Analysing the association of vitamin D status on selected cardiovascular risk markers using seasonal and genetic variations

A thesis presented for the degree of Doctor of Philosophy

UCL

Diane Joyce Berry

UCL Institute of Child Health

2012

Declaration

I, Diane Joyce Berry confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

…………………………….

Abstract

Vitamin D deficiency is common and has been proposed as a risk factor for cardiovascular disease (CVD), but much of this evidence is inconsistent. The aim of the thesis was to explore the associations between vitamin D status (25(OH)D) and selected risk biomarkers of CVD, in participants of the British 1958 birth cohort free from CVD. Different methodologies were used in an attempt to avoid confounded associations. Mediation analysis was used to infer an association between 25(OH)D and biomarkers from seasonal variations. A genome-wide association study (GWAS) was done to find single nucleotide polymorphisms (SNPs) associated with 25(OH)D. SNPs found by the GWAS and SNPs in candidate genes were evaluated as proxy markers for 25(OH)D, and used in Mendelian randomisation (MR) analysis with the biomarkers. Higher 25(OH)D concentrations were associated with lower levels of tissue plasminogen activator (tPA), after adjusting for lifestyle, socio-economic and adiposity covariates. An association between 25(OH)D and tPA was inferred using mediation analysis. In the GWAS, SNPs from genes involved in the synthesis, hydroxylation and transportation of vitamin D were associated with 25(OH)D. SNPs passing evaluation as proxy markers for 25(OH)D were classified as "synthesis" and "metabolism" based on gene function, with the former considered to be a more robust proxy than the latter. Statistical power to detect an association was limited in MR analysis. However, there was some evidence that 25(OH)D had a protective association with tPA, when metabolism SNPs were used as proxy for 25(OH)D in MR analysis. Findings from the different analyses were inconsistent for CRP, D-dimer and fibrinogen. In conclusion, the findings tentatively suggest that vitamin D has a beneficial influence on CVD through the mechanism of fibrinolysis.

However, more evidence is required from large MR studies and randomised controlled trials before the role of vitamin D in CVD is conclusively understood.

Table of Contents

Declaration **2**

Abstract **3**

Abbreviations and acronyms **14**

Acknowledgements **17**

1. Vitamin D: introduction, association with cardiovascular disease and aim of the thesis **19**

1.1 Background **19**

1.1.1 History of vitamin D 19

1.1.2 Vitamin D intake 20

1.1.3 Metabolism of vitamin D 21

1.1.4 Thresholds of 25-hydroxyvitamin D 24

1.1.5 Environmental determinants of vitamin D status 28

1.1.6 Personal determinants of vitamin D status 30

1.1.7 Genetic determinants of vitamin D status 32

1.1.8 Levels of 25-hydroxyvitamin D in the UK 33

1.2 Cardiovascular disease **36**

1.2.1 Risk markers of CVD 39

1.2.2 Haemostasis and thrombosis 40

1.3 Vitamin D and cardiovascular disease **46**

1.4 Rationale **55**

1.5 Aim **57**

1.5.1 Objectives 57

1.6 Overview of the thesis and related papers **58**

2 Data source **63**

2.1 Introduction **63**

2.2 Biomedical survey **64**

2.2.1 Phenotype data 66

2.2.2 Genetic data 73

3 The seasonal variation of inflammatory and haemostatic biomarkers due to the seasonal variation of 25-hydroxyvitamin D (Paper I) 91

3.1 Introduction 91

3.2 Data 93

3.2.1 Participants 93

3.2.2 Laboratory analyses 94

3.2.3 Lifestyle, socio-economic and adiposity factors 94

3.3 Statistical Methods 95

3.3.1 Descriptive methods 95

3.3.2 Multiple linear regression models 96

3.3.3 Mediation Analysis 101

3.3.4 The mediation product of coefficients test incorporating seasonal patterns 111

3.4 Results 121

3.4.1 Characteristics of study participants and multiple linear regression results 122

3.4.2 Mediation analysis results 132

3.5 Discussion 139

3.5.1 Comparisons with other studies 139

3.5.2 Explanations 142

3.5.3 Methodological considerations 144

4 Genome-wide association study on vitamin D status (Paper II) 148

4.1 Introduction 148

4.2 Methods 154

4.2.1 Participants 154

4.2.2 Procedures 154

4.2.3 Statistical methods 155

4.3 Results 163

4.3.1 Meta-analyses results 165

4.3.2 Associations of SNPs with 25(OH)D and 25(OH)D deficiency/insufficiency in the 1958BC 174

4.3.3 Candidate gene results 178

4.4 Discussion 179

5 A Mendelian randomisation study on 25-hydroxyvitamin D and inflammation/haemostatic biomarkers in the British1958 birth cohort (Papers III & IV) 186

5.1 Introduction 186

5.2 Genetic determinants of 25-hydroxyvitamin D for instruments in Mendelian randomisation 192

5.2.1 Candidate genes from hair and skin colour, tanning GWAS for instruments of 25(OH)D 202

5.2.2 Summary of genetic variants considered as instruments for 25(OH)D in MR 204

5.3 Methods 205

5.3.1 Participants 205

5.3.2 Procedures 205

5.3.3 Lifestyle and socio-economic covariates 206

5.3.4 Selection of SNPs, candidate genes and genotyping procedures 207

5.3.5 Statistical Methods 208

5.4 Results 222

5.4.1 Evaluation of the genetic instruments 222

5.4.2 Power calculation and MR simulation study 232

5.4.3 IV regression of 25(OH)D with haemostatic/inflammation markers using genetic instruments 234

5.5 Discussion 240

5.5.1 Methodological considerations 240

5.5.2 Explanations 245

5.5.3 Conclusions 250

6 Discussion 252

6.1 Conclusions 252

6.1.1 Judging the evidence 253

6.1.2 Future work 261

References 263

Appendix 1: Publications 303

Appendix 2: Additional tables for Chapter 3 349

Appendix 3: Additional figures for Chapter 4 362

Appendix 4: Additional methods for Chapter 5..366

Appendix 5: Additional figures for Chapter 5 377

Appendix 6: Additional tables for Chapter 5 383

List of tables

Table 2.1 Number of cohort members surveyed in 1958BC biomedical sweep and percentage of final response achieved for selected body measures and blood samples 65

Table 2.2 Biomarker data used in the analyses of this thesis assayed from blood samples taken for the biomedical survey 67

Table 2.3 Lifestyle, socio-economic and anthropometric covariates used in the analyses of this thesis 69

Table 2.4 De novo SNPs genotyped on the Taqman platform call rates, MAF and HWE test 75

Table 2.5 Genotyping, quality control and imputation of the 1958BC genome-wide sub-samples 77

Table 2.6 Scenarios of shared SNP genotypes between two individuals and their degree of IBS 86

Table 3.1 Frequencies of missing information for lifestyle and adiposity variables 98

Table 3.2 Additional variables included in the imputation models. 100

Table 3.3 The distribution of 25(OH)D concentrations by lifestyle and socio-economic variables. 123

Table 3.4 The individual amplitude and magnitude of the seasonal variation of 25(OH)D. 135

Table 3.5 The mediation product of coefficients test and 25(OH)D association in the Frequentist setting with the haemostatic markers adjusted for seasonal variation in mixed effects model 137

Table 3.6 The seasonal amplitude and change in amplitude associated with the mediator 25(OH)D in D-Dimer, Fibrinogen, tPA (Bayesian setting). 138

Table 4.1 Characteristics of 1958BC genetic sub-samples with 25(OH)D concentrations and covariates used in genome-wide analysis 164

Table 4.2 SNPs identified in the SUNLIGHT consortium GWAS with the results from the 1958BC 168

Table 5.1 25-hydroxyvitamin D candidate gene studies since 2008 193

Table 5.2 Genetic loci associated with skin, hair colour and tanning ability in GWAS 203

Table 5.3 The SNP associations with ln 25(OH)D concentrations adjusting for sex 223

Table 5.4 Allele counts/scores and haplotype association with ln 25(OH)D adjusted for sex... 228

Table 5.5 25-hydroxyvitamin D association with the inflammation/haemostatic biomarkers and the 25(OH)D association using the genetic proxies as instruments for inflammation/haemostatic biomarkers adjusted for sex, month of measure, geographical region and BMI... 237

List of figures

Figure 1.1 The pathway of the vitamin D metabolism....23

Figure 1.2. Physiological symptoms of vitamin D depletion and toxicity....25

Figure 1.3. Monthly variation of 25(OH)D in the 1958BC at age 45 years old. 29

Figure 1.4. Prevalence of 25(OH)D < 25 nmol/L in the UK. 33

Figure 1.5. The progression of an atherosclerosis lesion....38

Figure 1.6. Normal haemostatic response to vascular damage involves blood vessels, platelets and blood coagulation.. 41

Figure 1.7. Hazard ratios (HR) of CVD mortality by 25(OH)D nmol/L categories fully adjusted as reported in the studies....52

Figure 1.8 The proposed pathway of the exposure's (vitamin D status) influence on the outcome of cardiovascular health. 56

Figure 2.1 The intensity plot for a SNP taken from the T1DGC....81

Figure 2.2 T1DGC frequency of SNPs with missing data....82

Figure 2.3 T1DGC frequency of SNP HWE *p*-values 83

Figure 2.4 T1DGC frequency of SNPs MAF....84

Figure 2.5 T1DGC frequency of sample/subject with missing SNP data....85

Figure 2.6 T1DGC The heterozygosity score for each sample/subject 87

Figure 2.7 T1DGC ethnicity analysis....89

Figure 3.1 True histogram of 25(OH)D concentrations....96

Figure 3.2 Pathway model for a single mediator.. 102

Figure 3.3 The mean, period, amplitude and phase of a rhythmic pattern....105

Figure 3.4. Establishing whether mediation analysis can be applied for the seasonal variations in 25(OH)D and in the haemostatic/inflammation biomarkers....110

Figure 3.5 The logic taken to applied the mediation test of the product of coefficients in the Bayeisan and Frequentist setting. 119

Figure 3.6 Variation in the mean 25(OH)D concentration by (A) BMI and (B) waist circumference. 126

Figure 3.7 Variation in the mean sex specific standard deviation (SDS) concentrations of 1. CRP, 2. D-dimer, 3. fibrinogen, 4. tPA and 5. vWF by (A) body mass index and (B) waist circumference. 129

Figure 3.8 Variation in inflammatory and haemostatic markers by serum 25(OH)D concentration.. 131

Figure 3.9 Seasonal patterns of A. D-Dimer; B. fibrinogen; C. tPA; and D. vWF. 134

Figure 4.1 The variation explained in 25(OH)D concentration as a function of the power to detect an association of a SNP 156

Figure 4.2 QQ Plot of the *Z*-score meta-analysed 1958BC sub-samples (WTCCC and T1DGC) SNP associations with standardised residuals of 25(OH)D.. 165

Figure 4.3 Manhattan plot of the observed *p*-values from *Z*-score meta-analysed 1958BC sub-samples (WTCCC and T1DGC) over chromosomes. 166

Figure 4.4 QQ Plot of the *Z*-score meta-analysed SUNLIGHT discovery samples SNP associations with 25(OH)D standardised residuals. 167

Figure 4.5 LD blocks of SNPs for regions A) region 4q12 near *GC*, B) region 11q12 near *DHCR7/NADSYN1* and C) region 11p15 near *CYP2R1*. 173

Figure 4.6 The top SUNLIGHT consortium SNPS for *GC*, *DHCR7/NADSYN1* and *CYP2R1* meta-analysis results of the 1958BC samples. 175

Figure 4.7 The *GC*, *CYPR21* and *DHCR7* SNP OR with thresholds of 25(OH)D at < 25 nmol/l, < 50 nmol/l and < 75 nmol/l in the 1958BC.. 176

Figure 4.8 The genotype score odds ratio with thresholds of 25(OH)D at < 25 nmol/l, < 50 nmol/l and < 75 nmol/l in the 1958BC. 178

Figure 5.1 The similarities between a randomised control trial and a MR study. 188

Figure 5.2 The pathways of the MR framework. 189

Figure 5.3 Polymorphisms and candidate regions considered as instruments for 25(OH)D 208

Figure 5.4 The relationships between the genetic markers, exposure, and outcome in MR. 209

Figure 5.5 Proposed scenarios of pleiotropic SNP in a MR study.........................221

Figure 5.6 SNP associations with 25(OH)D after three model adjustments...........224

Figure 5.7 SNP odds ratios with selected dichotomous dietary, lifestyle and region variables. 226

Figure 5.8 Allele score, counts and haplotype associations with 25(OH)D after three model adjustments..230

Figure 5.9 Allele score, counts and haplotype odds ratios with selected dichotomous dietary, lifestyle and region variables.. 231

Figure 5.10 Power calculations for IV analysis with the single SNPs as instruments (A) and allele scores/counts/haplotypes as instruments (B) 233

Abbreviations and acronyms

1958BC	British 1958 birth cohort
1,25$(OH)_2$D	1, 25-Dihydroxyvitamin D, active form of vitamin D
25(OH)D	Total 25-Hydroxyvitamin D, nutritional measure of vitamin D status
2SLS	Two-stage least squares
ADP	Adenosine diphosphate
ALTM	All laboratory trimmed mean
BIC	Bayesian Information Criterion
BMI	Body mass index
CEU	Hapmap sample that consists of Utah residents with ancestry from northern and western Europe
CI	Confidence interval
CKD	Chronic kidney diease
CRP	C-reactive protein
CVD	Cardiovascular disease
DEQAS	International Vitamin D External Quality Assessment Scheme
DIC	Deviance Information Criterion
DNA	Deoxyribonucleic acid
DHC-7	7-Dehydrocholestrol
FEV_1	Forced expiratory volume in 1 second
FVC	Forced vital capacity
FDR	False discovery rate
GMM	Generalised methods of moments
GWAS	Genome-wide association study
HbA1c	Glycosylated haemoglobin
HDL	High density lipoproteins
HPFS	Health Professionals Follow-up Study
HR	Hazard ratio
HWE	Hardy-Weinberg equilibrium
IgE	Total circulating immunoglobulin
IBD	Identity by descent

IBS	Identity by state
IGF-1	Insulin-like growth factor
IOM	Institute of Medicine
IV	Instrumental variable
ln	Natural logarithm
LD	Linkage disequilibrium
LDL	Low density lipoproteins
LRT	Likelihood ratio test
JDRF/WT DIL	Juvenile Diabetes Research Foundation/Wellcome Trust Diabetes and Inflammation Laboratory
MAF	Minor allele frequency
MAR	Missing at random
MCAR	Missing completely at random
MET	Metabolic equivalent task
MDS	Multidimensional scaling
MI	Myocardial infarction
MI	Multiple imputation
ML	Maximum likelihood
MNAR	Missing not at random
MR	Mendelian randomisation
NDNS	National diet and nutrition survey
NHANES	National health and nutrition examination survey
OLS	Ordinary least squares
OR	Odds ratio
PAI-1	Plasmin activator inhibitor-1
PC	Personal computer
PTH	Parathyroid hormone
QC	Quality control
RCT	Randomised control trial
RXR	Retinoid-X receptor
SCD	Sudden cardiac death
SD	Standard deviation
SDS	Sex-specific standard deviation scores

SNP	Single nucleotide polymorphism
tPA	Tissue plasminogen activator
T1DGC	Type 1 Diabetes Genetics Consortium
T2D	Type 2 diabetes
TV	Television
ULSAM	Uppsala Longitudinal Study of Adult Men
UK	United Kingdom
US; USA	United States of America
UVB	Ultraviolet B
vWF	von Willebrand factor
VDBP	Vitamin D binding protein
VDR	Vitamin D receptor
VDRE	Vitamin D response elements
VIF	Variance inflation factor
WTCCC	Wellcome Trust Case Control Consortium

Acknowledgements

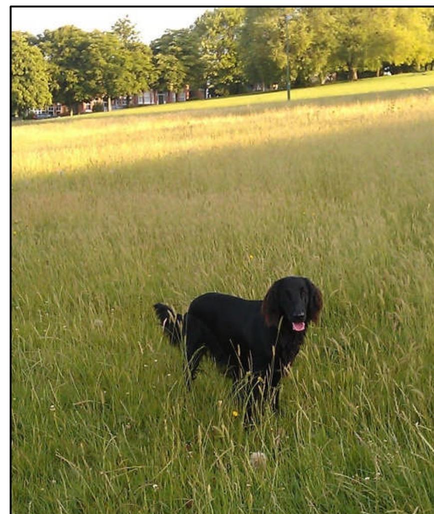
I wish to express my sincere gratitude to my supervisors Dr Elina Hyppönen and Dr Mario Cortina-Borja who both fulfilled invaluable roles in the thesis, and for their immeasurable encouragement, guidance and wisdom throughout.

I would also like to thank my work colleagues at the Institute of Child Health, in particular Chris Power and Karani Vimaleswaran, and external collaborators, especially those involved in the first genome-wide association study of the EGG consortium, for their advice along the way.

I am very grateful to my fellow present and past PhD students and room colleagues who have become friends, in particular Rachel Denholm, Naufil Alam, Eirini Koutoumanou, Sanja Stanojevic, Kirsty Little, whose shared experiences, advice, and encouragement has been invaluable. I would also like to thank my dear friend Judith Corbyn for her support, encouragement, and for reading my thesis.

Words cannot express my gratitude to all my dear friends and family, especially my mother Janet Berry and soon to be husband Daryl Henderson, for their belief and support.

Finally, I would also like to thank my dog Tui, her insistence that there is always time for a walk has helped keep me sane.
(Photo taken Hilly Fields, Lewisham 2011)



In the first year of this thesis, my research position was funded by BUPA Foundation (PI Dr Hyppönen), and then for later years by UK Medical Research Council (grant G0601653, PI Dr Hyppönen). The 25-hydroxyvitamin D assay in the 1958 British birth cohort was funded by the BUPA foundation (PI Dr Hyppönen). Genotype data from the British 1958 birth cohort DNA collection was funded by MRC grant G0000934 and Wellcome Trust grant 068545/Z/02. This thesis used resources provided by the Type 1 Diabetes Genetics Consortium, a collaborative clinical study sponsored by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), National Institute of Allergy and Infectious Diseases, National Human Genome Research Institute, National Institute of Child Health and Human Development, and Juvenile Diabetes Research Foundation International (JDRF) and supported by U01 DK062418. This thesis also makes use of data generated by the Wellcome Trust Case-Control Consortium. A full list of the investigators who contributed to the generation of the data is available from www.wtccc.org.uk. Funding for the project was provided by the Wellcome Trust under award 076113. The work was undertaken at the Centre for Paediatric Epidemiology and Biostatistics which benefits from funding support from the MRC in its capacity as the MRC Centre of Epidemiology for Child Health. Research at the University College London Institute of Child Health and Great Ormond Street Hospital for Children NHS Trust benefits from R&D funding received from the NHS Executive.

1. Vitamin D: introduction, association with cardiovascular disease and aim of the thesis

1.1 Background

Vitamin D is an essential nutrient for maintaining bone health. A severe vitamin D deficiency leads to rickets in children and osteomalacia in adults (Wassernam 2005). Emerging evidence suggests that hypovitaminosis D is a risk factor for many serious conditions including cardiovascular disease (CVD) (Swales & Wang 2010; Wallis *et al.* 2008; Zittermann *et al.* 2005). However much is still to be learnt, as there is still no consensus on the thresholds of vitamin D status defined as deficient, sufficient or optimal (Dawson-Hughes *et al.* 2005; Ginde *et al.* 2011; Heaney & Holick 2011; Hollis 2005; Institute of Medicine 2011; Norman & Bouillon 2010). More evidence is needed and has been called for to better understand the role vitamin D plays in health and diseases, beyond its classical role in calcium homeostasis (Institute of Medicine 2011).

1.1.1 History of vitamin D

The study of the paediatric disease rickets led to the discovery of the pro-hormone vitamin D (DeLuca 2005; Mellanby 1919). At the end of the 19th century, investigation was underway to discover why children residing in sunlight-deprived industrial cities developed skeletal deformities, unlike their contemporaries dwelling in the countryside or sunnier climates (Chick *et al.* 1923). By the early 20th century, clinicians had started to successfully treat children suffering from rickets with liver oil from fish, and later exposure to sunlight or mercury arc lamp (Chick *et al.* 1923; Hess 1922; Selkirk *et al.* 1928). During that period, it was also discovered that foods

irradiated with ultraviolet B (UVB) could both prevent the onset of rickets and also treat it (Hess & Weinstock 1924; Steenbock 1924). By mid-20th century, British children were given cod liver oil and dried milk fortified with vitamin D under the Welfare Foods Scheme (Department of Health 2002). In addition, food manufacturers were routinely fortifying infants' cereals, evaporated and dried milks (Bransby *et al.* 1964). However, reported cases of hypercalcaemia in infants and children were linked with excessive vitamin D quantities in infants' milk, and fortification of food was sharply curtailed (British Paediatric Association 1956; Ministry of Health 1957). A subsequent review of the cases found, that there was a lack of evidence to conclude whether excessive vitamin D fortification was responsible for the hypercalcaemia incidences (Fraser 1967). As it stands today in the UK, vitamin D fortification of margarine and infants' formula is mandatory (Ministry of Agriculture 1995a; Ministry of Agriculture 1995b).

1.1.2 Vitamin D intake

The nutrient comes in two forms, vitamin D_3 and vitamin D_2 (Haddad & Hahn 1973). Vitamin D_3, or cholecalcifierol, is synthesised in the skin after exposure to UVB radiation (typically from sunlight) of wavelengths 290-315 nm (Holick 1995). The synthesis converts 7-dehydrocholestrol (DHC-7) to pre-vitamin D_3, which is then quickly converted by heat induction to vitamin D_3. Over-exposure to UVB radiation does not cause vitamin D toxicity as excess pre-vitamin D_3 and vitamin D_3 is inactivated by radiation (Holick *et al.* 1981; Webb *et al.* 1989).

Vitamin D_2 or ergocalciferol, is the plant form of vitamin D irradiated from yeast and can occur naturally in some types of mushrooms (Lamberg-Allardt 2006). Vitamin D_3 can also be found in food from animal-based sources, such as oily fish e.g. from salmon, sardines and mackerel. Foods can be fortified with analogues of both forms of vitamin D and taken as dietary supplements (Holden *et al.* 2008).

Both forms of vitamin D (D_3 and D_2) will increase an individual's nutritional vitamin D status. There is some evidence that vitamin D_3 is more effective at increasing the nutritional status of vitamin D, than vitamin D_2 (Armas *et al.* 2004; Heaney *et al.* 2011; Trang *et al.* 1998). However, a biological preference for vitamin D_3 over D_2 has not been consistently observed in all studies (Biancuzzo *et al.* 2010; Holick *et al.* 2008). Regardless, the metabolism of both forms is similar and vitamin D_2 and D_3 will be referred as vitamin D hereafter.

1.1.3 Metabolism of vitamin D

The majority of vitamin D, whether obtained from skin synthesis or diet, is transported to the liver for the first hydroxylation to 25-hydroxyvitamin D (25(OH)D) (Figure 1.1). The 25-hydroxylation has also been shown to occur in some extra-hepatic tissues (DeLuca 2008). The hydroxylation of vitamin D to 25(OH)D in the liver has little or no regulation and circulating concentrations of 25(OH)D increase in proportion to vitamin D intake (Horst *et al.* 2005). The half-life of 25(OH)D is 15 days and is the preferred form for measuring nutritional vitamin D status (Horst *et al.* 2005; Jones 2000).

A second hydroxylation of 25(OH)D in the kidneys and to some extent in extra-renal tissues produces the active hormonal form of 1,25-dihydroxyvitamin D (1,25$(OH)_2$D, calcitriol) (Horst *et al.* 2005; Shultz *et al.* 1983). Circulating concentrations of 1,25$(OH)_2$D are tightly regulated, its half-life is 10-20 hours and is much shorter than 25(OH)D (Jones 2008). The renal synthesis of 1,25$(OH)_2$D is up-regulated by parathyroid hormone (PTH) to maintain calcium homeostasis (Bergwitz & Juppner 2010; Galitzer *et al.* 2008). Phosphorous homeostasis is maintained by down-regulation of 1,25$(OH)_2$D, triggered by fibroblast growth factor 23 synthesised from bone cells (osteocytes) (Bergwitz & Juppner 2010; Galitzer *et al.* 2008).

The vitamin D binding protein (VDBP, *GC*-group component, Gc-globulin) binds to circulating vitamin D metabolites (including vitamin D, 25(OH)D, and 1,25$(OH)_2$D) (Daiger *et al.* 1975; Speeckaert *et al.* 2006). Excess 25(OH)D and 1,25$(OH)_2$D are cleared by 24-hydroxylation, leading to inactive metabolites and eventually extraction from the system (DeLuca 2008). Genomic actions of 1,25$(OH)_2$D are mediated by binding with nuclear vitamin D receptors (VDR). The VDR forms a heterodimer structure with the retinoid-X receptor (RXR), and binds to vitamin D response elements (VDRE) in the regulatory element region of a vitamin D target gene (Pike & Shevde 2005). In turn, this can affect the expression of the target gene. There are over 2700 VDR binding sites that exist in the genome responding to the active 1,25$(OH)_2$D hormone, suggesting that hormonal vitamin D has a wide range of influence in the human metabolism (Ramagopalan *et al.* 2010).

