained from the Genetic Investigation of ANthropometric Traits (GIANT) consortium, in which they meta-analysed summary data from the GIANT consortium with the UK Biobank. The association of over 2 million directly genotyped or imputed SNPs were tested. Approximately 700,000 European participants from 101 cohorts were included in the analysis. Each SNP was tested for association with BMI under an additive model in a linear mixed model association, adjusted for age, sex, 10 principal components, recruitment centre, and batches (205).

364 SNPs were identified to be associated with BMI at GWAS threshold level of p-value< $1x10^{-8}$ with an LD< 0.001 (this threshold was used due to the large number of SNPs associated with BMI). These SNPs explained approximately 3.3% of the variance in BMI (**Table 5.2**). BMI data was downloaded from the following website

https://bit.ly/2CIpFrR. The beta estimates in this GWAS were inverse normal transformed and therefore BMI was interpreted as per one SD in which it was calculated to be 5.13 kg/m^2.

Statistical analysis

Causal association between serum 25-hydroxyvitamin D and metabolic syndrome components

A 2-sample MR approach was used to estimate the causal relationship between vitamin D and MetS components using the IVW approach. This was calculated using the six 25(OH)D-associated SNPs from the SUNLIGHT consortium and the corresponding SNP to MetS components associations from their respective GWAS. The results were expressed in effect estimate (β) per one SD increase in 25(OH)D lowering alleles. All six 25(OH)D-associated SNPs were used in a single analysis. To take multiple testing into account, a Bonferroni corrected significance level was computed, where the p-value of 0.05 was divided by 6 (p-value< 0.008).

Causal association between metabolic syndrome components and serum 25-hydroxyvitamin D

A 2-sample MR approach was also used for the causal inference between MetS components and vitamin D. The IVW estimate was calculated using beta estimates and SEs for each of the 364 BMI-associated SNPs, 211 HDL cholesterol-associated SNPs, 137 triglycerides-associated SNPs, 64 HbA1c-associated SNPs, 284 SBP-associated SNPs and 267 DBP-associated SNPs and their corresponding SNP-25(OH)D associations from the updated SUNLIGHT consortium. The results were expressed in effect estimate (β) per one SD increase/decrease in MetS component,

except for HbA1c in which the results were expressed as the effect estimate per one percent increase in HbA1c levels. The Bonferroni corrected significance threshold (p-value< 0.008) was considered as suggestive evidence for a potential association.

Mendelian randomisation assumptions

There are three assumptions that needs to be assessed to conduct a MR, 1) the IV is associated with the exposure of interest, 2) the IV is not associated with the confounders of the exposure and outcome relationship, and 3) the IV is associated with the outcome only through the exposure of interest. For the first assumption, I selected SNPs that were genome-wide associated with the exposure of interest (p-value< $5x10^{-8}$).

The second and third assumption is not fully testable, as not all confounders are known or measured (286). Since this is a summary MR analysis, I cannot assess the association between the SNPs with measured covariates, however this can be approximated by testing associations of the SNPs with other covariates from the GWAS catalog and Phenoscanner. Several methods were used to assess pleiotropy and therefore, test for compliance of the assumptions. The Cochran's Q and I-squared statistics were used to estimate the degree of heterogeneity. A high degree of heterogeneity between estimates of the exposure-outcome association could indicate violation of the no pleiotropy assumption in MR. Moreover, pleiotropy using MR-Egger regression, where the intercept provides a test for directional pleiotropy and the slope provides a potential causal estimate after adjusting for pleiotropic SNPs was investigated. Furthermore, a weighted median approach, which provides a potential causal effect estimate if at least 50% of the weight comes from valid SNPs (285), and a weighted modal test, which assumes that the causal effect estimate is consistent

with the true causal effect estimate, even if the majority of the instruments are invalid were also tested (301). MR-PRESSO test was used to identify horizontal pleiotropy based on outliers (302). All these sensitivity analyses measure the consistency of the effect estimates to the IVW method rather than the significance (286). If all sensitivity methods provide similar estimates, then a causal effect is more plausible (286).

Another sensitivity analysis was done using results from Phenoscanner to test whether the SNPs associated with MetS components were also associated with other phenotypes. A multivariable MR for the MetS components that were found to be causally associated with serum 25(OH)D was also done in a sensitivity analysis (303). Multivariable MR is an extension of the MR paradigm, in which IVs associated with more than one risk factor is used to estimate the causal association with the outcome of interest. This method is used when IVs are associated with several risk components, which can then evaluate the causal risk of the risk factor to the outcome, even if no variants were uniquely associated with it (303). Moreover, a power analysis was performed to estimate the minimum detectable magnitude of association for MetS components and vitamin D to obtain 80% with alpha set to 0.05 using a web-based application (https://sb452.shinyapps.io/power/).

Results

Potential causal association between vitamin D and metabolic syndrome components

Blood pressure

The present analysis had 80% power to detect an effect between vitamin D and blood pressure, assuming that 3% of the 25(OH)D variance was explained by the six SNPs (**Table 5.3**). There was no evidence of a causal association between serum 25(OH)D with either SBP or DBP (β: -0.34, 95%CI: -1.25 – 0.57 p-value: 0.46 and β: 0.03, 95%CI: -0.46 – 0.53, p-value: 0.90 respectively) (**Table 5.4**). There was no evidence of pleiotropy using either the MR-Egger method or MR-PRESSO (**Table 5.5**). However, the Cochran's Q-test reported some evidence of heterogeneity for SBP (β: -0.34, 95%CI: -1.25 – 0.57, p-value: 0.46, heterogeneity p-value: 0.002) (**Table 5.4**).

HDL cholesterol

This MR study also had 80% power to detect effect sizes greater/less than ±0.05 between HDL cholesterol and vitamin D (**Table 5.3**). There was weak evidence of a causal effect between low levels of vitamin D and HDL cholesterol levels (β: 0.09, 95%CI: -0.002 – 0.19, p-value: 0.06). No evidence of pleiotropy was found using any of the methods (**Table 5.4** and **Table 5.5**). Overall, the weighted median, weighted modal and MR-PRESSO all had similar estimates to the IVW, while MR-Egger did not (β: -0.005, 95%CI: -0.17 – 0.16, p-value: 0.95) (**Table 5.5 and Figure 5.1**).

Triglyceride

No evidence was found for a causal association between vitamin D and triglyceride levels (β: -0.004, 95%CI: -0.10 – 0.09, p-value: 0.93) (**Table 5.4**). No evidence of pleiotropy was found in any of the methods (**Table 5.4** and **Table 5.5**). However, the estimates were not consistent for the MR-Egger regression, weighted median, and weighted modal with the IVW method. This study analysis did not reach an effect estimate to obtain 80% power (**Table 5.3**).

HbA1c

The MR analysis of vitamin D and HbA1c levels achieved 68% power (**Table 5.3**). There was a 4% increase in HbA1c per 1% decrease in serum 25(OH)D (β: 0.04 95%CI: 0.003 – 0.08, p-value: 0.03) (**Table 5.4**). The MR Egger, weighted median, weighted modal and MR PRESSO all had consistent effect estimates to the IVW effect estimate (**Table 5.5 and Figure 5.1**). No evidence of pleiotropy or heterogeneity was found from the MR-Egger analysis or the Cochran's Q test (**Table 5.4** and **Table 5.5**). However, the result did not reach statistical significance at the Bonferroni threshold (p-value< 0.008)

BMI

The MR analysis of vitamin D and BMI achieved 80% power to detect a true association (**Table 5.3**). **Table 5.4** showed no evidence of a causal effect between vitamin D with BMI (β: 0.02, 95%CI: -0.02 – 0.06 p-value: 0.26). Moreover, there was no evidence of horizontal pleiotropy from MR-Egger, MR PRESSO (**Table 5.5**) or the Cochran's Q test for heterogeneity (**Table 5.4**).

Potential causal association between metabolic syndrome components and vitamin D

The MR analyses of MetS components with vitamin D had 80% power to detect effect sizes of ±0.06, assuming that 3% of the variance of 25(OH)D was explained by the six SNPs and the significance level was set to 0.05 (**Table 5.3**).

Blood pressure

No evidence was found for a causal association between DBP and SBP with serum 25(OH)D (β: 0.001, 95%CI: -0.001 – 0.002, p-value: 0.42 and β: 0.002, 95%CI: 0.000 – 0.001, p-value: 0.58) (**Table 5.6**). MR-Egger, weighted median weighted modal, and MR PRESSO all reported consistent effect estimates to the IVW effect estimate for SBP (**Table 5.7 and Figure 5.2**). Evidence of pleiotropy was found for SBP using the MR-Egger method (p-value of intercept <0.01). Evidence of heterogeneity was found for DBP (p-value: 0.03), but not for SBP (p-value: 0.23) (**Table 5.6**).

HDL cholesterol

Per one SD decrease in HDL cholesterol, vitamin D decreased by 1% (β: -0.01, 95%CI: -0.02 – -0.002, p-value: 0.01) (**Table 5.6**). However the results was not statistically significant at the Bonferroni correction threshold. No evidence of pleiotropy was found using the MR-Egger method, MR-PRESSO, or the Cochran's Q test (**Table 5.6** and **Table 5.7**). The effect estimates from the MR-Egger, weighted median, weighted modal and MR-PRESSO were all consistent with the IVW effect estimate (**Table 5.7 and Figure 5.2**). SNPs that were pleiotropic for secondary associations including waist-to-hip ratio, Alzheimers disease, T2D (N= 108) from the Phenoscanner were removed (Supplemental Table 5.2) and the MR analysis redone. Consistent results to the original IVW effect estimate was found (β: -0.01, 95%CI: -0.02 – 0.004, p-value: 0.18) (**Table 5.8**). The multivariable MR analysis was adjusted for low density

lipoprotein (LDL) cholesterol, triglycerides and total cholesterol and both HDL cholesterol adjusted for either LDL cholesterol or triglycerides reported consistent results (β: -0.01, 95%CI -0.02 – -0.001, p-value: 0.04 and β: -0.01, 95%CI: -0.02 – -0.002, p-value: 0.02, respectively) (**Table 5.9**). All adjustments were consistent with the IVW method in **Table 5.6**.

Triglyceride

No evidence was found for a causal effect between triglycerides and vitamin D (β: -0.001, 95%CI: -0.01 – 0.10, p-value: 0.88) (**Table 5.6**). There was no evidence of pleiotropy from the MR-Egger intercept; however, the Cochran's Q test reported evidence of heterogeneity (I^2=62%, heterogeneity p-value< 0.001) (**Table 5.6** and **Table 5.7**).

HbA1c

Table 5.6 reports that per 1% increase in HbA1c, vitamin D increased by 5% (β: 0.05, 95%CI: 0.01 – 0.09, p-value< 0.01). The results remained statistically significant after adjusting for multiple testing. The effect estimates from the MR-Egger, weighted median, weighted modal and MR-PRESSO were all consistent with the IVW effect estimate (**Table 5.7 and Figure 5.2**). There was no evidence of pleiotropy from the intercept of the MR-Egger analysis, MR-PRESSO, or the Cochran's Q test (**Table 5.6** and **Table 5.7**). SNPs that were found to be pleiotropic (N= 29 out of 64 SNP) (Supplemental Table 5.3) were removed and the MR analysis reported no evidence of a causal association between HbA1c and vitamin D (β: 0.02, 95%CI: -0.03 – 0.07, p-value: 0.47) (**Table 5.8**). Another sensitivity analysis was performed where a multivariable MR analysis was adjusted for fasting glucose and T2D and both

adjustments reported consistent results to the IVW method (β: 0.05, 95%CI: 0.005 – 0.10, p-value: 0.03 and β: 0.06, 95%CI: 0.01 – 0.10, p-value: 0.01, respectively) (**Table 5.10**).

BMI

A 1% decrease in vitamin D was found per one SD increase in BMI (β: -0.01, 95%CI: -0.03 – -0.002, p-value: 0.03), which was not statistically significant at the Bonferroni correction threshold. (**Table 5.6**). There was no evidence of pleiotropy in the intercept from the MR-Egger analysis or MR-PRESSO (**Table 5.7)**. There was also no evidence of heterogeneity from the Cochran Q statistics (**Table 5.6**). The potential causal association between BMI and vitamin D remained consistent after adjusting for pleiotropy in the MR-Egger analysis. The effect was also consistent in the weighted median, the weighted modal, and MR-PRESSO analyses (**Table 5.7 and Figure 5.2**).

After the removal of pleiotropic SNPs (N=73 (Supplemental Table 5.4), the results remained consistent (**Table 5.8**) with the effect estimates from **Table 5.6**.

Discussion

In this Chapter, I tested the bi-directional association between vitamin D and 5 MetS components. A potential causal association between vitamin D lowering alleles and HbA1c levels was reported, in which 1% decrease in 25(OH)D was associated with 4% higher levels of HbA1c. However, after Bonferroni corrections, the association was not significant. There was no evidence for a causal link between vitamin D with any of the remaining components. In the reverse association, high BMI and low levels of HDL cholesterol potentially reduced the levels of 25(OH)D. While high levels of HbA1c and high SBP potentially increased levels of 25(OH)D. No evidence for a causal association was found for DBP or triglyceride levels. Only the association between high levels of HbA1c and vitamin D remained significant at the Bonferroni correction threshold.

Blood Pressure

There was no evidence of a causal association between low levels of vitamin D with either SBP or DBP. The current MR study was powered to detect a minimum detectable effect size of 0.02 per 1 SD increase in either SBP or DBP. Furthermore, no evidence of a causal association was found between SBP or DBP increasing alleles with serum 25(OH)D in the other direction (p-value>0.05).

Conflicting results have been found between vitamin D and blood pressure across study types. Observational studies reported a significant association between low levels of vitamin D and the risk of hypertension (186). A MR study also reported a significant association between 25(OH)D with hypertension and DBP, but no evidence

of a causal association was found between 25(OH)D and SBP (180). The MR study by Vimaleswaran *et al.* used data from the D-CarDia study (180) and reported a significant causal association between 25(OH)D and DBP (per 10% increase in 25(OH)D β:-0.29, 95%CI: -0.52 – -0.07). Although the present study used a different GWAS for the MR analysis, in which it had a larger sample size (N=488,000 for the UK Biobank compared to N=108,173 D-CarDia) and used more IV (6 SNPs) compared to the study by Vimaleswaran *et al.* (3 SNPs), this present study reported no evidence of a causal relationship between vitamin D and blood pressure. One limitation to the study by Vimaleswaran *et al.* is that they did not test for pleiotropy using any of the methods used in this present study. Therefore, the potential causal association between vitamin D and DBP from Vimaleswaran *et al.* could be a false positive. A systematic review and meta-analysis on 46 RCTs on 4,541 participants for a minimum of 4 weeks, supported the results from this present study, which reported no association between vitamin D supplementation and blood pressure (ES: 0.0, 95%CI: -0.8 – 0.8 and ES: -0.1, 95%CI: -0.6 – 0.5 for SBP and DBP respectively) (187). This non-significant result in the meta-analysis could be due to the small number of hypertensive patients at baseline (10%) that could have benefited from the intervention of vitamin D supplementation. Moreover, some of these patients could have been taking antihypertensive medication, which can interact with vitamin D, causing hypercalcemia, which is associated with high levels of serum vitamin D, nullifying the results (187,304).

No studies have investigated the potential causal effects blood pressure has on vitamin D. Results from the present study reported no causal association between high levels of SBP and DBP with levels of vitamin D. These result supports previous RCTs and MR studies, but not observational studies. The difference in results between

observational studies and MR and RCT could be due to residual confounding that was not taken into account in observational studies including antihypertensive drugs.

HDL cholesterol and triglyceride levels

There was no evidence of a causal association between low levels of vitamin D with HDL cholesterol and triglyceride levels. This MR study was powered to detect a minimum detectable effect size of 0.06 per 25 nmol/L increase in 25(OH)D levels. On the other hand, a potential causal association was found between HDL cholesterol lowering alleles and low levels of 25(OH)D. No evidence of a causal association was found for high levels of triglyceride levels with 25(OH)D. In the opposite direction, the MR study was powered to detect a minimum detectable effect size of 0.05 per 1 SD increase in either HDL cholesterol or triglyceride levels.

A previous bi-directional MR study done by Ooi *et al.* investigated the association between vitamin D and HDL cholesterol and reported a positive causal association between low levels of serum 25(OH)D and HDL cholesterol (change in HDL cholesterol per 50% decrease in 25(OH)D: -6.0, 95%CI: -10 – -2.3, p-value: 0.001) (191). This study performed a 1-sample MR on vitamin D-associated synthesis SNPs only and suffered from weak instrument bias, in which their SNPs for vitamin D explained only 1.0% of the variance of vitamin D. Moreover, the study was done on 85,363 white Danish individuals in which serum 25(OH)D for approximately 6,000 participants were obtained from plasma that was collected between 1981-1983. This long duration between collection and laboratory analysis might have affected the level of serum vitamin D (305). There is a chance that the result by Ooi *et al.* could be false positive, since pleiotropy was not assessed in the study by Ooi *et al.* (191).

In the present analysis a larger sample size (N=95,000) was analysed with more variance explained for vitamin D (2.7%). Power analysis reached 80% power for HDL cholesterol. Furthermore, sensitivity analyses for pleiotropy were consistent with the IVW effect estimate. Therefore, it is more likely that the results in this present analysis of no causal association between low levels of vitamin D and HDL cholesterol to be more accurate than the study by Ooi *et al.* (191). A RCT on 422 participants that were given 20,000 IU/week of vitamin D supplementation or placebo for 4 months supported the results of this thesis and reported no difference in HDL cholesterol levels (306).

Ooi *et al.* also reported that halving of HDL cholesterol was associated with a 20% increase in plasma 25(OH)D (191), which contradicted the results from their observational analysis, where halving of HDL cholesterol was associated with a 1.5% decrease in 25(OH)D (191). The study by Ooi *et al.* speculated that the causal association they found with low HDL cholesterol and high serum 25(OH)D could have been driven by the association between high remnant cholesterol and low 25(OH)D levels (191) and therefore be a false positive result. The present result reported a 1% decrease in serum 25(OH)D per one SD decrease in HDL cholesterol. A multivariable MR analysis was adjusted for other lipoproteins in a sensitivity analyses and showed that the association between HDL cholesterol and 25(OH)D remained consistent, even after adjusting for LDL cholesterol, triglycerides, and total cholesterol. A mechanism explaining the relationship between vitamin D and HDL cholesterol is that low levels of vitamin D is associated with insulin resistance, which affects lipoprotein metabolism by decreasing HDL cholesterol levels (307). No studies investigated the mechanism as to how HDL cholesterol could affect vitamin D levels.

HbA1c

Results from the present study showed a bi-directional causal relationship between HbA1c and vitamin D. The findings showed that low levels of 25(OH)D was associated with high HbA1c levels (β: 0.04, 95%CI: 0.003 – 0.08). Moreover, high levels of HbA1c was also associated with high 25(OH)D levels (β: 0.05, 95%CI: 0.01 – 0.09). However, after correcting for multiple testing, only the association between high levels of HbA1c with 25(OH)D remained significant. This MR study was powered to detect a minimum detectable effect size of 0.06 per 25 nmol/L increase in 25(OH)D levels, while in the reverse direction, the MR study was powered to detect a minimum detectable effect size of 0.05 per 1% increase in HbA1c levels.

Only one MR study by Ye *et al.* has previously investigated the causal association between vitamin D and HbA1c, in which they reported no causal association (β:0.01, 95%CI: -0.04 – 0.05, p-value: 0.8, per one SD reduction in 25(OH)D) (276). Their null results could be due to the lack of power to detect small effects. The study by Ye *et al.* had approximately one third (N=46,368) the sample size of the present study (N=123,665). Moreover, they also used half the amount of IV (3 SNPs) compared to the present study (6 SNPs). No tests of pleiotropy was conducted in the study by Ye *et al.* using any of the methods used in this present study. After adjustments for multiple testing in this study, no potential association was found between low levels of vitamin D and HbA1c. A systematic review and meta-analysis of 23 RCT studies analysing the effects of vitamin D supplementation on HbA1c reported reduced HbA1c levels with vitamin D supplementation among T2D patients (SMD: -0.25, 95%CI: -0.45 – 0.05, I^2=75.5%) (182), supporting the result in the present study.

One mechanism explaining this association, is the relationship between vitamin D and insulin. Vitamin D is essential for normal insulin release in response to glucose through

the regulation of calcium. Low levels of vitamin D are associated with decreased pancreatic insulin secretion (108,109) and low levels of insulin are in turn associated with high levels of glucose and high levels of HbA1c .

No previous studies investigated the potential causal association between HbA1c and vitamin D in a MR framework. Results from the present study reported a potential causal association between high levels of HbA1c with increased levels of vitamin D. However, the result from this study did not support the direction of association in observational studies, where high HbA1c is associated with low levels of vitamin D (308,309). One explanation could be the presence of horizontal pleiotropy, which can induce false positive causal relationships (302). Sensitivity analysis, using MR-Egger (β: 0.05, 95%CI: -0.04 – 0.14) and MR-PRESSO (β: 0.05, 95%CI: 0.02 – 0.12) found no evidence of pleiotropy. Effect estimates remained consistent, and statistically significant, after adjusting for pleiotropy and multivariable adjustment (β: 0.06, 95%CI: 0.01 – 0.10, after adjusting for BMI). However, when pleiotropic SNPs, based on results from Phenoscanner, were removed from the analysis, no evidence of a causal association was found between HbA1c and vitamin D (β: 0.02, 95%CI: -0.03 – 0.07). Therefore, the potential causal association between high levels of HbA1c and vitamin D could possibly be a false positive. Furthermore, no mechanism has been found to support the relationship between HbA1c and vitamin D.

BMI

No evidence of a causal association was found between low levels of vitamin D and BMI. This study however, was underpowered to detect small effects (minimum detectable effect estimate: 0.02). In the reverse direction, a potential causal

association was found between BMI increasing alleles and low levels of 25(OH)D. This MR study was powered to detect a minimum detectable effect size of 0.06 per 25 nmol/L increase in 25(OH)D levels. However, the result was not significant after correcting for multiple testing.

A previous MR by Vimaleswaran *et al.* on the potential causal association between vitamin D and BMI supported the results of the present analysis of no causal association between low levels of vitamin D and BMI (176). Two RCTs also supported the findings in the present study when they assessed the effect of vitamin D supplementation on adiposity and reported no changes in body measurements with vitamin D_3 supplementations (310,311).

The study by Vimaleswaran *et al.* also assessed the potential causal association between BMI and vitamin D using 12 BMI-associated SNPs to create a weighted score for each of the 21 individual-level population-based studies from North America and Europe. Results from the IV ratio reported a 10% increase in BMI was causally associated with a 4.2% decrease in 25(OH)D concentration, which supports the results of the present study of an inverse association, which did not survive multiple testing. The present study is however based on a much larger set of IV, 364 BMI-associated SNPs which explain approximately 3.3% of the variance of BMI compared to 0.97% of variance explained in the study by Vimaleswaran *et al.* (176).

It has been suggested that obese individuals increase the sequestration of vitamin D in the adipose tissues, which decrease the concentration of 25(OH)D available in the blood (251). Moreover, differences in lifestyles can also contribute to lower 25(OH)D, where obese individuals may have lower intake of vitamin D rich foods and expose less skin to the sun (179).

The largest GWAS consortia were used for the MR analyses of MetS components and vitamin D to investigate the potential causal association between vitamin D and MetS components and vice versa, which is one of the main strengths of the present thesis. These consortia gave us enough power to detect causal associations for most of the MetS components. This is the first study that investigates the bi-directional association between vitamin D with HbA1c, triglyceride levels and blood pressure. Another strength of this study is the study design, where the direction of association between vitamin D and MetS components was disentangled using a bi-directional MR approach. Three assumptions needs to be met to have a valid MR result. The first is that the SNPs are associated with the exposure of interest. This was done by selecting SNPs that were significantly associated with the metabolic factor of interest at the genome-wide threshold of $< 5x10^{-8}$. In regards to the SNPs associated with vitamin D, these SNPs also supported a biological significance in which they are located nearby genes that are involved in the synthesis and metabolism of vitamin D.

The next two assumptions in MR cannot be truly tested. However, the range of analyses in this work include MR sensitivity analyses and investigations for pleiotropy which were not reported in the previous MR studies discussed above. Pleiotropy was tested using multiple methods including the Cochran's Q test, MR-Egger regression analysis, weighted median, weighted modal, and MR-PRESSO. Some evidence of pleiotropy was found for HbA1c, which could have given a false positive result.

To conclude, this study showed evidence of a causal association between vitamin D and HbA1c, in which low levels of vitamin D suggested an increase in HbA1c levels. The significant result obtained in the opposite direction can be interpreted as presence of pleiotropy on the associations which fall beyond the current MR methodologies to

uncover complex causal relationships. Evidence also suggest a potential causal association between high BMI and low HDL cholesterol with low levels of vitamin D. None of these three factors remained significant after Bonferroni corrections. Further studies need to be conducted to understand mechanisms of association between these MetS factors with vitamin D.

Table 5.1. Summary of previous Mendelian randomisation studies on causal association between vitamin D and metabolic syndrome components

Metabolic factor	Published MR	Method used	Number of SNPs/IV	Adjustment	Sample size/Population	Result
BMI	PMID: 23393431 **Bi-directional MR**	A weighted BMI score was used An unweighted allele score was used for VD IV ratio method was used for MR, where the meta-analysed BMI allele score with 25(OH)D was divided by the association of BMI allele score with BMI	12 SNPs BMI 4 SNPs VD R^2: 0.97% BMI R^2:1.9% VD	BMI and VD as outcomes were natural log transformed Models with BMI as outcome were adjusted for age, sex, geographical site and or PC from population stratification analysis Models with VD as outcome were adjusted for age, sex, geographical site, month of blood collection, laboratory batch and/or PC from population stratification analysis	42,024 (bi-directional) D-CarDia Replication 123,864 (BMI) GIANT (only in VD-BMI direction)	10% ↑ BMI caused 4.2%↓ of vitamin D (p=0.005) and not vice versa (p≥0.08). *Doesn't report association for all 4 SNPs of VD with BMI, but the metabolism and synthesis scores
SBP/DBP	PMID: 24974252	The IV ratio was used for the first part, where the meta-analysed VD allele score with SBP/DBP/HTN was divided by the association of VD allele score with VD	4 SNPs VD R^2:1.9%	VD was natural log transformed Additive models with SBP, DBP and HTN as outcome were adjusted for age, age-squared, BMI, sex, geographical region, and PC Additive models with VD as outcome were adjusted for age, age-squared, BMI, sex, geographical region, month of blood sample collection, laboratory batch and PC	99,582 D-CarDia 146,581 D-CarDia+ ICBP 142,255 D-CarDia + CHARGE + BPGen	Each 10% increase in genetically instrumented VD vitamin D, there was lower DBP (-0.29, 95%CI:-0.52 - -0.07, p:0.01) as well as a 8% decrease odds of HTN No significant association was found for SBP(β:-0.37, 95%CI: -0.73 – 0.003, p:0.05)
HDL/TG	PMID: 23460889	MR was done using the 2SLS Bonferroni corrected pval=0.005	Filaggrin mutation increases VD R^2:8% for VD	25(OH)D were log2 transformed for the first stage and HDL, and TG were log transformed Regressions were adjusted for gender, age, study cohort, season of	11,983 Monica10, Inter99, and Health2006 from Danish Central Personal Register	Per doubling of vitamin D reported **NO** association was found with HDL and TG AFTER adjusting for Bonferroni in 2SLS

Metabolic factor	Published MR	Method used	Number of SNPs/IV	Adjustment	Sample size/Population	Result
				blood sample, education, fish intake, PA, smoking, alcohol, and BMI		
HDL (2)	PMID: 25065375 **Bi-directional MR**	An allele score was constructed 2SLS MR was used	4 VD SNPs: rs7944926, rs11234047, rs10741657, rs12794714 R^2 for VD: 1% R^2 for HDL: 0.4%	The lipids were log2transformed and therefore it was per doubling/per halving of the lipoproteins For VD as an outcome they used a 50% decrease	25(OH)D:31,435 HDL: 17,756 Two Danish cohort studies	Halving of HDL was genetically associated with higher % plasma VD levels (20%, 95%CI: 7.4% - 34%, p:0.003) A 50% decrease in VD was genetically associated with lower HDL (-6%, 95%CI:-10% - -2.3%, p: 0.001)
HbA1c	PMID: 25281353 Outcome: T2D Secondary outcome: HbA1c	Assessed association of each SNP with VD and other baseline factors Examined association between SNPs and T2D MR between VD and T2D using the Bayesian Likelihood-based method	4 VD SNPs: rs12785878, rs10741657, rs4588, rs17217119 R^2 for VD: 3.6%	Additive model adjusted for age, sex, and season of blood draw for VD Additive model adjusted for age, sex, and BMI for T2D	104,488 T2D cases and controls Ely, EPIC-Norfolk, EPIC-InterAct, DIAGRAM consortium 5449 noncases for 25(OH)D measurement 46,368 for HbA1c from MAGIC consortia	VD SNPs were **not** causally associated with T2D (OR per 1SD reduction on 25(OH)D: 1.01, 95%CI: 0.75 – 1.36, p:0.94) or HbA1c (β:0.01, 95%CI: -0.04 – 0.05, p:0.8, per 1 SD reduction in 25(OH)D)

VD: vitamin D, MR: Mendelian randomisation, BMI: body mass index, PC: principal components, SBP: systolic blood pressure, DBP: diastolic blood pressure, HTN: hypertension, SNP: single nucleotide polymorphism, HDL: high density lipoprotein, TG: triglyceride, HbA1c, glycated haemoglobin A1c

Table 5.2. Variance explained for each of the metabolic syndrome components at linkage equilibrium thresholds of 0.1 (0.001 for BMI)

Metabolic component	LD threshold	Number of SNPs	Variance explained
Blood Pressure			
SBP	0.1	284	4.3%
DBP	0.1	267	3.8%
Lipids			
HDL-C	0.1	211	13.9%
TG	0.1	137	8.7%
HbA1c	0.1	64	3.5%
BMI	0.001	364	3.3%

SBP: Systolic blood pressure, DBP: Diastolic blood pressure, HDL: high density lipoprotein-cholesterol, TG: triglycerides, HbA1c: glycated haemoglobin A1c, BMI: body mass index

Table 5.3. Minimum detectable effect estimate to achieve 80% statistical power

Metabolic component	Study	Sample size	Minimum detectable β*
Blood Pressure			
SBP	UK Biobank	318,417	0.025
DBP	UK Biobank	488,000	0.023
Lipids			
HDL-C	GLGC	95,000	0.05
TG	GLGC	95,000	0.05
HbA1c	MAGIC	123,665	0.05
BMI	GIANT	700,000	0.02
Vitamin D	SUNLIGHT	79,366	0.06

*Assume 80% power, 5% alpha level, and 3% of the 25(OH)D variance is explained by the six SNPs.
SBP: Systolic blood pressure, DBP: Diastolic blood pressure, HDL: high density lipoprotein-cholesterol, TG: triglycerides, HbA1c: glycated haemoglobin A1c, BMI: body mass index

Table 5.4. Mendelian randomisation of the causal association between serum 25-hydroxyvitamin D and metabolic syndrome components

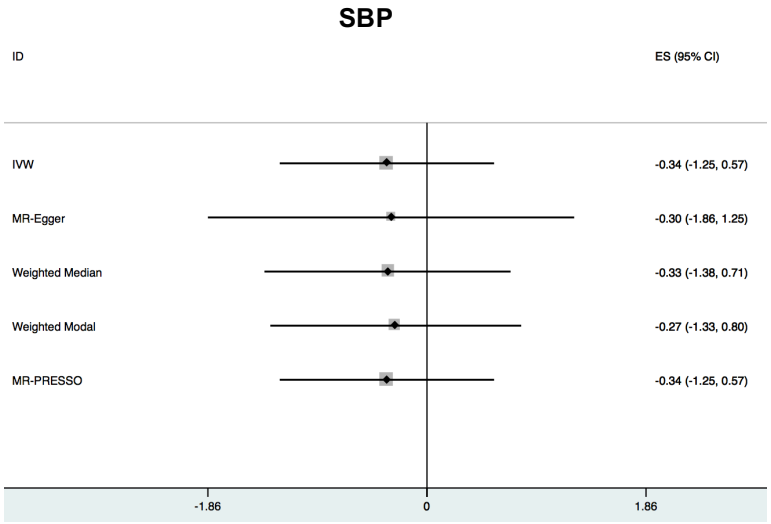
	Inverse-variance weighted estimate				
Metabolic component allele score	**β**	**95%CI**	**p-value**	**I^{2*}**	**heterogeneity p-value**
Blood Pressure					
SBP	-0.34	-1.25 – 0.57	0.46	77	**0.002**
DBP	0.03	-0.46 – 0.53	0.9	23	0.27
Lipids					
HDL-C	0.09	-0.002 – 0.19	0.06	0	0.51
TG	-0.004	-0.10 – 0.09	0.93	0	0.68
HbA1c	0.04	0.003 – 0.08	**0.03**	0	1.00
BMI	0.02	-0.02 – 0.06	0.26	0	0.65

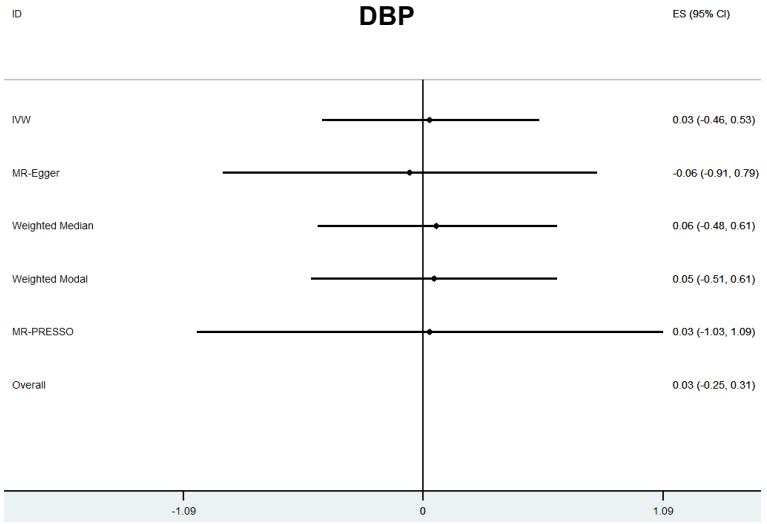
β per 1 SD increase in the natural log transformed serum 25-hydroxyvitamin D
SBP: systolic blood pressure, DBP: diastolic blood pressure, HDL: high density lipoprotein-cholesterol, TG: triglycerides, HbA1c, glycated haemoglobin, BMI: body mass index
*This test assesses the potential violation of the second assumption of Mendelian randomisation, and is performed by examining the null hypothesis that the association of each SNP with vitamin D is proportional to its association with each metabolic syndrome components.

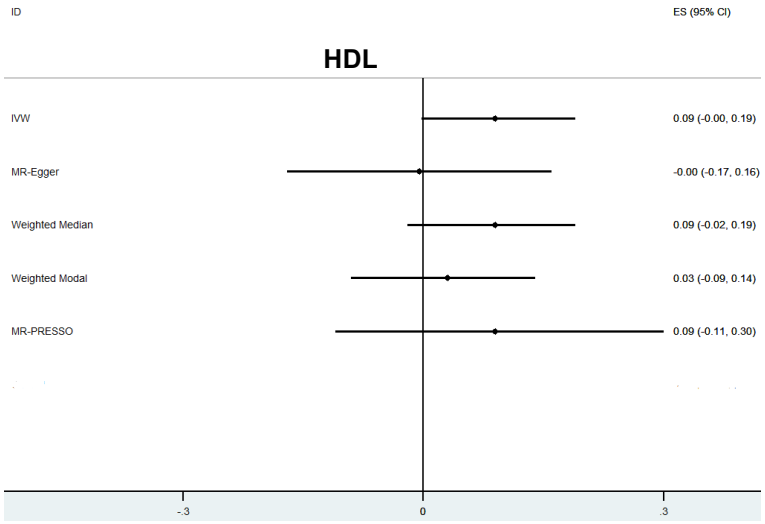
Table 5.5. Mendelian randomisation estimates between multi-SNP risk scores of continuous 25(OH)D and metabolic syndrome components using MR-Egger's regression, weighted median, weighted modal, and MR-PRESSO approach

		MR Egger		Weighted Median		Weighted Modal		MR PRESSO	
MetS component	**Intercept p-value**	**Slope β (95% CI)**	**p-value**	**β (95%CI)**	**p-value**	**β (95%CI)**	**p-value**	**β (95%CI)**	**p-value**
Blood Pressure									
SBP	0.96	-0.30 (-1.86 – 1.25)	0.70	-0.33 (-1.38 – 0.71)	0.53	-0.27 (-1.33 – 0.80)	0.62	-0.34 (-3.04 – 2.14)	0.51
DBP	0.80	-0.06 (-0.91 – 0.79)	0.89	0.06 (-0.48 – 0.61)	0.81	0.05 (-0.51 – 0.61)	0.61	0.03 (-1.03 – 1.09)	0.91
Lipid									
HDL-C	0.15	-0.005 (-0.17 – 0.16)	0.95	0.09 (-0.02 – 0.19)	0.12	0.03 (-0.09 – 0.14)	0.62	0.09 (-0.11 – 0.30)	0.12
TG	0.69	0.02 (-0.14 – 0.19)	0.78	0.001 (-0.11 – 0.11)	0.99	0.01 (-0.10 – 0.12)	0.85	-0.004 (-0.15 – 0.14)	0.91
HbA1c	0.62	0.05 (-0.01 – 0.12)	0.10	0.04 (0.005 – 0.08)	0.03	0.05 (0.004 – 0.09)	0.03	0.04 (0.001 – 0.08)	<0.01
BMI	0.27	-0.01 (-0.08 – 0.06)	0.81	0.004 (-0.04 – 0.05)	0.85	0.004 (-0.04 – 0.05)	0.83	0.02 (-0.04 – 0.08)	0.24

MetS: metabolic syndrome, SBP: Systolic blood pressure, DBP: Diastolic blood pressure, HDL: high density lipoprotein-cholesterol, TG: triglycerides, HbA1c: glycated haemoglobin A1c, BMI: body mass index


SBP
ID
ES (95% CI)
IVW
-0.34 (-1.25, 0.57)
MR-Egger
-0.30 (-1.86, 1.25)
Weighted Median
-0.33 (-1.38, 0.71)
Weighted Modal
-0.27 (-1.33, 0.80)
MR-PRESSO
-0.34 (-1.25, 0.57)
-1.86
0
1.86


DBP
ID
ES (95% CI)
IVW
0.03 (-0.46, 0.53)
MR-Egger
-0.06 (-0.91, 0.79)
Weighted Median
0.06 (-0.48, 0.61)
Weighted Modal
0.05 (-0.51, 0.61)
MR-PRESSO
0.03 (-1.03, 1.09)
Overall
0.03 (-0.25, 0.31)
-1.09
0
1.09


HDL
ID
ES (95% CI)
IVW
0.09 (-0.00, 0.19)
MR-Egger
-0.00 (-0.17, 0.16)
Weighted Median
0.09 (-0.02, 0.19)
Weighted Modal
0.03 (-0.09, 0.14)
MR-PRESSO
0.09 (-0.11, 0.30)
-.3
0
.3


TG
ID
ES (95% CI)
IVW
-0.00 (-0.10, 0.09)
MR-Egger
0.02 (-0.14, 0.19)
Weighted Median
0.00 (-0.11, 0.11)
Weighted Modal
0.01 (-0.10, 0.12)
MR-PRESSO
-0.00 (-0.15, 0.14)
Overall
0.00 (-0.05, 0.06)
-.19
0
.19

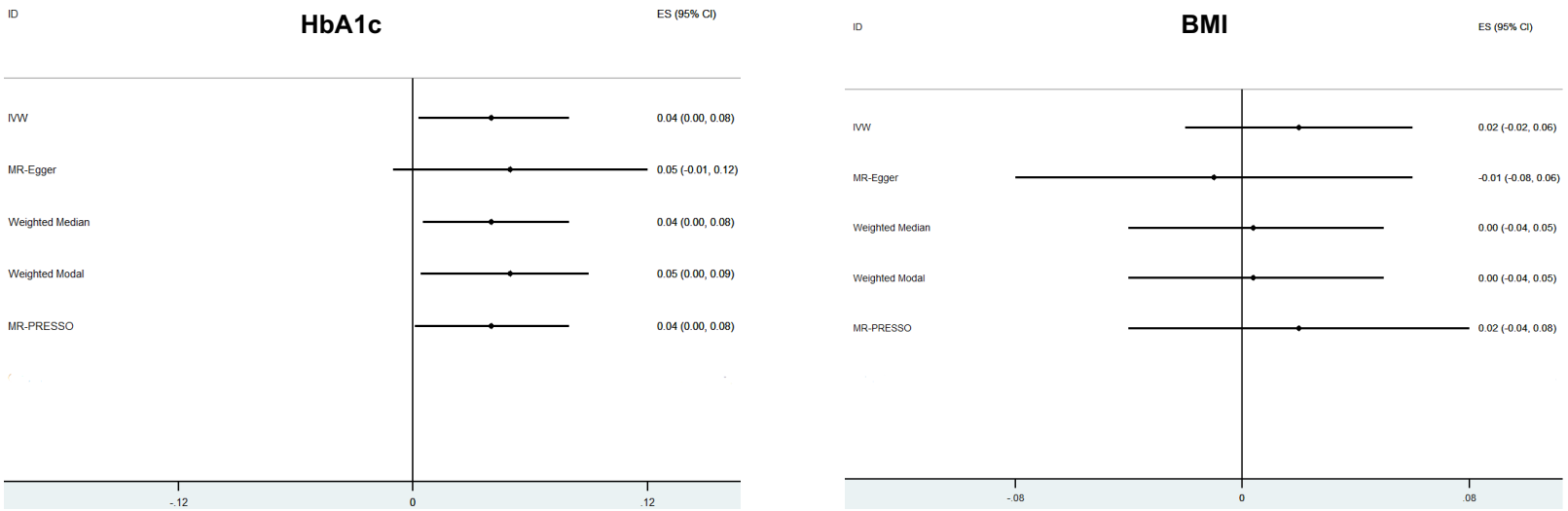


Figure 5.1. Mendelian randomisation plots between multi-SNP risk scores of 25-hydroxyvitamin D and metabolic syndrome components using inverse-variance weighted method, MR Egger, weighted median, weighted modal, and MR PRESSO approach