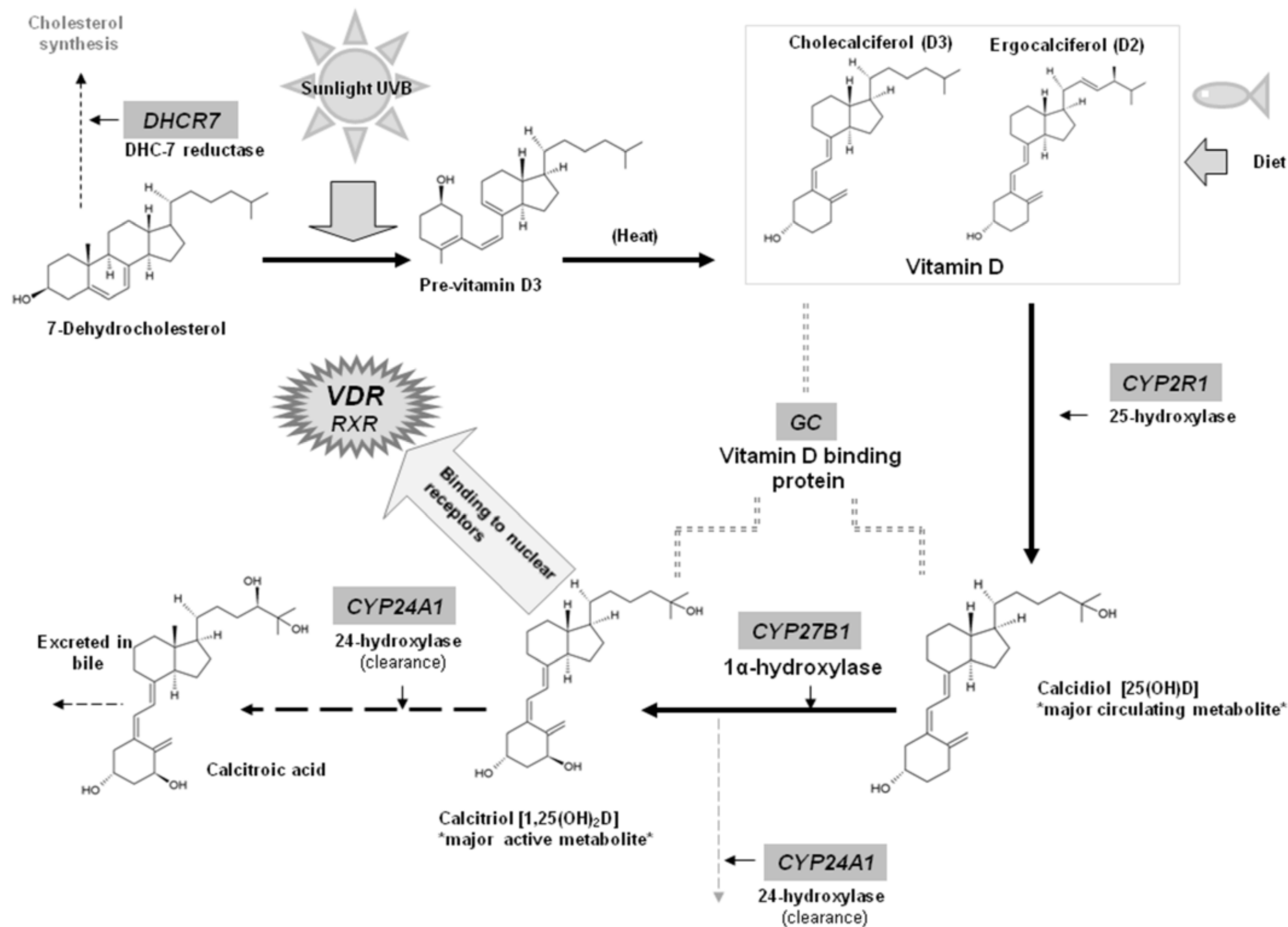


Figure 1.1 The pathway of the vitamin D metabolism. Taken from (Berry & Hyppönen 2011)

1.1.4 Thresholds of 25-hydroxyvitamin D

Thresholds of 25(OH)D can be defined in terms of biological and clinical outcomes associated with vitamin D metabolism (Lips 2005). At sufficient concentrations of 25(OH)D, it is uncorrelated with 1,25$(OH)_2$D and PTH (Lips 2005). However, below the sufficiency threshold, the synthesis of 1,25$(OH)_2$D becomes dependent on its pre-cursor, 25(OH)D (

Figure 1.2). Secretion of PTH will increase to maintain normocalcemia as less calcium is absorbed in the intestine. Triggered by the up-regulation of PTH, more 25(OH)D is synthesised to 1,25(OH)D further aggravating the vitamin D deficiency (Holick 2006). In a deficient state, calcium and phosphorous absorption will fall to 10-20% and 50-60%, respectively, of what it is during periods of sufficiency and growth (Ozkan 2010). Calcium and phosphorous will be released from the bone mediated by PTH responding to depressed calcium levels. A severe vitamin D deficiency can be characterised by defective bone mineralisation and stunted growth that can deform bones, especially in growing children (rickets) (Kanis 2002). In adults, the disease osteomalacia can present with musculoskeletal pain and muscle weakness.

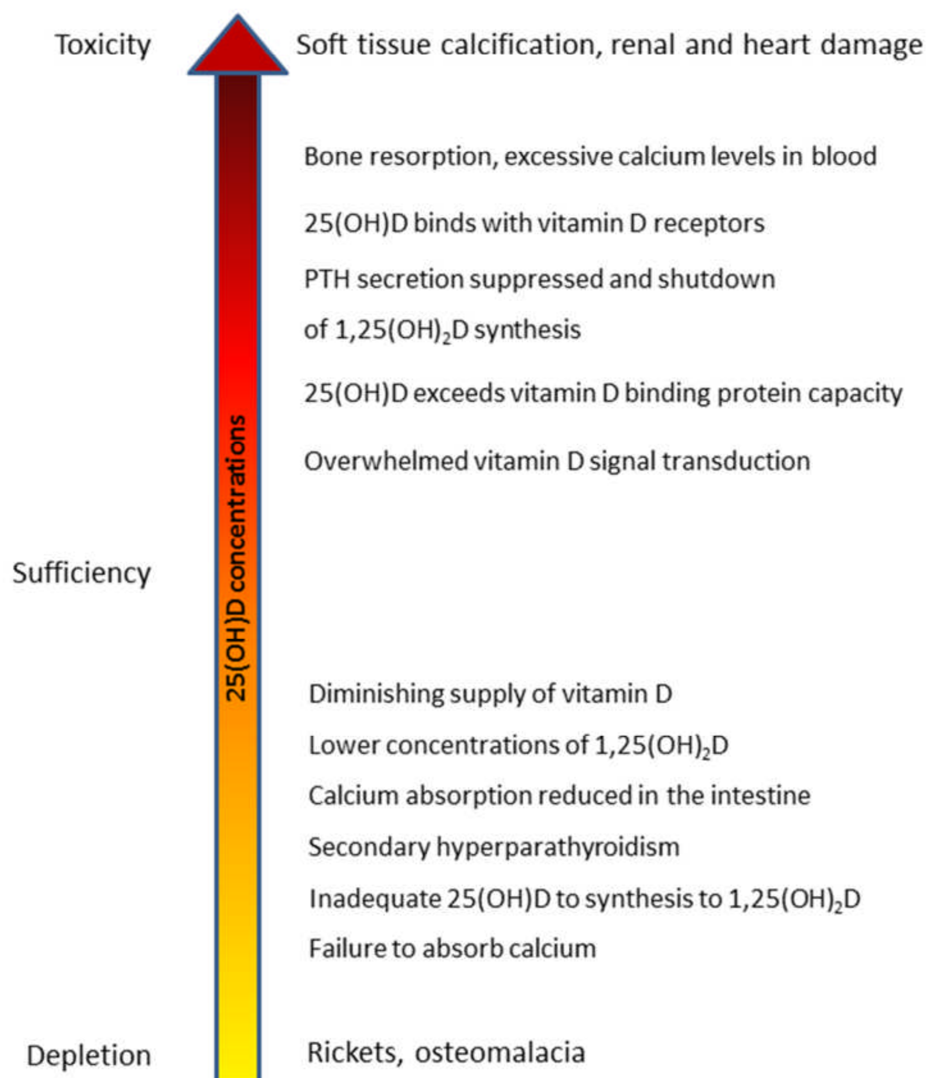


Figure 1.2. Physiological symptoms of vitamin D depletion and toxicity.

Conversely, too much vitamin D will be toxic and leads to hypercalcaemia, calcification of soft tissue, and bone resorption causing renal and heart damage (Jones 2008). Excessive vitamin D intake via food or supplements will undergo synthesis and elevate 25(OH)D concentrations (Jones 2008). Evidence suggests that excessive 25(OH)D quantities exceeds the capabilities of VDBP and will bind directly with VDR in target cells (DeLuca *et al.* 2011). The synthesis to 1,25$(OH)_2$D does not appear to take place due to its tight regulation by PTH responding to the increase of plasma calcium levels (Shephard & DeLuca 1980). The vitamin D signal transduction process becomes overwhelmed and results in an inability to effectively clear the excess quantities of vitamin D metabolites (Jones 2008).

Difficulties arise in defining exact thresholds of 25(OH)D concentrations for a number of reasons. One issue has been the ability to reliably compare results across studies due to the assay variability of 25(OH)D concentrations. To this end, the International Vitamin D External Quality Assessment Scheme (DEQAS) was established in 1989 with the aim to "monitor the performance of individual laboratories" (Carter *et al.* 2010). Participating laboratories receive a certificate if 80% of their results from the 5 quarterly samples issued by DEQAS are within 30% of the DEQAS All-Laboratory Trimmed Mean (ALTM) (Carter *et al.* 2010). However, the use of ALTM as the standard has its own issues. The ALTM has drifted away from the gold standard (gas chromatography-mass spectrometry) as more methods are used to measure 25(OH)D (Carter 2009). A passing result by DEQAS does not deem a method as accurate but as being relatively accurate compared with other laboratories that used that method (Carter 2009; Lensmeyer *et al.* 2006). Recently, US agencies and government departments have developed a

serum based reference material and it is hope that this will improve assay comparability (Wallace *et al.* 2010).

There is vigorous debate on 25(OH)D thresholds needed for health. One summation of the arguments surrounding the debate was the two positions held at roundtable event of the 14th Vitamin D Workshop in 2009 (Henry *et al.* 2010). The first position was that 25(OH)D concentrations should be at least 50 nmol/L for maintaining bone health. This threshold was akin to the IOM's , but the IOM further quantified that the 50 nmol/L threshold was sufficient to meet 97.5% of the population needs for bone health (Institute of Medicine 2011). The second position was that 25(OH)D levels should be at least 75-100 nmol/L based on newer evidence of vitamin D associations with non-skeletal diseases (Henry *et al.* 2010). Some scientists have also argued for the 75-100 nmol/L threshold on the basis of bone health (Heaney & Holick 2011; Hollis 2005) .

Convened in 2009, the Rank Forum on Vitamin D concluded that "optimal" should be defined in terms of vitamin D requirements with health outcomes before recommendations are made (Lanham-New *et al.* 2011). An earlier consensus statement from six scientists suggested that the optimal 25(OH)D threshold was between 50-80 nmol/L based on evidence about bone health, but five of the authors had specified the lower end should be higher at 70 nmol/L (Dawson-Hughes *et al.* 2005). Based on interpretation of PTH levels, a review the same year defined deficiency as 0-25 nmol/L, insufficient as 25-50 nmol/L, hypovitaminosis D as 50-70 to 100 nmol/L and adequacy as 70-100 to 250 nmol/L (Zittermann *et al.* 2005). The Rank Forum was in agreement that everyone (including infants and children)

should have 25(OH)D concentrations above 25 nmol/L (Lanham-New *et al.* 2011). The IOM report had the 25(OH)D deficient threshold slightly higher at 30 nmol/L, and concluded that below this threshold there is an increased risk of bone disease (Institute of Medicine 2011).

Toxicity thresholds are a little less controversial than (sub-) optimal thresholds, and the general agreement has been that 25(OH)D levels above 250 nmol/L may be unsafe and lead to hypercalcemia (Hollis 2005; Jones 2008; Zittermann *et al.* 2005). The IOM has the 25(OH)D safety level lower at 125-150 nmol/L (Institute of Medicine 2011). These safety levels were based on reports of adverse associations with health outcomes when 25(OH)D levels were below 125 nmol/L, and reports of individuals' 25(OH)D levels remaining around 125-150 nmol/L after maximal sun exposure (Institute of Medicine 2011).

Estimates of the prevalence of vitamin D deficiency in the general population vary given the differences in thresholds. Holick has estimated that, based on the 25(OH)D threshold of 75 nmol/L, vitamin D insufficiency is at epidemic proportions and affects up to one billion people worldwide (Holick 2008).

1.1.5 Environmental determinants of vitamin D status

The majority of vitamin D intake is from causal sun exposure that initiates the cutaneous synthesis rather than diet (Haddad & Hahn 1973; Webb *et al.* 2010). Since the nutrient can be obtained from UVB exposure it is a pro-hormone rather than a vitamin (DeLuca 2008). In the UK, 25(OH)D levels are lowest during late winter and highest during late summer and follows a yearly cycle (Figure 1.3).

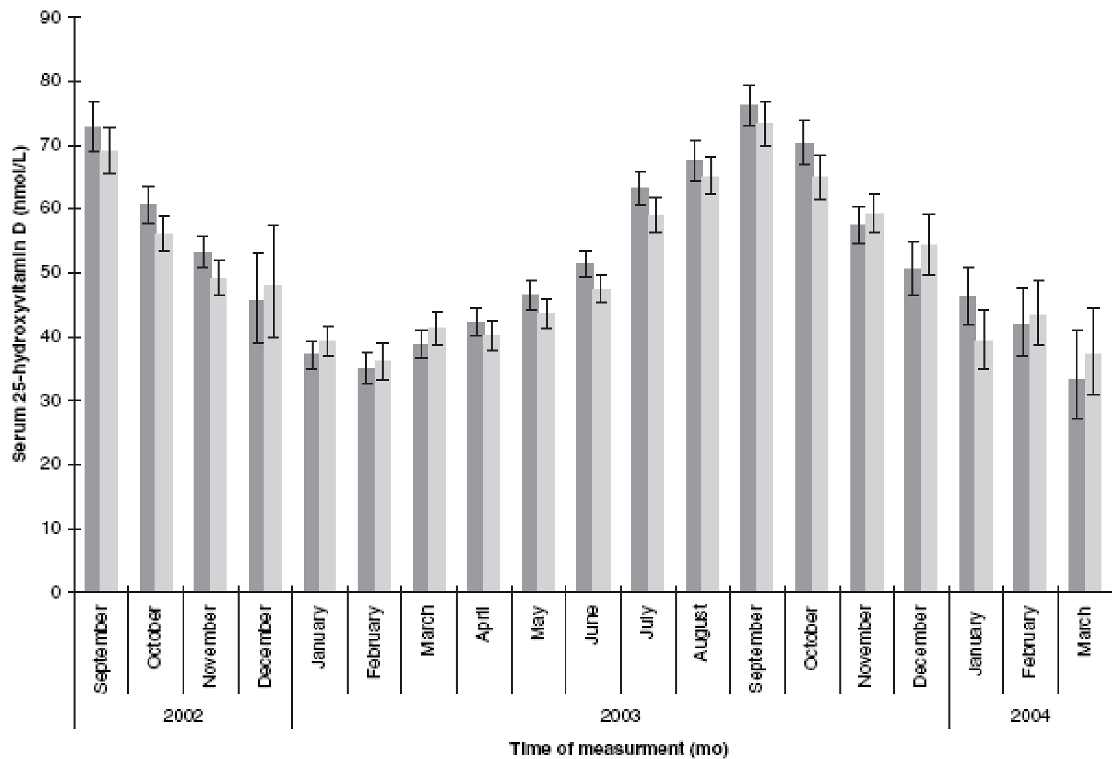


Figure 1.3. Monthly variation of 25(OH)D in the 1958BC at age 45 years old. Data are geometric means (95% CI) monthly variation in serum 25-hydroxyvitamin D 25(OH)D concentrations in men (*n*=3725 dark grey) and women (*n*=3712 light grey). Taken from (Hyppönen & Power 2007).

The seasonal variation in 25(OH)D is caused by changes in the solar zenith angle, i.e. the angle at which the sun's rays enter the earth's atmosphere (Holick 1995). The solar zenith angle is measured from the position of the sun to the zenith sky (directly overhead). The lower the sun is in the sky (as it is in wintertime), the larger the solar zenith angle, and hence the greater the distance travelled by the sun's rays. During wintertime in the UK, most of the UVB rays needed to initiate skin synthesis of vitamin D are absorbed by the ozone, thereby reducing levels of 25(OH)D (Engelsen *et al.* 2005). Likewise, time of day, latitude and altitude will affect the solar zenith angle and whether the sun's rays will reach the surface of the earth (Holick 1995; Kimlin 2004). In the Northern Hemisphere, mean concentrations of

25(OH)D follow a strong north-to-south gradient (and vice-versa for the Southern Hemisphere), so individuals dwelling in the northern most latitudes tend to have lower 25(OH)D concentrations than their southern counterparts (Chapuy *et al.* 1997; Hyppönen & Power 2007). Heavy (stratus) cloud cover and thick pollution can attenuate the necessary UVB rays to initiate the skin synthesis of vitamin D (Webb 2006).

1.1.6 Personal determinants of vitamin D status

There are several personal factors that affect an individual's vitamin D status. Lifestyle, culture and working patterns impact on the time spent outside, choice of clothing, and use of sun protection. Similarly, skin colour (as can be denoted by ethnicity) and aging influence the levels of vitamin D synthesised (Chen *et al.* 2007; MacLaughlin & Holick 1985). Obesity has been consistently found to be associated with reduced 25(OH)D levels (Cheng *et al.* 2010; Forrest & Stuhldreher 2011; Jorde *et al.* 2010b; Martins *et al.* 2007).

Individuals with outdoor occupations will have greater exposure to sunlight than individuals working indoors (Ward *et al.* 2011). For example, lifeguards will have higher 25(OH)D concentrations than submarine sailors (Vieth 1999). Clothing cover limits the amount of skin exposed to UVB, and this is particularly evident in the Middle East, where it is common for women to be dressed conservatively. In a Lebanese study, it was found that 61.8% of veiled women had 25(OH)D concentrations < 12.5 nmol/L, whereas the prevalence among non-veiled women was 23.5% and in men it was 7.1% (Gannage-Yared *et al.* 2000). Compared with the rest of the world, the incidence of rickets is also highest in the Middle East and

Africa despite an abundance of sunshine (Mithal *et al.* 2009). In these regions, the reluctance of pregnant women to expose skin to sunlight due to cultural practises is a contributing factor of rickets in newborn infants (Baroncelli *et al.* 2008).

The use of sunscreen is also another personal determinant of vitamin D_3, as depending on the sunscreen's sun protection factor, it can absorb the UVB preventing skin synthesis of vitamin D (Sayre & Dowdy 2007). Melanin in the skin gives it its pigmentation and can act as a sunscreen against UVB (Holick *et al.* 1981). Compared with individuals with lighter skin, individuals with darker skin will form less vitamin D_3 under the same UVB conditions (Chen *et al.* 2007). Furthermore, differences between ethnic groups' 25(OH)D levels are readily seen in populations; individuals of ethnicities with darker skin have lower 25(OH)D concentrations, compared with lighter skin contemporaries (Looker *et al.* 2002; Macdonald *et al.* 2011; Rockell *et al.* 2005).

Age can also influence the levels of pre-vitamin D_3 as the quantity of DHC-7 in skin reduces with age. In a UVB controlled environment, skin samples from older subjects (aged 77-82 years) produced less pre-vitamin D_3, than samples from younger subjects (aged 8-18 years) (MacLaughlin & Holick 1985). At a population mean level, older adults have a lower 25(OH)D concentrations than that of middle-aged and young adults (Ginde *et al.* 2009a).

In the general population, obesity is a risk marker of hypovitaminosis D (Cheng *et al.* 2010; Forrest & Stuhldreher 2011; Jorde *et al.* 2010b; Martins *et al.* 2007). Analysis of the National Health and Nutrition Examination survey (NHANES) 2005-

2006, found obese participants (body mass index (BMI) ≥ 30 kg/m^2) had an increased prevalence of having 25(OH)D < 50nmol/L, compared with non-obese participants (Forrest & Stuhldreher 2011). An adverse body fat (as measured by BMI, visceral and subcutaneous adiposity fat) relationship with 25(OH)D has been consistently found in adults independent of season (Cheng *et al.* 2010; Jorde *et al.* 2010b). It has also been observed that individuals of non-European ancestry have a stronger adverse BMI association with 25(OH)D < 37.5nmol/L, than individuals of European ancestry residing in northern latitudes (Yanoff *et al.* 2006). Research is so far inconclusive on whether obesity reduces 25(OH)D concentrations, however it is probable that adiposity sequesters vitamin D from circulation thereby lowering 25(OH)D concentrations (Earthman *et al.* 2011).

1.1.7 Genetic determinants of vitamin D status

Research thus far suggests a strong genetic influence on vitamin D status. Heritability estimates of 25(OH)D range from 29% from the analysis of the Framingham Offspring Study (Shea *et al.* 2009), 43% from the Twins UK Study (Hunter *et al.* 2001), to 80% from the German Asthma Study (Wjst *et al.* 2006). A Swedish twin study found that the variation explained in 25(OH)D concentrations by genetic factors changed across season (Snellman *et al.* 2009). Nearly half of the variation explained in 25(OH)D during summer was due to genetic factors whereas this was negligible during winter, which may also suggest an interaction between genetic preposition and environment. At the initiation of this thesis, candidate gene studies had used genetic variants of genes expressed in the metabolism vitamin D pathway to assess their relationship with diseases (Bailey *et al.* 2007; Ramos-López *et al.* 2007).

1.1.8 Levels of 25-hydroxyvitamin D in the UK

The prevalence of 25(OH)D levels lower than 25 nmol/L is common in UK adults, and at least 10% of adults of all ages have been found to be under this threshold (Figure 1.4). The National Diet & Nutrition Survey (NDNS) was carried out between July 2000 to June 2001, and revealed that the overwhelming majority of the UK adult population have 25(OH)D levels lower than < 50 nmol/L (Ruston *et al.* 2004). The subsequent Health for England survey 2005 also found that the near majority of adults aged at least 65 years had 25(OH)D concentrations < 50 nmol/L (49% of men and 58% of women), and confirmed that 25(OH)D concentrations < 25 nmol/L were common (8% of men and 14% of women) (Hirani *et al.* 2010).

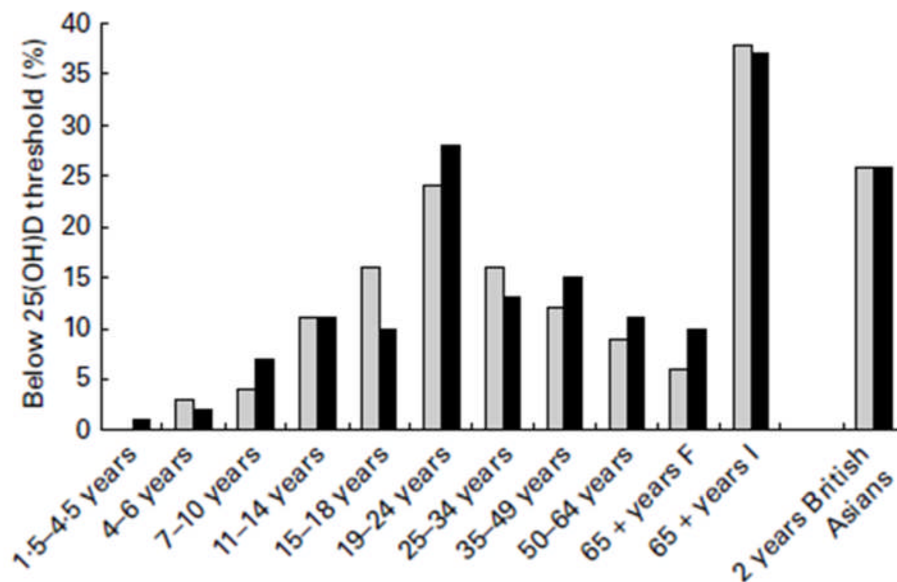


Figure 1.4. Prevalence of 25(OH)D < 25 nmol/L in the UK. F, free-living; I, institutionalised, 2 year British Asians were defined as of those as Pakistani, Indian and Bangladeshi that had resided in the UK aged 2 years old. Light grey bars are males and dark grey bars are females. Taken from (Lanham-New *et al.* 2011)

Several sub-populations within the UK are particularly susceptible to low vitamin D intake/status. The NDNS found that of free-living adults, young adults (19-24 years old) had the highest prevalence of 25(OH)D < 25 nmol/L (Lanham-New *et al.* 2011). A study of healthy teenaged girls living in Manchester found that 73% of the girls had 25(OH)D levels < 30 nmol/L (Das *et al.* 2006). The differences in 25(OH)D levels across ethnicities were striking in this study. The teenage girls of European ancestry had a much higher median levels of 25(OH)D at 37.3 nmol/L, compared with girls of non-European ancestry at 14.8 nmol/L.