Table 5.6. Mendelian randomisation estimates between multi-SNP risk scores of continuous metabolic syndrome components and serum 25-hydroxyvitamin D using the inverse-variance weighted approach

	Inverse-variance weighted estimate				
MetS component allele score	**β**	**95%CI**	**p-value**	**I^{2*}**	**heterogeneity p-value**
Blood Pressure					
SBP	0.0002	0.000 - 0.001	0.58	7	0.18
DBP	0.001	-0.001 - 0.002	0.42	15	**0.03**
Lipids					
HDL-C	-0.010	-0.020 - -0.002	**0.01**	18	0.16
TG	-0.001	-0.010 - 0.100	0.88	62	**<0.001**
HbA1c	0.050	0.010 - 0.090	**<0.01**	17	0.20
BMI	-0.010	-0.030 - -0.002	**0.03**	6	0.21

R^2 for BMI was <0.001

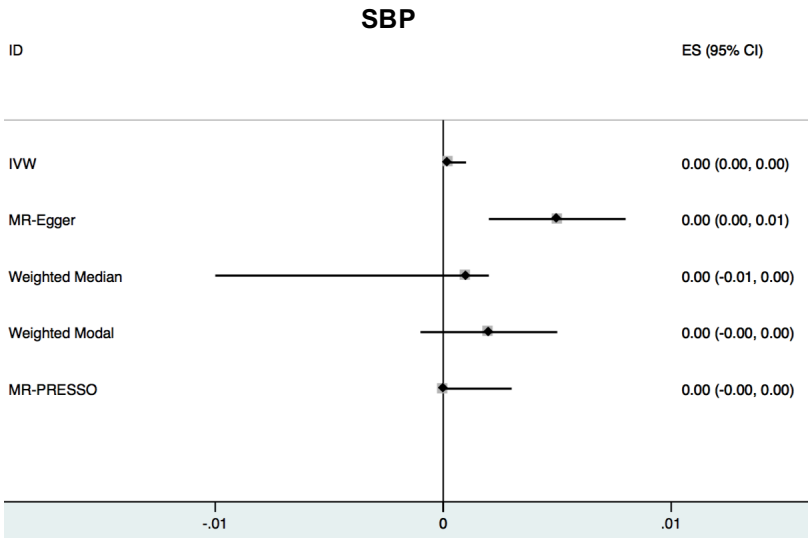
MetS: metabolic syndrome, SBP: systolic blood pressure, DBP: diastolic blood pressure, HDL: high density lipoprotein-cholesterol, TG: triglycerides, HbA1c, glycated haemoglobin, BMI: body mass index

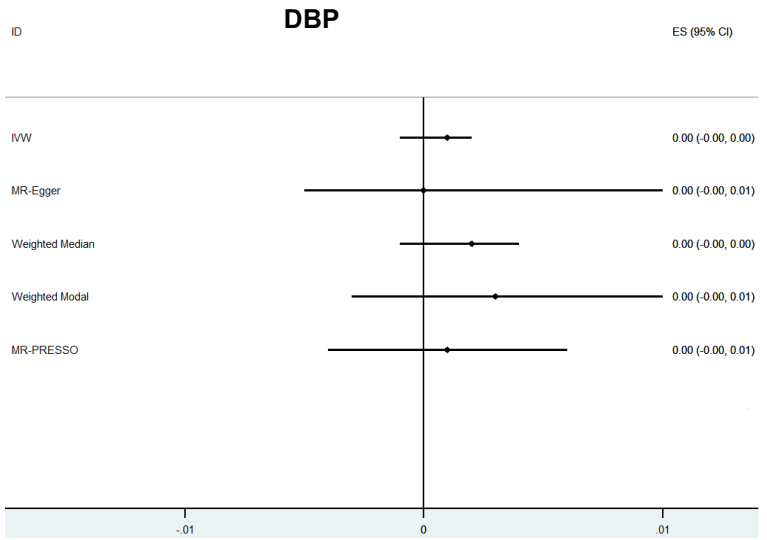
*This test assesses the potential violation of the second assumption of Mendelian randomisation, and is performed by examining the null hypothesis that the association of each SNP with vitamin D is proportional to its association with each metabolic syndrome components.

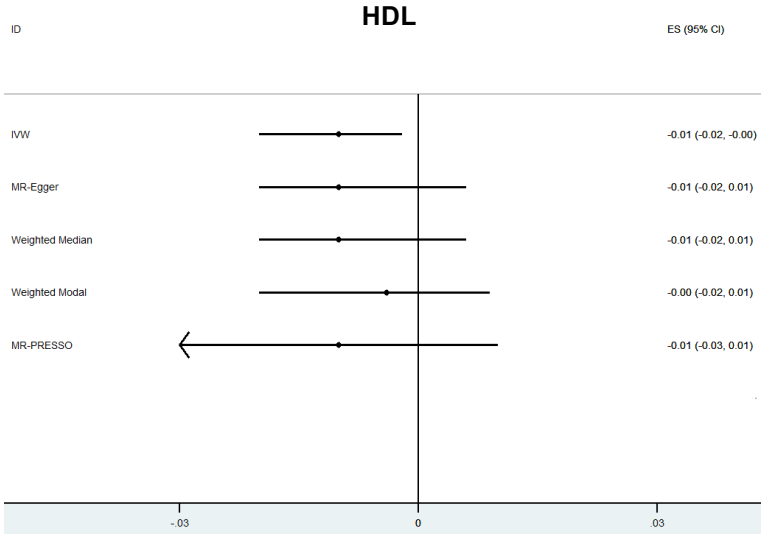
Table 5.7. Mendelian randomisation estimates between multi-SNP risk scores of metabolic syndrome components and 25-hydroxyvitamin D calculated using MR-Egger's regression, weighted median, weighted modal, and MR-PRESSO approach

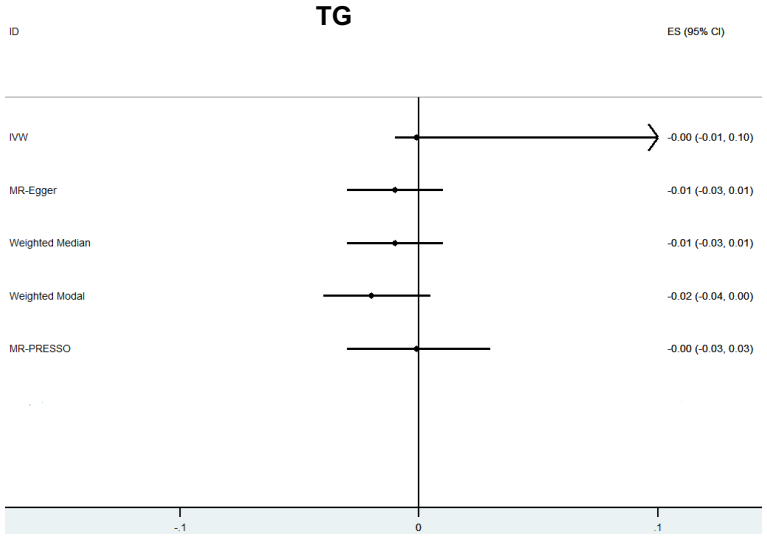
		MR Egger		Weighted Median		Weighted Modal		MR PRESSO	
MetS components	**Intercept p-value**	**Slope β* (95% CI)**	**p-value**	**β*(95%CI)**	**p-value**	**β (95%CI)**	**p-value**	**β (95%CI)**	**p-**
Blood Pressure									
SBP (284 SNPs)	<0.01	0.005 (0.002 – 0.008)	<0.001	0.001 (-0.01 – 0.002)	0.30	0.002 (-0.001 – 0.005)	0.28	0.000 (0.000 – 0.000)	
DBP (267 SNPs)	0.92	0.00 (-0.005 – 0.01)	0.91	0.002 (-0.001 – 0.004)	0.15	0.003 (-0.003 – 0.01)	0.33	0.001 (-0.004 – 0.006)	
Lipids									
HDL-C (210 SNPs)	0.71	-0.01 (-0.02 – 0.006)	0.27	-0.01 (-0.02 – 0.006)	0.26	-0.004 (-0.02 – 0.009)	0.52	-0.01 (-0.03 – 0.01)	
TG (137SNPs)	0.25	-0.01 (-0.03 – 0.01)	0.28	-0.01 (-0.03 – 0.01)	0.23	-0.02 (-0.04 – 0.005)	0.11	-0.001 (-0.03 – 0.03)	
HbA1c (64 SNPs)	0.98	0.05 (-0.04 – 0.14)	0.27	0.06 (0.005 – 0.12)	0.03	0.06 (-0.03 – 0.15)	0.20	0.05 (0.02 – 0.12)	
BMI (364 SNPs)	0.82	-0.02 (-0.07 – 0.03)	0.39	-0.02 (-0.04 – 0.004)	0.13	-0.02 (-0.08 – 0.04)	0.61	-0.015 (-0.04 – 0.01)	

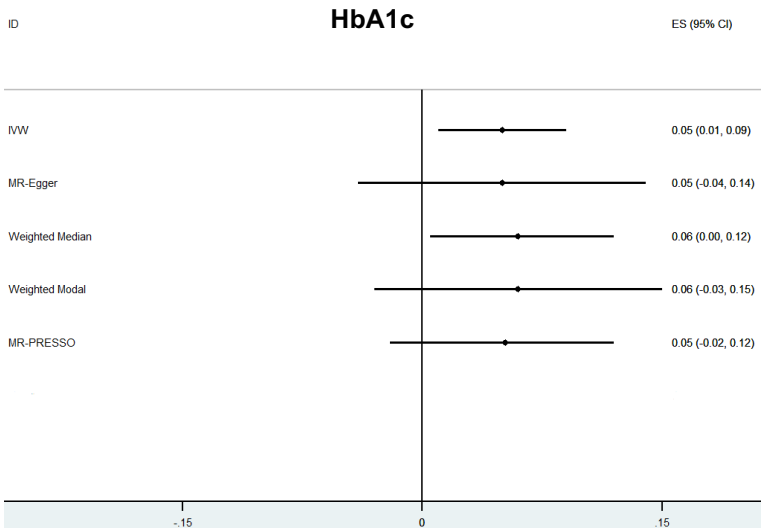
MetS: metabolic syndrome, SBP: Systolic blood pressure, DBP: Diastolic blood pressure, HDL: high density lipoprotein-cholesterol, TG: triglycerides, HbA1c: glycated haemoglobin A1c, BMI: body mass index


SBP
ID
ES (95% CI)
IVW
0.00 (0.00, 0.00)
MR-Egger
0.00 (0.00, 0.01)
Weighted Median
0.00 (-0.01, 0.00)
Weighted Modal
0.00 (-0.00, 0.00)
MR-PRESSO
0.00 (-0.00, 0.00)
-.01
0
.01


DBP
ID
ES (95% CI)
IVW
0.00 (-0.00, 0.00)
MR-Egger
0.00 (-0.00, 0.01)
Weighted Median
0.00 (-0.00, 0.00)
Weighted Modal
0.00 (-0.00, 0.01)
MR-PRESSO
0.00 (-0.00, 0.01)
-.01
0
.01


HDL
ID
ES (95% CI)
IVW
-0.01 (-0.02, -0.00)
MR-Egger
-0.01 (-0.02, 0.01)
Weighted Median
-0.01 (-0.02, 0.01)
Weighted Modal
-0.00 (-0.02, 0.01)
MR-PRESSO
-0.01 (-0.03, 0.01)
-.03
0
.03


TG
ID
ES (95% CI)
IVW
-0.00 (-0.01, 0.10)
MR-Egger
-0.01 (-0.03, 0.01)
Weighted Median
-0.01 (-0.03, 0.01)
Weighted Modal
-0.02 (-0.04, 0.00)
MR-PRESSO
-0.00 (-0.03, 0.03)
-.1
0
.1



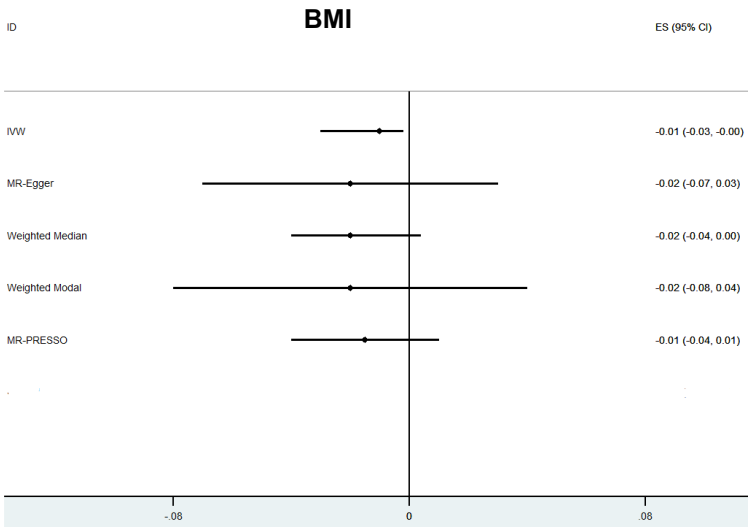


Figure 5.2. Mendelian randomisation plots between multi-SNP risk scores of metabolic syndrome components with 25-hydroxyvitamin D using inverse-variance weighted method, Egger's regression, weighted median, weighted modal, and MR-PRESSO approach

Table 5.8. Mendelian randomisation estimates between multi-SNP risk scores of metabolic syndrome components and 25-hydroxyvitamin D calculated using the inverse-variance weighted approach after removing pleiotropic SNPs from Phenoscanner

	Inverse-variance weighted	
MetS components	**β (95%CI)**	**P-value**
HDL-C (102 SNPs)	-0.01 (-0.02 – 0.004)	0.18
HbA1c (35 SNPs)	0.02 (-0.03 – 0.07)	0.47
BMI (291 SNPs)	-0.01 (-0.03 – 0.003)	0.11

MetS: metabolic syndrome, HDL: high density lipoprotein-cholesterol, HbA1c: glycated haemoglobin A1c, BMI: body mass index

Table 5.9. Multivariable Mendelian randomisation for high density lipoprotein cholesterol adjusting for low density lipoprotein cholesterol, triglycerides and total cholesterol

	Inverse-variance weighted	
	β (95%CI)	**P-value**
HDL-C adjusted for LDL-C (210 SNPs)*	**-0.01 (-0.02 – -0.001)**	**0.04**
HDL-C adjusted for TG (210)*	**-0.01 (-0.02 – -0.002)**	**0.02**
HDL-C adjusted for TC (210)*	-0.01 (-0.02 – 0.002)	0.11

*SNPs for HDL, TG, and TC were obtained from the GLGC consortium

HDL: high density lipoprotein-cholesterol, LDL: low density lipoprotein-cholesterol, TG: triglycerides, TC: total cholesterol

Table 5.10. Multivariable Mendelian randomisation for glycated haemoglobin A1c adjusting for fasting glucose and Type 2 Diabetes

	Inverse-variance weighted	
	β (95%CI)	**P-value**
HbA1c adjusted for fasting glucose (64 SNPs)*	**0.05 (0.005 – 0.10)**	**0.03**
HbA1c adjusted for T2D (53 SNPs)*	**0.06 (0.01 – 0.10)**	**0.01**
HbA1c adjusted for BMI (60 SNPs)*	**0.06 (0.01 – 0.10)**	**0.01**

*11 SNPs were not found in the DIAGRAM consortium for T2D

*SNPs associated with fasting glucose was obtained from MAGIC consortium

*SNPs associated with BMI was obtained from the GIANT consortium

HbA1c: glycated haemoglobin, T2D: Type 2 Diabetes, BMI: body mass index

Table 5.11. Multivariable Mendelian randomisation for body mass index adjusting for body fat percentage, hip circumference, waist circumference, and height

	Inverse-variance weighted	
	β (95%CI)	**P-value**
BMI-adjusted for body fat percentage* (362 SNPs)	-0.02 (-0.04 – 0.01)	0.18
BMI-adjusted for hip circumference* (364 SNPs)	0.02 (-0.05 – 0.02)	0.33
BMI-adjusted for height* (361 SNPs)	**-0.01 (-0.03 – -0.001)**	**0.04**
BMI-adjusted for waist circumference* (364 SNPs)	-0.01 (-0.05 – 0.02)	0.41

*SNPs for anthropometric factors were obtained from the GIANT consortium. SNPs for body fat percentage was obtained from a study of a meta-analysis of GWAS done by Lu *et al.* (312)

BMI: body mass index

CHAPTER 6 CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE STUDIES

In this chapter I will conclude the results of the present thesis and recommend further studies that could be done to advance in this field.

6.1 Conclusion

Previous studies have reported associations between vitamin D with MetS and CRC. The purpose of this thesis was to disentangle the relationship between vitamin D, MetS, and CRC.

Using a cross-sectional study, I first investigated the association between vitamin D and MetS in the QBB and found that individuals with low levels of vitamin D were more likely to have MetS. I then repeated the analysis in a nested case-control study in the prospective EPIC cohort and found similar results. I also investigated the association between vitamin D and CRC in the EPIC cohort and found that high levels of vitamin D was associated with a lower risk of CRC. Using the same nested case control study, I investigated whether the association between vitamin D and CRC was mediated by MetS or its components and found that waist circumference was the main component in MetS that partially mediated the association between vitamin D and CRC.

I then assessed whether results from observational studies of an association between vitamin D and CRC was potentially causal. This was done using a MR approach on three different datasets, EPIC, the UK Biobank and GECCO, each one larger than the previous. A meta-analysis of the MR results from the individual studies reported no evidence of a causal relationship between vitamin D and CRC. The inconsistency between observational and MR studies suggests that the results from observational studies may be false positives due to confounders that were not accounted for including medication and family history of CRC, that might affect the levels of vitamin D. The inconsistency can also be due to a lack of sufficient power in the present MR study to detect small associations. It can also indicate that there is no causal

association between vitamin D and CRC. Instead vitamin D is likely a marker of good health and not necessarily a causal factor for CRC.

The associations that were found between vitamin D and MetS components in the QBB and the EPIC cohort were then assessed for potentially causality. I also assessed the direction of association by performing bi-directional MR analyses of vitamin D and MetS components. Evidence of a causal association between vitamin D and HbA1c was found in both directions, in which low levels of vitamin D was associated with an increase in HbA1c levels, and high levels of HbA1c was associated with an increase in vitamin D levels. However, the association between low levels of 25(OH)D and HbA1c did not survive multiple testing. Moreover, these conflicting results are possibly attributes to the presence of pleiotropy with current MR methodologies unable to untangle these complex relationships. Results from this analysis also suggested potential causal associations between high BMI and low HDL cholesterol levels with low levels of vitamin D. However, none of these factors remained significant after Bonferroni corrections.

Based on the current evidence from this study, there is no evidence supporting vitamin D supplementation will reduce the risk of CRC. Evidence from this study does suggest that waist circumference partially mediates the association between vitamin D and CRC in an observational study. Waist circumference could be a potential target in reducing the risk of CRC. Evidence does not support a reduced risk in MetS factors with vitamin D supplementation, however, reducing BMI and HbA1c and increasing HDL may increase vitamin D levels. Larger studies with stronger genetic variants for vitamin D are needed to confirm these potential causal associations. Moreover,

additional MR methodologies needs to be understood to uncover complex causal relationships.

6.2 Recommendation

Previous observational studies have reported consistent result of an association between vitamin D and CRC (19). Although I did not find a statistically significant causal effect between vitamin D and CRC, this may have been due to lack of sufficient power to detect small causal associations. GECCO will soon be releasing data with more CRC cases, approximately 60,000 cases and 60,000 controls, and with the updated summary statistics data from the SUNLIGHT consortium, this 2-sample MR analysis could give us greater power to detect a small potential causal association. To further achieve greater power, more IVs that explain the variance of 25(OH)D are needed. Thus far, only 6 IVs have been found to be associated with 25(OH)D which explain approximately 3% of the variance.

There has been one GWAS consortia done for MetS on the European population comprising of approximately 22,000 participants. However, this data was not publically available and therefore could not be used to run in the bi-directional MR analyses in this thesis with vitamin D. Since the UK Biobank has genotype data and will be releasing serum biomarkers soon, a GWAS on MetS on approximately half a million participants could be done. A bi-directional MR analysis could be performed to see whether MetS potentially causes a change in vitamin D levels or vice versa and compare the results with the individual components.

Many mechanisms have been uncovered for the association between vitamin D and obesity and lipids (179,251,307). However, little information is known on how HbA1c or SBP affect the levels of vitamin D that will support the results in this study. Further

pathway analysis needs to be conducted to understand the mechanism of how these components affect vitamin D levels.

Moreover, this study found that MetS and two of its components mediates the association between vitamin D and CRC. Therefore, MR studies on potential causal association between MetS and its components with CRC could be done to further understand the pathway between vitamin D, MetS and CRC.