Differences in 25(OH)D levels among ethnic groups have been seen across all ages in the UK. Analysis of adults aged at least 65 years in the Health for England survey found that compared with individuals of European ancestry, individuals of non-European ancestry had nearly an eight fold increase in risk of being 25(OH)D < 50 nmol/L (Hirani *et al.* 2010). Ethnic differences in vitamin D status were also seen in post-menopausal women from Blackburn (Lancashire, UK) (Lowe *et al.* 2010). Almost all (97%) of the women of South Asian ancestry had 25(OH)D levels < 50 nmol/L, compared with 60% of women of European ancestry (Lowe *et al.* 2010). At the lower threshold of 25(OH)D < 25 nmol/L, the difference in prevalence across the two ethnic groups was even bigger; 85% of women of South Asian ancestry, compared with 10% of those of European ancestry (Lowe *et al.* 2010). Similarly, children (aged 9-10 years old) of South Asian or Black African-Caribbean ancestry have been observed to have a lower intake of vitamin D, compared with those of European ancestry (Donin *et al.* 2010).

Regional differences of 25(OH)D concentrations have also been readily observed in the UK. An analysis of the British 1958 birth cohort (1958BC) found that the prevalence of being 25(OH)D < 40 nmol/L was much greater for individuals residing in Scotland, compared with those residing at southern latitudes in Britain (Hyppönen & Power 2007). That said, findings from a cohort of women aged 45-54 years suggested that regardless of season, women of Asian ancestry residing in Surrey (South of England) had a higher prevalence of lower vitamin D intake/status than women of European ancestry residing in Scotland (Macdonald *et al.* 2011). The regional differences in 25(OH)D levels were still evident in women of European ancestry, those residing in Scotland had a greater prevalence of 25(OH)D < 25 nmol/L, than those residing in the South of England.

In recognition of low vitamin D intake/status in the UK, the Department of Health recommends that all pregnant and breastfeeding women, infants and children from 6 months up to 5 years, people who have limited sun exposure, and people who are 65 years and older supplement their diet with vitamin D (Department of Health 2010). Since 2006, the Healthy Start statutory scheme in the UK has offered pregnant women, new mothers and children in low income families free vitamin D supplements amongst other vitamin supplements and food items (Department of Health 2011). However, the attendees of the Rank Forum on Vitamin D think that there had been low compliance with the Healthy Scheme recommendations and uptake of vitamins (Lanham-New *et al.* 2011).

1.2 Cardiovascular disease

CVD is the leading cause of death in the UK, and during 2008 it accounted for a third of deaths across all ages (Scarborough *et al.* 2010). CVD is also one of the main causes of premature death, killing 28% of men and 20% of women before reaching the age of 75 years (Scarborough *et al.* 2010). Since the 1970s there has been a steady decline in mortality rates for CVD and this is attributed to medical treatment and changes in lifestyle, especially in smoking habits (Unal *et al.* 2004). There is a seasonal aspect to CVD-related mortality, with an excess of deaths occurring during the coldest months of the year (Crawford *et al.* 2003a)

The progression of CVD typically starts with fatty streaks in the intima layer of arteries (Hansson 2005). Arteries have a three layer structure of externa, media and intima (Aaronson & Ward 2007). The external layer (externa) of the artery contains fibroblasts, collagenous tissue, and if the artery is large enough it will also have small blood vessels to support it. The middle layer (media) contains vascular smooth muscle cells that are held within an extracellular matrix consisting of collagen and elastin. The internal layer or intima is covered by a monolayer of tightly connected endothelial cells that restrict larger molecules from crossing it.

In situations of hypercholesterolemia, low density lipoproteins (LDL) will pass across the endothelium cells and be retained in the intima of the artery (Ross 1993) (Figure 1.5). Modified LDL particles caused by oxidation or enzymatic activity are picked by the scavenger receptors of macrophages and are turned into cholesterol laden foam cells, or fatty streaks. Fatty streaks start forming during childhood and in adulthood some of them will develop into atherosclerosis lesions (Ross 1993).

Inflammation induced by macrophages in the lesions activate T cells, which differentiate to T helper cell 1 and maintain a state of inflammation (Hansson 2005). Vascular smooth muscles cells in this state are either drawn across from the media into intima or go on to produce and excrete extracellular matrix components. Part of the lesion core contains crystalline calcium deposits, which accumulate as the lesion grows and may lead to calcification of the plaque (Wexler *et al.* 1996). As the lesion becomes larger the internal space of the artery (lumen) is maintained to some extent by stretching (Glagov *et al.* 1987). The lesion forms a fibrous cap created from the smooth muscle cells and macrophages. Foam cells can die in the lesion causing a release of cellular debris that may go on to block smaller vessels. A rupture or tear in the fibrous cap can lead to on-going thrombi that cause intermitted blood flow (unstable angina) or a large thrombus that blocks the artery. Depending on the artery blocked, this causes a stroke (cerebral artery) or myocardial infarction (MI) (coronary artery). However, when the cap remains intact the lesion will continue to grow and may cause obstruction in the artery (ischaemic). Alternatively, the media under the lesion can break, weakening and eventually rupturing the vascular wall (aneurysm).

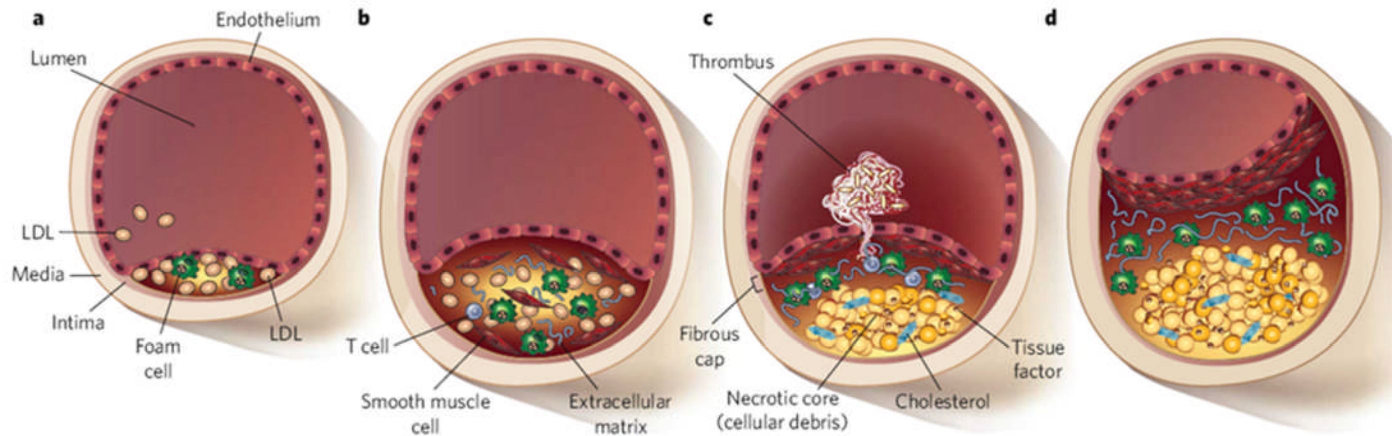


Figure 1.5. The progression of an atherosclerosis lesion. a) The lesion is the accumulation of foam cells (originally LDL particles) in the intima layer of the artery. b) As the lesion develops, inflammation leads to other cells accumulating at the lesion site. c) The lesion ruptures causing a thrombus that may block the artery. d) Alternatively the lesion continues to grow and may obstruct the blood flow or weaken the external wall of the artery leading to an aneurysm. Taken from (Rader & Daugherty 2008).

1.2.1 Risk markers of CVD

Inflammation in the arteries occurs at the creation of the fatty streak (Hansson 2005). The immune cells guarding the arteries can be triggered by the oxidised LDL particles in the lesions. The monocytes differentiated into macrophages release inflammatory cytokines, and likewise, dendritic cells, mast cells and endothelial cells also release toll-like receptors initiating inflammation.

C-reactive protein

C-reactive protein (CRP) is mainly produced by hepatocytes cells in the liver as part of the acute-phase response to inflammation (Pepys & Hirschfield 2003). Its production is triggered by most inflammation, infection and tissue damage and is not specific to CVD. The production of CRP rises rapidly under the transcriptional control of interleukin 6 and peaks around 48 hours after a stimulus (Vigushin *et al.* 1993). Once the stimulus from the damage has ceased so does the production of CRP (Vigushin *et al.* 1993). Baseline CRP levels tend to be stable over a long period of time (Ockene *et al.* 2001). The population median of circulating plasma CRP rises with age and is indicative of an increased incidence of sub-clinical diseases such as CVD (Hutchinson *et al.* 2000). There is on-going research as to whether CRP is partly causal of CVD or a risk marker (Casas *et al.* 2008). Regardless, elevated CRP is a good predictor of future vascular events (Ridker 2003). The inclusion of CRP to risk prediction models improves individuals’ risk classification for developing coronary heart disease (Buckley *et al.* 2009). CRP concentrations have an adverse relationship with the risk of coronary heart disease,

ischaemic stroke and non-specified stroke, vascular and non-vascular mortality (Kaptoge *et al.* 2010).

1.2.2 Haemostasis and thrombosis

The haemostatic response to vascular damage is a complex and tightly controlled system that involves several components, from the control of bleeding; adhesion of the blood platelets in the vessel at the damaged site; the activation of the platelets; the aggregation of platelets; coagulation of the blood; to the formation of a stable plug (Figure 1.6) (Aaronson & Ward 2007; Goodnight & Hathaway 2001; Hoffbrand & Moss 2011a). The mechanisms for creating the stable plug and restricting the bleeding are the same for a thrombus that can block a vessel (Macfarlane 1977). Thrombosis has been described as "haemostasis in the wrong place" (Macfarlane 1977).

Blood vessel damage exposes von Willebrand factor attached to collagen that lies under the endothelial cells monolayer in the intima (Aaronson & Ward 2007; Goodnight & Hathaway 2001; Hoffbrand & Moss 2011a). Glycoprotein receptors on platelet membranes bind to von Willebrand factor, triggered glycoprotein receptors also bind to the exposed collagen, further attaching the platelet to the damaged site. The adhered platelets change shape and make thromoboxane, which excretes serotonin and adenosine diphosphate (ADP), and leads to vasoconstriction, reducing the flow of blood to the area. The platelets expose lipids on their surface and activate receptors for fibrinogen (GPIIb/IIIa). Activated platelets further activate and attract more platelets, and plasma fibrinogen binds to their receptors on the platelets forming an initial soft plug.

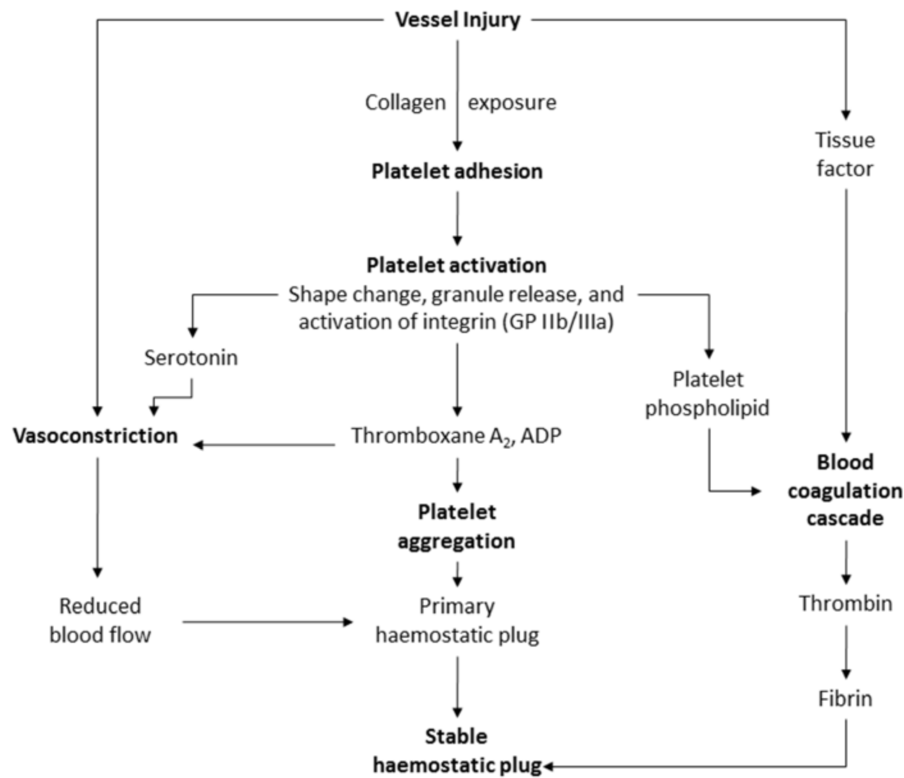


Figure 1.6. Normal haemostatic response to vascular damage involves blood vessels, platelets and blood coagulation. Taken from (Hoffbrand & Moss 2011a).

The blood coagulation cascade is initiated by tissue factor exposed on the damaged site binding with factor VIIa in plasma (Aaronson & Ward 2007; Goodnight & Hathaway 2001; Hoffbrand & Moss 2011a). The cascade is a sequential stepwise activation of coagulation factor enzymes that requires calcium to generate thrombin. Thrombin, in turn, converts the soluble fibrinogen that binds the anchored platelets into insoluble fibrin, thereby transforming the soft platelet plug to a firm stable haemostatic plug.

The haemoastasis response has a positive feedback to prevent thromobosis and involves a series of inhibitors (Aaronson & Ward 2007; Goodnight & Hathaway

2001; Hoffbrand & Moss 2011a). The first inhibitor is synthesised in the endothelial cells, which leads to the inhibition of tissue factor binding with factor VIIa. Circulating antithrombin inactivates thrombin alongside other inhibitors of thrombin and coagulation factors. Fibrinolysis is part of the feedback to haemostasis response, where activated plasmin breaks down fibrin into fibrin degradation products such as small D-dimer particles. The activation of plasmin comes from the release of tissue plasminogen activator from the disrupted endothelial cells.

von Willebrand factor (vWF)

The synthesis by vascular endothelial cells is main source of vWF (Ruggeri & Ware 1992), and levels of circulating vWF can be related to the individual's blood type (Hoffbrand & Moss 2011b). In plasma, vWF forms a complex with coagulation factor VIII, which stabilises the factor and protects it from degradation. A crucial role of vWF in the sub-endothelial matrix is forming the initial bridges with platelets and subsequent aggregation in the haemostasis response to vessel injury (Goodnight & Hathaway 2001).

Elevated circulating plasma vWF is a common marker of endothelial disturbance and can be suggestive of vascular clot formation (Lowe 1997). Plasma vWF can increase rapidly over the short term due to strenuous muscular exercise or in response to medication, that mediates the exercise-induced increase (Mannucci 1998). Plasma vWF levels can also rise slowly over the long term due to conditions such as liver cirrhosis, renal failure, diabetes and CVD (Blann 2006; Mannucci 1998). Furthermore, plasma vWF levels increase with age (Rumley *et al.* 2006; Tofler *et al.* 2005) and in high-risk populations of coronary disease, high plasma levels of vWF

predict coronary events and mortality as reviewed in Vischer 2006. More recently, an adverse relationship between vWF and risk of stroke has been found in the general population (Wieberdink *et al.* 2010).

Fibrinogen

Plasma fibrinogen is synthesised in the liver by hepatocytes and is an acute phase reactant simulated by interleukin 6 as part of the inflammation response (Folsom 1995). It is a key glycoprotein in the blood coagulation cascade, as its transformation to fibrin mediated by thrombin is the basis for forming a stable clot. Due to its mass, fibrinogen is a strong determinant of plasma viscosity (Lowe 1997), which is related to peripheral vascular resistance (Koenig & Ernst 1992).

Elevated plasma fibrinogen is adversely associated with several cardiovascular risk factors including old age (Blann 2006; Mannucci 1998) and smoking; fibrinogen levels can remain elevated for years after quitting smoking (Sinha *et al.* 2005). Conversely, there is evidence that moderate alcohol consumption can decrease fibrinogen levels (Brien *et al.* 2011). Evidence suggests that plasma fibrinogen levels are a predictor of cardiovascular events (Lowe 1997), independent of several cardiovascular risk factors (Danesh *et al.* 2005). Furthermore, fibrinogen is independently associated with all-cause mortality (Woodward *et al.* 2003).

Tissue plasminogen activator (tPA)

The glycoprotein of tPA is secreted after the inactivation of plasmin activator inhibitor-1 (PAI-1) initiated by thrombin bound on the endothelial cell surface as

part of the fibrinolysis process (Hoffbrand & Moss 2011a). Manufactured tPA is used as part of the immediate treatment for acute ischemic stroke to initiate the breakdown of the cerebral blockage (Albers *et al.* 2000; Hacke *et al.* 2008). The measurement of plasma tPA is mostly a marker of the inactive complex formed of tPA and PAI-1, and can be a marker of impaired fibrinolytic activity (Lowe 1997).

In the general population, as with the other haemostatic factors, tPA levels increase with age (Rumley *et al.* 2006) and has been associated with other cardiovascular risk factors such as smoking, lipids and blood pressure (Lowe *et al.* 2004). Findings from a 2001 meta-analysis were that elevated tPA levels were associated with the risk of coronary heart disease in the general population (Lowe *et al.* 2004). However in studies where the participants have had vascular disease, tPA levels does not remain associated with coronary heart disease after adjustment for classical cardiovascular risk factors (Lowe *et al.* 2004). In some more recent studies, elevated tPA has been adversely associated with cardiovascular events exclusive of other cardiovascular risk markers (May *et al.* 2007; Smith *et al.* 2005; Woodward *et al.* 2007) .

D-dimer

Fibrin degradation is the sequential action of three enzymes; thrombin acting on fibrinogen, followed by factor XIIIa to create a stable clot of cross-linked fibrin mesh, and plasmin degrading the cross-linked fibrin to release and expose D-dimer (Adam *et al.* 2009). Elevated plasma D-dimer levels are present in patients with deep venous thrombosis, pulmonary embolism or MI and are a measure of fibrin turnover (Bounameaux *et al.* 1991). It has been suggested that D-dimer should be

measured as part of the diagnostic tools to screen for pulmonary embolism and venous thromboembolism (Bounameaux *et al.* 1991; Di Nisio M. *et al.* 2007). There is evidence that raised D-dimer levels are adversely associated with the severity of pulmonary embolism (Galle *et al.* 2001).

Plasma D-dimer levels increase with age and this might be related to an underlying inflammation response to atherosclerosis (Cohen *et al.* 2003; Rumley *et al.* 2006). There is evidence that elevated D-dimer levels are adversely associated with the risk of coronary heart disease (Danesh *et al.* 2001; Woodward *et al.* 2007). However, not all studies have reported a D-dimer association with classical CVD risk factors, such as smoking, lipids and blood pressure (Danesh *et al.* 2001). Some evidence suggests that the use of D-dimer in prediction models can increase the probability of predicting cardiovascular adverse events in addition to classical risk factors (Smith *et al.* 2005). Conversely, other studies have not found a relationship between D-dimer and CVD, and subsequent mortality has not been observed or remained after adjustment for risk factors (Empana *et al.* 2008; May *et al.* 2007; Wannamethee *et al.* 2009).

1.3 Vitamin D and cardiovascular disease

Vitamin D deficiency has been proposed as a risk factor for CVD (Swales & Wang 2010; Wallis *et al.* 2008; Zittermann & Gummert 2010), with much of this evidence coming from observational studies.

CVD mortality has been correlated with determinants of vitamin D status, including time of year, latitude and altitude. Increases in CVD mortality have been reported during wintertime when serum vitamin D status is at its lowest level (Scragg 1981; Zipes 1999). In the past, higher CVD mortality rates have been seen in Scotland, Finland and Northern Ireland, compared with Eastern Mediterranean countries that lie at lower latitudes, following patterns of low-high sunlight exposure (Smith & Tunstall-Pedoe 1984). Similarly, in areas lying at higher altitudes CVD mortality rates in men are lower, compared with rates from areas lying at lower altitudes (Fabsitz & Feinleib 1980; Mortimer, Jr. *et al.* 1977; Pickle & Gillum 1999). Patterns in CVD mortality rates become somewhat more inconsistent along the south-to-north gradient of the USA. Sun exposure is less in northern States compared with southern (Holick *et al.* 2007); however CVD mortality rates are highest in southern States of Mississippi and Oklahoma for all ethnicities (Pickle & Gillum 1999), suggesting that other risk factors of CVD are prominent in these areas.

Nutritional vitamin D status has been associated with risk factors of CVD and subsequent death caused by it. Vitamin D status has also been found to have a protective association with hypertension and to a lesser extent with type 2 diabetes (T2D), which are risk factors of CVD (Forman *et al.* 2007; Forman *et al.* 2008; Hyppönen *et al.* 2008; Martins *et al.* 2007; Zhao *et al.* 2010).

Hypertension, defined as systolic blood pressure over 140 mmHg, diastolic blood pressure over 90 mmHg, or taking anti-hypertensive medication, has been negatively related to vitamin D status in several cross-sectional and prospective studies. Three large cross-sectional studies have reported a reduced prevalence of hypertension in participants with high concentrations of 25(OH)D, compared to those with low 25(OH)D concentrations (Hyppönen *et al.* 2008; Martins *et al.* 2007; Zhao *et al.* 2010). Analysis of the data collected in NHANES III of some 15,000 adults, found that participants with 25(OH)D concentrations <52.5 nmol/L had 30% increased prevalence of hypertension, compared to participants with 25(OH)D concentrations ≥ 92.5 nmol/ L (Martins *et al.* 2007). The later NHANES 2003-2006 study showed a similar protective 25(OH)D association in the prevalence of hypertension (Zhao *et al.* 2010).

Further evidence of protective 25(OH)D relationship with the risk of developing hypertension has come from prospective studies. A prospective study that followed both sexes, found that individuals with 25(OH)D concentrations ≤ 37.5 nmol/L had a much greater risk of developing hypertension, compared to those with ≥ 75 nmol/L, and the risk of hypertension was greater in men, compared with women (Forman *et al.* 2007). A nested case-control study of a prospective female cohort observed a negative trend in the risk of incident hypertension by increasing 25(OH)D concentrations (Forman *et al.* 2008). A cross-sectional analysis in elderly men aged 70 years, found an increased prevalence of hypertension in men with 25(OH)D concentrations < 37.5 nmol/L, compared to men with 25(OH)D between 50 and 75 nmol/L; but compared to men with 25(OH)D > 75 nmol/L no association with

hypertension was observed (Burgaz *et al.* 2011). Not all studies have reported a statistically significant relationship between 25(OH)D and hypertension. Results from the Tromsø Study (Norway) found no association between 25(OH)D and the risk of developing hypertension or being put on anti-hypertensive medication after following some 2,400 subjects for 14 years (Jorde *et al.* 2010a). Likewise, a relationship between 25(OH)D and the prevalence of hypertension was not significant in the study of some 2,500 adults of European ancestry living in the USA that had a mean age of 65 years (Brock *et al.* 2011).

Observational evidence of an protective 25(OH)D association with T2D (often defined by the results of fasting glucose tests) is strongest when comparing the prevalence of T2D between study participants with 25(OH)D concentrations at the extremes of the range, and trend in prevalence across 25(OH)D categories (Brock *et al.* 2011; Gagnon *et al.* 2011; Martins *et al.* 2007; Pittas *et al.* 2010b). A study on the incidence of T2D based on a 7 year follow up of 1728 participants, predicted a protective 25(OH)D association with the risk of T2D incidence (Liu *et al.* 2010). However, two large prospective studies that followed participants for 7 to 11 years, respectively, reported a null 25(OH)D association with the risk of incident T2D after adjusting for BMI (Grimnes *et al.* 2010; Robinson *et al.* 2011). Another large prospective study of two Finnish cohorts pooled together, only found an inverse 25(OH)D relationship with the risk of T2D incidence in men, but not in women (Knekt *et al.* 2008).

Mortality from CVD has been linked to serum vitamin D status. An early case-control study of 179 patients presenting with MI, found participants with $25(OH)D_3$

concentrations ≥ 33 nmol/L had a lower relative risk of MI, than participants with $25(OH)D_3$ < 25 nmol/L (Scragg *et al.* 1990). The more recent nested case-control study from the male Health Professionals Follow-up Study (HPFS), that had a mean recruitment age of 64 years and followed the men for 10 years, also found a protective 25(OH)D association with the risk of MI (Giovannucci *et al.* 2008). In the follow-up period, 454 men reported having MI out of the initial 18,225 men recruited. It was found that those with higher 25(OH)D concentrations (≥ 75 mol/L) had a reduced risk of MI, compared to men with lower 25(OH)D concentrations (≤ 37.5 mol/L). Likewise, a similar aged USA based study of men and women at baseline, found a reduction in incidence risk of the first cardiovascular event when comparing participants with 25(OH)D≥ 37.5 nmol/L, to those with 25(OH)D ≤ 25 nmol/L (Wang *et al.* 2008). However, after the data were stratified by the presence of hypertension in participants, the association between 25(OH)D and risk of cardiovascular event incidence was only present in those with hypertension.

Results from prospective mortality studies are somewhat more mixed and appear to depend on age of recruitment and length of follow up (Figure 1.7). The largest prospective study and the longest was based on the Mini-Finland Health Survey carried out in 1978-1980, and analysed a random sample of the Finnish adult population (Kilkkinen *et al.* 2009). Adults recruited in the study at baseline were aged at least 30 years and were free from CVD. Concentrations of 25(OH)D (assayed retrospectively from frozen blood samples) were divided into sex specific quintiles. For all cardiovascular deaths, there was a significant reduction in the risk of dying in individuals with 25(OH)D concentrations in the top 20%, compared to those with concentrations in the bottom 20% after adjustment for several

confounders. After reclassifying CVD deaths into cerebrovascular or coronary, the reduction in mortality risk remained significant between the highest and lowest quintiles of 25(OH)D for cerebrovascular death, but was not so for coronary death.

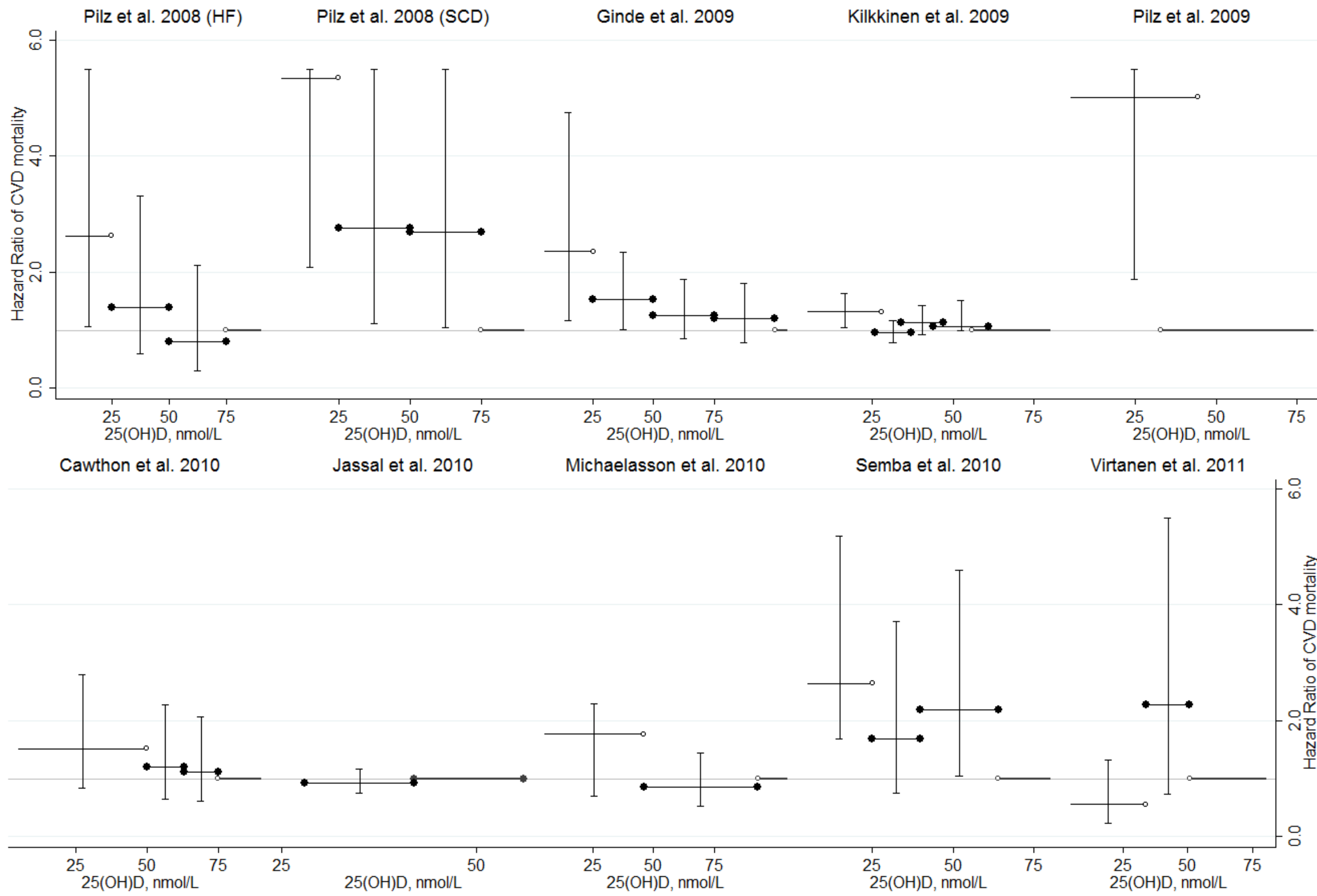

Pilz et al. 2008 (HF)
Pilz et al. 2008 (SCD)
Ginde et al. 2009
Kilkkinen et al. 2009
Pilz et al. 2009
Cawthon et al. 2010
Jassal et al. 2010
Michaelasson et al. 2010
Semba et al. 2010
Virtanen et al. 2011
Hazard Ratio of CVD mortality
0.0
2.0
4.0
6.0
25
50
75
25(OH)D, nmol/L

Figure 1.7. Hazard ratios (HR) of CVD mortality by 25(OH)D nmol/L categories fully adjusted as reported in the studies. Kilkkinen et al 2009 and Virtanen et al 2011 hazard ratios were flipped to make it comparable with the other studies. Kilkkinen et al 2009 used sex-specific 25(OH)D quintiles. The horizontal lines indicate 25(OH)D categories, and thin grey line indicates the 25(OH)D reference with an open circle at the left end of the line. The lowest 25(OH)D category is indicated by an open circle on the right end of the line. The vertical lines are the 95% CI for the HR and where a dash is absent indicates that the upper 95% CI is beyond the boundaries of the y-axis in the graph.

In prospective studies of older people, a protective 25(OH)D association with the risk of mortality caused by CVD has not been consistently observed. For the Uppsala Longitudinal Study of Adult Men (ULSAM), older men were recruited and at baseline the mean age was 71 years (Michaelsson *et al.* 2010). Nearly half of the men had passed away during the follow-up period that was a median length of 12.7 years. After adjustment for several lifestyle, socio-economic and health factors including illness and serum biomarkers, a significant 25(OH)D association with the risk of CVD mortality did not remain across the 25(OH)D categories. This null finding was also observed in a USA based prospective study in men aged at least 65 years, where after adjusting for similar factors to the ULSAM study, there was no risk reduction in the incidence of CVD mortality across the 25(OH)D categories (Cawthon *et al.* 2010). Likewise, a null finding was also observed in a slightly younger mixed sexes population based prospective study of eastern Finnish individuals aged at baseline between 53-73 years (Virtanen *et al.* 2010).

Two prospective studies based on populations of older individuals from sunny climates, with comparable sample sizes, and follow up periods have reported different results. The InCHIANTI study was based in Tuscany, Italy, and

participants at baseline were aged at least 65 years old (Semba *et al.* 2010). For those who had 25(OH)D concentrations in the bottom quartile, there was an increased risk of CVD mortality, compared to participants with 25(OH)D in the top quartile. In contrast with these positive findings, were the null findings from the Rancho Bernardo Study (Jassal *et al.* 2010). This Southern California study was based on individuals of European ancestry with baseline median age of 76 years. There was no significant difference in the risk of CVD mortality per standard deviation increase of 25(OH)D concentrations.

Data from the multi-ethnic USA based NHANES III have been analysed with respect to 25(OH)D and CVD mortality in two separate studies, and a cross-sectional study on CVD. In the prospective analysis of NHANES III adults aged at least 20 years indicated that the rate ratios of CVD mortality did not vary across 25(OH)D categories after adjusting for several lifestyle, socio-economic and health factors (Melamed *et al.* 2008). In contrast, were the findings from the restricted analysis that included only older adults aged 65 years or more from NHANES III (Ginde *et al.* 2009b). During a median follow up of 7.3 years, nearly half of the participants had passed away. Participants with 25(OH)D concentrations ≥ 100 nmol/L had a lower incidence of CVD death, than those with 25(OH)D < 50 nmol/L, after adjusting for key variables including illness. A broader cross-sectional analysis of NHANES III adults aged at least 18 years, found a protective 25(OH)D association with self-reported CVD for participants with 25(OH)D ≥ 50 nmol/L, compared with participants with lower concentrations (Kendrick *et al.* 2009).

The largest reduction in incidence risk of CVD mortality has been found in the analysis of the individuals in the German LURIC study (Pilz *et al.* 2008). Study participants were patients referred to coronary angiography and had a heightened risk of coronary events at the time of recruitment. After adjustment for lifestyle and health variables, for sudden cardiac death (SCD) a HR of 5.35 (95% CI 2.09, 13.67) was found when comparing participants with 25(OH)D < 25 nmol/L, to those with 25(OH)D ≥ 75 nmol/L. A very similar reduction in risk of CVD mortality incidence was also found in the analysis of a population cohort study recruited from the Dutch town of Hoorn of adults aged 50 to 75 years (Pilz *et al.* 2009). The 25(OH)D concentrations were divided into sex-specific quartiles, and after adjusting for lifestyle and health variables, a HR of 5.02 (95% CI 1.88, 13.42) for CVD mortality was observed when comparing participants with 25(OH)D concentrations in the upper three quartiles, with those with levels in the lowest quartile.

The majority of these studies were part of a meta-analysis on cardiometabolic outcomes that also included randomised control trials (RCT) (Pittas *et al.* 2010a). For hypertension, the studies of three cohorts (Forman *et al.* 2007; Forman *et al.* 2008) were deemed to be of sufficient quality to be meta-analysed together. A significant 25(OH)D association was found with incidence of hypertension when comparing meta-analysed results in the lowest 25(OH)D category (37.1-51 nmol/L) with the highest (75.1-81 nmol/L). However, in the meta-analysis of RCTs that had intervened with vitamin D supplements, no significant change was found in systolic or diastolic blood pressure. Meta-analysis was not performed on the 25(OH)D association with the risk of incident T2D or cardiovascular mortality or events, due to the variation in assessment, outcomes and definition of risk. For RCTs on T2D or

CVD, interventions and outcomes greatly differed and were likewise not meta-analysed together. A further systematic review of RCTs for cardiovascular events found two trials that reported supplementing with vitamin D and three trials with vitamin D and calcium (Wang *et al.* 2010a). Combined, there was no risk difference in cardiovascular events between groups not supplemented, supplemented with either vitamin D alone or vitamin D and calcium together.

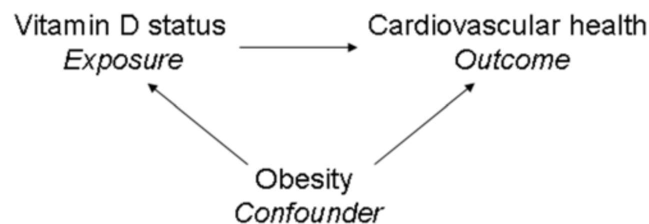
1.4 Rationale

Evidence from several observational studies suggests that vitamin D might have a relationship with risk factors of CVD and subsequent death from the disease. However, better understanding of this potential role is required as the accumulated evidence is far from conclusive. Knowledge of whether relationships between serum vitamin D status and novel biomarkers of CVD in sub-clinical CVD individuals are needed, as these may elucidate pathways in the progression of the disease. Regardless of whether associations will be discovered, this investigation will provide knowledge towards the role of vitamin D in cardiovascular health.

There are challenges with observational research in vitamin D. The inter-relationships between adiposity, cardiovascular health and vitamin D complicate the study. Obesity is a determinant of current vitamin D status (Hyppönen & Power 2007), and large adiposity stores are a proposed risk factor of reduced vitamin D levels (Cheng *et al.* 2010). Obesity is a direct risk factor of CVD and many of the conditions that are also risk factors for CVD, such as diabetes mellitus (Brown *et al.* 2009). High BMI is adversely associated with the risk of CVD (Ni *et al.* 2004). Individuals have a 40% higher risk of vascular mortality for each 5 kg/m^2 increase in

BMI above 25 kg/m^2, compared to individuals with a BMI between 22.5-25 kg/m^2 (Whitlock *et al.* 2009). Obesity is a confounding factor between vitamin D status and cardiovascular health (Figure 1.8 A), as 25(OH)D and cardiovascular health remained linked by obesity after removing the direct path between them (Figure 1.8 B) (Greenland *et al.* 1999).

A



B

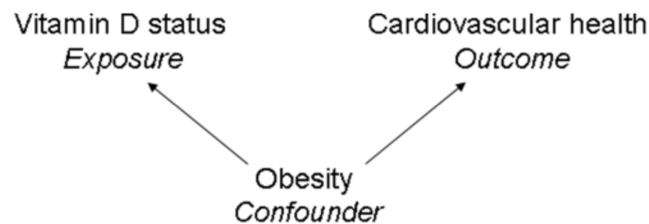


Figure 1.8 The proposed pathway of the exposure's (vitamin D status) influence on the outcome of cardiovascular health. Figure A shows the direct path of vitamin D status on cardiovascular health. Figure B excludes the direct pathway, but due to role of obesity on both the exposure and outcome there is still a link.

Control for confounding in a cross-sectional observational study can be done by adjusting for the confounder in the statistical model, but this approach is not without issues (Glymour & Greenland 2008). It is possible that after adjusting for a

confounder, its effect will not be completely controlled and residual confounding will still remain. Furthermore, statistical power to detect a 25(OH)D association with a cardiovascular biomarker will be reduced after adjusting for adiposity, as the 25(OH)D association is expected to be much smaller than the adiposity association (Cohen 1988a). In this thesis, different methodologies are used in an attempt to control the confounding effect of adiposity on 25(OH)D status and cardiovascular health biomarkers.

1.5 Aim

To explore the associations between vitamin D status (25(OH)D) and novel biomarkers of cardiovascular health from inflammation and haemostasis, in participants of the 1958BC free from clinical CVD using different methodologies.

1.5.1 Objectives

1. To explore the associations between 25(OH)D and CRP, D-dimer, fibrinogen, tPA, vWF by using multiple linear regression and consideration of confounders such as adiposity.
2. To model seasonal patterns of CRP, D-dimer, fibrinogen, tPA, vWF unadjusted and adjusted for 25(OH)D. To use the framework of mediation to infer an association between 25(OH)D and the biomarkers based on the seasonal patterns observed, since it would be expected that adiposity would have little seasonal variation.
3. To find genetic variants associated with 25(OH)D using the approach of genome-wide association.

4. To review genetic variants associated with 25(OH)D and to explore their suitability as instruments for 25(OH)D in the framework of Mendelian randomisation.
5. To explore the associations between 25(OH)D and CRP, D-dimer, fibrinogen, tPA, vWF by using genetic variants as instruments for 25(OH)D in the framework of Mendelian randomisation, in order to avoid confounding typically found in cross-sectional studies.

1.6 Overview of the thesis and related papers

The rest of this thesis is presented in the following order:

Data analysed in this thesis is from the 1958BC, specifically the biomedical survey that took place when the cohort members were aged 44-46 years old. Data collected in the survey included blood samples that were used in assays for the biomarkers and genotyping of the DNA. Chapter 2 provides information on the 1958BC and data used in this thesis. A description of specific DNA quality control procedures that were applied to the genetic data is also given in preparation for genome-wide association study.

Various methodologies and analytical approaches were used in this thesis in an attempt to control confounding the relationships between 25(OH)D and the inflammation/haemostatic biomarkers. In Chapter 3 to meet objectives 1 and 2, methods used are: multiple linear regression; a harmonic function inside random effects regression to model seasonal patterns; and mediation to infer a 25(OH)D relationship with the inflammation/haemostatic biomarkers from the seasonal patterns. Findings from the techniques are presented and discussed.

Genetic variants that are associated with 25(OH)D are required in order to use the framework of Mendelian randomisation to infer a relationship between 25(OH)D and the inflammation/haemostatic biomarkers. Chapter 4 presents a genome-wide association study (GWAS) on 25(OH)D to meet objective 3. In Chapter 5, genetic variants associated with 25(OH)D are reviewed. The genes suitability as instruments for vitamin D status in a Mendelian randomisation study is explored, along with candidate variants from GWAS on skin/hair pigmentation and tanning ability. Instrumental variable regression is used to model the framework of Mendelian randomisation using suitable genetic variants as instruments for the 25(OH)D association with the inflammation/haemostasis biomarkers. Chapter 5 meets objectives 4 and 5.

The final discussion of the thesis is presented in Chapter 6.

Papers published in peer-reviewed journals directly related to work in this thesis are:

Paper I

Hyppönen, E., **Berry, D**., Cortina-Borja, M. & Power, C. 2010. 25-Hydroxyvitamin D and pre-clinical alterations in inflammation and hemostatic markers. *PLoS One*, 5, (5) e10801

Paper II

Wang, T.J.*, Zhang, F.*, Richards, J.B.*, Kestenbaum, B.*, van Meurs, J.B.*, **Berry, D**.*, Kiel, D.P., Streeten, E.A., Ohlsson, C., Koller, D.L., Peltonen, L., Cooper, J.D., O'Reilly, P.F., Houston, D.K., Glazer, N.L., Vandenput, L., Peacock, M., Shi, J., Rivadeneira, F., McCarthy, M.I., Anneli, P., de, B., I, Mangino, M.,

Kato, B., Smyth, D.J., Booth, S.L., Jacques, P.F., Burke, G.L., Goodarzi, M., Cheung, C.L., Wolf, M., Rice, K., Goltzman, D., Hidiroglou, N., Ladouceur, M., Wareham, N.J., Hocking, L.J., Hart, D., Arden, N.K., Cooper, C., Malik, S., Fraser, W.D., Hartikainen, A.L., Zhai, G., Macdonald, H.M., Forouhi, N.G., Loos, R.J., Reid, D.M., Hakim, A., Dennison, E., Liu, Y., Power, C., Stevens, H.E., Jaana, L., Vasan, R.S., Soranzo, N., Bojunga, J., Psaty, B.M., Lorentzon, M., Foroud, T., Harris, T.B., Hofman, A., Jansson, J.O., Cauley, J.A., Uitterlinden, A.G., Gibson, Q., Jarvelin, M.R., Karasik, D., Siscovick, D.S., Econs, M.J., Kritchevsky, S.B., Florez, J.C., Todd, J.A., Dupuis, J.*, Hyppönen, E.*, & Spector, T.D.* 2010. Common genetic determinants of vitamin D insufficiency: a genome-wide association study. *Lancet*, 376, (9736) 180-188
* Authors contributed equally

<u>Paper III</u>

Berry, D & Hyppönen, E. 2011. Determinants of vitamin D status: focus on genetic variations. *Curr.Opin.Nephrol.Hypertens.*, 20, (4), 331-336

<u>Paper IV</u>

Berry, D.J*., Vimaleswaran, K.S.*, Whittaker, J.C., Hingorani, A.D., & Hypponen, E. 2012. Evaluation of genetic markers as instruments for mendelian randomization studies on vitamin d. *PLoS.ONE.*, 7, (5) e37465
* Authors contributed equally

The papers are shown in Appendix 1.

First author abstracts and poster presentations directly related to work in this thesis are:

Berry, D.J., Hingorani A.D., Hyppönen E. “Vitamin D: selecting instruments for Mendelian randomisation” Poster presentation at UCL Institute of Child Health open day, Nov 2010

Berry, D.J., Hingorani AD, Hyppönen E. “Vitamin D: selecting instruments for Mendelian randomisation” Poster presentation at Human Genome Organisation (HUGO), May 2010

Berry, D., Cortina-Borja M, Hyppönen E. “Modelling the effect of vitamin D status on seasonal patterns in haemostatic markers” Poster presentation at UCL Graduate School open day, Mar 2010

Berry, D., Cortina-Borja M, Hyppönen E. “Modelling the effect of vitamin D status on seasonal patterns in haemostatic markers” Poster presentation at The 14th Vitamin D workshop Brugge, Beligum, Oct 2009.

Berry, D., Cortina-Borja M, Hyppönen E. “Modelling the effect of vitamin D status on seasonal patterns in haemostatic markers” Poster presentation at Royal Statistical Society conference, Nottingham, Sep 2008.

I have contributed to several papers whilst working towards the thesis and am highlighting a selection of studies that have influenced this thesis. I am drawing attention to these particular papers as I gained skills and knowledge from my involvement that I have directly applied and used in this work. These peer reviewed papers are:

Berry, D.J., Hesketh, K., Power, C., & Hyppönen, E. 2011. Vitamin D status has a linear association with seasonal infections and lung function in British adults. Br.J.Nutr., 106, (9) 1433-1440

Freathy, R.M.*, Mook-Kanamori, D.O.*, Sovio, U.*, Prokopenko, I.*, Timpson, N.J.*, **Berry, D.J.***, Warrington, N.M.*, Widen, E., Hottenga, J.J., Kaakinen, M., Lange, L.A., Bradfield, J.P., Kerkhof, M., Marsh, J.A., Magi, R., Chen, C.M., Lyon, H.N., Kirin, M., Adair, L.S., Aulchenko, Y.S., Bennett, A.J., Borja, J.B., Bouatia-Naji, N., Charoen, P., Coin, L.J., Cousminer, D.L., de Geus, E.J., Deloukas, P.,

Elliott, P., Evans, D.M., Froguel, P., Glaser, B., Groves, C.J., Hartikainen, A.L., Hassanali, N., Hirschhorn, J.N., Hofman, A., Holly, J.M., Hyppönen, E., Kanoni, S., Knight, B.A., Laitinen, J., Lindgren, C.M., McArdle, W.L., O'Reilly, P.F., Pennell, C.E., Postma, D.S., Pouta, A., Ramasamy, A., Rayner, N.W., Ring, S.M., Rivadeneira, F., Shields, B.M., Strachan, D.P., Surakka, I., Taanila, A., Tiesler, C., Uitterlinden, A.G., van Duijn, C.M., Wijga, A.H., Willemsen, G., Zhang, H., Zhao, J., Wilson, J.F., Steegers, E.A., Hattersley, A.T., Eriksson, J.G., Peltonen, L., Mohlke, K.L., Grant, S.F., Hakonarson, H., Koppelman, G.H., Dedoussis, G.V., Heinrich, J., Gillman, M.W., Palmer, L.J., Frayling, T.M., Boomsma, D.I., Davey, S.G., Power, C., Jaddoe, V.W., Jarvelin, M.R., & McCarthy, M.I. 2010. Variants in ADCY5 and near CCNL1 are associated with fetal growth and birth weight. Nat Genet., 42, (5) 430-435

* Authors contributed equally

Ward, M., **Berry, D.J.**, Power, C., & Hyppönen, E. 2011. Working patterns and vitamin D status in mid-life: a cross-sectional study of the 1958 British birth cohort. *Occup.Environ.Med.*, 68, (12) 902-907

Hyppönen, E., **Berry, D.J**., Wjst, M., & Power, C. 2009. Serum 25-hydroxyvitamin D and IgE - a significant but nonlinear relationship. *Allergy*, 64, (4) 613-620

Hyppönen, E., Boucher, B.J., **Berry, D.J.**, & Power, C. 2008. 25-hydroxyvitamin D, IGF-1, and metabolic syndrome at 45 years of age: a cross-sectional study in the 1958 British Birth Cohort. *Diabetes*, 57, (2) 298-305

2 Data source

2.1 Introduction

The British 1958 birth cohort (1958BC) originated as the Perinatal Mortality study (Bulter & Bonham 1963). All births in England, Wales and Scotland for one week in March 1958 were surveyed, and deaths were followed through to the end of May that year. There were 17,638 infants recruited in total and data were collected from their mothers, midwives and medical records. Childhood sweeps took place when the cohort members were aged 7, 11 and 16 years old. As part of the childhood sweeps, children born in the reference week that had been previously missed by earlier surveys or had subsequently immigrated into the country since the last survey were recruited (n=920), giving a total of 18,558 cohort members (Centre for Longitudinal Studies 2009; Shepherd 1995). Identification of the additional children recruited was done through school registers (Shepherd 1995). Data collected at the childhood sweeps were from parents, schools (teachers or head teacher), the cohort members and medical examinations. The subsequent adulthood sweeps have had no further recruitment and have taken place when the cohort members have been aged 23, 33, 42, 45, 46, 50 years old (Centre for Longitudinal Studies 2009; Power & Elliott 2006). At adulthood, the primary method of data collection has mostly been in the form of questionnaires. The data used in this thesis comes from the first biomedical sweep when the cohort members were aged 44-46 years old.