REFERENCES

1. Ferlay J, Ervik M, Lam F, Colombet M, Mery L, Piñeros M, Znaor A, Soerjomataram I BF. Global Cancer Observatory: Cancer Today. Lyon, France: International Agency for Research on Cancer [Internet]. International Agency for Research on Cancer. 2018 [cited 2018 Sep 28]. Available from: https://gco.iarc.fr/today

2. Center MM, Jemal A, Ward E. International trends in colorectal cancer incidence rates. Cancer Epidemiol Biomarkers Prev. 2009;18(6):1688–94.

3. Haggar FA, Boushey RP. Colorectal cancer epidemiology: Incidence, mortality, survival, and risk factors. Clin Colon Rectal Surg. 2009;22:191–7.

4. Boyle P, Langman JS. Epidemiology. BMJ Br Med J [Internet]. 2000 Sep 30;321(7264):805–8. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1118620/

5. Arnold M, Sierra MS, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global patterns and trends in colorectal cancer incidence and mortality. Gut [Internet]. 2017 Apr 1;66(4):683 LP-691. Available from: http://gut.bmj.com/content/66/4/683.abstract

6. International Agency for Research on Cancer WHO. Colorectal cancer estimates [Internet]. 2015 [cited 2015 Feb 26]. Available from: http://globocan.iarc.fr/Pages/fact_sheets_cancer.aspx

7. Moghadamyeghaneh Z, Alizadeh RF, Phelan M, Carmichael JC, Mills S, Pigazzi A, et al. Trends in colorectal cancer admissions and stage at presentation: impact of screening. Surg Endosc [Internet]. 2015/11/05. 2016 Aug;30(8):3604–10. Available from: https://www.ncbi.nlm.nih.gov/pubmed/26541735

8. Dukes CE. The classification of cancer of the rectum. J Pathol Bacteriol [Internet]. 1932 Jan 1;35(3):323–32. Available from: https://doi.org/10.1002/path.1700350303

9. Bruening W, Sullivan N, Paulson EC et al. Imaging Tests for the Staging of Colorectal Cancer. (142).

10. Sack J, Rothman JM. Colorectal cancer: natural history and management. Hosp Physician. 2000;36:64–73.

11. Hamilton SR, Aaltonen LA. Tumours of the Digestive System. 2000;

12. Thanki K, Edward Nicholls M, Gomez G, Gajjar A, James Senagore A, Rashidi L, et al. Consensus Molecular Subtypes of Colorectal Cancer and their Clinical Implications. Ibbj. 2017;3(3):105–11.

13. O'Brien MJ, Winawer SJ, Zauber AG, Gottlieb LS, Sternberg SS, Diaz B, et al. The National Polyp Study. Patient and polyp characteristics associated with high-grade dysplasia in colorectal adenomas. Gastroenterology. 1990 Feb;98(2):371–9.

14. Grady WM, Markowitz SD. The molecular pathogenesis of colorectal cancer and its potential application to colorectal cancer screening. Dig Dis Sci [Internet]. 2014/12/10. 2015 Mar;60(3):762–72. Available from: https://www.ncbi.nlm.nih.gov/pubmed/25492499

15. Fedirko V, Tramacere I, Bagnardi V, Rota M, Scotti L, Islami F, et al. Alcohol drinking and colorectal cancer risk: an overall and dose–response meta-analysis of published studies. Ann Oncol [Internet]. 2011 Sep 1;22(9):1958–72. Available from: http://annonc.oxfordjournals.org/content/22/9/1958.abstract

16. Cross AJ, Ferrucci LM, Risch A, Graubard BI, Ward MH, Park Y, et al. A Large Prospective Study of Meat Consumption and Colorectal Cancer Risk: An Investigation of Potential Mechanisms Underlying this Association. Cancer Res [Internet]. 2010 Mar 15;70(6):2406–14. Available from: http://cancerres.aacrjournals.org/content/70/6/2406.abstract

17. Limsui D, Vierkant RA, Tillmans LS, Wang AH, Weisenberger DJ, Laird PW, et al. Cigarette Smoking and Colorectal Cancer Risk by Molecularly Defined Subtypes. J Natl Cancer Inst [Internet]. 2010 Jul 21;102(14):1012–22. Available from: http://jnci.oxfordjournals.org/content/102/14/1012.abstract

18. Ma Y, Yang Y, Wang F, Zhang P, Shi C, Zou Y, et al. Obesity and risk of colorectal cancer: a systematic review of prospective studies. PLoS One. 2013;8(1):e53916.

19. World cancer research fund. Diet, nutrition, physical activity and colorectal cancer. 2017;

20. Rennert G, Rennert HS, Pinchev M, Lavie O, Gruber SB. Use of Hormone Replacement Therapy and the Risk of Colorectal Cancer. J Clin Oncol [Internet]. 2009 Sep 20;27(27):4542–7. Available from: http://jco.ascopubs.org/content/27/27/4542.abstract

21. Chan AT, Ogino S, Fuchs CS. Aspirin and the Risk of Colorectal Cancer in Relation to the Expression of COX-2. N Engl J Med [Internet]. 2007 May 24;356(21):2131–42. Available from: http://dx.doi.org/10.1056/NEJMoa067208

22. Wu K, Feskanich D, Fuchs CS, Willett WC, Hollis BW, Giovannucci EL. A Nested Case–Control Study of Plasma 25-Hydroxyvitamin D Concentrations and Risk of Colorectal Cancer. J Natl Cancer Inst [Internet]. 2007 Jul 18;99(14):1120–9. Available from: http://jnci.oxfordjournals.org/content/99/14/1120.abstract

23. Boyle T, Keegel T, Bull F, Heyworth J, Fritschi L. Physical Activity and Risks of Proximal and Distal Colon Cancers: A Systematic Review and Meta-analysis. J Natl Cancer Inst [Internet]. 2012 Aug 22; Available from:

http://jnci.oxfordjournals.org/content/early/2012/08/22/jnci.djs354.abstract

24. Erdrich J, Zhang X, Giovannucci E, Willett W. Proportion of colon cancer attributable to lifestyle in a cohort of US women. Cancer Causes Control. 2015 Sep;26(9):1271–9.

25. Platz EA, Willett WC, Colditz GA, Rimm EB, Spiegelman D, Giovannucci E. Proportion of Colon Cancer Risk That Might Be Preventable in a Cohort of Middle-Aged US Men. Cancer Causes Control [Internet]. 2000;11(7):579–88. Available from: http://www.jstor.org/stable/3554161

26. de la Chapelle A. Genetic predisposition to colorectal cancer. Nat Rev Cancer [Internet]. 2004 Oct;4(10):769–80. Available from: http://dx.doi.org/10.1038/nrc1453

27. Johnson CM, Wei C, Ensor JE, Smolenski DJ, Amos CI, Levin B, et al. Meta-analyses of colorectal cancer risk factors. Cancer Causes Control [Internet]. 2013/04/06. 2013 Jun;24(6):1207–22. Available from: https://www.ncbi.nlm.nih.gov/pubmed/23563998

28. Zhou L, Xie J, Gu E-L, Huang Y, Qu Y, Xu A-P, et al. Common genetic variant on BMP4 contributes to colorectal adenoma and cancer: A meta-analysis based on 15 studies. Cytokine [Internet]. 2015 Jan 31 [cited 2015 Mar 9];72(2):154–9. Available from: http://www.sciencedirect.com/science/article/pii/S1043466614006541

29. Ellulu MS, Patimah I, Khaza'ai H, Rahmat A, Abed Y. Obesity and inflammation: the linking mechanism and the complications. Arch Med Sci. 2016/03/31. 2017 Jun;13(4):851–63.

30. Leufkens AM, Van Duijnhoven FJB, Siersema PD, Boshuizen HC, Vrieling A, Agudo A, et al. Cigarette smoking and colorectal cancer risk in the European Prospective Investigation into Cancer and Nutrition study. Clin Gastroenterol Hepatol. 2011 Feb;9(2):137–44.

31. Gram IT, Braaten T, Lund E, Le Marchand L, Weiderpass E. Cigarette smoking and risk of colorectal cancer among Norwegian women. Cancer Causes Control. 2009 Aug;20(6):895–903.

32. Chyou PH, Nomura AM, Stemmermann GN. A prospective study of colon and rectal cancer among Hawaii Japanese men. Ann Epidemiol. 1996 Jul;6(4):276–82.

33. Paskett ED, Reeves KW, Rohan TE, Allison MA, Williams CD, Messina CR, et al. Association between cigarette smoking and colorectal cancer in the Women's Health Initiative. J Natl Cancer Inst. 2007 Nov;99(22):1729–35.

34. Cheng J, Chen Y, Wang X, Wang J, Yan Z, Gong G, et al. Meta-analysis of prospective cohort studies of cigarette smoking and the incidence of colon and rectal cancers. Eur J Cancer Prev. 2015 Jan;24(1):6–15.

35. O'Byrne KJ, Dalgleish AG, Browning MJ, Steward WP, Harris AL. The relationship between angiogenesis and the immune response in carcinogenesis and the progression of malignant disease. Eur J Cancer. 2000 Jan;36(2):151–69.

36. Seitz H, Stickel F. Acetaldehyde as an underestimated risk factor for cancer development: role of genetics in ethanol metabolism. Genes Nutr [Internet]. 2010;5(2):121–8. Available from: http://dx.doi.org/10.1007/s12263-009-0154-1

37. Green J, Czanner G, Reeves G, Watson J, Wise L, Roddam A, et al. Menopausal hormone therapy and risk of gastrointestinal cancer: nested case-control study within a prospective cohort, and meta-analysis. Int J cancer. 2012 May;130(10):2387–96.

38. Luan N-N, Wu L, Gong T-T, Wang Y-L, Lin B, Wu Q-J. Nonlinear reduction in risk for colorectal cancer by oral contraceptive use: a meta-analysis of epidemiological studies. Cancer Causes Control. 2015 Jan;26(1):65–78.

39. Murphy N, Strickler HD, Stanczyk FZ, Xue X, Wassertheil-Smoller S, Rohan TE, et al. A Prospective Evaluation of Endogenous Sex Hormone Levels and Colorectal Cancer Risk in Postmenopausal Women. J Natl Cancer Inst. 2015 Oct;107(10).

40. Johnson JR, Lacey Jr J V, Lazovich D, Geller MA, Schairer C, Schatzkin A, et al. Menopausal hormone therapy and risk of colorectal cancer. Cancer Epidemiol Biomarkers Prev. 2009 Jan;18(1):196–203.

41. Gunter MJ, Hoover DR, Yu H, Wassertheil-Smoller S, Rohan TE, Manson JE, et al. Insulin, Insulin-like Growth Factor-I, Endogenous Estradiol, and Risk of Colorectal Cancer in Postmenopausal Women. Vol. 68, Cancer Research. 2008. p. 329–37.

42. Qiao L, Feng Y. Intakes of heme iron and zinc and colorectal cancer incidence: a meta-analysis of prospective studies. Cancer Causes Control. 2013 Jun;24(6):1175–83.

43. Bastide NM, Pierre FHF, Corpet DE. Heme Iron from Meat and Risk of Colorectal Cancer: A Meta-analysis and a Review of the Mechanisms Involved. Cancer Prev Res [Internet]. 2011 Feb 1;4(2):177–84. Available from: http://cancerpreventionresearch.aacrjournals.org/content/4/2/177.abstract

44. Esposito K, Chiodini P, Capuano A, Bellastella G, Maiorino MI, Rafaniello C, et al. Colorectal cancer association with metabolic syndrome and its components: a systematic review with meta-analysis. Endocrine. 2013 Dec;44(3):634–47.

45. Esposito K, Chiodini P, Colao A, Lenzi A, Giugliano D. Metabolic Syndrome and Risk of Cancer: A systematic review and meta-analysis . Diabetes Care [Internet]. 2012 Nov 1;35(11):2402–11. Available from: http://care.diabetesjournals.org/content/35/11/2402.abstract

46. Xu J, Ye Y, Wu H, Duerksen-Hughes P, Zhang H, Li P, et al. Association between markers of glucose metabolism and risk of colorectal cancer. BMJ Open. 2016;6(6).

47. Rinaldi S, Rohrmann S, Jenab M, Biessy C, Sieri S, Palli D, et al. Glycosylated hemoglobin and risk of colorectal cancer in men and women, the European prospective investigation into cancer and nutrition. Cancer Epidemiol Biomarkers Prev [Internet]. 2008;17(11):3108–15. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18990751

48. Yao X, Tian Z. Dyslipidemia and colorectal cancer risk: a meta-analysis of prospective studies. Cancer Causes Control [Internet]. 2015;26(2):257–68. Available from: http://link.springer.com/10.1007/s10552-014-0507-y

49. McCullough ML, Bostick RM, Mayo TL. Vitamin D gene pathway polymorphisms and risk of colorectal, breast, and prostate cancer. Annu Rev Nutr. 2009;29:111–32.

50. Garland CF, Gorham ED. Dose-response of serum 25-hydroxyvitamin D in association with risk of colorectal cancer: A meta-analysis. J Steroid Biochem Mol Biol [Internet]. 2017;168:1–8. Available from: http://www.sciencedirect.com/science/article/pii/S0960076016303429

51. Wactawski-Wende J, Kotchen JM, Anderson GL, Assaf AR, Brunner RL, O'Sullivan MJ, et al. Calcium plus Vitamin D Supplementation and the Risk of Colorectal Cancer. N Engl J Med [Internet]. 2006 Feb 16;354(7):684–96. Available from: http://dx.doi.org/10.1056/NEJMoa055222

52. Ross AC, Taylor CL, Yaktine AL, Del HB. Dietary Reference Intakes for Calcium and Vitamin D. Nutrition. 2011.

53. Lappe J, Watson P, Travers-Gustafson D, Recker R, Garland C, Gorham E, et al. Effect of Vitamin D and calcium supplementation on cancer incidence in older women: A randomized clinical trial. JAMA - J Am Med Assoc [Internet]. 2017;317(12):1234–43. Available from: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85016597461&doi=10.1001%252Fjama.2017.2115&partnerID=40&md5=e3d992c633a05174efae237d344269ba

54. Ekmekcioglu C, Haluza D, Kundi M. 25-Hydroxyvitamin D Status and Risk for Colorectal Cancer and Type 2 Diabetes Mellitus: A Systematic Review and Meta-Analysis of Epidemiological Studies. Tchounwou PB, editor. Int J Environ Res Public Health [Internet]. 2017 Feb 28;14(2):127. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC5334681/

55. Davis DM, Marcet JE, Frattini JC, Prather AD, Mateka JJL, Nfonsam VN. Is It Time to Lower the Recommended Screening Age for Colorectal Cancer? J Am Coll Surg [Internet]. 2011;213(3):352–61. Available from: http://www.sciencedirect.com/science/article/pii/S1072751511003802

56. Ponz de Leon M, Percesepe A. Pathogenesis of colorectal cancer. Dig Liver Dis [Internet]. 2000 Dec [cited 2015 May 15];32(9):807–21. Available from: http://www.sciencedirect.com/science/article/pii/S1590865800803618

57. CANAVAN C, ABRAMS KR, MAYBERRY J. Meta-analysis: colorectal and small bowel cancer risk in patients with Crohn's disease. Aliment Pharmacol Ther [Internet]. 2006 Apr 1;23(8):1097–104. Available from: http://dx.doi.org/10.1111/j.1365-2036.2006.02854.x

58. Winawer SJ. Natural history of colorectal cancer. Am J Med. 1999 Jan;106(1A):3S–6S; discussion 50S–51S.

59. Hisabe T, Hirai F, Matsui T. Development and progression of colorectal cancer based on follow-up analysis. Dig Endosc. 2014 Apr;26 Suppl 2:73–7.

60. Lichtenstein P, Holm N V, Verkasalo PK, Iliadou A, Kaprio J, Koskenvuo M, et al. Environmental and heritable factors in the causation of cancer--analyses of cohorts of twins from Sweden, Denmark, and Finland. N Engl J Med. 2000 Jul;343(2):78–85.

61. Fortini BK, Tring S, Plummer SJ, Edlund CK, Moreno V, Bresalier RS, et al. Multiple functional risk variants in a SMAD7 enhancer implicate a colorectal cancer risk haplotype. PLoS One. 2014;9(11):e111914.

62. Tan SC. Low penetrance genetic polymorphisms as potential biomarkers for colorectal cancer predisposition. J Gene Med. 2018 Apr;20(4):e3010.

63. Abulí A, Fernández-Rozadilla C, Alonso-Espinaco V, Muñoz J, Gonzalo V, Bessa X, et al. Case-control study for colorectal cancer genetic susceptibility in EPICOLON: previously identified variants and mucins. BMC Cancer [Internet]. 2011;11(1):339. Available from: https://doi.org/10.1186/1471-2407-11-339

64. Aarnio M, Mecklin J, Aaltonen LA, Nyström-Lahti M, Järvinen HJ. Life-time risk of different cancers in hereditary non-polyposis colorectal cancer (hnpcc) syndrome. Int J cancer. 1995;64(6):430–3.

65. de Jong AE, Morreau H, Nagengast FM, Mathus-Vliegen EMH, Kleibeuker JH, Griffioen G, et al. Prevalence of adenomas among young individuals at average risk for colorectal cancer. Am J Gastroenterol. 2005 Jan;100(1):139–43.

66. Patnala R, Clements J, Batra J. Candidate gene association studies: a comprehensive guide to useful in silico tools. BMC Genet [Internet]. 2013 May 9;14:39. Available from: https://www.ncbi.nlm.nih.gov/pubmed/23656885

67. de Voer RM, Hahn M-M, Weren RDA, Mensenkamp AR, Gilissen C, van Zelst-Stams WA, et al. Identification of Novel Candidate Genes for Early-Onset Colorectal Cancer Susceptibility. PLOS Genet [Internet]. 2016 Feb 22;12(2):e1005880. Available from: https://doi.org/10.1371/journal.pgen.1005880

68. Amos W, Driscoll E, Hoffman JI. Candidate genes versus genome-wide associations: which are better for detecting genetic susceptibility to infectious disease? Proceedings Biol Sci [Internet]. 2010/10/06. 2011 Apr 22;278(1709):1183–8. Available from: https://www.ncbi.nlm.nih.gov/pubmed/20926441

69. Schumacher FR, Schmit SL, Jiao S, Edlund CK, Wang H, Zhang B, et al. Genome-wide association study of colorectal cancer identifies six new susceptibility loci. Nat Commun. 2015 Jul;6:7138.

70. Schmit SL, Edlund CK, Schumacher FR, Gong J, Harrison TA, Huyghe JR, et al. Novel Common Genetic Susceptibility Loci for Colorectal Cancer. J Natl Cancer Inst. 2018 Jun;

71. Office for National Statistics. Cancer survival by stage at diagnosis for England (experimental statistics): Adults diagnosed 2012, 2013 and 2014 and followed up to 2015 [Internet]. 2016. Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/cancersurvivalbystageatdiagnosisforenglandexperimentalstatistics/adultsdiagnosed20122013and2014andfollowedupto2015

72. American Cancer Society. Cancer Prevention & Early Detection Facts & 2015. AACS. Cancer prevention.

73. Koo S, Neilson LJ, Von Wagner C, Rees CJ. The NHS Bowel Cancer Screening Program: current perspectives on strategies for improvement. Risk Manag Healthc Policy [Internet]. 2017 Dec 4;10:177–87. Available from: https://www.ncbi.nlm.nih.gov/pubmed/29270036

74. Holick MF. Vitamin D Deficiency. N Engl J Med [Internet]. 2007 Jul 19;357(3):266–81. Available from: http://www.nejm.org/doi/abs/10.1056/NEJMra070553

75. Prabhu A V, Luu W, Sharpe LJ, Brown AJ. Cholesterol-mediated Degradation of 7-Dehydrocholesterol Reductase Switches the Balance from Cholesterol to Vitamin D Synthesis. J Biol Chem [Internet]. 2016 Apr 15;291(16):8363–73. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4861412/

76. Hossein-nezhad A, Holick MF. Vitamin D for Health: A Global Perspective. Mayo Clin Proc [Internet]. 2015 Jun 27;88(7):720–55. Available from: http://dx.doi.org/10.1016/j.mayocp.2013.05.011

77. Le Goff C, Cavalier E, Souberbielle J-C, González-Antuña A, Delvin E. Measurement of circulating 25-hydroxyvitamin D: A historical review. Pract Lab Med [Internet]. 2015;2:1–14. Available from: http://www.sciencedirect.com/science/article/pii/S2352551715000104

78. Salimterics. Calculating Inter- and Intra-Assay Coefficients of Variability. 2019.

79. Hart PH, Gorman S, Finlay-Jones JJ. Modulation of the immune system by UV

radiation: more than just the effects of vitamin D? Nat Rev Immunol [Internet]. 2011 Sep;11(9):584–96. Available from: http://dx.doi.org/10.1038/nri3045

80. Mathieu C, Gysemans C, Giulietti A, Bouillon R. Vitamin D and diabetes. Diabetologia. 2005 Jul;48(7):1247–57.

81. Bikle DD. Vitamin D Metabolism, Mechanism of Action, and Clinical Applications. Chem Biol [Internet]. 2014 Mar 20;21(3):319–29. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3968073/

82. Institute of Medicine (US) Standing Committee on the Scientific Evaluation of Dietary Reference Intakes. Dietary Reference Intakes for Calcium, Phosphorus, Magnesium, Vitamin D, and Fluoride. Washington (DC): National Academies Press (US); 1997. 7, Vitami.