2.2 Biomedical survey

The period of data collection for the biomedical survey was from September 2002 until March 2004 (Fuller *et al.* 2006). Contact was attempted with cohort members that had participated directly (not by proxy) in the previous sweep and were still considered eligible (Fuller *et al.* 2006). The eligibility criteria for biomedical sweep was: that the cohort member was alive and residing in the United Kingdom (UK) with a valid address; was not serving overseas as part of the armed forces; had not permanently refused or refused via proxy; was not considered a difficult case (on the basis of inappropriate behaviour at previous sweeps); and had not participated in the third pilot study for the biomedical sweep (Fuller *et al.* 2006). Cohort members were asked to complete questionnaires and each member was visited by a nurse who collected the questionnaires. If the cohort member had given consent, then the visiting nurse recorded health and anthropometric measures, took up to four non-fasting blood samples, and a saliva sample. The blood samples were sent off by the nurses to the participating laboratories in London, Newcastle and Glasgow. Separate consent was sought from the cohort member for four different uses of their blood samples - that being: blood collection and specified analysis; storage of blood and future analysis; deoxyribonucleic acid (DNA) extraction (storage and analysis); and storage and creation of immortalised cell lines. The response and numbers achieved for particular aspects of the survey can be seen in Table 2.1

Table 2.1 Number of cohort members surveyed in 1958BC biomedical sweep and percentage of final response achieved for selected body measures and blood samples (Fuller *et al.* 2006)

	Number (%*)
Issued sample†	11964
Final Response**	9377
At least one blood sample achieved	8207 (87.5%)
Blood pressure and pulse	9269 (98.8%)
Height/sitting height	9248 (98.6%)
Weight	9218 (98.3%)
Waist/hip circumference	9272 (98.9%)
25(OH)D concentrations	7591 (81.0%)

* The percentages are taken from the final response (*n*=9377)

† Excludes 91 ineligible subjects who had either emigrated or died

** Excluded 10 subjects whose data were lost in transmission or later asked to be removed from the cohort and includes 29 subjects who took part in the pilot study prior to the main survey.

The biomedical survey did suffer from some sample loss, however a third of those lost to the cohort were considered unavoidable due to death or emigration (Atherton *et al.* 2008). Particular groups underrepresented in the biomedical survey were ethnic minorities and the disadvantaged. Disadvantaged participants were those who had been born in a household with no male head present and who at 7 years of age had a period of social care, problematic behavioural or poor cognitive characteristics, or were short for their age. At age 45 years, the total surviving cohort was mostly of European ancestry (96.8%), and participants of the biomedical survey with European ancestry had increased slightly to 98%. The survey is considered broadly

representative of the surviving cohort and of the general middle-aged population in key adult lifestyle, social-economic and biomedical characteristics (Atherton *et al.* 2008).

2.2.1 Phenotype data

A description of the phenotype data from the survey that is used in the analyses for this thesis is described in two tables. The serum biomarkers laboratory techniques from the collected blood samples are described in Table 2.2, and a description of lifestyle and socio-economic covariates with categories used in the statistical analyses in Table 2.3.

Table 2.2 Biomarker data used in the analyses of this thesis assayed from blood samples taken for the biomedical survey

Description of biomarker	Units of Measure	Laboratory technique
25-Hydroxyvitamin D, 25(OH)D	nmol/L	The IDS OCTEIA enzyme-linked immunosorbent assays (ELISA) automated application on the Dade-Behring BEP2000 analyser. The concentrations were then validated and adjusted according to "All-Laboratory Trimmed Mean" of the Vitamin D External Quality Assessment Scheme (DEQAS) (Hyppönen *et al.* 2007)
C-reactive Protein (CRP)	g/L	High sensitivity CRP assay using BN Prospec protein analyser (Dabe Behing), which passes a plane-polarised light through a solution mixed from the plasma sample to determine the CRP antigen concentrations.
D-Dimer (fibrin)	ng/ml	Measured using stored blood samples by ELISA assay over two kits. Replicate measures done in both kits and standardized for interbatch variation (Johnson 2007).
Fibrinogen	g/L	A Biomerieux MDA 180 coagulometer with the Claus method on citrated plasma.
Tissue plasminogen activator (tPA) and von Willebrand factor (vWF)	ng/ml, IU/dl	A double sandwich technique on citrated plasma by a Biopool antigen ELISA.

Description of biomarker	Units of Measure	Laboratory technique
Triglycerides, total cholesterol, high density lipoprotein (HDL), low density lipoprotein (LDL)	mmol/L	Measured using an Olympus AU640 autoanalyser and LDL was calculated from total cholesterol, HDL and Triglycerides.
Glycosylated haemoglobin, (Hba1c)	% total	The Tosoh Haemoglobin A_{1c} method using a high ion exchange high performance liquid chromatography (HPLC) on whole blood.
Insulin-like growth factor (IGF-1)	(nmol/L)	Measured by a Nichols Advantage IGF-1 chemiluminescence immunoassay
Total circulating immunoglobulin (IgE)	kU/L	Measured by a HYTEC enzyme immunoassay

Table 2.3 Lifestyle, socio-economic and anthropometric covariates used in the analyses of this thesis

Covariates from the 1958BC biomedical survey	
Name	**Description of question used**
Time spent outside	Length of time per day usually spent outside during daylight hours last month recorded as no time, less than 15 minutes, 15-30 minutes, 30 minutes- 1 hour, 1-2 hours, 3-4 hours and 4 or more hours. For the analyses it is combined into categories of < 1 hour, 1-2 hours and 3 or more hours.
Skin colour	The colour of the inner arm, with the recorded choice of light (white, fair, or ruddy), medium (olive, light/medium brown) and dark (dark brown or black).
Sun cover	The use of skin protection with either sunscreen or clothing in sunny weather in the UK or abroad, reported as often, sometimes, rarely or never.
Oily fish consumption	The consumption of oily fish (salmon, trout, mackerel, sardines and fresh tuna) recorded as never, occasionally, < 1 day/week, 1-2 days/week, 3-6 days/week, once a day, 2-4 times/day, 4 or more times/day. For the analyses it is combined into categories of daily, weekly and < weekly.
Alcohol consumption	Drinks per week are calculated using Alcohol Use Disorders Identification Test questionnaire, that records frequency of drinking and standard drinks per session. Frequency was recorded as not in the last month, ≤once/month, 2–4 times/month, 2–3 times/week, and >4 times/week. Quantity recorded was converted to standard units and categorised as none, 1–2 units, 3–4 units, 5–6 units, 7–9 units, and ≥10 units per drinking session. For the analyses it was categorised into non-drinker, light <7, moderate 7-14 units vs heavy 14-21 and very heavy > 21.

Covariates from the 1958BC biomedical survey	
Name	**Description of question used**
PC/TV time	Average leisure time spent per day television (TV) watching or personal computer (PC) usage in the last year reported as none, < 1 hour/day, 1-2 hours/day, 2-3 hours/day, 3-4 hours/day and 4 or more hours. For the analyses it was categorised as < 1 hour/day, 1-2 hours/day and 3 or more hours/day.
Recreation MET hours	The Recreation Metabolic Equivalent of Task (MET) hours per week was derived from up to 37 activities with reported frequencies and usual durations and published MET scores. An additional category was assigned for those subjects with implausibly high values where the weekly recreation hours were 3 standard deviations (SD) above the sex mean (Parsons *et al.* 2009) .
Vigorous activity	Calculated from activity where the MET scores derived from recreation MET hours was greater than six (Atherton *et al.* 2009).
Geographical location	Current region of residence, which is categorised into 13 regions and for the analyses is further reduced to Scotland, northern England, middle England and Wales and southern England.
Respiratory infections	A respiratory infection (influenza, pneumonia, bronchitis or severe cold) within the last 3 weeks recorded as yes or no.
Vitamin D supplements	The use of cod liver or fish oil or other supplements containing vitamin D reported, recorded as yes or no.
Body mass index, waist circumference	Weight and standing height were measured without shoes and in light clothing by a trained nurse. Body mass index was calculated from weight and standing height (kg/m^2). The waist circumference

Covariates from the 1958BC biomedical survey	
Name	**Description of question used**
	was measured by the nurse midway between the costal margin and the iliac crest.
Blood pressure; systolic and diastolic	Using an Omron 705CP automated sphygmomanometer after participants had been in a seated position for an initial 5 minute rest. The measurement was repeated three times and the average was calculated from successful recordings.
Lung function; forced expiratory volume in 1 second (FEV_1) and forced vital capacity (FVC)	Done in accordance to the criteria of the American Thoracic Society 1987(American Thoracic Society 1987), with the participant in the standing position, without nose-clips. Using the Vitalograph Mirco spirometer a minimum of three (up to five) spirograms were recorded until three satisfactory blows were obtained (assessed by best-test variation where the sum of FEV_1 and FVC was $\leq$5%). Measures with a best-test variation $\geq$10% or values with standardized residuals > ±3 SDs were excluded. The highest technically satisfactory values for FEV_1 and FVC from each set of spirograms were used.
Derived covariates from other 1958BC sweeps	
Smoking	Recorded at age 42 and supplemented with data from previous adult sweeps coded as never, quit before age 33 years, quit from ages 33-42 years, smokes 1-19 cigarettes/day and smokes 20 or more cigarettes/day.
Social class at child- and adult-hood	Childhood social class is based on paternal occupation at birth and if missing supplemented with data from sweep in 1965 (at age 7). For adulthood, it was based on the subject's most recent occupation at age 42 yrs, and supplemented with data at age 33 if missing. The following Register General's groups

Covariates from the 1958BC biomedical survey	
Name	**Description of question used**
	were used: I professional, II managerial and technical, IIInm skilled non-manual, IIIm skilled manual, IV partly skilled and V unskilled manual. For childhood social class, the final category was those with no male head present in the household. For adulthood social class, the final category was those who were institutionalised, retired, or unemployed over a long period.

2.2.2 Genetic data

Sub-samples of the cohort have been used as control populations for genome-wide association studies on various diseases with the WTCCC (Wellcome Trust Case Control Consortium 2007a), the second phase of the WTCCC and the Type 1 Diabetes Genetics Consortium (T1DGC) (Barrett *et al.* 2009). DNA was extracted from blood and the immortalised cell lines, which were created from the collected serum samples. This thesis utilises these genome-wide data resources as they have become available. In addition, there has been opportunity to use selected de-novo genetic variants tagged from the consenting participants in the biomedical survey (Cooper *et al.* 2011).

The genetic variants that are used in the analyses of this thesis are single nucleotide polymorphisms.

A ***single nucleotide polymorphism*** (SNP) is an allele variant or point mutation (Ingles 2004). The most common type of SNP is a transition, where purines substitute each other (adenine A or guanine G) or pyrimidines substitute each other (thymine T or cytosine C). Less common is where a purine is substituted for a pyrimidine or vice versa, and is referred to as a transversion. SNPs within the coding region of a gene (exons) can be classified as synonymous or non-synonymous. SNPs located outside of exons are either in introns (inside the region of a gene) or intergenic (outside the region of a gene) areas of the chromosome. Synonymous SNPs do not change amino acid sequence, unlike non-synonymous SNPs that do. Furthermore, a sequence change may cause the codon to alter the amino acid, which is to refer as a missense, or the change might be to a stop codon in the sequence, which is a nonsense SNP.

De novo genotyping

The SNPs were genotyped using the Taqman platform (Applied Biosystems, Warrington, UK) at the Juvenile Diabetes Research Foundation/Wellcome Trust Diabetes and Inflammation Laboratory (JDRF/WT DIL), centred in the Cambridge Institute for Medical Research (Cooper *et al.* 2011). The de novo SNPs that were included into this thesis passed the exclusion criteria of Hardy-Weinberg equilibrium (HWE) p-value > 0.01, minor allele frequency (MAF) > 0.01 and SNP call rate > 80 % (Table 2.4).

Table 2.4 De novo SNPs genotyped on the Taqman platform call rates, MAF and HWE test

Gene symbol	Chromosome position	Gene names	SNPs	Call rate (%)	MAF	HWE *p*-value
MC1R	16q24.3	Melanocortin 1 receptor	rs1805005	98.4	0.12	0.27
MC1R	16q24.3	Melanocortin 1 receptor	rs2228479	98.3	0.10	0.26
OCA2	15q	Oculocutaneous albinism II	rs4778138	97.4	0.12	0.09
OCA2	15q	Oculocutaneous albinism II	rs7495174	96.9	0.06	0.42
OCA2	15q	Oculocutaneous albinism II	rs4778241	95.9	0.17	0.48
SLC45A2	5p13.2	Solute carrier family 45, member 2	rs13289	97.4	0.38	0.90
CYP24A1	20q13	Cytochrome P450, family 24, subfamily A, polypeptide 1	rs6013897	98.9	0.80	0.64
CYP27B1	12q13.1-q13.3	Cytochrome P450, family 27, subfamily B, polypeptide 1	rs10877012	95.1	0.33	0.17
CYP2R1	11p15.2	Cytochrome P450, family 2, subfamily R, polypeptide 1	rs10741657	82.5	0.40	0.39
GC	4q12-q13	Vitamin D binding protein	rs4588	83.4	0.30	0.24
GC	4q12-q13	Vitamin D binding protein	rs7041	97.8	0.45	0.07
DHCR7/ NADSYN1	11q13.4	7-dehydrocholesterol reductase/ NAD synthetase 1	rs12785878	100.0	0.22	0.80
CYP27A1	2q33	Cytochrome P450, family 27, subfamily A, polypeptide 1	rs17470271	80.6	0.44	0.98

Genome-wide genotyping

The genome-wide data came from three studies (Table 2.5). The first WTCCC project used 1500 subjects randomly chosen from participants of the biomedical survey (referred to as the WTCCC), that were genotyped on Affymetrix Genechip 500K and Illumina Infinium 550K chip (Wellcome Trust Case Control Consortium 2007a; Wellcome Trust Case Control Consortium 2011). The second phase of the WTCCC has re-genotyped the initial 1500 subjects and additional 1500 subjects (*n*=3000) on the newer genechip technology Affymetrix 6.0 and Illumina 1.2 (referred to as the WTCCC2) (Sawcer *et al.* 2011). The T1DGC used 2596 subjects of the cohort (referred to as the T1DGC), exclusive of the WTCCC subjects, and were genotyped on the Illumina Infinium 550K (Barrett *et al.* 2009).

Table 2.5 Genotyping, quality control and imputation of the 1958BC genome-wide sub-samples

Sub-samples	**WTCCC**	**WTCCC2**	**T1DGC**
Time of availability for this thesis	2008	2010	2009
Genotyping			
Platform and SNP panel used in analysis	Affymetrix Genechip 500K* (Santa Clara, CA, USA)	Affymetrix v6.0* (Santa Clara, CA, USA)	Illumina 550K Infinium (San Diego, CA, USA)
Centre	Affymetrix Services Lab, California, USA	Wellcome Trust Sanger Institute, Cambridge, UK	JDRF/WT DIL Lab in Cambridge, UK
Calling algorithm	Chiamo	Chiamo	ILLUMINUS
Quality control prior to imputation			
Centre/Project	Done as part of the WTCCC	Done as part of the WTCCC2	Replicated as part of this thesis (**see next section Quality Control and Imputation with application to T1DGC**) and completed in collaboration with Oxford Centre for Diabetes, Endocrinology and Metabolism,

Sub-samples	**WTCCC**	**WTCCC2**	**T1DGC**
			University of Oxford
Sample QC Filters			
Call Rate	> 0.03/*n*=9	Outlying individuals were excluded on the basis of relatedness, non-European ancestry, sex discrepancy, channel differences and identity	> 0.03 / *n*=2
Heterozygosity	> 0.3 or < 0.225/*n*=0		> 0.34 or <0.29 / *n*=0
Ethnic outliers	Non-European ancestry *n*=6		Non-European ancestry *n*=54
Other exclusions	External discordance *n*=4, 1st/2nd degree relatives *n*=1		Potential sex discrepancy *n*=6
Subjects for analysis (n)	1480	2703	2530
SNP QC Filters			
MAF	> 1%	> 1%	> 1%
HWE p-value	> 5.7e-7	> 1e-20	> 1e-7
Call Rate	Studywise missing data proportion > 0.05 or (studywise MAF < 0.05 and studywise missing data proportion > 0.01)	> 98%	≥95%
Other	None	Statistical info rate ≥ 0.975, plate association > 1e-5	None
SNP's for imputation (*n*)	383,325	721,428	520,413

Sub-samples	WTCCC	WTCCC2	T1DGC
Imputation Statistics			
Centre	Done in collaboration with the Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford	Done in collaboration with the Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford	Done in collaboration with the Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford
Software	Impute	Impute	Impute
Quality metrics	Proper-info > 0.4	Proper-info > 0.4	Proper-info > 0.4
SNP QC filters	MAF > 1%	MAF > 1%	MAF > 1%
SNP's in analysis	2,357,290	2,446,857	3,140,837

* The WTCCC studies have genotyped subjects on both Affymetrix and Illumina platforms, for this thesis the genetic information was taken from the Affymetrix chips.

SNP, subject/sample quality control with application to T1DGC

The quality control (QC) tests and their metrics provide assurance that the data is of satisfactory quality. The QC process highlights problems with the genotyping, duplicates, related individuals, ethnic outliers and contaminated samples (Sullivan & Purcell 2008). The genetic data for the T1DGC sample were provided in the format of one SNP per file with sample identifier, call of genotype, probability of the call, normalised X and Y intensities (Figure 2.1). The clusters in the intensity plot indicate the genotype calls for a SNP, with the middle cluster typically corresponding to the homozygous genotype, and side clusters corresponding to the minor and major heterozygous genotypes. The call of genotype and probability of call were estimated by the Illuminus algorithm (Teo *et al.* 2007) from the signal intensities from the genotyping process. The genotype calling algorithm uses a three component bivariate mixture model where the components correspond to the genotype call of a SNP. A threshold of 0.95 for the probability of the call was used to define a null component/genotype call.

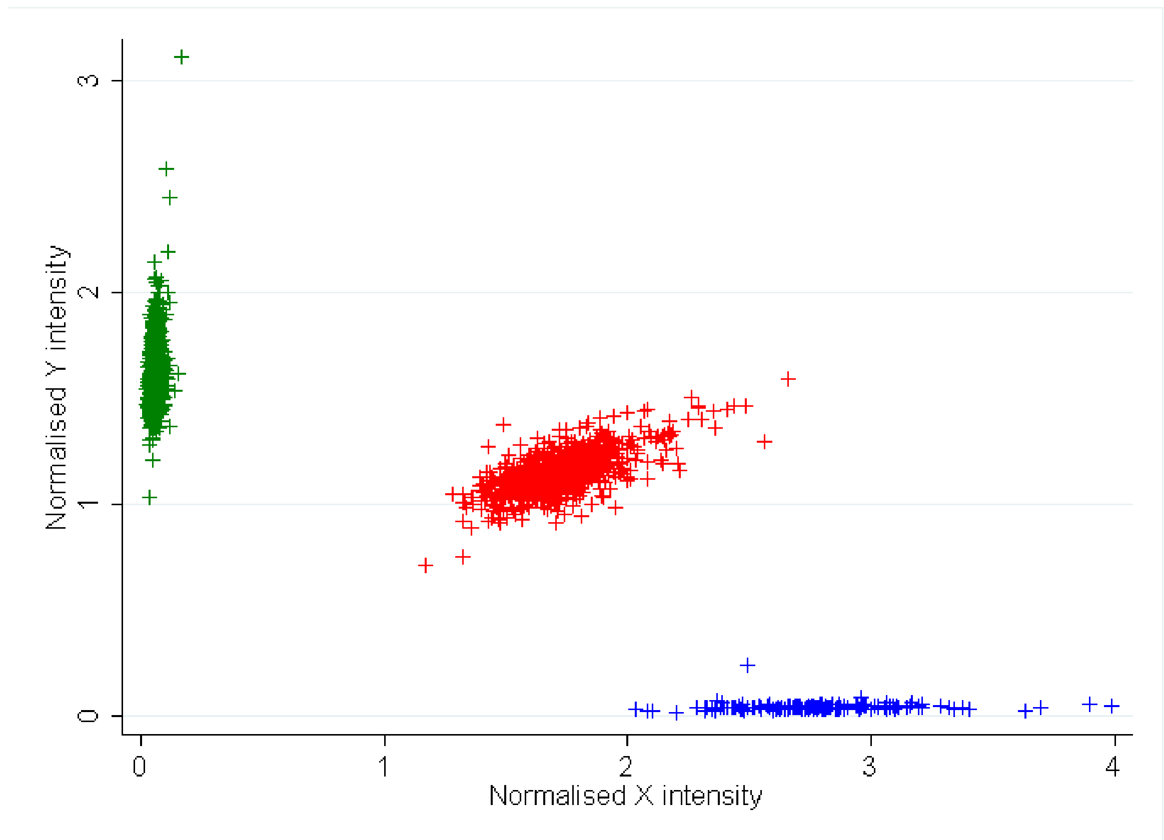


Figure 2.1 The intensity plot for a SNP taken from the T1DGC. Each point represents an individual's X and Y intensity for a signal from the genotyping process. The heterozygous genotypes are in red, the homozygous genotype correspond to the green cluster, and the other, the blue cluster. The intensities were used to estimate the genotype call for the individual.

The QC criteria of the T1DGC sample was based on the criteria in the first phase for the WTCCC (Wellcome Trust Case Control Consortium 2007a). The methods and results from the T1DGC sample QC thresholds are presented in the order of SNP call rate, HWE, MAF, sample call rate, sex discrepancy, cryptic relatedness, heterozygosity, ethnicity/population stratification. The genetic statistical software PLINK (Purcell *et al.* 2007; Purcell 2009) was used for the QC procedures.

SNP call rate

The SNP call rate is the percentage of null genotype calls given the total number of samples for a SNP. In the T1DGC, 95.6% of the SNPs had no missing genotype calls (Figure 2.2). When a SNP had data with 5% or more null genotype calls it was excluded from further analysis.

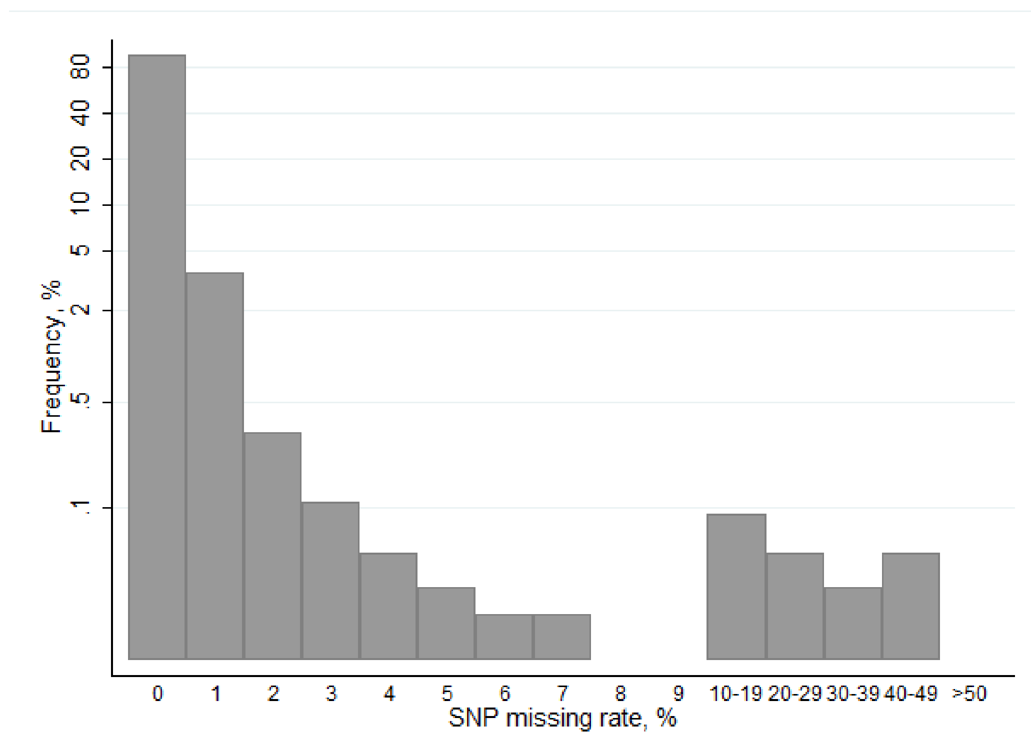


Figure 2.2 T1DGC frequency of SNPs with missing data.

HWE

The HWE is the (ideal) constant state of the allele and genotype frequencies from an autosomal bi-allelic genetic marker that occurs in a large population due to random mating (Thomas 2004c). Copies of the alleles should be inherited independently and follow a Binomial distribution (Cleeves 1999). If the marker has alleles *A* and *B* with probabilities *p* and *q*, where q=1-p, then the genotypes of the marker (*AA, AB,*

BB) should follow the Hardy-Weinberg law and have the probabilities of (p^2, $2qp$, q^2). For SNP QC purposes, any evidence against the HWE can be taken as an indication of sample contamination or consanguinity in a population, with the SNPs in question excluded from further analyses. The exclusion threshold in the T1DGC was SNPs with a HWE *p*-value ≤ 1e-7 (Figure 2.3).

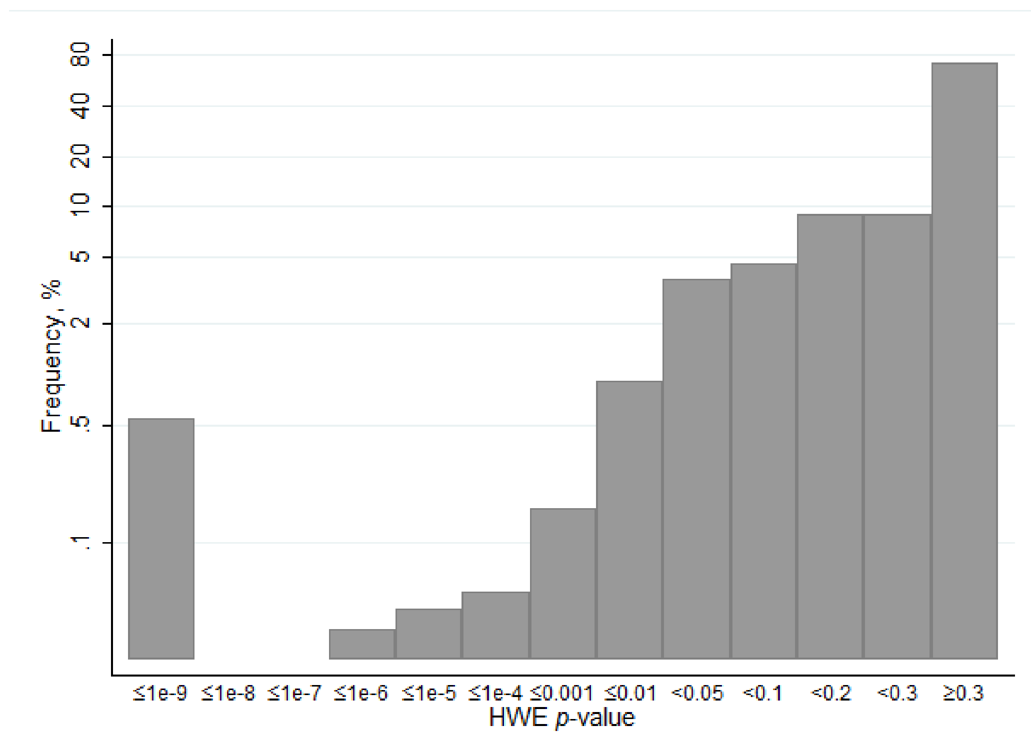


Figure 2.3 T1DGC frequency of SNP HWE *p*-values

MAF

SNPs with low MAF were excluded in genome-wide association analysis as there is limited statistical power for testing (Sullivan & Purcell 2008). In total 24,415 SNPs in the T1DGC had a MAF < 1% (Figure 2.4).

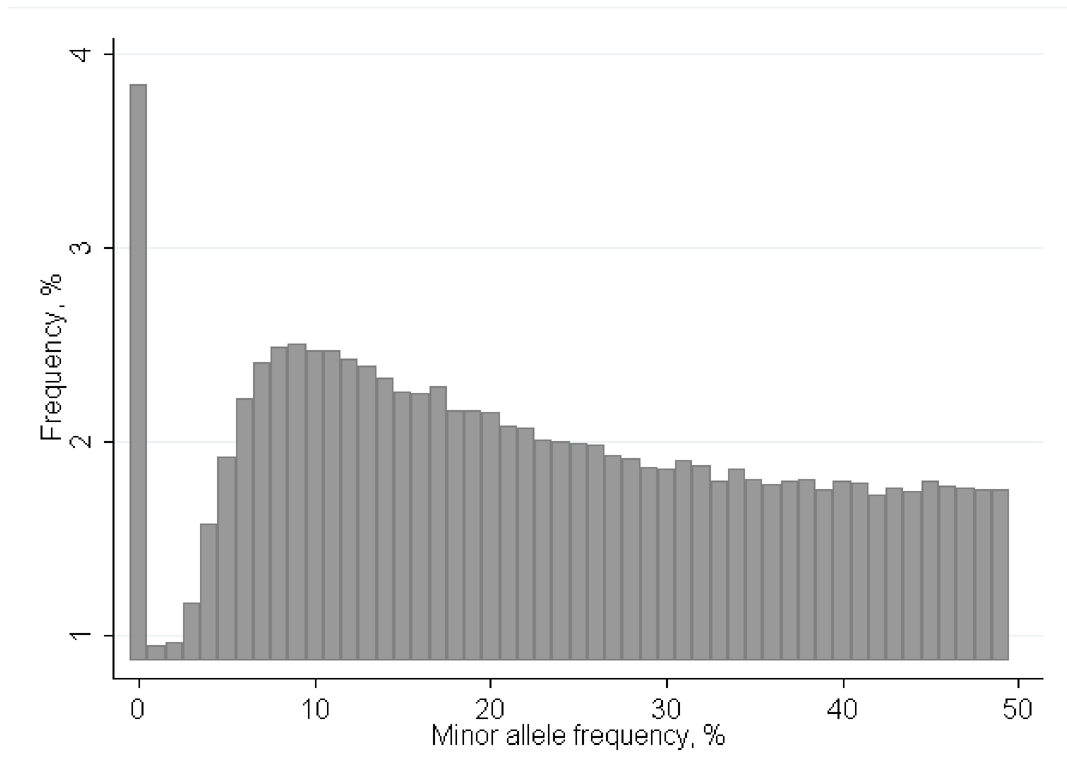


Figure 2.4 T1DGC frequency of SNPs MAF

Sample call rate

The sample call rate is the proportion of SNPs that have a null genotype call for an individual. Two samples/individuals had more than 3% of SNP data missing in the T1DGC (Figure 2.5) and were excluded.

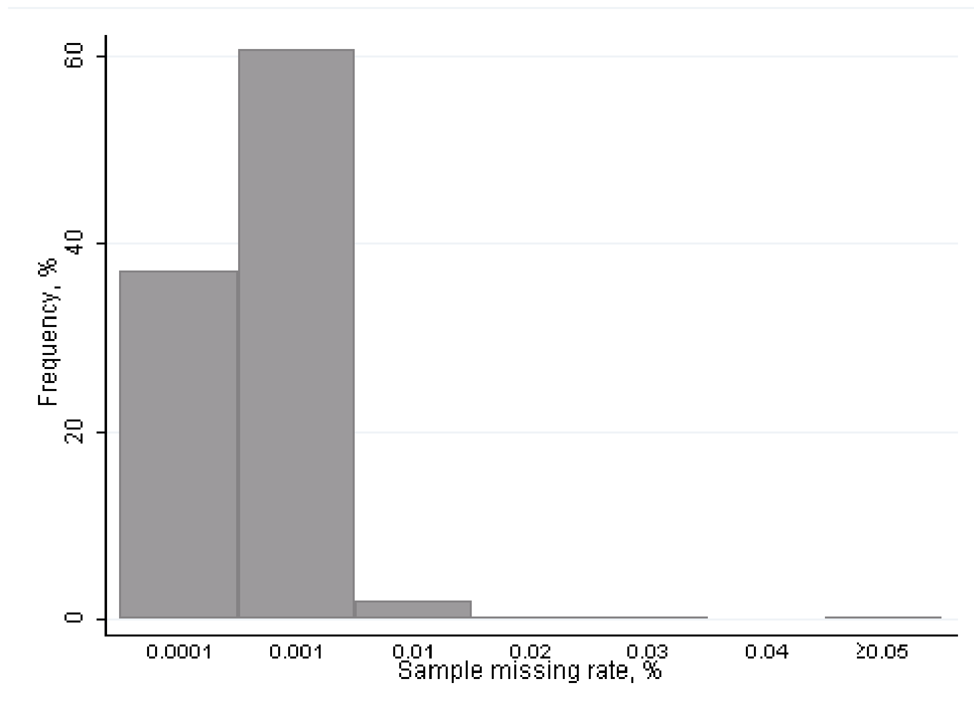


Figure 2.5 T1DGC frequency of sample/subject with missing SNP data

Sex

The sex reported was validated against the X chromosome. This may indicate that sex for the individual was mis-reported, however, for QC purposes it may also indicate that the sample had been contaminated in some way. In the T1DGC n=6 samples were excluded on this basis.

Cryptic Relatedness

In the 1958BC it is assumed that the samples are unrelated and this was tested by estimating the degree of identity by state (IBS) and then inferring the identity by descent (IBD). The concept of IBS is the amount of allele commonality for a given SNP that occurs between individuals (Table 2.6).

Table 2.6 Scenarios of shared SNP genotypes between two individuals and their degree of IBS

Individual 1	Individual 2	IBS
AA	AA	2
BB	BB	2
AB	AB	2
AB	AA	1
AB	BB	1
AA	BB	0

When alleles are inherited from a common ancestor of the two individuals, it is termed IBD, which can only be inferred when it is assumed that the samples are unrelated. The proportion of relatedness (π) between subjects is estimated for each SNP given the frequency of the alleles to find the probability of sharing 0, 1 or 2 alleles IBD (Purcell *et al.* 2007). For instance, monozygotic twins would have a probability of sharing two alleles IBD near 1, and likewise if the sample has been contaminated during genotyping. No subjects were excluded due to relatedness in the T1DGC ($\hat{\pi} > 0.2$).

As part of estimating the relatedness of individuals in PLINK, genetically similar individuals are clustered into subgroups by the pairwise population concordance test. Pairwise population concordance tests that a pair of individuals in a random-mating population should follow a 2:1 ratio of IBS 2 SNPs (alleles are the same) to IBS 0 SNPs (alleles are different). A pair of individuals from different populations are expected to have more IBS 0 SNPs. The IBS distance between two individuals is the sum of the number of IBS 2 SNPs and half the number of IBS 1 SNPs given the

number of SNPs and is used as the basis for multidimensional scaling (MDS) to understand the population stratification of the study.

Heterozygosity

A measure of inbreeding within an individual is assessed by the proportion of observed heterozygous SNPs given the number of SNPs. In a non-related sample, the measure should be relatively constant over individuals, and furthermore, it can be used as indication of sample contamination. There was no evidence of inbreeding (or contamination) in the T1DGC as the subjects had a heterozygosity score between 0.29 and 0.34 (Figure 2.6).

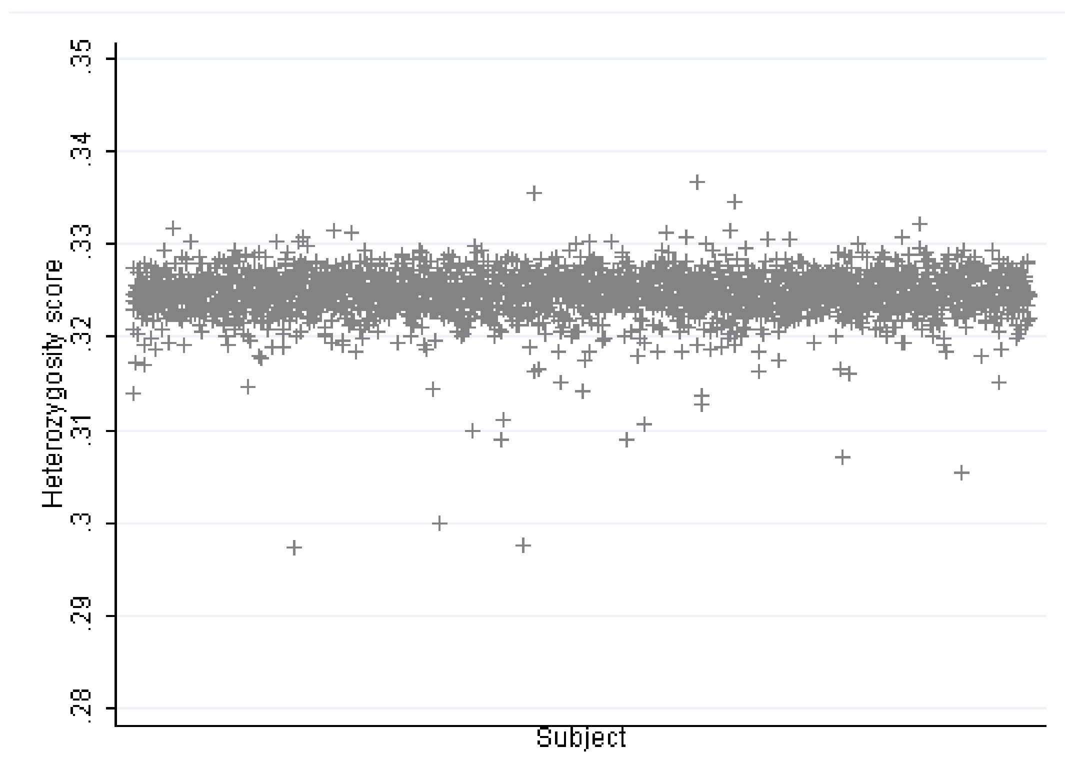


Figure 2.6 T1DGC The heterozygosity score for each sample/subject

Ethnicity/Population stratification

Underlying population stratification in genome-wide association study can cause false positive results and may distort findings (Tian *et al.* 2008). Population stratification can be assessed through MDS (Purcell *et al.* 2007), which uses the pair-wise IBS distances on subjects. The classical metric MDS implemented in PLINK, reduces the subject relatedness data to a configuration on a specified number of dimensions. Typically this representation will have 2 or 3 dimensions. Distances between individuals are defined in such a way that they match as close as possible to the observed similarity/dissimilarity. The distances are based on the Euclidean IBS distance between individuals. To limit the amount of information lost when reducing the data to a fixed number of dimensions, a goodness-of-fit measure is minimised between the distances and the observed similarity/dissimilarity (Everitt 1987). Classical MDS assumes a metric structure in the data, which is preserved to some extent in the lower-dimensional configuration.

For QC purposes, four HapMap sample populations (Han Chinese and Japanese, Yoruba Nigeria and Caucasian/European) were used as a reference to detect possible ethnic outliers in the T1DGC. The HapMap populations were merged into the study and the SNPs are thinned by removing those in high linkage disequilibrium to reduce the time taken and bias (Price et al. 2008) in estimating the pair-wise IBS distances. The MDS coordinates are estimated on the merged study. The first two coordinates in the T1DGC study were sufficient to detect the ethnic outliers from the four HapMap populations. The emphasis to detect the outliers was based on visual inspection (Figure 2.7). It was deemed that 54 subjects were of non-European

ancestry as they were outside of the main cluster on the two dimensional MDS representations.

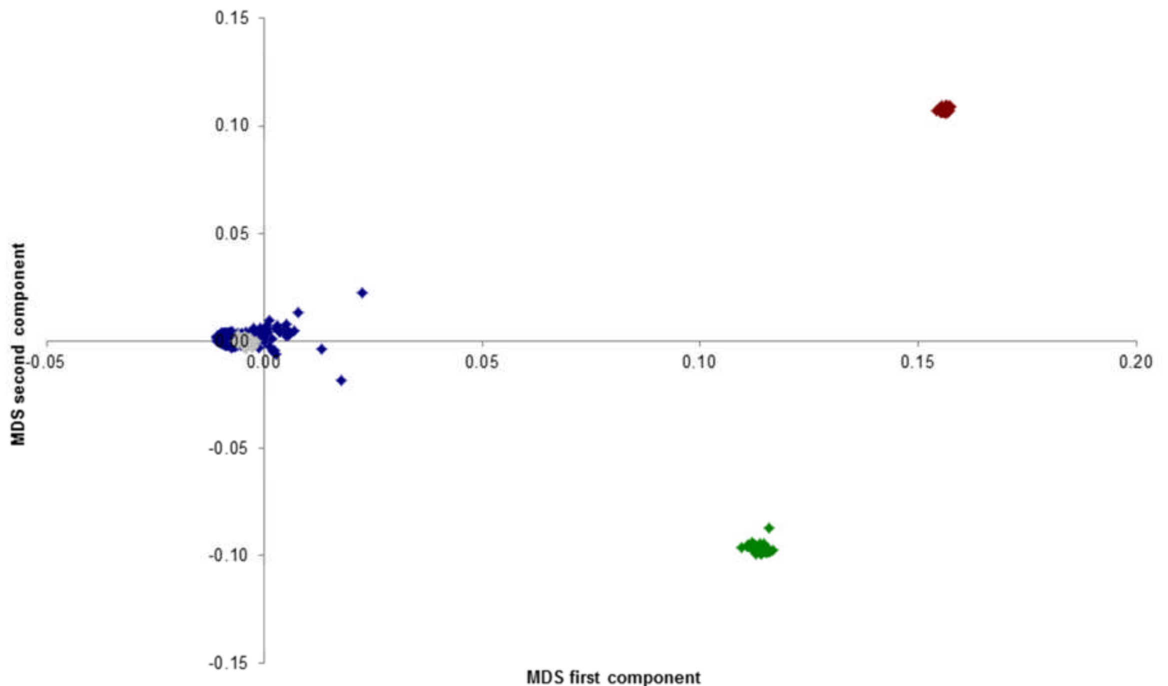


Figure 2.7 T1DGC ethnicity analysis. The T1DGC (grey) in comparison to HapMap US samples with ancestors from Northern and Western Europe (blue), HapMap Han Chinese and Japanese sample (green) and HapMap Yoruba Nigeria sample (red).

Imputation

Imputing SNPs that are not covered by the gene-chip can achieve greater coverage and finer detail of the genome (McCarthy *et al.* 2008). For the 1958BC genetic studies, imputation was done either as part of the WTCCC or in collaboration with the team in Oxford. The imputation software used for the 1958BC was `Impute` (Marchini *et al.* 2007), and uses the appropriate reference population data (for the 1958BC it is the Hapmap Northern and Western European (CEU) sample

population) that contains recombination rates for the genetic regions and the genotyped SNPs. The reference population data is in the form of haplotype data, and is defined as a set of haplotypes, $H = \{H_1, ..., H_N\}$ where N is the number of haplotypes. There are three scenarios for missing SNP data for subjects: 1) the SNP is not genotyped and information exists in the reference population; 2) it is genotyped and exists in the reference population; 3) it is genotyped and does not exist in the reference population. For the *i*-th individual, the genotype set for the *j*-th SNP is defined as $G_{ij} \in \{0, 1, 2, \text{ missing}\}$ and partitioned according to its missing status, so $G = \{G_0, G_M\}$. The probability of an individual's genotype vector $\Pr(G_i / H)$ is conditional on the reference haplotypes H and is estimated by a Hidden Markov Model, where the hidden states are sequence pairs of known haplotypes H. In theory, an imputed SNP uses information across the chromosome, however the further away a SNP is from the position of the reference haplotype, the less likely that information will be relevant. The marginal probability distribution for each individual's missing SNP is estimated, so $P(\Omega) = 1$ is satisfied, where the sample space Ω is the set of the three genotype combinations

3 The seasonal variation of inflammatory and haemostatic biomarkers due to the seasonal variation of 25-hydroxyvitamin D (Paper I)

3.1 Introduction

Evidence suggests that vitamin D has a role in cardiovascular health (Swales & Wang 2010; Wallis *et al.* 2008; Zittermann *et al.* 2005). Much of this evidence has come from observational studies reporting a protective 25-hydroxyvitamin D (25(OH)D – nutritional marker for vitamin D status) association with classical risk factors of cardiovascular disease (CVD), such as hypertension and type 2 diabetes (T2D), and with the risk of mortality from CVD as introduced in Chapter 1. However, this evidence is far from conclusive. At the initiation of the thesis it was uncertain whether vitamin D status was associated with the underlying progression of CVD, as indicated by biomarkers of inflammation and haemostasis in a middle-aged adult population free from clinical CVD.

In order, to explore the 25(OH)D associations with novel cardiovascular biomarkers of CVD (C-reactive protein (CRP), D-dimer, fibrinogen, tissue plasminogen activator (tPA) and von-Willebrand factor (vWF)) in participants free from CVD in the British 1958 birth cohort (1958BC), two analytical approaches were used. The first fitted multiple linear regression models incrementally controlled for lifestyle, socio-economic and adiposity variables. Adiposity is a strong confounder between 25(OH)D and the inflammation/haemostatic biomarkers, and as such, is adjusted for in the models. Missing data in the lifestyle factors were imputed in multiple datasets

by a multiple imputation by chained equations (MICE) method (Royston *et al.* 2009), and used in the multiple linear regressions models that accounted for uncertainty across the imputed datasets.

The second analytical approach uses the framework of mediation (Mackinnon *et al.* 2007) to explore the relationship between seasonal variations of 25(OH)D and of haemostatic/inflammatory biomarkers. The seasonal variations of the biomarkers were used as adiposity is expected to have no or limited seasonal variation, and it would be unexpected for adiposity to influence the seasonal variations of 25(OH)D and the biomarkers. To establish whether mediation could be applied in this situation, seasonal patterns of CRP, D-dimer, fibrinogen, tPA and vWF, and likewise, of 25(OH)D must be established. The seasonal patterns of the biomarkers were modelled through a linear transformation of a harmonic function and adjusted for seasonal lifestyle variables. In the models where the seasonal patterns of the biomarkers were explored, random effects were used for the laboratory assay batches to avoid variance inflation caused by multi-collinearity with the batches in the coefficients of the harmonic components. Mediation analysis was applied in a Frequentist setting, where it is well established, and in a Bayesian setting, where it is less so. Using the flexibility afforded by the latter in terms of approximating the sampling distributions of the derived quantities, the change of seasonal variation observed in the outcomes due to 25(OH)D was established via the posterior probability distributions.

The remainder of the chapter is divided into two parts. The first part describes the methods and data used (including the number of participants), an overview of

laboratory measures, and covariates (which were covered in greater depth in Chapter 2). The statistical methods mentioned in the previous two paragraphs are also described in detail. The second part presents and discusses the results of the analyses. A paper related to these findings has been published in a peer-reviewed journal (Hyppönen *et al.* 2010) and is shown in Appendix I.

The work presented in this chapter has extended the work presented in the paper. The mediation analysis was applied in the Bayesian setting to validate the findings from the Frequentist setting. Furthermore, in the Bayesian setting the magnitude of the change in seasonal variation of the biomarkers due to 25(OH)D was quantified.

3.2 Data

3.2.1 Participants

As outlined in Chapter 2, data from participants of the 1958BC biomedical survey were analysed in this work. Of those participants who responded to the biomedical survey (*n*=9377), 80.9% had measures of 25(OH)D concentrations (*n*=7591). Most cohort participants are of European ancestry (98%) (Atherton *et al.* 2008) and the 154 participants who reported belonging to any other ethnic groups were excluded. One participant was pregnant at the time of the survey and was also excluded from the analysis. Since the focus of this work was on underlying sub-clinical CVD, participants who reported use of any type of medication for cardiovascular health were excluded (*n*=532, British National Formulary code 2: Cardiovascular medication). Finally, for participants who had missing data on any of the

inflammatory/haemostatic biomarkers (*n*=366) were excluded, leaving in total *n*=6538 subjects in the analyses.

3.2.2 Laboratory analyses

The laboratory assays used for 25(OH)D, CRP, D-dimer, fibrinogen, tPA, and vWF are described in Chapter 2. The study time period comprised 18 months, with nurses visiting survey participants throughout the time period to take and collect, among other measures, serum blood samples. The nurses sent the samples to the participating laboratories as and when collected. This meant that the serum markers were assayed over several laboratory batches. For the inflammatory/haemostatic biomarkers there was a potential assay batch effect that may have influence the seasonal patterns of the biomarkers. A description of the mixed effects model used to account for the between assay batch variation is provided in the statistical methods.

3.2.3 Lifestyle, socio-economic and adiposity factors

A fuller description of the lifestyle, socio-economic and adiposity variables appears in Chapter 2, along with their categorization where applicable. The self-reported lifestyle and socio-economic variables included in this work were: recreation metabolic equivalent of task (MET) hours per week, engaging in vigorous physical activity, time spent watching TV (television) or using a PC (personal computer) in leisure time, alcohol consumption, chest infections reported in the last three weeks, smoking history, respiratory infections, socio-economic position at birth and

adulthood are used in the analysis. Body mass index (BMI) and waist circumference are used as measures for adiposity.

3.3 Statistical Methods

3.3.1 Descriptive methods

The natural logarithmic transformation was used to calculate geometric means, as the distribution of 25(OH)D concentration was slightly skewed (Figure 3.1). To describe the extremes of this distribution, dichotomous indicators were created for concentrations below 25 nmol/L and above 125 nmol/ L. The inflammatory and haemostatic biomarkers' values were transformed to within-sample sex-specific standard deviation scores (SDS) to compare variation across models. Initial analyses included validation and graphical examination of data, statistical evaluation of linear and quadratic terms for 25(OH)D, adiposity measures (BMI and waist circumference), and single and joint effects of these measures on the inflammatory/haemostatic outcomes. Interactions between continuous variables of adiposity measures and 25(OH)D concentrations were tested with each of the outcomes. The 25(OH)D geometric means were standardized by gender and season categorised as: spring (March-May); summer (June-August); autumn (September-November); winter (December-February). The adiposity measures were categorised to illustrate their relationship with mean concentrations of 25(OH)D and the outcomes. The BMI variable was categorised as below 20 kg/m^2, then by 2.5 kg/m^2 divisions, and above 35 kg/m^2, and the waist circumference variable was categorised into below < 75 cm, then by 5cm divisions and above 100cm.