83. Mokry LE, Ross S, Ahmad OS, Forgetta V, Smith GD, Leong A, et al. Vitamin D and risk of multiple sclerosis: a Mendelian randomization study. PLoS Med. 2015;12(8):e1001866.

84. Hewison M, Zehnder D, Bland R, Stewart PM. 1alpha-Hydroxylase and the action of vitamin D. J Mol Endocrinol [Internet]. 2000 Oct 1;25(2):141–8. Available from: http://jme.endocrinology-journals.org/content/25/2/141.abstract

85. Alvarez JA, Ashraf A. Role of vitamin d in insulin secretion and insulin sensitivity for glucose homeostasis. Int J Endocrinol. 2010;2010:351385.

86. Holick MF. VITAMIN D STATUS: MEASUREMENT, INTERPRETATION AND CLINICAL APPLICATION. Ann Epidemiol [Internet]. 2009 Feb 10;19(2):73–8. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2665033/

87. Gallone G, Haerty W, Disanto G, Ramagopalan S V, Ponting CP, Berlanga-Taylor AJ. Identification of genetic variants affecting vitamin D receptor binding and associations with autoimmune disease. Hum Mol Genet [Internet]. 2017 Jun 1;26(11):2164–76. Available from: http://dx.doi.org/10.1093/hmg/ddx092

88. Liberato SC, Pinheiro-Sant'Ana HM. Fortification of industrialized foods with vitamins . Vol. 19, Revista de Nutrição . scielo ; 2006. p. 215–31.

89. Calvo MS, Whiting SJ, Barton CN. Vitamin D fortification in the United States and Canada: current status and data needs. Am J Clin Nutr [Internet]. 2004 Dec 1;80(6):1710S–1716S. Available from: http://ajcn.nutrition.org/content/80/6/1710S.abstract

90. Palacios C, Gonzalez L. Is vitamin D deficiency a major global public health problem? J Steroid Biochem Mol Biol [Internet]. 2013/11/12. 2014 Oct;144 Pt A:138–45. Available from: https://www.ncbi.nlm.nih.gov/pubmed/24239505

91. Vieth R. Vitamin D supplementation, 25-hydroxyvitamin D concentrations, and safety. Am J Clin Nutr [Internet]. 1999;69(5):842–56. Available from: http://www.ncbi.nlm.nih.gov/pubmed/10232622

92. Dawson-Hughes B, Heaney RP, Holick MF, Lips P, Meunier PJ, Vieth R. Estimates of optimal vitamin D status. Vol. 16, Osteoporosis International. 2005. p. 713–6.

93. Vieth R, Holick MF. Chapter 57B - The IOM—Endocrine Society Controversy on Recommended Vitamin D Targets: In Support of the Endocrine Society Position. In: Feldman DBT-VD (Fourth E, editor. Academic Press; 2018. p. 1091–107. Available from: http://www.sciencedirect.com/science/article/pii/B9780128099650000598

94. Thacher TD, Clarke BL. Vitamin D Insufficiency. Mayo Clin Proc [Internet]. 2011 Jan;86(1):50–60. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3012634/

95. Bischoff-Ferrari HA, Giovannucci E, Willett WC, Dietrich T, Dawson-Hughes B. Estimation of optimal serum concentrations of 25-hydroxyvitamin D for multiple health outcomes. Am J Clin Nutr [Internet]. 2006 Jul 1;84(1):18–28. Available from: http://ajcn.nutrition.org/content/84/1/18.abstract

96. LIPs P, MEUNIER PJ, VIETH Reinh. Vitamin D Round Table". Nutr Asp Osteoporos. 2004;263.

97. Chapuy MC, Preziosi P, Maamer M, Arnaud S, Galan P, Hercberg S, et al. Prevalence of vitamin D insufficiency in an adult normal population. Osteoporos Int. 1997;7(5):439–43.

98. Sharifi F, Mousavinasab N, Mellati AA. Defining a cutoff point for vitamin D deficiency based on insulin resistance in children. Diabetes Metab Syndr. 2013;7(4):210–3.

99. Holick MF, Bischoff-Ferrari H, Gordon C, Hanley D. Guidelines Evaluation , Treatment , and Prevention of Vitamin D Deficiency : J Clin Endocrinol Metab. 2011;96(7):1911–30.

100. Gandini S, Boniol M, Haukka J, Byrnes G, Cox B, Sneyd MJ, et al. Meta-analysis of observational studies of serum 25-hydroxyvitamin D levels and colorectal, breast and prostate cancer and colorectal adenoma. Int J Cancer [Internet]. 2011 Mar 15;128(6):1414–24. Available from: http://dx.doi.org/10.1002/ijc.25439

101. Vieth R. Vitamin D Toxicity, Policy, and Science. J Bone Miner Res [Internet]. 2007 Dec 1;22(S2):V64–8. Available from: http://dx.doi.org/10.1359/jbmr.07s221

102. Houghton LA, Vieth R. The case against ergocalciferol (vitamin D2) as a vitamin supplement. Am J Clin Nutr [Internet]. 2006 Oct 1;84(4):694–7. Available from: http://dx.doi.org/10.1093/ajcn/84.4.694

103. Reid IR. What diseases are causally linked to vitamin D deficiency? Arch Dis Child. 2015 Jul;

104. Theodoratou E, Tzoulaki I, Zgaga L, Ioannidis JPA. Vitamin D and multiple health outcomes: umbrella review of systematic reviews and meta-analyses of observational studies and randomised trials. BMJ. 2014;348.

105. Zhang Y, Leung DYM, Richers BN, Liu Y, Remigio LK, Riches DW, et al. Vitamin D inhibits monocyte/macrophage proinflammatory cytokine production by targeting MAPK phosphatase-1. J Immunol. 2012;188(5):2127–35.

106. Giovannucci E. Vitamin D and cardiovascular disease. Curr Atheroscler Rep [Internet]. 2009;11(6):456. Available from: http://dx.doi.org/10.1007/s11883-009-0068-9

107. Nemerovski CW, Dorsch MP, Simpson RU, Bone HG, Aaronson KD, Bleske BE. Vitamin D and cardiovascular disease. Pharmacotherapy [Internet]. 2009 Jun [cited 2017 Feb 15];29(6):691–708. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19476421

108. Seshadri KG, Tamilselvan B, Rajendran A. Role of vitamin D in diabetes. J Endocrinol Metab. 2011;1(2):47–56.

109. Palomer X, González-Clemente JM, Blanco-Vaca F, Mauricio D. Role of vitamin D in the pathogenesis of type 2 diabetes mellitus. Diabetes Obes Metab. 2008;10:185–97.

110. Finklea JD, Grossmann RE, Tangpricha V. Vitamin D and Chronic Lung Disease: A Review of Molecular Mechanisms and Clinical Studies. Adv Nutr An Int Rev J [Internet]. 2011 May 1;2(3):244–53. Available from: http://advances.nutrition.org/content/2/3/244.abstract

111. Food Standards Agency. National Diet and Nutrition Survey : Results from Years 1-4 (combined) of the Rolling Programme About Public Health England. 2014;4:1–24.

112. Spiro A, Buttriss JL. Vitamin D: An overview of vitamin D status and intake in Europe. Nutr Bull [Internet]. 2014;39(4):322–50. Available from: http://dx.doi.org/10.1111/nbu.12108

113. Golbahar J, Al-Saffar N, Altayab Diab D, Al-Othman S, Darwish A, Al-Kafaji G. Predictors of vitamin D deficiency and insufficiency in adult Bahrainis: a cross-sectional study . Public Health Nutr. 2014;17(04):732–8.

114. Alhamad HK, Nadukkandiyil N, El-Menyar A, Abdel Wahab L, Sankaranarayanan A, Al Sulaiti EM. Vitamin D deficiency among the elderly: insights from Qatar. Curr Med Res Opin. 2014 Jun;30(6):1189–96.

115. Hussain A, Alkhenizan A, El Shaker M, Raef H, Gabr A. Increasing trends and significance of hypovitaminosis D: a population-based study in the Kingdom of Saudi Arabia. Arch Osteoporos [Internet]. 2014;9(1):1–5. Available from: http://dx.doi.org/10.1007/s11657-014-0190-3

116. Alyahya K, Lee WTK, Al-Mazidi Z, Morgan J, Lanham-New S. Risk factors of low vitamin D status in adolescent females in Kuwait: implications for high peak bone mass attainment. Arch Osteoporos. 2014;9:178.

117. Sadiya A, Ahmed SM, Skaria S, Abusnana S. Vitamin D status and its relationship with metabolic markers in persons with obesity and type 2 diabetes in the UAE: a cross-sectional study. J Diabetes Res. 2014;2014.

118. Mahdy S, Al-Emadi SA, Khanjar IA, Hammoudeh MM, Sarakbi HA, Siam AM, et al. Vitamin D status in health care professionals in Qatar. Saudi Med J. 2010;31(1):74–7.

119. Moussavi M, Heidarpour R, Aminorroaya A, Pournaghshband Z, Amini M. Prevalence of Vitamin D Deficiency in Isfahani High School Students in 2004. Horm Res Paediatr [Internet]. 2005;64(3):144–8. Available from: http://www.karger.com/DOI/10.1159/000088588

120. Siddiqui AM, Kamfar HZ. Prevalence of vitamin D deficiency rickets in adolescent school girls in Western region, Saudi Arabia. Saudi Med J. 2007;28:441–4.

121. Karohl C, Su S, Kumari M, Tangpricha V, Veledar E, Vaccarino V, et al. Heritability and seasonal variability of vitamin D concentrations in male twins. Am J Clin Nutr [Internet]. 2010 Dec 13;92(6):1393–8. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2980965/

122. Hunter D, De Lange M, Snieder H, MacGregor AJ, Swaminathan R, Thakker R V, et al. Genetic Contribution to Bone Metabolism, Calcium Excretion, and Vitamin D and Parathyroid Hormone Regulation. J Bone Miner Res [Internet]. 2001;16(2):371–8. Available from: http://dx.doi.org/10.1359/jbmr.2001.16.2.371

123. Shea MK, Benjamin EJ, Dupuis J, Massaro JM, Jacques PF, D'Agostino RB, et al. Genetic and non-genetic correlates of vitamins K and D. Eur J Clin Nutr [Internet]. 2009 Apr 21;63(4):458–64. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2681093/

124. Snellman G, Melhus H, Gedeborg R, Olofsson S, Wolk A, Pedersen NL, et al. Seasonal genetic influence on serum 25-hydroxyvitamin D levels: A twin study. PLoS One. 2009;4(11):e7747.

125. Wang TJ, Zhang F, Richards JB, Kestenbaum B, van Meurs JB, Berry D, et al. Common genetic determinants of vitamin D insufficiency: a genome-wide association study. Lancet (London, England). 2010 Jul;376(9736):180–8.

126. Ahn J, Yu K, Stolzenberg-Solomon R, Simon KC, McCullough ML, Gallicchio L, et al. Genome-wide association study of circulating vitamin D levels. Hum Mol Genet [Internet]. 2010 Apr 23; Available from: http://hmg.oxfordjournals.org/content/early/2010/05/07/hmg.ddq155.abstract

127. Hiraki LT, Major JM, Chen C, Cornelis MC, Hunter DJ, Rimm EB, et al. Exploring

the genetic architecture of circulating 25-hydroxyvitamin D. Genet Epidemiol [Internet]. 2013 Jan 7;37(1):92–8. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3524394/

128. Jiang X, O'Reilly PF, Aschard H, Hsu YH, Richards JB, Dupuis J, et al. Genome-wide association study in 79,366 European-ancestry individuals informs the genetic architecture of 25-hydroxyvitamin D levels. Nat Commun. 2018;9(1).

129. Nylen H, Bjorkhem-Bergman L, Ekstrom L, Roh H-K, Bertilsson L, Eliasson E, et al. Plasma levels of 25-hydroxyvitamin D3 and in vivo markers of cytochrome P450 3A activity in Swedes and Koreans: effects of a genetic polymorphism and oral contraceptives. Basic Clin Pharmacol Toxicol. 2014 Oct;115(4):366–71.

130. Moller UK, Streym S vieth, Jensen LT, Mosekilde L, Schoenmakers I, Nigdikar S, et al. Increased plasma concentrations of vitamin D metabolites and vitamin D binding protein in women using hormonal contraceptives: a cross-sectional study. Nutrients. 2013 Sep;5(9):3470–80.

131. Marks R, PA F, Jolley D, KR K, Harrison J, SC T. The effect of regular sunscreen use on vitamin d levels in an australian population: Results of a randomized controlled trial. Arch Dermatol [Internet]. 1995 Apr 1;131(4):415–21. Available from: http://dx.doi.org/10.1001/archderm.1995.01690160043006

132. Jayaratne N, Russell A, van der Pols JC. Sun protection and vitamin D status in an Australian subtropical community. Prev Med (Baltim) [Internet]. 2012 Aug [cited 2015 Jun 9];55(2):146–50. Available from: http://www.sciencedirect.com/science/article/pii/S0091743512002113

133. Engelman CD, Fingerlin TE, Langefeld CD, Hicks PJ, Rich SS, Wagenknecht LE, et al. Genetic and environmental determinants of 25-hydroxyvitamin D and 1,25-dihydroxyvitamin D levels in Hispanic and African Americans. J Clin Endocrinol Metab. 2008;93:3381–8.

134. Zhen D, Liu L, Guan C, Zhao N, Tang X. High prevalence of vitamin D deficiency among middle-aged and elderly individuals in northwestern China: its relationship to osteoporosis and lifestyle factors. Bone. 2015 Feb;71:1–6.

135. Heikkinen A, Parviainen MT, Tuppurainen MT, Niskanen L, Komulainen MH, Saarikoski S. Effects of postmenopausal hormone replacement therapy with and without vitamin D3 on circulating levels of 25-hydroxyvitamin D and 1,25-dihydroxyvitamin D. Vol. 62, Calcified tissue international. 1998.

136. Looker A., Dawson-Hughes B, Calvo M., Gunter E., Sahyoun N. Serum 25-hydroxyvitamin D status of adolescents and adults in two seasonal subpopulations from NHANES III. Bone [Internet]. 2002 May [cited 2015 Sep 15];30(5):771–7. Available from: http://www.sciencedirect.com/science/article/pii/S8756328202006920

137. Zadshir A, Tareen N, Pan D, Norris K, Martins D. The prevalence of hypovitaminosis D among US adults: data from the NHANES III. Ethn Dis.

2005;15(4):S5.

138. Hilger J, Friedel A, Herr R, Rausch T, Roos F, Wahl DA, et al. A systematic review of vitamin D status in populations worldwide. Br J Nutr [Internet]. 2013/08/09. 2014;111(1):23–45. Available from: https://www.cambridge.org/core/article/systematic-review-of-vitamin-d-status-in-populations-worldwide/69657BC57AF7A214271655C5463F5293

139. Chan J, Jaceldo-Siegl K, Fraser GE. Determinants of serum 25 hydroxyvitamin D levels in a nationwide cohort of blacks and non-Hispanic whites. Cancer Causes Control [Internet]. 2010 Apr 11;21(4):501–11. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3427006/

140. Manios Y, Moschonis G, Lambrinou C-P, Tsoutsoulopoulou K, Binou P, Karachaliou A, et al. A systematic review of vitamin D status in southern European countries. Eur J Nutr. 2017 Oct;

141. Tønnesen R, Hovind PH, Jensen LT, Schwarz P. Determinants of vitamin D status in young adults: influence of lifestyle, sociodemographic and anthropometric factors. BMC Public Health [Internet]. 2016 May 11;16:385. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4863340/

142. Daraghmeh AH, Bertoia ML, Al-Qadi MO, Abdulbaki AM, Roberts MB, Eaton CB. Evidence for the vitamin D hypothesis: The NHANES III extended mortality follow-up. Atherosclerosis [Internet]. 2016;255:96–101. Available from: http://www.sciencedirect.com/science/article/pii/S0021915016301381

143. Hibler EA, Sardo Molmenti CL, Dai Q, Kohler LN, Warren Anderson S, Jurutka PW, et al. Physical activity, sedentary behavior, and vitamin D metabolites. Bone [Internet]. 2016;83:248–55. Available from: http://www.sciencedirect.com/science/article/pii/S8756328215004172

144. Skaaby T, Husemoen LLN, Thuesen BH, Pisinger C, Hannemann A, Jorgensen T, et al. Longitudinal associations between lifestyle and vitamin D: A general population study with repeated vitamin D measurements. Endocrine. 2016 Feb;51(2):342–50.

145. Dong Y, Pollock N, Stallmann-Jorgensen IS, Gutin B, Lan L, Chen TC, et al. Low 25-Hydroxyvitamin D Levels in Adolescents: Race, Season, Adiposity, Physical Activity, and Fitness. Pediatr [Internet]. 2010 Jun 1;125(6):1104–11. Available from: http://pediatrics.aappublications.org/content/125/6/1104.abstract

146. Shirazi L, Almquist M, Malm J, Wirfält E, Manjer J. Determinants of serum levels of vitamin D: a study of life-style, menopausal status, dietary intake, serum calcium, and PTH. BMC Womens Health [Internet]. 2013 Aug 15;13:33. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3765227/

147. Voipio AJW, Pahkala KA, Viikari JSA, Mikkila V, Magnussen CG, Hutri-Kahonen N, et al. Determinants of serum 25(OH)D concentration in young and middle-

aged adults. The Cardiovascular Risk in Young Finns Study. Ann Med. 2015 May;47(3):253–62.

148. Harmon QE, Umbach DM, Baird DD. Use of Estrogen-Containing Contraception Is Associated With Increased Concentrations of 25-Hydroxy Vitamin D. J Clin Endocrinol Metab. 2016 Sep;101(9):3370–7.

149. van Hoof HJ, van der Mooren MJ, Swinkels LM, Sweep CG, Merkus JM, Benraad TJ. Female sex hormone replacement therapy increases serum free 1,25-dihydroxyvitamin D3: a 1-year prospective study. Clin Endocrinol (Oxf). 1999 Apr;50(4):511–6.

150. Dick IM, Prince RL, Kelly JJ, Ho KK. Oestrogen effects on calcitriol levels in post-menopausal women: a comparison of oral versus transdermal administration. Clin Endocrinol (Oxf). 1995 Aug;43(2):219–24.

151. Edwards JE, Moore RA. Statins in hypercholesterolaemia: a dose-specific meta-analysis of lipid changes in randomised, double blind trials. BMC Fam Pract [Internet]. 2003 Dec 1;4:18. Available from: https://www.ncbi.nlm.nih.gov/pubmed/14969594

152. Yavuz B, Ertugrul DT, Cil H, Ata N, Akin KO, Yalcin AA, et al. Increased levels of 25 hydroxyvitamin D and 1,25-dihydroxyvitamin D after rosuvastatin treatment: a novel pleiotropic effect of statins? Cardiovasc drugs Ther. 2009 Aug;23(4):295–9.

153. Ertugrul DT, Yavuz B, Cil H, Ata N, Akin KO, Kucukazman M, et al. STATIN-D study: comparison of the influences of rosuvastatin and fluvastatin treatment on the levels of 25 hydroxyvitamin D. Cardiovasc Ther. 2011 Apr;29(2):146–52.

154. Brot C, Jorgensen NR, Sorensen OH. The influence of smoking on vitamin D status and calcium metabolism. Eur J Clin Nutr [Internet]. 1999;53(12):920–6. Available from: http://europepmc.org/abstract/MED/10602348

155. Morabia A, Bernstein MS, Antonini S. Smoking, dietary calcium and vitamin D deficiency in women: a population-based study. Eur J Clin Nutr. 2000 Sep;54(9):684–9.

156. McKibben RA, Zhao D, Lutsey PL, Schneider ALC, Guallar E, Mosley TH, et al. Factors Associated With Change in 25-Hydroxyvitamin D Levels Over Longitudinal Follow-Up in the ARIC Study. J Clin Endocrinol Metab [Internet]. 2015 Oct 28;101(1):33–43. Available from: http://dx.doi.org/10.1210/jc.2015-1711

157. Rees JR, Mott LA, Barry EL, Baron JA, Bostick RM, Figueiredo JC, et al. Lifestyle and Other Factors Explain One-Half of the Variability in the Serum 25-Hydroxyvitamin D Response to Cholecalciferol Supplementation in Healthy Adults. J Nutr. 2016 Nov;146(11):2312–24.

158. Ward KD, Klesges RC. A meta-analysis of the effects of cigarette smoking on

bone mineral density. Calcif Tissue Int [Internet]. 2001;68(5):259–70. Available from: https://doi.org/10.1007/BF02390832

159. Hutchinson MS, Grimnes G, Joakimsen RM, Figenschau Y, Jorde R. Low serum 25-hydroxyvitamin D levels are associated with increased all-cause mortality risk in a general population: the Tromsø study. Eur J Endocrinol [Internet]. 2010 May 1;162(5):935–42. Available from: http://www.eje-online.org/content/162/5/935.abstract

160. Touvier M, Deschasaux M, Montourcy M, Sutton A, Charnaux N, Kesse-Guyot E, et al. Determinants of vitamin D status in Caucasian adults: influence of sun exposure, dietary intake, sociodemographic, lifestyle, anthropometric, and genetic factors. J Invest Dermatol. 2015 Feb;135(2):378–88.