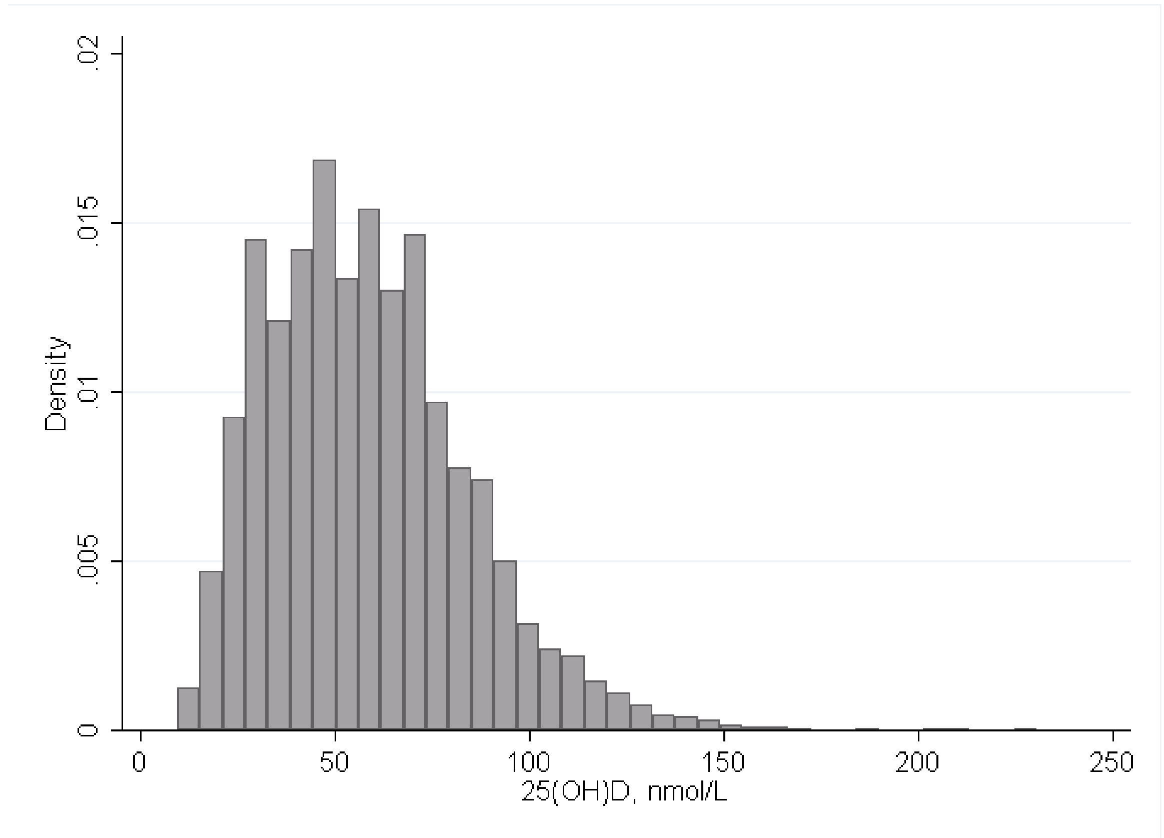


Figure 3.1 True histogram of 25(OH)D concentrations.

3.3.2 Multiple linear regression models

The first analytical approach used was multiple linear regression models to explore the role of 25(OH)D as exposure with the inflammation and haemostatic biomarkers as outcomes. A multiple linear regression model has the form:

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_{p-1} x_{ip-1} + \varepsilon_i \qquad (1)$$

where $i = 1,\ldots,n$ denotes the individuals in the model, $\varepsilon_i \sim N\left(0,\sigma^2\right)$, and $\beta_0,\ldots,\beta_{p-1}$ are the unknown parameters corresponding to the independent variables $x_1, x_2, \ldots, x_{p-1}$.

Three multiple linear regression models were fitted sequentially for each of the SDS haemostatic/inflammatory outcomes. In these models, the exposure of 25(OH)D was a categorized in 10 nmol/L divisions, with tails below 25 nmol/L and above 125 nmol/L, making 12 categories in total. The models included incremental adjustments for the variables of lifestyle, socio-economic and adiposity factors. Model one was adjusted for sex and the month of blood sample as a categorical variable. Model two, in addition, adjusted for recreation MET hours, vigorous physical activity, time spent watching TV or using PC, alcohol consumption, smoking history, and socio-economic position at childhood and adulthood. Finally, model three adjusted for all previous variables and adiposity measures. In model three, an additional 25(OH)D categorical variable (defined as < 25, 25-74.9 and ≥ 75 nmol/L) was used to summarize and aggregate the effect size.

Missing data

Some participants had missing data in the questions used to derived the variables for recreation MET hours, vigorous physical activity, time spent watching TV or using PC, alcohol consumption, smoking history and waist circumference (Table 3.1). The proportion of missing entries on one of more variables was 7.9% (*n*=514) of the study's eligible participants. Missing data for the multiple linear regression models were imputed using multiple imputation (MI) by chained equations implemented in Stata, version 10 (Royston *et al.* 2009; StataCorp 2010). MI is not recommended when 50% or more of the data is missing (Royston 2004), however since the proportion of missing data was much less in this work, it was not considered to be an issue.

Table 3.1 Frequencies of missing information for lifestyle and adiposity variables

	Lifestyle and adiposity covariates					
	Recreation MET hours	Vigorous physical activity	TV watching/ use of PC	Alcohol consumption	Smoking history	Waist circum-ference
Number missing (%)	96 (1.5)	96 (1.5)	282 (4.3)	18 (0.3)	222 (3.4)	12 (0.2)

MI relies on the assumption that missing values are missing at random (MAR), i.e. that conditional on the observed data, the missing data mechanism is independent of the unobserved data (Kenward & Carpenter 2007). Alternatives to the MAR assumption are missing completely at random (MCAR) or missing not at random (MNAR). MCAR assumes that the missing mechanism is independent of the observed or unobserved data, meaning that analysis on the complete data will be consistent with analysis on multiple imputed datasets (although there would be some loss of information in the complete data). When the missing data is MNAR, the missing mechanism depends on the unobserved data and the mechanism is largely unknown. It is impossible to use the observed data alone to distinguish between MAR and MNAR, but sensitivity analysis can be used to determine whether MI introduced any bias in the model's estimates.

In MI, the imputation model is different from the model used for the statistical analysis on the participants (the multiple linear regression models described above) (Kenward & Carpenter 2007). The imputation model should have at least the outcome and the covariates used in the multiple linear regression model, and MI has the advantage that additional variables can be considered in the imputation model

(Allison 2002). The additional variables assessed for use in the imputation model were time spent outside, vitamin D supplements, geographical region of residence and marital status. These factors were selected due to the results from two studies on the 1958BC participants. The first study reported that non-response in the biomedical survey of the 1958BC showed a significant bias to an earlier covariate not too dissimilar from marital status (Atherton *et al.* 2008). The second (Hyppönen & Power 2007) concluded that 25(OH)D was significantly related to time spent outside, vitamin D supplements and geographical region. A missing indicator for each of the covariates adjusted for in the multiple linear regression models was used as an outcome in a logistic regression model. Backwards stepwise variable selection with an exclusion threshold of p-value>0.05 was used in the logistic regression model with the missing indicator, to justify the inclusion of any of the additional variables for the imputation models (Table 3.2). To test the sensitivity of the MI procedure, the procedure was re-run without the addition of the four variables from the imputation models. In total there were two sets of imputed datasets created; the first set came from imputation models that adjusted for the additional variables of time spent outside, vitamin D supplements, geographical region of residence, marital status as well as the covariates used in multiple linear regression models that were missing data, and the second set came from imputation models that only adjusted for the covariates missing data.

Table 3.2 Additional variables included in the imputation models. For imputation of the lifestyle and adiposity covariates used in the multiple linear regression models (✓ in the box).

Missing data covariates	Additional variables for the imputation models*			
	Time spent outside	Vitamin D supplements	Geographical region of residence	Marital status
Waist Circumference		✓		
Recreation MET hours		✓		✓
Vigorous activity		✓		✓
TV watching/using of PC	✓	✓		✓
Smoking	✓			
Alcohol consumption			✓	✓

*The imputation models also adjusted for; sex, month of blood sample, recreation MET hours, vigorous physical activity, time spent watching TV or using PC, alcohol consumption, smoking history, and socio-economic position at birth and adulthood, BMI and waist circumference

Imputation was run over 100 cycles and trace plots of the imputed covariates were checked for convergence. The trace plots showed the mean value of each imputed variable over the cycle and highlighted any extreme values or trends. Ten datasets were imputed in each set, although five datasets are considered a "conservative choice" for situations where up to 20% of data is missing (Royston 2004). The three multiple linear regression models for each outcome were run on the complete dataset (restricted to participants with no missing data) and on the two sets of 10 imputed

datasets, generated from imputation models with and without the additional variables. The regression coefficients of 25(OH)D were identical to two decimal places from the results between the two sets of imputed datasets. Results similar in direction and approximate size were found between the imputed dataset and the complete dataset. The final results are presented from the imputation model without the additional variables.

3.3.3 Mediation Analysis

The framework of mediation analysis was applied to assess whether the seasonal variations (if they exist) of the haemostatic/inflammation biomarkers (CRP, D-dimer, fibrinogen, tPA, and vWF) are mediated by the seasonal variations of vitamin D intake/status, thereby inferring a 25(OH)D association with the biomarkers. Mediation analysis conceptualises how an independent variable is related to an outcome, and whether this relationship is partially, fully, or not at all mediated through an intermediate factor (termed the mediator) (Mackinnon *et al.* 2007). In this case, the independent variable is the distinct seasonal patterns of the biomarkers and their mediator is current vitamin D status (25(OH)D). The pathway of the seasonal variation of the biomarker mediated by 25(OH)D may avoid the effect of adiposity, as adiposity is expected to have little or no influence on external environmental factors that cause seasonal variation, and little or no seasonal variation itself.

The definition of a single mediator model implies a pathway between an independent variable and outcome, and an additional pathway from the independent variable

through the mediator, to the outcome (Figure 3.2); this is defined in the three model equations (Mackinnon *et al.* 2007) which can be written as:

$$Y = \beta_{0(1)} + \tau_X X + \varepsilon_{(1)} \tag{2}$$

$$Y = \beta_{0(2)} + \tau'_X X + \beta_M M + \varepsilon_{(2)} \tag{3}$$

$$M = \beta_{0(3)} + \alpha_X X + \varepsilon_{(3)}, \tag{4}$$

where Y is the outcome, X the independent variable and M the mediator.

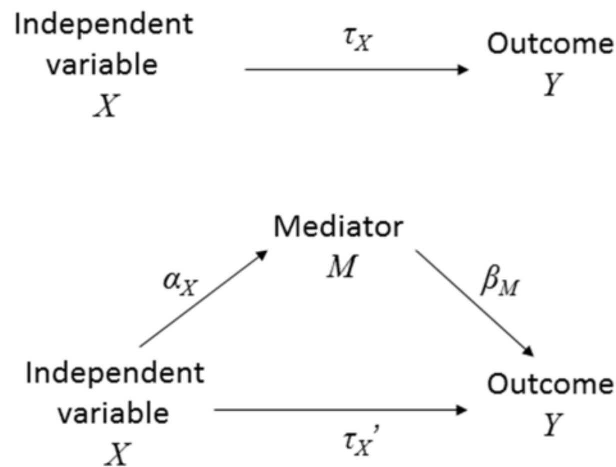


Figure 3.2 Pathway model for a single mediator. The independent variable is assumed to be associated with the outcome, and the mediator is within the sequence between the two. Adapted from (Mackinnon *et al.* 2002).

At least two conditions must be met in order to apply mediation analysis (Mackinnon *et al.* 2007). The independent variable X must be associated with the outcome Y, i.e. $H_0 : \tau_X = 0$ (as modelled in Equation (2)) Likewise, the independent variable X must have an association with the mediator M, i.e. $H_0 : \alpha_X = 0$ (as modelled in Equation (4)). In the unusual situation of complete

mediation, where the association of the independent variable *X* with the outcome *Y* is entirely through the mechanism of the mediator *M,* the coefficient of τ'_X will be zero as the adjustment for the mediator will fully account for the association of the independent variable (Mackinnon *et al.* 2002). In the situation of no mediation, the independent variable *X*'s association with the outcome *Y*, the coefficient of τ'_X, will be unaffected by the adjustment for the mediator i.e. $H_0 : \tau_X = \tau'_X$. In the third situation of partial mediation, it is argued that the inequality $|\tau'_X| < |\tau_X|$ should hold for the mediator's effect to be valid and a test for direct mediation is thus $H_0 : \tau_X - \tau'_X = 0$, i.e. the difference in the independent variable's association with the outcome, unadjusted and adjusted for the mediator (Mackinnon *et al.* 2002). However, as the independent variable is related to the mediator they will tend to be correlated and multi-collinearity may occur when adjusting for both in the same model (Baron & Kenny 1986). Multi-collinearity can reduce the power for testing the direct mediation by inflating the effect size of τ'_X. It is plausible in this situation to have a valid mediator and $|\tau'_X| > |\tau_X|$, making it conceptually difficult to test for direct mediation.

An alternative test to $\tau_X - \tau'_X$ is an indirect test on the product of the independent variable's association with mediator (as modelled in Equation (4)), and the mediator's association with the outcome adjusted for the independent variable (as modelled in Equation (3)) (Mackinnon *et al.* 2002). The relationship between the independent variable and mediator, denoted by α_X, must exist $H_0 : \alpha_X = 0$ otherwise the conditions of mediation do not hold (Mackinnon *et al.* 2007). The

second is the strength of the mediator's association with the outcome, denoted by β_M from the model that accounts for the independent variable. In the situation of no mediation, β_M will not be statistically significant i.e. alternative hypothesis will hold $H_A : \beta_M = 0$. Therefore, the statistical significance of the indirect mediation test based on the statistic $\hat{\beta}_M \hat{\alpha}_X$ is largely determined by the statistical significance of the mediator-outcome association β_M. The product of coefficients test, $H_0 : \beta_M \alpha_X = 0$ is algebraically equivalent to the direct test under maximum likelihood and linear regression model conditions (Mackinnon *et al.* 1995), but is without the inequality restriction of $|\tau'_X| < |\tau_X|$ (Mackinnon *et al.* 2002).

Harmonic function

The seasonal patterns of the haemostatic/inflammatory outcomes and 25(OH)D concentrations in the 1958BC were modelled using the harmonic function. The function was used to determine whether the outcomes had a significant seasonal patterns (represented by the τ_X association from Equation (2)) and the validity of 25(OH)D as a mediator (represented by the α_X association from Equation (4)), and so whether the mediation test could be applied. The harmonic function was the most attractive choice of method with the data in the 1958BC for the following reasons (Refinetti *et al.* 2007).

1) The data were stationary; the sample time frame of the biomedical survey was 18 months from Sept 2002 to March 2004, which gave the opportunity for investigating a seasonal pattern of the biomarkers. However, there was limited information for investigating a seasonal trend or a change in pattern over time.

2) Simple seasonal patterns of the biomarkers are expected. The frequency proposed for the patterns were yearly, half yearly (semestral) and quarterly, and a combination of the frequencies within a time length of a year. Semestral and quarterly components are unusual in a biological context, but they may reduce the residual variance in the model.
3) The observations were fairly uniformly distributed over the sampling time period and did not occur in response to an event or intervention.

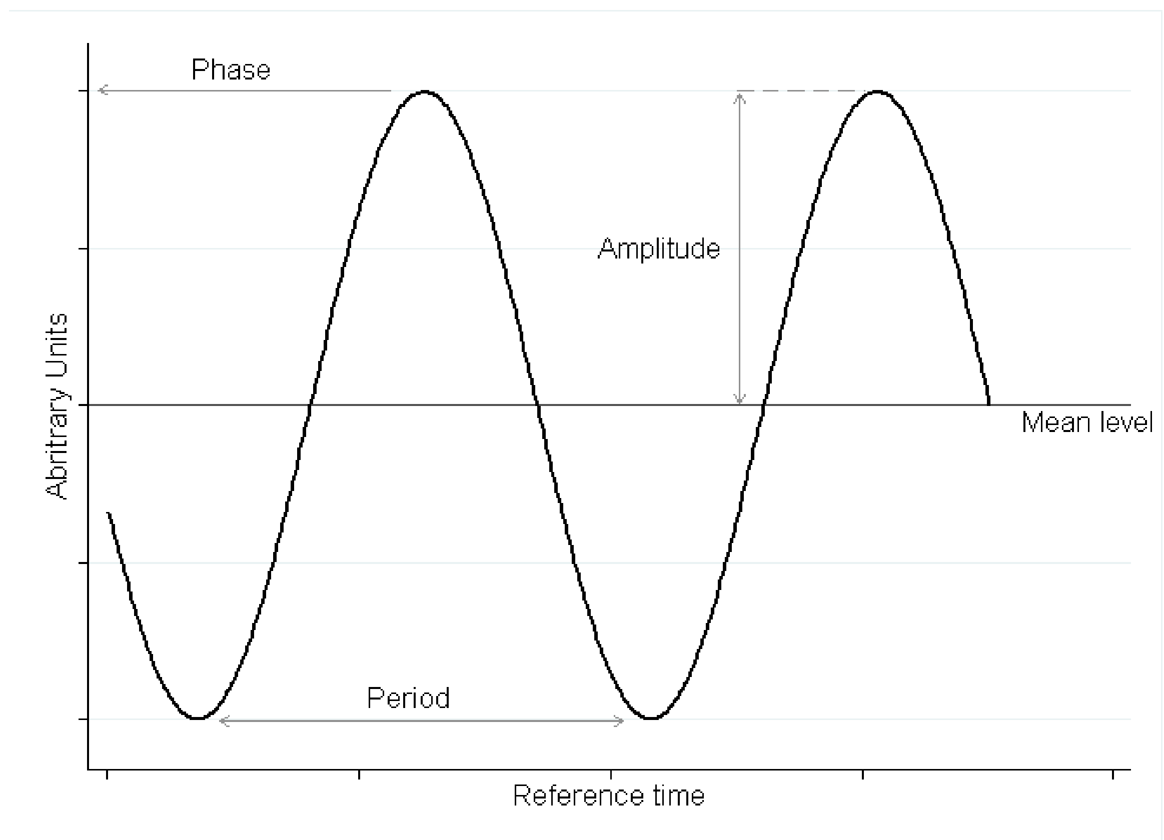


Figure 3.3 The mean, period, amplitude and phase of a rhythmic pattern

A rhythmic pattern comprises of mean level, period (or frequency), amplitude and phase (Figure 3.3) (Halberg *et al.* 1967). The mean level L is the overall mean central to the variation of the pattern. The period (or frequency), ω is the length of

the full cycle, i.e. the time from one maximum to the next. The amplitude A (or magnitude, when there is more than one frequency in the pattern), is the variation from the mean level to the extrema. Finally, the phase ϕ determines the time to the maximum in the cycle measured in radians (Bingham *et al.* 1982). A singular frequency harmonic function, *h,* in a cross-sectional setting, given its properties of *L, A*, ω and ϕ, and where the measurements are not necessarily equally spaced can be expressed in the following form (Tong 1976):

$$\begin{aligned} h(t_i) &= L + A\cos(\omega\, t_i - \phi) \\ &= L + \gamma_1 \sin(\omega\, t_i) + \gamma_2 \cos(\omega\, t_i) \end{aligned} \qquad (5)$$

where t_i is a time point for the individual *i* within a length of time *T*. The parameters γ_1, γ_2 are estimated from linear regression and can be used to estimate the amplitude $A = \sqrt{\gamma_1^2 + \gamma_2^2}$ and phase $\phi = B + C\tan^{-1}(\gamma_1/\gamma_2)$ †. The frequency is defined as $\omega = 2\pi p/T$, where *p* is the number of times the cycle repeats itself within the time *T*.

The rhythmic pattern can be extended to account for different, simultaneous frequencies within time *T*. The length of time considered for the haemostatic/inflammation biomarkers and 25(OH)D was a year (365 days) with frequencies of yearly, semestral and quarterly. The harmonic function in a linear model can be defined as:

$$h(t) = \sum_{k=1}^{m_c} \left[\gamma_{1k} \sin\left(\frac{2\pi p_k t}{T}\right) + \gamma_{2k} \cos\left(\frac{2\pi p_k t}{T}\right) \right] \qquad (6)$$

† if $\gamma_1 > 0,\ \gamma_2 \geq 0$ then $B = 0$ and C=-1, if $\gamma_1 \geq 0,\ \gamma_2 < 0$ then $B = -\pi$ and C=1
if $\gamma_1 < 0,\ \gamma_2 \leq 0$ then $B = -\pi$ and C=-1, if $\gamma_1 \leq 0,\ \gamma_2 > 0$ then $B = -2\pi$ and C=1

where m_c is the number of frequency components. As the frequency $p_{k,}$ is number of times the cycle repeats itself, for yearly, semestral and quarterly frequency it is $p_k = (1, 2, 4)$, respectively, Further, $1 \leq t \leq T$; and γ_{1k}, γ_{2k} are the regression parameters from the linear model. The criterion used to include a specific frequency term in the harmonic function differed between the Frequentist and Bayesian settings.

Seasonal confounders

To account for a lifestyle that may potentially vary or have an influence on the seasonality of the haemostatic/inflammatory markers, the covariates adjusted for in the harmonic function models were: TV watching/PC usage in leisure time; recreation MET hours; self-reported chest infection within three weeks prior to survey; and social class at adulthood. Missing data were not imputed for the models in the mediation analyses, due to the additional complexity of estimating the standard error of the harmonic function from a multiple imputed datasets.

Harmonic mixed effects model

The logistics of a large survey meant that the venous blood samples were taken and sent to the laboratories throughout the survey period. The laboratories assayed the measures in batches as the samples were received. Therefore, measures of the inflammation and haemostatic biomarkers within laboratory batches may be correlated with the time of measure due to when it was assayed. In the initial seasonal variation models for each of the outcomes (CRP, D-dimer, fibrinogen, tPA, vWF), their laboratory batch was adjusted for as a fixed factor in the linear

regression models. The predicted means from the linear models adjusting for batch with fibrinogen and vWF had marked inflation compared with the sex-standardized mean concentrations by month, with some of the predicted means being several orders of magnitude larger than the standardized means. However, in this context the laboratory batch is not of intrinsic interest though it needs to be controlled for, and as such, the laboratory batch was considered as a random effect (Vonesh & Chinchilli 1997).

Fitting mixed effects models and treating laboratory batch as a random effect for fibrinogen, vWF, tPA and D-dimer, the between-batch variation for the random effect was significant (p-value <0.05), which met the criteria for including the laboratory batch as a random effect. Furthermore, using the random effect term eliminated multi-collinearity between laboratory batch and the harmonic coefficients. The mean variance inflation factor (VIF) was > 3.4 in the linear models with laboratory batch and without it, VIF=1. The between-batch variation for CRP was not significant in the mixed effects model, and for CRP a linear regression model was used without adjusting for assay batch. The regression model with the cardiovascular health markers including the harmonic function was of the form:

$$Y_{ij} = \left(\beta_0 + \mu_{0j}\right) + \sum_{k=1}^{m_c}\left[\gamma_{1k}\sin\left(\frac{2\pi\, p_k\, t}{T}\right) + \gamma_{2k}\cos\left(\frac{2\pi\, p_k\, t}{T}\right)\right] + \mathbf{X}_{i3}\,\boldsymbol{\beta}_3 + \varepsilon_{ij} \qquad (7)$$

where i is the observation within laboratory batch j, T, m_c and p_k are as in Equation (6) , $\mathbf{X}_{i3}$ is a design matrix for the vector of parameters $\boldsymbol{\beta}_3$ (inclusive of the mediator variable and seasonal confounders), and $\varepsilon_{ij} \sim N\left(0,\sigma_\varepsilon^2\right)$; the random effect term is assumed $\mu_{0j} \sim N\left(0,\sigma_\mu^2\right)$.

In summary, the steps taken to establish whether mediation analysis can be used for the seasonal variations of 25(OH)D and haemostatic/inflammation biomarkers are outlined in Figure 3.4. Different criteria were applied to determine whether there was seasonal variation in 25(OH)D and the biomarkers, depending on whether the inference was performed in the Frequentist or Bayesian setting (as expanded on in the following sub-sections).