161. Gascon-Barre M. Interrelationships between vitamin D3 and 25-hydroxyvitamin D3 during chronic ethanol administration in the rat. Vol. v. 31, Metabolism, clinical and experimental. 1982.

162. Xiang F, Lucas R, de Gruijl F, Norval M. A systematic review of the influence of skin pigmentation on changes in the concentrations of vitamin D and 25-hydroxyvitamin D in plasma/serum following experimental UV irradiation. Photochem Photobiol Sci. 2015 Dec;14(12):2138–46.

163. Hagenau T, Vest R, Gissel TN, Poulsen CS, Erlandsen M, Mosekilde L, et al. Global vitamin D levels in relation to age, gender, skin pigmentation and latitude: an ecologic meta-regression analysis. Osteoporos Int. 2009 Jan;20(1):133–40.

164. Norval M, Wulf HC. Does chronic sunscreen use reduce vitamin D production to insufficient levels? Br J Dermatol [Internet]. 2009 Oct 1;161(4):732–6. Available from: http://dx.doi.org/10.1111/j.1365-2133.2009.09332.x

165. Faurschou A, Beyer DM, Schmedes A, Bogh MK, Philipsen PA, Wulf HC. The relation between sunscreen layer thickness and vitamin D production after ultraviolet B exposure: a randomized clinical trial. Br J Dermatol. 2012 Aug;167(2):391–5.

166. Mishal AA. Effects of Different Dress Styles on Vitamin D Levels in Healthy Young Jordanian Women. Osteoporos Int [Internet]. 2001;12(11):931–5. Available from: http://dx.doi.org/10.1007/s001980170021

167. Osmancevic A, Gillstedt M, Landin-Wilhelmsen K, Larko A-MW, Larko O, Holick MF, et al. Size of the exposed body surface area, skin erythema and body mass index predict skin production of vitamin D. J Photochem Photobiol B-BIOLOGY. 2015 Aug;149:224–9.

168. Klingberg E, Oleröd G, Konar J, Petzold M, Hammarsten O. Seasonal variations in serum 25-hydroxy vitamin D levels in a Swedish cohort. Endocrine [Internet]. 2015;49(3):800–8. Available from: http://dx.doi.org/10.1007/s12020-015-0548-3

169. Heidari B, Haji Mirghassemi MB. Seasonal variations in serum vitamin D according to age and sex. Casp J Intern Med [Internet]. 2012 May 15;3(4):535–40. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3755860/

170. Cinar N, Harmanci A, Yildiz BO, Bayraktar M. Vitamin D status and seasonal changes in plasma concentrations of 25-hydroxyvitamin D in office workers in Ankara, Turkey. Eur J Intern Med [Internet]. 2014 Mar [cited 2015 Oct 29];25(2):197–201. Available from: http://www.sciencedirect.com/science/article/pii/S0953620513009904

171. Forouhi NG, Luan J, Cooper A, Boucher BJ, Wareham NJ. Baseline serum 25-hydroxy vitamin d is predictive of future glycemic status and insulin resistance the medical research council ely prospective study 1990-2000. Diabetes. 2008;57:2619–25.

172. Hyppönen E, Boucher BJ, Berry DJ, Power C. 25-hydroxyvitamin D, IGF-1, and metabolic syndrome at 45 years of age A cross-sectional study in the 1958 british birth cohort. Diabetes. 2008;57(2):298–305.

173. Ford ES, Ajani UA, McGuire LC, Liu S. Concentrations of Serum Vitamin D and the Metabolic Syndrome Among U.S. Adults. Diabetes Care [Internet]. 2005 May 1;28(5):1228–30. Available from: http://care.diabetesjournals.org/content/28/5/1228.short

174. Khan H, Kunutsor S, Franco OH, Chowdhury R. Vitamin D, type 2 diabetes and other metabolic outcomes: a systematic review and meta-analysis of prospective studies. Proc Nutr Soc. 2013;72(01):89–97.

175. Pereira-Santos M, Costa PRF, Assis AMO, Santos CAST, Santos DB. Obesity and vitamin D deficiency: a systematic review and meta-analysis. Obes Rev [Internet]. 2015 Apr 1;16(4):341–9. Available from: http://dx.doi.org/10.1111/obr.12239

176. Vimaleswaran KS, Berry DJ, Lu C, Tikkanen E, Pilz S, Hiraki LT, et al. Causal Relationship between Obesity and Vitamin D Status: Bi-Directional Mendelian Randomization Analysis of Multiple Cohorts. PLoS Med. 2013;10(2).

177. Chandler PD, Wang L, Zhang X, Sesso HD, Moorthy M V, Obi O, et al. Effect of vitamin D supplementation alone or with calcium on adiposity measures: a systematic review and meta-analysis of randomized controlled trials. Nutr Rev [Internet]. 2015 Sep 15;73(9):577–93. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4537389/

178. Zemel MB. Regulation of adiposity and obesity risk by dietary calcium: mechanisms and implications. J Am Coll Nutr. 2002 Apr;21(2):146S–151S.

179. Vanlint S. Vitamin D and Obesity. Nutrients [Internet]. 2013 Mar 20;5(3):949–56. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3705328/

180. Vimaleswaran KS, Cavadino A, Berry DJ, Jorde R, Dieffenbach AK, Lu C, et al.

Association of vitamin D status with arterial blood pressure and hypertension risk: a mendelian randomisation study. lancet Diabetes Endocrinol [Internet]. 2014 Sep [cited 2015 Feb 2];2(9):719–29. Available from: http://www.sciencedirect.com/science/article/pii/S2213858714701135

181. Ye Z, Sharp SJ, Burgess S, Scott RA, Imamura F, Consortium I, et al. Association between circulating 25-hydroxyvitamin D and incident type 2 diabetes : a mendelian randomisation study. LANCET Diabetes Endocrinol [Internet]. 3(1):35–42. Available from: http://dx.doi.org/10.1016/S2213-8587(14)70184-6

182. Wu C, Qiu S, Zhu X, Li L. Vitamin D supplementation and glycemic control in type 2 diabetes patients: A systematic review and meta-analysis. Metabolism [Internet]. 2017;73:67–76. Available from: http://www.sciencedirect.com/science/article/pii/S0026049517301476

183. Pittas AG, Lau J, Hu FB, Dawson-Hughes B. The role of vitamin D and calcium in type 2 diabetes. A systematic review and meta-analysis. J Clin Endocrinol Metab. 2007;92:2017–29.

184. Chiu KC, Chu A, Go VLW, Saad MF. Hypovitaminosis D is associated with insulin resistance and beta cell dysfunction. Am J Clin Nutr. 2004;79:820–5.

185. Sung C-C, Liao M-T, Lu K-C, Wu C-C. Role of vitamin D in insulin resistance. J Biomed Biotechnol [Internet]. 2012;2012:634195. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3440067&tool=pmcentrez&rendertype=abstract

186. Kunutsor SK, Apekey TA, Steur M. Vitamin D and risk of future hypertension: meta-analysis of 283,537 participants. Eur J Epidemiol. 2013 Mar;28(3):205–21.

187. Beveridge LA, Struthers AD, Khan F, Jorde R, Scragg R, Macdonald HM, et al. Effect of Vitamin D Supplementation on Blood Pressure: A Systematic Review and Meta-analysis Incorporating Individual Patient Data. JAMA Intern Med. 2015 May;175(5):745–54.

188. Li YC, Qiao G, Uskokovic M, Xiang W, Zheng W, Kong J. Vitamin D: a negative endocrine regulator of the renin-angiotensin system and blood pressure. J Steroid Biochem Mol Biol. 2004 May;89–90(1–5):387–92.

189. Liu S, Song Y, Ford ES, Manson JE, Buring JE, Ridker PM. Dietary Calcium, Vitamin D, and the Prevalence of Metabolic Syndrome in Middle-Aged and Older U.S. Women. Diabetes Care [Internet]. 2005 Dec 1;28(12):2926–32. Available from: http://care.diabetesjournals.org/content/28/12/2926.abstract

190. Skaaby T, Husemoen LLN, Martinussen T, Thyssen JP, Melgaard M, Thuesen BH, et al. Vitamin D status, filaggrin genotype, and cardiovascular risk factors: a Mendelian randomization approach. PLoS One. 2013;8(2):e57647.

191. Ooi EM, Afzal S, Nordestgaard BG. Elevated Remnant Cholesterol in 25-Hydroxyvitamin D Deficiency in the General Population Mendelian Randomization Study. Circ Cardiovasc Genet. 2014;7(5):650–8.

192. Wang H, Xia N, Yang Y, Peng D-Q. Influence of vitamin D supplementation on plasma lipid profiles: a meta-analysis of randomized controlled trials. Lipids Heal Dis. 2012;11:42.

193. Li S, He Y, Lin S, Hao L, Ye Y, Lv L, et al. Increase of circulating cholesterol in vitamin D deficiency is linked to reduced vitamin D receptor activity via the Insig-2/SREBP-2 pathway. Mol Nutr Food Res. 2016 Apr;60(4):798–809.

194. Alberti KGMM, Zimmet P, Shaw J. The metabolic syndrome—a new worldwide definition. Lancet [Internet]. 2005 Sep 24;366(9491):1059–62. Available from: https://doi.org/10.1016/S0140-6736(05)67402-8

195. Hwang I-C, Kim K-K, Jee S-H, Kang H-C. A comparison of predictability of cardiovascular events between each metabolic component in patients with metabolic syndrome based on the revised National Cholesterol Education Program criteria. Yonsei Med J. 2011 Mar;52(2):220–6.

196. Samson SL, Garber AJ. Metabolic Syndrome. Endocrinol Metab Clin North Am. 2014;43(1):1–23.

197. Ngai H-Y, Yuen K-KS, Ng C-M, Cheng C-H, Chu S-KP. Metabolic syndrome and benign prostatic hyperplasia: An update. Asian J Urol [Internet]. 2017/05/25. 2017 Jul;4(3):164–73. Available from: https://www.ncbi.nlm.nih.gov/pubmed/29264226

198. International Diabetes Federation. The IDF consensus worldwide definition of the metabolic syndrome [Internet]. Vol. 28, The IDF consensus worldwide definition of the metabolic syndrome. 2006. p. 1–7. Available from: http://www.idf.org/webdata/docs/MetS_def_update2006.pdf

199. Eckel RH, Grundy SM, Zimmet PZ. The metabolic syndrome. Lancet [Internet]. 2005;365(9468):1415–28. Available from: http://www.sciencedirect.com/science/article/pii/S0140673605663787

200. Ford ES, Giles WH, Dietz WH. Prevalence of the metabolic syndrome among US adults: findings from the third National Health and Nutrition Examination Survey. JAMA. 2002 Jan;287(3):356–9.

201. Moore JX, Chaudhary N, Akinyemiju T. Metabolic Syndrome Prevalence by Race/Ethnicity and Sex in the United States, National Health and Nutrition Examination Survey, 1988-2012. Prev Chronic Dis. 2017 Mar;14:E24.

202. Gao W. Does the constellation of risk factors with and without abdominal adiposity associate with different cardiovascular mortality risk? Int J Obes (Lond). 2008 May;32(5):757–62.

203. Mabry RM, Reeves MM, Eakin EG, Owen N. Gender differences in prevalence of the metabolic syndrome in Gulf Cooperation Council Countries: a systematic review. Diabet Med. 2010 May;27(5):593–7.

204. Kaur J. A Comprehensive Review on Metabolic Syndrome. Hindawi Publ Corp Cardiol Res Pract. 2014;2014:21.

205. Yengo L, Sidorenko J, Kemper KE, Zheng Z, Wood AR, Weedon MN, et al. Meta-analysis of genome-wide association studies for height and body mass index in ~700,000 individuals of European ancestry. bioRxiv [Internet]. 2018 Jan 1; Available from: http://biorxiv.org/content/early/2018/03/02/274654.abstract

206. Health B, Guideline PC, Management P. Vitamin D and Bone Health : A Practical Clinical Guideline for a breakfree future.

207. Lawlor DA, Harbord RM, Sterne JAC, Timpson N, Davey Smith G. Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. Stat Med. 2008 Apr;27(8):1133–63.

208. Burgess S, Thompson SG. Mendelian randomization: methods for using genetic variants in causal estimation. Chapman and Hall/CRC; 2015.

209. Hingorani A, Humphries S. Nature's randomised trials. Lancet (London, England). 2005 Dec;366(9501):1906–8.

210. Hemani G, Bowden J, Davey Smith G. Evaluating the potential role of pleiotropy in Mendelian randomization studies. Hum Mol Genet [Internet]. 2018/05/16. 2018 Aug 1;27(R2):R195–208. Available from: https://www.ncbi.nlm.nih.gov/pubmed/29771313

211. Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. Genet Epidemiol. 2013 Nov;37(7):658–65.

212. Al Kuwari H, Al Thani A, Al Marri A, Al Kaabi A, Abderrahim H, Afifi N, et al. The Qatar Biobank: background and methods. BMC Public Health [Internet]. 2015;15(1):1–9. Available from: http://dx.doi.org/10.1186/s12889-015-2522-7

213. Association AD. 2. Classification and Diagnosis of Diabetes. Diabetes Care [Internet]. 2015 Jan 1;38(Supplement 1):S8–16. Available from: http://care.diabetesjournals.org/content/38/Supplement_1/S8.short

214. Klipstein-Grobusch K, Georg T, Boeing H. Interviewer variability in anthropometric measurements and estimates of body composition. Int J Epidemiol. 1997;26 Suppl 1:S174-80.

215. Jenab Mazda, Bueno-de-Mesquita H Bas, Ferrari Pietro, van Duijnhoven Franzel J B, Norat Teresa PT et al. Associationbetween pre-diagnostic circulating vitamin D concentration and risk of colorectal cancer in European populations a nested case-control study.

216. Al-Delaimy WK, Jansen EHJM, Peeters PHM, van der Laan JD, van Noord PAH, Boshuizen HC, et al. Reliability of biomarkers of iron status, blood lipids, oxidative stress, vitamin D, C-reactive protein and fructosamine in two Dutch cohorts. Biomarkers Biochem Indic Expo response, susceptibility to Chem. 2006;11(4):370–82.

217. Aleksandrova K, Boeing H, Jenab M, Bas Bueno-de-Mesquita H, Jansen E, van Duijnhoven FJB, et al. Metabolic Syndrome and Risks of Colon and Rectal Cancer: The European Prospective Investigation into Cancer and Nutrition Study. Cancer Prev Res [Internet]. 2011 Nov 1;4(11):1873 LP-1883. Available from: http://cancerpreventionresearch.aacrjournals.org/content/4/11/1873.abstract

218. Schulze MB, Kroke A, Saracci R, Boeing H. The effect of differences in measurement procedure on the comparability of blood pressure estimates in multi-centre studies. Blood Press Monit. 2002 Apr;7(2):95–104.

219. Peters U. Quality control report for genotypic data. 2012.

220. Biobank U. About the UK Biobank. 2019.

221. Biobank UK. UK Biobank ethics and governance framework. Adswood; 2007.

222. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. Genome-wide genetic data on ~500,000 UK Biobank participants. bioRxiv [Internet]. 2017 Jan 1; Available from: http://biorxiv.org/content/early/2017/07/20/166298.abstract

223. Lips P. Relative Value of 25(OH)D and 1,25(OH)2D Measurements. J Bone Miner Res [Internet]. 2007 Nov 1;22(11):1668–71. Available from: http://dx.doi.org/10.1359/jbmr.070716

224. Gorham ED, Garland CF, Garland FC, Grant WB, Mohr SB, Lipkin M, et al. Optimal vitamin D status for colorectal cancer prevention: a quantitative meta analysis. Am J Prev Med [Internet]. 2007 Mar [cited 2015 Jun 10];32(3):210–6. Available from: http://www.sciencedirect.com/science/article/pii/S0749379706004983

225. Anderson JL, May HT, Horne BD, Bair TL, Hall NL, Carlquist JF, et al. Relation of vitamin D deficiency to cardiovascular risk factors, disease status, and incident events in a general healthcare population. Am J Cardiol [Internet]. 2010 Oct 1 [cited 2015 Aug 28];106(7):963–8. Available from: http://www.sciencedirect.com/science/article/pii/S0002914910011318

226. Forouhi NG, Ye Z, Rickard a. P, Khaw KT, Luben R, Langenberg C, et al. Circulating 25-hydroxyvitamin D concentration and the risk of type 2 diabetes: Results from the European Prospective Investigation into Cancer (EPIC)-Norfolk cohort and updated meta-analysis of Prospective studies. Diabetologia. 2012;55:2173–82.

227. Eliassen AH, Spiegelman D, Hollis BW, Horst RL, Willett WC, Hankinson SE. Plasma 25-hydroxyvitamin D and risk of breast cancer in the Nurses' Health Study II. Breast Cancer Res. 2011;13(3):R50.

228. Kayaniyil S, Harris SB, Retnakaran R, Vieth R, Knight JA, Gerstein HC, et al. Prospective association of 25(OH)D with metabolic syndrome. Clin Endocrinol (Oxf) [Internet]. 2014 Apr 1 [cited 2016 Jan 11];80(4):502–7. Available from: http://doi.wiley.com/10.1111/cen.12190

229. Lu Y, Liu M, Pei Y, Li J, Tian H, Cheng X, et al. Low levels of serum 25-hydroxyvitamin D and risk of metabolic syndrome in China. Int J Clin Exp Med [Internet]. 2015 Aug 15;8(8):13790–6. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4613012/

230. Fung GJ, Steffen LM, Zhou X, Harnack L, Tang W, Lutsey PL, et al. Vitamin D intake is inversely related to risk of developing metabolic syndrome in African American and white men and women over 20 y: the Coronary Artery Risk Development in Young Adults study. Am J Clin Nutr [Internet]. 2012 Jul 30;96(1):24–9. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3374731/

231. Vitezova A, Zillikens MC, van Herpt TTW, Sijbrands EJG, Hofman A, Uitterlinden AG, et al. Vitamin D status and metabolic syndrome in the elderly: the Rotterdam Study. Eur J Endocrinol [Internet]. 2015 Mar 1;172(3):327–35. Available from: http://www.eje-online.org/content/172/3/327.abstract

232. Amirbaigloo A, Hosseinpanah F, Sarvghadi F, Tohidi M, Eskandary PS, Azizi F. Absence of association between vitamin D deficiency and incident metabolic syndrome: Tehran Lipid and Glucose Study. Metab Syndr Relat Disord. 2013 Aug;11(4):236–42.

233. Chon SJ, Yun BH, Jung YS, Cho SH, Choi YS, Kim SY, et al. Association between vitamin D status and risk of metabolic syndrome among Korean postmenopausal women. PLoS One. 2014;9(2).

234. Pekkanen MP, Ukkola O, Hedberg P, Piira OP, Lepojärvi S, Lumme J, et al. Serum 25-hydroxyvitamin D is associated with major cardiovascular risk factors and cardiac structure and function in patients with coronary artery disease. Nutr Metab Cardiovasc Dis [Internet]. 2015 May [cited 2016 Jan 11];25(5):471–8. Available from: http://www.sciencedirect.com/science/article/pii/S0939475315000484

235. Park HY, Lim Y-H, Kim JH, Bae S, Oh S-Y, Hong Y-C. Association of Serum 25-Hydroxyvitamin D Levels with Markers for Metabolic Syndrome in the Elderly: A Repeated Measure Analysis. J Korean Med Sci [Internet]. 2012 Jun 26;27(6):653–60. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3369452/

236. Bener A, Zirie M, Musallam M, Khader YS, Al-Hamaq AOAA. Prevalence of Metabolic Syndrome According to Adult Treatment Panel III and International

Diabetes Federation Criteria: A Population-Based Study. Metab Syndr Relat Disord [Internet]. 2009 Mar 25;7(3):221–30. Available from: http://dx.doi.org/10.1089/met.2008.0077

237. Al Zenki S, Al Omirah H, Al Hooti S, Al Hamad N, Jackson RT, Rao A, et al. High prevalence of metabolic syndrome among Kuwaiti adults--a wake-up call for public health intervention. Int J Environ Res Public Health. 2012 May;9(5):1984–96.

238. Al-Lawati JA, Jousilahti P. Prevalence of metabolic syndrome in Oman using the International Diabetes Federation's Criteria. Saudi Med J. 2006;27(12):1925–6.

239. Hajat C, Shather Z. Prevalence of metabolic syndrome and prediction of diabetes using IDF versus ATPIII criteria in a Middle East population. Diabetes Res Clin Pract. 2012 Dec;98(3):481–6.

240. Al-Daghri NM, Al-Attas OS, Alokail MS, Alkharfy KM, Sabico SLB, Chrousos GP. Decreasing prevalence of the full metabolic syndrome but a persistently high prevalence of dyslipidemia among adult Arabs. PLoS One. 2010;5(8):e12159.

241. Wang TJ, Pencina MJ, Booth SL, Jacques PF, Ingelsson E, Lanier K, et al. Vitamin D Deficiency and Risk of Cardiovascular Disease. Circ [Internet]. 2008 Jan 29;117(4):503–11. Available from: http://circ.ahajournals.org/content/117/4/503.abstract

242. Pittas AG, Dawson-Hughes B, Li T, Van Dam RM, Willett WC, Manson JE, et al. Vitamin D and Calcium Intake in Relation to Type 2 Diabetes in Women. Diabetes Care [Internet]. 2006 Mar 1;29(3):650–6. Available from: http://care.diabetesjournals.org/content/29/3/650.abstract

243. Hoteit M, Al-Shaar L, Yazbeck C, Bou Sleiman M, Ghalayini T, Fuleihan GE-H. Hypovitaminosis D in a sunny country: time trends, predictors, and implications for practice guidelines. Metabolism [Internet]. 2014 Jul [cited 2015 Nov 9];63(7):968–78. Available from: http://www.sciencedirect.com/science/article/pii/S0026049514001176

244. Gagnon C, Lu ZX, Magliano DJ, Dunstan DW, Shaw JE, Zimmet PZ, et al. Low serum 25-hydroxyvitamin D is associated with increased risk of the development of the metabolic syndrome at five years: results from a national, population-based prospective study (The Australian Diabetes, Obesity and Lifestyle Study: AusDiab). J Clin Endocrinol Metab. 2012 Jun;97(6):1953–61.

245. Reis JP, von Mühlen D, Miller ER. Relation of 25-hydroxyvitamin D and parathyroid hormone levels with metabolic syndrome among US adults. Eur J Endocrinol [Internet]. 2008 Jul 1;159(1):41–8. Available from: http://www.eje-online.org/content/159/1/41.abstract

246. Maki KC, Fulgoni VL, Keast DR, Rains TM, Park KM, Rubin MR. Vitamin D Intake and Status Are Associated with Lower Prevalence of Metabolic Syndrome in U.S. Adults: National Health and Nutrition Examination Surveys 2003–2006.

Metab Syndr Relat Disord [Internet]. 2012 Jul 13;10(5):363–72. Available from: http://dx.doi.org/10.1089/met.2012.0020

247. Vitezova A, Voortman T, Zillikens MC, Jansen PW, Hofman A, Uitterlinden AG, et al. Bidirectional associations between circulating Vitamin D and cholesterol levels: The Rotterdam Study. Maturitas. 2015;82(4):411–7.

248. Majumdar V, Nagaraja D, Christopher R. Vitamin D status and metabolic syndrome in Asian Indians. Int J Obes (Lond). 2011 Aug;35(8):1131–4.

249. Reis JP, von Mühlen D, Kritz-Silverstein D, Wingard DL, Barrett-Connor E. Vitamin D, Parathyroid Hormone Levels, and the Prevalence of Metabolic Syndrome in Community-Dwelling Older Adults. Diabetes Care [Internet]. 2007 Jun 1;30(6):1549–55. Available from: http://care.diabetesjournals.org/content/30/6/1549.abstract

250. Zhang M, Li P, Zhu Y, Chang H, Wang X, Liu W, et al. Higher visceral fat area increases the risk of vitamin D insufficiency and deficiency in Chinese adults. Nutr Metab (Lond). 2015;12:50.

251. Wortsman J, Matsuoka LY, Chen TC, Lu Z, Holick MF. Decreased bioavailability of vitamin D in obesity. Am J Clin Nutr. 2000 Sep;72(3):690–3.

252. Awad AB, Alappat L, Valerio M. Vitamin D and Metabolic Syndrome Risk Factors: Evidence and Mechanisms. Crit Rev Food Sci Nutr [Internet]. 2012 Feb 1;52(2):103–12. Available from: http://dx.doi.org/10.1080/10408391003785458

253. Ardawi M-SM, Sibiany AM, Bakhsh TM, Qari MH, Maimani AA. High prevalence of vitamin D deficiency among healthy Saudi Arabian men: relationship to bone mineral density, parathyroid hormone, bone turnover markers, and lifestyle factors. Osteoporos Int [Internet]. 2011;23(2):675–86. Available from: http://dx.doi.org/10.1007/s00198-011-1606-1

254. Lu L, Yu Z, Pan A, Hu FB, Franco OH, Li H, et al. Plasma 25-Hydroxyvitamin D Concentration and Metabolic Syndrome Among Middle-Aged and Elderly Chinese Individuals. Diabetes Care [Internet]. 2009 Jul 1;32(7):1278–83. Available from: http://care.diabetesjournals.org/content/32/7/1278.abstract

255. Moy F-M, Bulgiba A. High prevalence of vitamin D insufficiency and its association with obesity and metabolic syndrome among Malay adults in Kuala Lumpur, Malaysia. BMC Public Health [Internet]. 2011;11(1):1–7. Available from: http://dx.doi.org/10.1186/1471-2458-11-735

256. Oosterwerff MM, Eekhoff EMW, Heymans MW, Lips P, van Schoor NM. Serum 25-hydroxyvitamin D levels and the metabolic syndrome in older persons: a population-based study. Clin Endocrinol (Oxf). 2011 Nov;75(5):608–13.

257. World Health Organization. Qatar Stepwise Report [Internet]. 2012. Available from: http://www.who.int/chp/steps/Qatar_2012_STEPwise_Report.pdf?ua=1

258. SNOJ J. Population of Qatar by nationality - 2017 report. Priya Dsouza communication. 2017.

259. Al-Dabhani K, Tsilidis KK, Murphy N, Ward HA, Elliott P, Riboli E, et al. Prevalence of Vitamin D deficiency and association with metabolic syndrome in a Qatari population. Nutr Diabetes. 2017;7(4).

260. Stocks T, Lukanova A, Bjørge T, Ulmer H, Manjer J, Almquist M, et al. Metabolic factors and the risk of colorectal cancer in 580,000 men and women in the metabolic syndrome and cancer project (Me-Can). Cancer [Internet]. 2011 Jun 1;117(11):2398–407. Available from: http://dx.doi.org/10.1002/cncr.25772

261. Pelucchi C, Negri E, Talamini R, Levi F, Giacosa A, Crispo A, et al. Metabolic syndrome is associated with colorectal cancer in men. Eur J Cancer [Internet]. 2010 Jul [cited 2015 Sep 16];46(10):1866–72. Available from: http://www.sciencedirect.com/science/article/pii/S0959804910002418

262. Sturmer T, Buring JE, Lee I-M, Gaziano JM, Glynn RJ. Metabolic abnormalities and risk for colorectal cancer in the physicians' health study. Cancer Epidemiol Biomarkers Prev. 2006 Dec;15(12):2391–7.

263. Ahmed RL, Schmitz KH, Anderson KE, Rosamond WD, Folsom AR. The metabolic syndrome and risk of incident colorectal cancer. Cancer [Internet]. 2006 Jul 1;107(1):28–36. Available from: http://dx.doi.org/10.1002/cncr.21950

264. Sharma P. Inflammation and the metabolic syndrome. Indian J Clin Biochem. 2011/11/01. 2011 Oct;26(4):317–8.

265. MacKinnon DP, Lockwood CM, Hoffman JM, West SG, Sheets V. A comparison of methods to test mediation and other intervening variable effects. Psychol Methods [Internet]. 2002;7(1):83–104. Available from: http://europepmc.org/abstract/MED/11928892

266. MacKinnon DP, Fairchild AJ, Fritz MS. Mediation analysis. Annu Rev Psychol. 2007;58:593–614.

267. VanderWeele TJ, Vansteelandt S. Odds ratios for mediation analysis for a dichotomous outcome. Vol. 172, American Journal of Epidemiology. 2010. p. 1339–48.

268. Jenab M, Bueno-de-Mesquita HB, Ferrari P, van Duijnhoven FJB, Norat T, Pischon T, et al. Association between pre-diagnostic circulating vitamin D concentration and risk of colorectal cancer in European populations:a nested case-control study. BMJ [Internet]. 2010 Jan 22;340. Available from: http://www.bmj.com/content/340/bmj.b5500.abstract

269. McKeown-Eyssen G. Epidemiology of colorectal cancer revisited: are serum triglycerides and/or plasma glucose associated with risk? Cancer Epidemiol Biomarkers & Prev [Internet]. 1994 Dec 1;3(8):687 LP-695. Available from: http://cebp.aacrjournals.org/content/3/8/687.abstract

270. Touvier M, Chan DSM, Lau R, Aune D, Vieira R, Greenwood DC, et al. Meta-analyses of vitamin D intake, 25-hydroxyvitamin D status, vitamin D receptor polymorphisms, and colorectal cancer risk. Cancer Epidemiol Biomarkers Prev. 2011 May;20(5):1003–16.

271. Theodoratou E, Palmer T, Zgaga L, Farrington SM, McKeigue P, Din FVN, et al. Instrumental variable estimation of the causal effect of plasma 25-hydroxy-vitamin D on colorectal cancer risk: A Mendelian randomization analysis. PLoS One. 2012;7(6).

272. He Y, Timofeeva M, Farrington SM, Vaughan-Shaw P, Svinti V, Walker M, et al. Exploring causality in the association between circulating 25-hydroxyvitamin D and colorectal cancer risk: a large Mendelian randomisation study. BMC Med [Internet]. 2018;16(1):142. Available from: https://doi.org/10.1186/s12916-018-1119-2

273. Barzilay JI, Abraham L, Heckbert SR, Cushman M, Kuller LH, Resnick HE, et al. The relation of markers of inflammation to the development of glucose disorders in the elderly: the Cardiovascular Health Study. Diabetes. 2001 Oct;50(10):2384–9.

274. Hamoya T, Fujii G, Miyamoto S, Takahashi M, Totsuka Y, Wakabayashi K, et al. Effects of NSAIDs on the risk factors of colorectal cancer: a mini review. Genes Environ [Internet]. 2016;38(1):6. Available from: https://doi.org/10.1186/s41021-016-0033-0

275. Brøndum-Jacobsen P, Benn M, Afzal S, Nordestgaard BG. No evidence that genetically reduced 25-hydroxyvitamin D is associated with increased risk of ischaemic heart disease or myocardial infarction: a Mendelian randomization study. Int J Epidemiol [Internet]. 2015 May 16; Available from: http://ije.oxfordjournals.org/content/early/2015/05/16/ije.dyv078.abstract

276. Ye Z, Sharp SJ, Burgess S, Scott RA, Imamura F, Consortium I, et al. Association between circulating 25-hydroxyvitamin D and incident type 2 diabetes: a mendelian randomisation study. Lancet Diabetes Endocrinol [Internet]. 2015 Jan;3(1):35–42. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4286815/

277. Manousaki D, Mokry LE, Ross S, Goltzman D, Richards JB. Mendelian randomization studies do not support a role for vitamin D in coronary artery disease. Circ Cardiovasc Genet. 2016;CIRCGENETICS-116.

278. Taylor AE, Burgess S, Ware JJ, Gage SH, Richards JB, Smith GD, et al. Investigating causality in the association between 25 (OH) D and schizophrenia. Sci Rep. 2016;6.

279. Mokry LE, Ross S, Morris JA, Manousaki D, Forgetta V, Richards JB. Genetically decreased vitamin D and risk of Alzheimer disease. Neurology. 2016;87(24):2567–74.

280. Rhead B, Bäärnhielm M, Gianfrancesco M, Mok A, Shao X, Quach H, et al. Mendelian randomization shows a causal effect of low vitamin D on multiple sclerosis risk. Neurol Genet. 2016;2(5):e97.

281. Ong J-S, Cuellar-Partida G, Lu Y, Fasching PA, Hein A, Burghaus S, et al. Association of vitamin D levels and risk of ovarian cancer: a Mendelian randomization study. Int J Epidemiol. 2016 Oct;45(5):1619–30.

282. Burgess S, Dudbridge F, Thompson SG. Combining information on multiple instrumental variables in Mendelian randomization: Comparison of allele score and summarized data methods. Stat Med. 2016;35(11):1880–906.

283. Hung RJ, Ulrich CM, Goode EL, Brhane Y, Muir K, Chan AT, et al. Cross cancer genomic investigation of inflammation pathway for five common cancers: Lung, ovary, prostate, breast, and colorectal cancer. J Natl Cancer Inst. 2015;107(11).

284. Cheng I, Kocarnik JM, Dumitrescu L, Lindor NM, Chang-Claude J, Avery CL, et al. Pleiotropic effects of genetic risk variants for other cancers on colorectal cancer risk: PAGE, GECCO and CCFR consortia. Gut. 2014;63(5):800–7.

285. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. Int J Epidemiol [Internet]. 2015 Jun 6; Available from: http://ije.oxfordjournals.org/content/early/2015/06/06/ije.dyv080.abstract

286. Burgess S, Bowden J, Fall T, Ingelsson E, Thompson SG. Sensitivity analyses for robust causal inference from mendelian randomization analyses with multiple genetic variants. Epidemiology. 2017.

287. Altman DG, Bland JM. How to obtain the P value from a confidence interval. BMJ [Internet]. 2011 Aug 8;343. Available from: http://www.bmj.com/content/343/bmj.d2304.abstract

288. Newman DL, Abney M, McPeek MS, Ober C, Cox NJ. The Importance of Genealogy in Determining Genetic Associations with Complex Traits. Am J Hum Genet [Internet]. 2001 Nov;69(5):1146–8. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1274359/

289. Lawlor DA. Commentary: Two-sample Mendelian randomization: opportunities and challenges. Int J Epidemiol [Internet]. 2016 Jun 17;45(3):908–15. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC5005949/

290. Manson JE, Bassuk SS, Lee I-M, Cook NR, Albert MA, Gordon D, et al. The VITamin D and OmegA-3 TriaL (VITAL): Rationale and design of a large randomized controlled trial of vitamin D and marine omega-3 fatty acid supplements for the primary prevention of cancer and cardiovascular disease. Contemp Clin Trials. 2012;33(1):159–71.

291. Manson JE, Cook NR, Lee I-M, Christen W, Bassuk SS, Mora S, et al. Vitamin D Supplements and Prevention of Cancer and Cardiovascular Disease. N Engl

J Med. 2018 Nov;380(1):33–44.

292. Wild SH, Byrne CD. The global burden of the metabolic syndrome and its consequences for diabetes and cardiovascular disease. Metab Syndr. 2005;1–41.

293. Soranzo N, Sanna S, Wheeler E, Gieger C, Radke D, Dupuis J, et al. Common variants at 10 genomic loci influence hemoglobin A(1)(C) levels via glycemic and nonglycemic pathways. Diabetes. 2010 Dec;59(12):3229–39.

294. Willer CJ, Schmidt EM, Sengupta S, Peloso GM, Gustafsson S, Kanoni S, et al. Discovery and refinement of loci associated with lipid levels. Nat Genet. 2013 Nov;45(11):1274–83.

295. Lab N. Updated GWAS analysis of the UK Biobank. 2018.

296. Manousaki D, Dudding T, Haworth S, Hsu Y-H, Liu C-T, Medina-Gomez C, et al. Low-Frequency Synonymous Coding Variation in CYP2R1 Has Large Effects on Vitamin D Levels and Risk of Multiple Sclerosis. Am J Hum Genet. 2017 Aug;101(2):227–38.

297. Carter AR, Gill D, Davies NM, Taylor AE, Tillmann T, Vaucher J, et al. What explains the effect of education on cardiovascular disease? Applying Mendelian randomization to identify the consequences of education inequality. bioRxiv [Internet]. 2018; Available from: https://www.biorxiv.org/content/early/2018/12/11/488254

298. Lab N. UKBiobank Imputed V3 Phenotype [Internet]. 2017. Available from: https://github.com/Nealelab/UK_Biobank_GWAS#imputed-v3-phenotypes

299. Shim H, Chasman DI, Smith JD, Mora S, Ridker PM, Nickerson DA, et al. A multivariate genome-wide association analysis of 10 LDL subfractions, and their response to statin treatment, in 1868 Caucasians. PLoS One. 2015;10(4):e0120758.

300. Wheeler E, Leong A, Liu C-T, Hivert M-F, Strawbridge RJ, Podmore C, et al. Impact of common genetic determinants of Hemoglobin A1c on type 2 diabetes risk and diagnosis in ancestrally diverse populations: A transethnic genome-wide meta-analysis. PLoS Med. 2017 Sep;14(9):e1002383.

301. Hartwig FP, Davey Smith G, Bowden J. Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. Int J Epidemiol. 2017 Dec;46(6):1985–98.

302. Verbanck M, Chen C-Y, Neale B, Do R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. Nat Genet [Internet]. 2018;50(5):693–8. Available from: https://doi.org/10.1038/s41588-018-0099-7

303. Burgess S, Thompson SG. Multivariable Mendelian Randomization: The Use of

Pleiotropic Genetic Variants to Estimate Causal Effects. Am J Epidemiol [Internet]. 2015 Feb 15;181(4):251–60. Available from: http://dx.doi.org/10.1093/aje/kwu283

304. Robien K, Oppeneer SJ, Kelly JA, Hamilton-Reeves JM. Drug-vitamin D interactions: a systematic review of the literature. Nutr Clin Pract. 2013/01/10. 2013 Apr;28(2):194–208.

305. Gislefoss RE, Grimsrud TK, Mørkrid L. Long-term stability of serum components in the Janus Serum Bank. Scand J Clin Lab Invest [Internet]. 2008 Jan 1;68(5):402–9. Available from: https://doi.org/10.1080/00365510701809235

306. Kubiak J, Thorsby PM, Kamycheva E, Jorde R. Vitamin D supplementation does not improve CVD risk factors in vitamin D-insufficient subjects. Endocr Connect [Internet]. 2018 May 15;7(6):840–9. Available from: https://www.ncbi.nlm.nih.gov/pubmed/29764903

307. Wang Y, Si S, Liu J, Wang Z, Jia H, Feng K, et al. The Associations of Serum Lipids with Vitamin D Status. PLoS One. 2016;11(10):e0165157.

308. Hutchinson MS, Figenschau Y, Njølstad I, Schirmer H, Jorde R. Serum 25-hydroxyvitamin D levels are inversely associated with glycated haemoglobin (HbA1c). The Tromsø Study. Scand J Clin Lab Invest [Internet]. 2011 Sep 1;71(5):399–406. Available from: https://doi.org/10.3109/00365513.2011.575235

309. Kositsawat J, Freeman VL, Gerber BS, Geraci S. Association of A1C Levels With Vitamin D Status in U.S. Adults. Diabetes Care [Internet]. 2010 Jun 1;33(6):1236 LP-1238. Available from: http://care.diabetesjournals.org/content/33/6/1236.abstract

310. Zittermann A, Frisch S, Berthold HK, Gotting C, Kuhn J, Kleesiek K, et al. Vitamin D supplementation enhances the beneficial effects of weight loss on cardiovascular disease risk markers. Am J Clin Nutr. 2009 May;89(5):1321–7.

311. Sneve M, Figenschau Y, Jorde R. Supplementation with cholecalciferol does not result in weight reduction in overweight and obese subjects. Eur J Endocrinol. 2008 Dec;159(6):675–84.

312. Lu Y, Day FR, Gustafsson S, Buchkovich ML, Na J, Bataille V, et al. New loci for body fat percentage reveal link between adiposity and cardiometabolic disease risk. Nat Commun. 2016 Feb;7:10495.

APPENDIX

Appendix A

Supplemental Table 5.1. Removed systolic blood pressure-associated SNPs and their association with other phenotypes from Phenoscanner

SNP	Phenotype	SNP	Phenotype
rs900144	Age at menarche	rs10493818	Height
rs692155	Age at menopause	rs1057040	Height
rs7914912	Age-related macular degeneration	rs10781976	Height
rs204883	Age-related macular degeneration, height, rheumatoid arthritis	rs11669336	Height
rs12339434	Body mass index	rs1173771	Height
rs2222544	Body mass index	rs11783683	Height
rs2977334	Body mass index	rs13131350	Height
rs6734118	Body mass index	rs13359423	Height
rs9636202	Body mass index	rs13412750	Height
rs12714414	Body mass index, age at menarche	rs17540044	Height
rs732998	Body mass index, CAD, Schizophrenia	rs2014615	Height
rs3933469	Circulating sex hormone binding globulin levels	rs2016520	Height
rs10770612	Coronary artery disease	rs2048098	Height
rs10781976	Coronary artery disease	rs2301597	Height
rs10781976	Coronary artery disease	rs3211995	Height
rs11079849	Coronary artery disease	rs4792819	Height
rs1250258	Coronary artery disease	rs6749818	Height
rs12643599	Coronary artery disease	rs6856448	Height
rs13062241	Coronary artery disease	rs887258	Height
rs1458038	Coronary artery disease	rs979532	Height
rs1867624	Coronary artery disease	rs981037	Height
rs2107595	Coronary artery disease	rs1872167	Height, HDL
rs2493291	Coronary artery disease	rs10269774	Height, waist circumference
rs360156	Coronary artery disease	rs10224210	Hematocrit
rs4691707	Coronary artery disease	rs1290784	Hematocrit
rs604723	Coronary artery disease	rs1470260	Hemoglobin concentration
rs8102876	Coronary artery disease	rs2071303	Hemoglobin concentration
rs2571445	Coronary artery disease, FEV	rs1052486	IgA deficiency, height, weight, T1 rheumatoid arthritis, age at menopause, schizophrenia
rs2681492	Coronary artery disease, MI	rs9349379	Migraine, CAD, MI
rs10774625	Coronary artery disease, myocardial infarction BMI, LDL, TC, DBP, SBP, HbA1c	rs10092781	Neuroticism
rs9875617	Crohns disease, IBD, ulcerative colitis age at menarche, years of education	rs2853950	Psoriasis, ulcerative colitis
rs7070797	Diastolic blood pressure	rs2507975	Rheumatoid arthritis
rs1408945	Eosinophil count	rs3749953	Rheumatoid arthritis
rs2291433	Eosinophil percentage of white cells	rs1763839	Schizophrenia

rs7255	Esophageal adenocarcinoma	rs10858938	Serum magnesium
rs10501320	Fasting blood glucose, T2D	rs1557765	Type II diabetes, BMI
rs10409243	Granulocyte count	rs3134798	Ulcerative colitis
rs1757915	Granulocyte count	rs747472	Waist circumference adjusted for BMI
rs12801636	HDL, CAD	rs2295680	Years of educational attainment
		rs9320747	Years of educational attainment i females

Supplemental Table 5.2. Removed high density lipoprotein-cholesterol-associated SNPs and their association with other phenotypes from Phenoscanner

SNP	Phenotype	SNP	Phenotype
rs2454722	adiponectin	rs247616	LDL, total cholesterol, triglycerides, CVD
rs731839	adiponectin, triglycerides, fasting insulin, BMI	rs102275	LDL, total cholesterol, triglycerides, heart rate, fasting glucose crohns
rs863750	adiponectin, waist-hip-ratio	rs6711016	LDL, total cholesterol, triglycerides, MetS,
rs7973683	adiponectin, waist-hip-ratio, triglycerides	rs12748152	LDL, triglycerides
rs7203984	age-related macular degeneration, low density lipoprotein, total cholesterol, triglycerides	rs2250802	LDL, triglycerides, total cholesterol
rs5167	Alzheimer's	rs1980493	rheumatoid arthritis, T1D
rs1877031	asthma	rs2606736	SBP
rs13107325	BMI, Crohn's, schizophrenia	rs1515110	T2D, BMI, triglycerides
rs205262	BMI, hip circumference	rs1121980	T2D, obesity
rs11057405	BMI, hip circumference, adiponectin	rs7607980	T2D, waist-hip-ratio, triglycerides, fasting insulin, HOMA-IR
rs9956279	BMI, hip circumference, waist circumference	rs10773003	total cholesterol
rs4752801	BMI, height, proinsulin	rs11789603	total cholesterol
rs2075650	BMI, waist circumference, low density lipoprotein, total cholesterol, CRP	rs12965544	total cholesterol
rs12801636	CAD	rs1787328	total cholesterol
rs6567160	CAD, BMI, height, hip circumference,	rs2148489	total cholesterol
rs12740374	CAD, MI, LDL, total cholesterol,	rs2230808	total cholesterol
rs653178	CAD, MI, LDL, total cholesterol, blood pressure	rs2293889	total cholesterol
rs301	CAD, triglycerides	rs2899624	total cholesterol
rs2013208	Crohn's, years of education	rs3780543	total cholesterol
rs6450176	fasting insulin	rs3890182	total cholesterol
rs3822072	fasting insulin, BMI	rs4121823	total cholesterol
rs3847502	height	rs6507945	total cholesterol
rs7947811	height	rs686030	total cholesterol
rs11758426	height, total cholesterol,	rs7117842	total cholesterol
rs2844513	height, total cholesterol, rheumatoid arthritis	rs7229377	total cholesterol
rs4775039	LDL	rs7241596	total cholesterol
rs1800978	LDL, total cholesterol	rs930991	total cholesterol
rs2642438	low density lipoprotein, total cholesterol	rs970548	total cholesterol
rs17135399	low density lipoprotein, total cholesterol,	rs103294	total cholesterol, prostate cancer
rs4803760	low density lipoprotein, total cholesterol, Alzheimer's	rs17092642	total cholesterol, triglycerides
rs9987289	LDL, total cholesterol, fasting glucose, CRP	rs17120244	total cholesterol, triglycerides
rs12279373	LDL, total cholesterol, triglycerides,	rs486394	total cholesterol, triglycerides
rs3825041	LDL, total cholesterol, triglycerides,	rs4938353	total cholesterol, triglycerides
rs157580	LDL, total cholesterol, triglycerides, Alzheimer's	rs7012891	total cholesterol, triglycerides
rs445925	LDL, total cholesterol, triglycerides, atherosclerosis, CVD, CAD	rs9644636	triglycerides
rs12908474	total cholesterol, LDL		
rs7170361	total cholesterol, triglycerides, LDL	rs10503666	triglycerides

SNP	Phenotype	SNP	Phenotype
rs687339	triglycerides, fibrinogen	rs10761771	triglycerides
rs6065906	triglycerides, low density lipoprotein	rs11204072	triglycerides
rs17145738	triglycerides, serum urate	rs16842	triglycerides
rs3741414	triglyceride, serum urate	rs1866956	triglyceride
rs180360	triglyceride, total cholesterol, crohns, IBD	rs2144300	triglyceride
rs2925979	waist-hip-ratio, adiponectin,	rs2652834	triglyceride
rs459193	waist-hip-ratio, T2D, fasting insulin	rs283	triglyceride
rs1936800	waist-hip-ratio, triglyceride	rs3111576	triglyceride
rs998584	waist-hip-ratio, triglyceride, adiponectin,	rs378114	triglyceride
rs10808546	low density lipoprotein, total cholesterol, triglyceride, CAD	rs3936511	triglyceride
rs10199768	low density lipoprotein, total cholesterol, triglyceride, CVD	rs4244457	triglyceride
rs603441	total cholesterol, fasting glucose	rs442177	triglyceride
rs12185072	total cholesterol, low density lipoprotein	rs492571	triglyceride
rs1531517	total cholesterol, low density lipoprotein, alzheimers	rs7005359	triglyceride
rs1800961	total cholesterol, low density lipoprotein, CRP	rs7010610	triglyceride
rs4917014	Lupus Erythematosus Systemic	rs7014168	triglyceride
rs3861397	Mean corpuscular hemoglobin concentration MChip circumference, triglyceride	rs7016529	triglyceride
		rs7942717	triglyceride

Supplemental Table 5.3. Removed glycated haemoglobin A1c-associated SNPs and their association with other phenotypes from Phenoscanner

SNP	Phenotype	SNP	Phenotype
rs11708067	birthweight	rs2210366	RBC,
rs17747324	BMI	rs2110073	Red blood cell fatty acid levels
rs9935401	BMI	rs10946800	Rheumatoid arthritis
rs198846	BP	rs267738	serum creatinine
rs10774625	CAD, MI	rs10454142	SHBG
rs579459	CAD, MI	rs11248914	Mean corpuscular volume
rs9818758	crohns disease, IBD	rs11964178	Mean corpuscular volume
rs837763	Mean corpuscular hemoglobin concentration	rs1547247	Mean corpuscular volume
rs2246434	Mean corpuscular hemoglobin concentration	rs1387153	MetS
rs4737009	Mean corpuscular hemoglobin concentration	rs4745982	hematocrit
rs5009712	Mean corpuscular hemoglobin concentration	rs17509001	height
rs2160906	Mean corpuscular hemoglobin	rs3778279	height
rs9389272	Mean corpuscular hemoglobin	rs855791	iron status
rs592423	HDL, triglyceride	rs7616006	LDL, TC
		rs1800562	LDL, TC, iron

Supplemental Table 5.4. Removed BMI-associated SNPs and their association with other Phenotypes from Phenoscanner

SNP	Phenotype	SNP	Phenotype
rs13665	Age at menarche	rs16973520	Low density lipoprotein, TC,
rs7258722	Age at menarche	rs34341	Low density lipoprotein, TC, BMI
rs900144	Age at menarche	rs12936083	Lymphocyte count
rs9515448	Age at menarche	rs3803286	Lymphocyte count
rs1064213	Allergic disease asthma hay fever or eczema, chronotype	rs7578575	Lymphocyte count
rs12735595	BMI, age at menarche	rs2183947	Mean corpuscular hemoglobin
rs13130484	BMI, age at menarche	rs741959	Mean corpuscular hemoglobin
rs1488830	BMI, age at menarche	rs7626079	Mean corpuscular hemoglobin
rs3784710	BMI, age at menarche	rs799449	Mean corpuscular hemoglobin
rs6604866	BMI, age at menarche	rs6831020	Monocyte count
rs7138803	BMI, age at menarche	rs4240673	Neuroticism
rs891387	BMI, educational attainment	rs215634	Nicotine dependence smoking cigarettes per day
rs11672660	BMI, fasting glucose	rs912690	Obesity class 2
rs7124681	BMI, HDL, proinsulin	rs1711171	Plasma fibrinogen, CAD
rs244415	BMI, height, age at menarche	rs2744957	Platelet count, BMI, HDL
rs7903146	BMI, T2D, fasting glucose	rs6725931	Platelet distribution width
rs2072858	BMI, uterine fibroids	rs8008285	Plateletcrit
rs10899736	Childhood acute lymphoblastic leukaemia	rs2281819	Rheumatoid arthritis
rs2472297	Coffee consumption	rs7925214	Schizophrenia
rs11079849	Coronary artery disease	rs3814883	Schizophrenia, age at menarche
rs2681780	Crohn's disease, Inflammatory bowel disease, ulcerative colitis, education	rs11675464	Serum butyrylcholinesterase activity
rs11611246	Eosinophil count	rs11976018	Serum dehydroepiandrosterone sulphate DHEAS
rs4237643	Forced vital capacity	rs157582	Triglycerides, C-reative protein, LDL, Alzheimers disease, MetS, coronary artery disease, posterior cortical atrophy
rs4148155	Gout, serum urate	rs6479905	Triglycerides, educational attainment
rs10878946	Granulocyte percentage of myeloid white cells	rs2242189	White blood cell count
rs7158822	Granulocyte percentage of myeloid white cells	rs6502482	White blood cell count
rs10790519	Granulocyte percentage of myeloid white cells, lymphocyte count, monocyte count, TC	rs11161044	Years of educational attainment
rs11084553	Height	rs12512994	Years of educational attainment
rs2450444	Height	rs6542924	Years of educational attainment
rs2791643	Height	rs6905544	Years of educational attainment
rs4942924	Height	rs903959	Years of educational attainment
rs4952843	Height	rs17379561	Years of educational attainment, schizophrenia
rs10515237	Height, fasting glucose, proinsulin	rs2694047	High light scatter percentage of red cells,
rs2284746	Height, wc, FEV,	rs1320251	Immature fraction of reticulocytes
rs4660443	High density lipoprotein	rs2253310	Intelligence
rs7084454	High grade serous ovarian cancer	rs1421334	Intelligence multi trait analysis

SNP	Phenotype	SNP	Phenotype
rs12369179	High light scatter percentage of red cells	rs11659764	Intraocular pressure

Appendix B

Difference of coefficient method

1- The SE_{both} was calculated

$$sqrt(\sigma^2_{WM} * \beta_M) + (\sigma^2_M * \beta_{WM}))$$

- WM: without mediator
- M: with mediator

2- Calculate the ratio (WM-M)/SE_{both}
3- This ratio was then compared to the t-distribution with df ~2000, t=-1.96 & 1.96 (similar to normal distribution due to large sample size)
4- P-value was calculated from the t-distribution

Product of coefficient method

1- Extract β (M~ βx) and Θ (Y~ βx + Θm…) from the separate regression models as well as the SE
2- Multiply βΘ
3- Calculate the SE_{both}

$$sqrt(\sigma^2_{WM} * \beta_M) + (\sigma^2_M * \beta_{WM}))$$

4- Divide βΘ/ SE_{both}
5- This was then compared with the z-distribution
6- P-value was calculated from the z-distribution

Counterfactual method using Stata

xi: paramed Caseclrt_frst, avar(vitd_cat3) mvar(casemets) cvars(i.smok Age_blood i.Sex i.alc_all) a0(0) a1(1) m(1) yreg(logistic) mreg(logistic) case

Appendix C

Permissions summary table for third party copyright works

Page No.	Type of work:	Name of work	Source of work	Copyright holder and contact	permission requested on	I have permission yes /no	Permission note
37	Figure	**Figure 1.1** Vitamin D production and metabolism	Vitamin D and risk of multiple sclerosis: a Mendelian randomization study.	https://openclipart.org/. https://doi.org/10.1371/journal.pmed.1001866.g002	September 10th 2018	Yes	Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited
39	Table	Table 1.1 Dietary and supplemental sources of Vitamin D	Vitamin D Deficiency	Massachusetts Medical Society (MMS), Publisher of the *New England Journal of Medicine*.	September 10th 2018	Yes	Reproduced with permission from (scientific reference citation), Copyright Massachusetts Medical Society.
	Table 1.4	WHO, EGIR, NCEP-ATP III and IDF definitions of metabolic syndrome	Metabolic syndrome and benign prostatic hyperplasia: An update	Editorial Office of Asian Journal of Urology. Production and hosting by Elsevier B.V	April 8th 2019	Yes	Creative Commons Attribution 4.0 International License
62	Figure	**Figure 1.2**. Comparison of a randomised controlled trial and Mendelian Randomisation.	Nature's randomised trials.	Copyright © 2005 Elsevier Ltd	September 10th 2018	Yes	The Lancet, Vol. number: 366, Author(s): Aroon Hingorani and Steve Humphries, Title of article: Nature's randomised trials, Pages No.: 1906 - 1908, Copyright (2005) License number: 4425430361847

Page No.	Type of work:	Name of work	Source of work	Copyright holder and contact	permission requested on	I have permission yes /no	Permission note
64	Figure	**Figure 1.4.** Linkage disequilibrium that leads to violation of the instrumental variable analysis. Z:G_1: genetic variant that is being used as the instrumental variable; G_2: genetic variant in linkage disequilibrium with $G_1$and related to Y ; X: exposure of interest; Y: outcome of interest; C: confounders	Mendelian randomization: Using genes as instruments for making causal inferences in epidemiology	Wiley Materials	September 11th 2018	Yes	Debbie A. Lawlor, Roger M. Harbord, Jonathan A. C. Sterne, et al., Title of article: Mendelian randomisation: Using genes as instruments for making causal inferences in epidemiology, Publishers: John Wily and Sons, License number: 4425990212638
84	Table	**Table 2.1.** The 66 clinical biomarkers routinely measured in the Qatar Biobank	The Qatar Biobank: background and methods	© Al Kuwari et al. 2015	November 28th 2018	Yes	Creative Commons Attribution 4.0 International License
85	Table	**Table 2.2.** Main characteristics of Qatar Biobank study participants stratified by sex	Prevalence of vitamin D deficiency and association with metabolic syndrome in a Qatari population		December 12th 2018	Yes	Creative Commons Attribution 4.0 International License
86	Table	**Table 2.3.** Unadjusted and adjusted mean 25-hydroxyvitamin D levels by participant characteristics in nmol/L in the Qatar Biobank	Prevalence of vitamin D deficiency and association with metabolic syndrome in a Qatari population		December 12th 2018	Yes	Creative Commons Attribution 4.0 International License
87	Table	**Table 2.4.** Linear regression analyses between vitamin D and	Prevalence of vitamin D deficiency and association with		December 12th 2018	Yes	Creative Commons Attribution 4.0 International License

Page No.	Type of work:	Name of work	Source of work	Copyright holder and contact	permission requested on	I have permission yes /no	Permission note
		metabolic syndrome and its components in the Qatar Biobank	metabolic syndrome in a Qatari population				
88	Table	**Table 2.5.** Logistic regression analyses between -- and its components with vitamin D deficiency	Prevalence of vitamin D deficiency and association with metabolic syndrome in a Qatari population		December 12th 2018	Yes	Creative Commons Attribution 4.0 International License

Appendix D

Core Questions on Physical Activity in EPIC Baseline Questionnaires

1. Work

We would like to know the type and amount of physical activity involved in your work. Please check what the best corresponds with your present occupation from the following four possibilities:

- Sedentary occupation ____

You spend most of your time sitting (such as in an office)

- Standing occupation____

  You spend most of your time standing and walking. However, your work does not require intense physical effort (e.g. shop assistant, hairdresser, guard, etc.)
- Manual work____

  This involves some physical effort including handling of heavy objects and use of tools (e.g. plumber, electrician, carpenter, etc.)
- Heavy manual work____

  This implies very vigorous physical activity including handling of very heavy objects (e.g. docker, miner, bricklayer, construction worker, etc.)

2. In a typical week during the past year, how many hours did you spend per week on each of the following activities:

- walking, including walking to work, shopping and leisure time
  in summer ____ hours per week
  in winter ____ hours per week
- cycling, including cycling to work, shopping and leisure time
  in summer ____ hours per week
  in winter ____ hours per week
- gardening
  in summer ____ hours per week
  in winter ____ hours per week
- do-it-yourself activities at home
  ____ hours per week
- physical exercise such as fitness, aerobics, swimming, jogging, tennis, etc.
  in summer ____ hours per week
  in winter ____ hours per week
- housework, such as cleaning, washing, cooking, child care, etc.
  ____ hours per week

3. In a typical week during the past year, did you engage in any of these activities vigorously enough to cause sweating or faster heartbeat?

No ___Yes ___

If yes, for how many hours per week in total did you perform vigorous activity?

____ hours per week

4. In a typical week during the past year, how many flights of stairs did you climb per day?

___ floors per day

CORE QUESTIONNAIRE ON TOBACCO SMOKING

1. Do you currently smoke?
1.1 Yes, I smoke cigarettes
1.2 Yes, I smoke cigars
1.3 Yes, I smoke a pipe
1.4 No, I have never smoked
1.5 No, I smoked in the past, but I no longer smoke

CURRENT SMOKERS
If you smoke cigarettes:
1. How many cigarettes do you smoke per day?

2. Do you smoke cigarettes
predominantly with filter
predominantly without filter
in between, some with and some without filter

3. Do you usually inhale the cigarette smoke?
Yes, deeply
I inhale a little
I do not inhale

4.1 Do you usually smoke cigarettes made with:
blond tobacco
black tobacco
mixture
I do not know

4.2 Do you usually smoke low tar (light or ultra light cigarettes)?
Yes
No

5. At what age did you start smoking cigarettes?

6 How many cigarettes per day did you usually smoke at the following ages. Indicate with a cross whether you smoked mainly filter or non-filter cigarettes?:
6.1 When you were about 20 years old
6.2 When you were about 30 years old
6.3 When you were about 40 years old
6.4 When you were about 50 years old

7. Did you give up smoking at any period of your life?
Yes
No

8. If yes, could you indicate from what age to what age you interrupted smoking?
8.1 1st interruption of smoking from age to age
8.2 2nd interruption of smoking from age to age
8.3 3rd interruption of smoking from age to age

9. How many cigars do you smoke per week?
   Do you usually smoke? Large cigars
   Medium cigars
   Small cigars

10. Do you usually inhale the cigar smoke?
    Yes, deeply
    I inhale a little
    I do not inhale

11. How many full pipes do you smoke per week?
    How much tobacco (ounces per week or packs per week) do you smoke?

12. Do you inhale the pipe smoke?
    Yes, deeply
    I inhale a little
    I do not inhale

EX-SMOKERS
Cigarettes
1. At what age did you start smoking cigarettes?

2. At what age did you give up?

3. How many cigarettes per day did you usually smoke at the following ages. Indicate with a cross whether you smoked mainly filter or non-filter cigarettes?:
   3.1 When you were about 20 years old
   3.2 When you were about 30 years old
   3.3 When you were about 40 years old
   3.4 When you were about 50 years old

Cigars
1. At what age did you start smoking cigars?

2. If you no longer smoke cigars, at what age did you give up?

3. How many cigars per day did you usually smoke when you were:
   About 20 years old
   About 30 years old
   About 40 years old
   About 50 years old

Pipe
1. At what age did you start smoking a pipe?

2. If you no longer smoke a pipe, at what age did you give up?

3. How many pipes per day did you usually smoke when you were:
   About 20 years old

About 30 years old
About 40 years old
About 50 years old

OCCASIONAL SMOKERS

1. Did you ever try to smoke?
   Yes
   No

2. Have you ever smoked cigarettes, cigars, pipe, even occasionally on social occasions and/or during a particular period of your life?
   Yes
   No

3. For how many years did you smoke occasionally?

4. Did you inhale tobacco smoke?

4.1 Yes, deeply
4.2 I inhaled a little
4.3 I did not inhale

EPIC-core questions on consumption of alcoholic beverages in the past (questionnaire of non-dietary aspects)

We would like to know your alcohol consumption at certain periods of your life

(Fill in only one box per line):

Mark with cross or if consumption: If you drank regularly indicate usual number of glasses:

None or Occasional (less than one Per week or Per day

Glass per week)

Wine

When you were about 20 years old

When you were about 30 years old

When you were about 40 years old

Beer or cider

When you were about 20 years old

When you were about 30 years old

When you were about 40 years old

Fortified wine (e.g. port, sherry,

vermouth, martini)

When you were about 20 years old

When you were about 30 years old

When you were about 40 years old

Spirits or liqueur

When you were about 20 years old

When you were about 30 years old

When you were about 40 years old