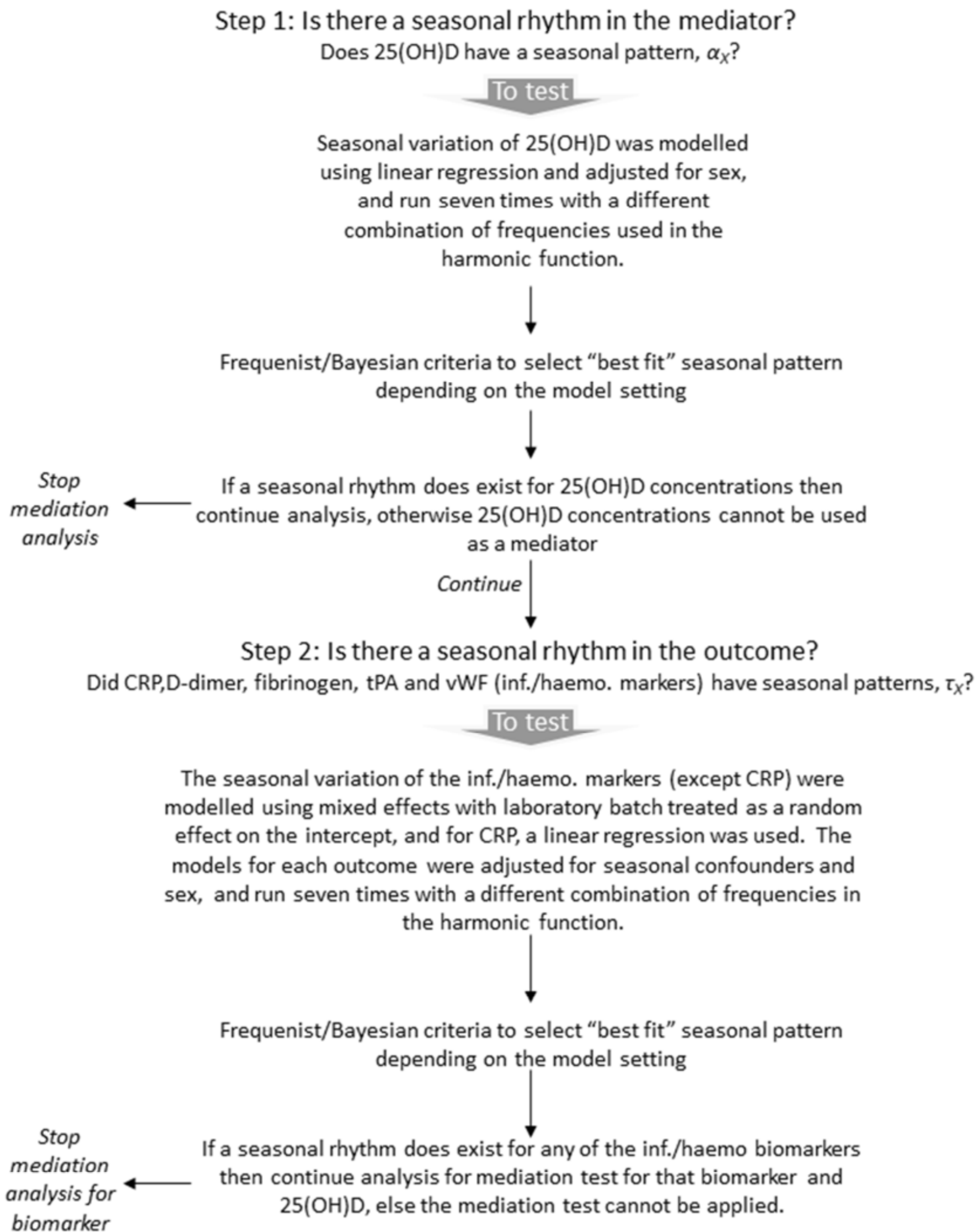


Figure 3.4. Establishing whether mediation analysis can be applied for the seasonal variations in 25(OH)D and in the haemostatic/inflammation biomarkers

3.3.4 The mediation product of coefficients test incorporating seasonal patterns

The seasonal variation of the mediator (as represented by the coefficient α_X from Equation (4)) had corresponding sine and cosine regression coefficients for each frequency in the harmonic function (Equation (6)). The Frequentist and Bayesian inference approaches were used to complement each other with the adapted mediation test for the seasonal patterns. Each setting used the same models to apply the mediation test of the product of coefficients.

In the Frequentist setting, the true parameter value (often denoted by ϑ) is fixed but unknown (Casella & Berger 1990). Inference is defined by the characteristics of the decision function used for estimating or testing the parameter ϑ . In this work, the maximum likelihood (ML) method was used to estimate the parameter ϑ and its confidence intervals (CI). ML procedures focus on the probability of the data actually observed, so the likelihood function $L(x;\vartheta)$ is conditional on values of the parameter ϑ and the observed data x derived from the sampling distribution.

In contrast, the Bayesian setting also incorporates the likelihood, but considers the parameter ϑ to be random variable with a probability distribution reflecting belief about possible values (Casella & Berger 1990). Often, there is prior knowledge about the parameter even before the data are observed, and this is modelled by a probability distribution on its possible values, denoted by $\pi(\vartheta)$. Using Bayes's Theorem, the prior distribution $\pi(\vartheta)$ is updated with the data observed, as defined by the likelihood function $L(x;\vartheta)$, to give a posterior distribution $\pi(\vartheta|x)$ i.e.

$\pi(\vartheta \mid x) \propto L(x:\vartheta)\pi(\vartheta)$. In this chapter, non-informative prior distributions were used as there was no prior knowledge on the parameters.

The product of coefficients mediation test in the Bayesian approach makes use of the property that a function of the parameter posterior samples is the posterior sample of the function of the parameter's (Yuan & Mackinnon 2009). Furthermore, due to the flexibility of the Bayesian setting, a quantitative measure of the change of seasonal variation of the haemostatic/inflammation outcomes due to 25(OH)D was derived.

Frequentist inference on the mediation product of coefficients test

The seasonal pattern (if one existed at all) providing the best fit needed to be established with CRP, D-dimer, fibrinogen, tPA and vWF as defined in equation (7), before mediation analysis could be applied. The model selection strategy was based on the Schwarz's Bayesian Information Criterion (BIC), and for nested models the likelihood ratio test (LRT) of the seasonal patterns was used. The BIC between models 1 and 2 is defined as:

$$\text{BIC} = 2\left[l\left(\vartheta_2\right) - l\left(\vartheta_1\right)\right] - \log n\left(p_2 - p_1\right)$$

where $l\left(\vartheta_2\right) - l\left(\vartheta_1\right)$ is difference between the log-likelihood functions, $p_2 - p_1$ is the difference between the number of parameters for n number of observations (Kuha 2004). For an outcome, if in two non-nested models the p-values of the harmonic function's LRT were both <0.05 then the model with the smallest BIC took precedence in an effort to penalise the most complicated model (Kuha 2004). Once a final seasonal pattern was determined for the outcome, the mediation test was applied.

The standard error for the mediation test statistic $\hat{\beta}_M \hat{\alpha}_X$ is typically obtained by the first order approximation of the multivariate delta method, $\sigma_{\alpha\beta} = \sqrt{\alpha^2 \sigma_\beta^2 + \beta^2 \sigma_\alpha^2}$ (Mackinnon *et al.* 2007). This standard error gives an adequate trade-off for Type 1 error and power across a range of coefficient sizes, compared with other estimators used for the product, such as the second order approximation (Mackinnon *et al.* 2002). The sampling distribution of the product $\hat{\beta}_M \hat{\alpha}_X$ can be highly skewed and is not normally distributed with respect to the distribution's kurtosis. The distribution of the product of two Normal densities is found by noting that under $H_0 : \beta_M \alpha_X = 0$ and using the fact that asymptotically, maximum likelihood estimates are normally distributed. The standardised quantities $z_\alpha = \alpha_X / \sigma_\alpha$ and $z_\beta = \beta_M / \sigma_\beta$ define the distribution of $Z_{\alpha\beta} = z_\alpha z_\beta$ so that an equivalent null hypothesis is $H_0 : Z_{\alpha\beta} = 0$ (Craig 1936). The R function `PRODCLIN` was used to derive the 95% CI and significance test for the product of coefficients (Mackinnon & Fritz 2007; R Development Core Team 2008).

The seasonal pattern of the mediator as represented by α_X modelled by the harmonic function in the linear form had two parameters for each frequency present. In order to have a single point measure of the fluctuation in the seasonal pattern to use in the product of coefficients test, the amplitude (when one frequency was present) or magnitude (when more than one frequency was present) was considered. However, using the magnitude meant that an estimator for its variance needed to be derived. As stated earlier under the harmonic function, the amplitude of a single

frequency (equation (5)) is defined as $A = \sqrt{\gamma_1^2 + \gamma_2^2}$ and the variance of its ML estimator is $\mathrm{var}(\hat{A}) = \sigma_A^2 = \frac{1}{A^2}\left(\sigma_{\gamma_1}^2 \gamma_1^2 + \sigma_{\gamma_2}^2 \gamma_2^2 + 2\sigma_{\gamma_1\gamma_2} \gamma_1 \gamma_2\right)$, where $\sigma_{\gamma_j}^2$ is the variance of γ_j, $\sigma_{\gamma_1\gamma_2}$ is the covariance between $\hat{\gamma}_1$ and $\hat{\gamma}_2$ (Tong 1976). The estimated amplitude $\hat{A}$ asymptotically follows a normal distribution, as it is a function of ML estimates (Rice 1995), since $\sqrt{n}(\hat{A} - A) \sim N(0, \sigma_A^2)$ as $n \to \infty$ (Tong 1976).

The magnitude in the instance of multiple frequencies, as in Equation (6) $m_c > 1$, depends on when the extrema of the harmonic function *h(t)* occur. The time of which the maxima/minima occur can be predicted from the model over the time period $t = \{1, \cdots, T\}$ or analytically. The size of the amplitude/magnitude from harmonic function is half the difference of the levels reached at the extrema time points. The analytical solution was derived using symbolic algebra in `Mathematica v6` (Wolfram Research 2007) to confirm the solution predicted from the harmonic function with yearly, semestral and quarterly frequencies over $T = 365$ days.

For estimating the variance of the magnitude, the contribution of each frequency to the harmonic function was derived. The cosine function in the first line of Equation (5) oscillates between [-1,1], and each frequency's amplitude contributes to the magnitude up to a maximum absolute value of 1. The fraction f_k that the kth frequency's amplitude contributes to the magnitude can be obtained as the difference

of each frequency's cosine function fixed on the magnitude's extrema time points divided by two. The frequencies are incorporated to give the magnitude, A_C :

$$\begin{aligned} A_C &= \frac{1}{2}\left[\sum_{k=1}^{m_c} A_k \cos\left(\frac{2\pi\, p_k\, t_{\max}}{T} - \phi_k\right) - \sum_{k=1}^{m_c} A_k \cos\left(\frac{2\pi\, p_k\, t_{\min}}{T} - \phi_k\right)\right] \\ &= \frac{1}{2}\left[\sum_{k=1}^{m_c} A_k\, f_{k,t_{\max}} - \sum_{k=1}^{m_c} A_k\, f_{k,t_{\min}}\right] \\ &= \sum_{k=1}^{m_c} A_k \frac{1}{2}\left(f_{k,t_{\max}} - f_{k,t_{\min}}\right) \\ &= \sum_{k=1}^{m_c} A_k\, f_k \end{aligned} \tag{8}$$

where $-1 \le f_k \le 1$ and $t_{\max}, t_{\min}$ are the times in $\{1,..T\}$ where the extrema occur.

The variance of $\hat{A}_C$ is then obtained using the first order approximation from the delta method (Rice 1995), since the frequencies' amplitudes are orthogonal. The magnitude can be thought of as approximately a sum of independent amplitudes. The maximum likelihood estimates of the amplitudes are asymptotically normally distributed, and as such, the sum is also asymptotically normally distributed, with variance;

$$\text{var}(\hat{A}_C) = \sum_{k=1}^{m_C} \frac{1}{A_k^2}\left(\sigma_{\gamma_{1k}}^2 \gamma_{1k}^2 + \sigma_{\gamma_{2k}}^2 \gamma_{2k}^2 + 2\sigma_{\gamma_{1k}\gamma_{2k}} \gamma_{1k}\gamma_{2k}\right) f_k^2 \tag{9}$$

where $\sigma_{\gamma_{jk}}^2$ is the variance of $\hat{\gamma}_{jk}$, and $\sigma_{\gamma_{1k}\gamma_{2k}}$ is the covariance between $\hat{\gamma}_{1k}$ and $\hat{\gamma}_{2k}$. The maximum likelihood estimates of A_C and var *(* $\hat{A}_C$ *)* represent α_x and σ_x in the mediation test of $H_0 : \beta_M \alpha_X = 0$.

Parametric Bootstrap

As alternative to using $\mathrm{var}(\hat{A}_C)$ in the Frequentist setting, a parametric bootstrap approach was taken to derive the magnitude and confidence intervals of the mediation test from the estimated coefficients (Pyrhonen et al. 2012). The harmonic terms of the mediator β_M and association of the mediator α_X with the outcomes were calculated by maximum likelihood estimation. In the parametric bootstrap, the point estimate and confidence interval for the magnitude was derived by sampling from a multivariate normal distribution 10,000 times with mean and covariance as taken from the estimated model for the seasonal pattern of the mediator. For each simulated sample, the magnitude was half the difference of the minimum and maximum achieved in the year period. The point estimate of the magnitude was the 50% centile of the simulated samples' magnitudes and the 5th and 95th centiles were the 95% confidence interval.

For the mediation product of coefficient test, in each dataset the association of the mediator α_X with the outcome was sampled from the normal distribution. The simulated product of $\beta_M \alpha_X$ centiles (50th, 5th and 95th) was used as the point estimate and 95% confidence interval. Evidence for the null hypothesis of the mediation test was concluded when the 95% confidence intervals of all of the test parameters for a given outcome were overlapping with zero.

Bayesian inference on the mediation product of coefficients test

Mediation analysis was also performed in a Bayesian setting as it is possible to create a parameter from a function (say that from the product of coefficient test) without formally defining the parameter's posterior distribution. The models fitted

were the same as those used in the Frequentist setting, assuming non-informative normal distribution priors ($N\left(0,\sigma_{\beta_i}^2\right)$ with the precision defined as $1/\sigma_{\beta_i}^2 = 10^{-5}$) for the covariates. The laboratory batches were fitted as random effects on the intercept, and assumed to be $N\left(0,\sigma_{\rho}^2\right)$ with the precision $1/\sigma_{\rho}^2$ following a standard non-informative gamma distribution $\Gamma(0.001, 0.001)$ prior (Goldstein 2010).

The Bayesian mediation analysis in this chapter was run using the software of WinBUGS (version 1.4.3). WinBUGS uses Markov chain Monte Carlo (MCMC) techniques as implemented through Gibbs sampling (Lunn *et al.* 2000). Gibbs sampling exploits the conditional independence assumptions of the random variables, so a posterior distribution of a random variable may not exist in analytic form but it can still be simulated. The chain algorithm iteratively samples in a top-down approach the probability distributions as defined in the model script, which are dependent on previous values. Convergence checks of the random values are required to ensure that the simulated posterior distributions are stable.

For each of the mediation models implemented in WinBUGS, the following checks were done (Spiegelhalter & Lunn 2009);

- For each model, two independent chains were initiated from different starting points. Sample monitors were set up on the parameters in the models and after 1000 iterations, the Brooks-Gelman-Rubin diagnostics for each parameter were inspected to confirm whether the ratio was near 1 and the chains were stable.
- The auto-correlation plots were also inspected for serial dependence of the iterations, so with increasing lags there should be no or only small correlation between samples.

- Once convergence had been confirmed, an additional 5000 iterations were run to give accurate posterior estimates. This was confirmed by following the rule of thumb that the Monte Carlo standard error should be at least less than 5% of the posterior standard deviation.

Once a successful model fit of the harmonic components in terms of the model's chain convergence was achieved, the best fit seasonal pattern for the each of the outcomes was based the Deviance Information Criterion (DIC) (Spiegelhalter *et al.* 2002) and the association of the seasonal pattern with the outcome by checking that the parameter's 95% credible interval did not overlap zero. A credible interval is an interval which includes the true value of the parameter with $1-\alpha$ probability, as opposed to a CI, which is the range containing the unknown value of the parameter $1-\alpha$ of the time sampled from the observed data.

The mediation test was performed in the Bayesian setting essentially in the same way as in the Frequentist setting, though with one difference. The harmonic function's vector of parameters modelling the seasonal variation of 25(OH)D (as represented by α_X) was multiplied by the mediator's association with the outcome (as represented by β_M) as adjusted for the best fit harmonic function. When the credible interval of the product of parameters did not overlap with zero, then this was taken as evidence of an association of 25(OH)D with the outcome. The median of the parameter's posterior distribution was used as the point estimate to represent the value of the parameter.

A summary of the logic taken to applied the mediation test of the product of coefficients in the Bayesian and Frequentist setting is outlined in Figure 3.5.

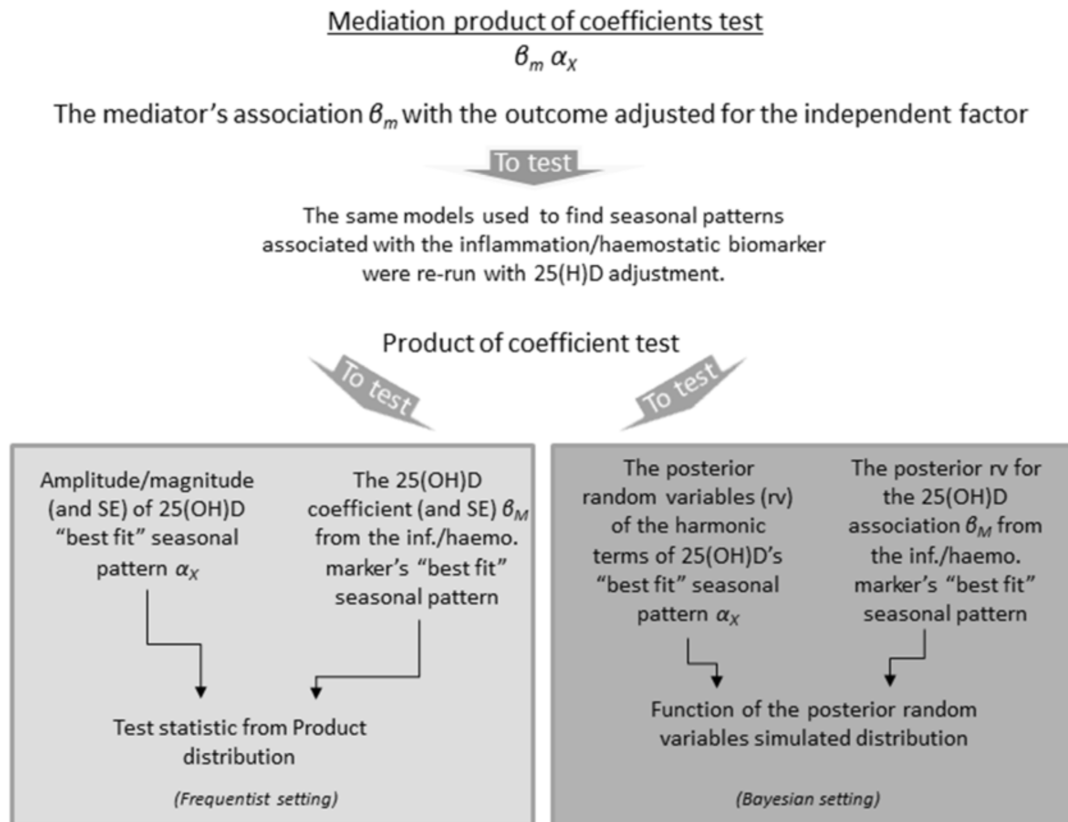


Figure 3.5 The logic taken to applied the mediation test of the product of coefficients in the Bayesian and Frequentist setting.

The proportion of seasonal variation of the inflammation/haemostatic biomarker attributed to 25(OH)D was derived by the change in amplitude of the harmonic function of the biomarker adjusted and unadjusted for 25(OH)D. The difference in seasonal variations of the biomarker from the two models (adjusted and unadjusted for 25(OH)D) was calculated by subtracting the same harmonic function frequency sine and cosine parameters. The amplitude/magnitude of harmonic function was simulated by predicting and ranking the outcome over a year period, $t = 1, .., 365$. The amplitude was then obtained as half the difference in the outcome on the days

that the extrema had occurred. When there was only a single frequency in the harmonic function for an outcome, the amplitude could be estimated directly from the parameters, as previously defined by $\hat{A} = \sqrt{\hat{\gamma}_1^2 + \hat{\gamma}_2^2}$.

3.4 Results

The results are presented in the order of the methods outlined in the statistical methods section. The first section presents: the distribution of 25(OH)D concentrations by lifestyle and socio-economic covariates; the relationship of BMI and waist circumference with 25(OH)D and the inflammation/haemostatic biomarkers; the association of 25(OH)D with the inflammation/haemostatic biomarkers, adjusted in stages for lifestyle, socio-economic and adiposity covariates using multiple linear regression.

The second section presents the results of the mediation analysis applied in Frequentist and Bayesian settings. Three mediation models were run: the first model established whether seasonal variation as modelled by the harmonic function existed for the inflammation/haemostatic biomarkers. The second model established how the biomarker's seasonal patterns were altered after adjusting for the mediator (25(OH)D), and the third model was to establish the seasonal pattern of 25(OH)D. Frequentist and Bayesian inference was applied to the mediation product of coefficients test.

3.4.1 Characteristics of study participants and multiple linear regression results

Compared with women, men had slightly higher 25(OH)D concentrations and a lower prevalence of 25(OH)D deficiency (< 25 nmol/L) (Table 3.3). Variations of 25(OH)D concentrations and trends in frequencies of 25(OH)D deficiency were similar across several lifestyle covariates. Non-smokers, participants who watched TV/used a PC < 1 hour/day, and who engaged in vigorous activity had higher 25(OH)D levels and a lower prevalence of 25(OH)D < 25 nmol/L, compared with participants who smoked ≥ 20 cigarettes/day, spent the most time watching TV/using a PC (≥ 3 hrs/day), and did not engage in vigorous exercise. For participants with high 25(OH)D levels (≥ 125 nmol/L), trends in prevalence were not significant between the sexes, types of smoking behaviour, or engagement of vigorous physical activity.

Table 3.3 The distribution of 25(OH)D concentrations by lifestyle and socio-economic variables.

	25-hydroxyvitamin D, nmol/l			
	Number (%)	Geometric Mean* (95% CI)	<25 nmol/l %* (*n*)	>125 nmol/l %* (*n*)
Sex				
Men	3270 (50.0)	53.6 (52.8, 54.5)	6.2 (203)	1.4 (46)
Women	3268 (50.0)	51.9 (51.1, 52.8)	8.4 (273)	1.4 (45)
p-values		0.003	0.0007	0.9
BMI				
<25	2361 (36.1)	55.1 (54.0, 56.2)	7.5 (178)	2.2 (52)
25-30	2737 (41.9)	54.2 (53.3, 55.1)	5.8 (159)	1.2 (32)
>30	1440 (22.0)	46.8 (45.7, 47.9)	9.7 (139)	0.5 (7)
p-values		≤0.0001	0.04	≤0.0001
Waist circumference**				
Quartile 1	1642 (25.1)	57.4 (56.0, 58.7)	6.8 (111)	2.8 (46)
Quartile 2	1639 (25.1)	55.2 (54.0, 56.4)	6.2 (102)	1.8 (29)
Quartile 3	1627 (24.9)	52.6 (51.4, 53.7)	6.1 (100)	0.7 (11)
Quartile 4	1618 (24.8)	46.6 (45.6, 47.6)	9.9 (160)	0.2 (4)
Unknown	12 (0.2)	44.8 (29.4, 68.3)	25.0 (3)	8.3 (1)
p-values		≤0.0001	0.003	≤0.0001
Vigorous physical activity				
No	3206 (49.0)	49.6 (48.8, 50.4)	9.4 (300)	1.2 (37)
Yes	3236 (49.5)	56.3 (55.4, 57.2)	5.0 (163)	1.7 (54)
Unknown	96 (1.5)	48.0 (43.2, 53.2)	13.5 (13)	0.0 (0)
p-values		≤0.0001	≤0.0001	0.07
TV watching/ use of PC				
< 1 hours/day	745 (11.4)	56.3 (54.5, 58.2)	5.8 (43)	2.3 (17)
1-2 hours/day	3455 (52.8)	54.9 (54.1, 55.7)	5.5 (191)	1.6 (54)
≥ 3 hours/day	2056 (31.4)	48.9 (47.9, 49.9)	9.9 (204)	0.9 (19)
Unknown	282 (4.3)	47.7 (44.9, 50.5)	13.5 (38)	0.4 (1)
p-values		≤0.0001	≤0.0001	0.01
Smoking				
None	3039 (46.5)	54.2 (53.3, 55.1)	6.1 (185)	1.3 (41)
Ex-smoker	1795 (27.5)	54.8 (53.7, 55.9)	5.1 (91)	1.3 (23)
1-19 per day	762 (11.7)	50.3 (48.5, 52.1)	10.4 (79)	1.7 (13)

≥20 per day	720 (11.0)	45.5 (43.8, 47.2)	14.4 (104)	1.9 (14)
Unknown	222 (3.4)	51.8 (48.8, 55.1)	7.7 (17)	0.0 (0)
p-values		≤0.0001	≤0.0001	0.2
Alcohol consumption				
Non-drinker	378 (5.8)	46.1 (44.0, 48.4)	11.4 (43)	0.0 (0)
Light < 7 drinks/ wk	3155 (48.3)	52.1 (51.3, 52.9)	7.5 (237)	1.0 (31)
Moderate 7-13 drinks/wk	1651 (25.3)	55.6 (54.4, 56.9)	5.6 (92)	1.8 (30)
Heavy 14-21 drinks/wk	746 (11.4)	55.9 (54.1, 57.8)	4.8 (36)	2.5 (19)
Very heavy >21 drinks/wk	590 (9.0)	49.7 (47.7, 51.8)	10.7 (63)	1.7 (10)
Unknown	18 (0.3)	41.5 (31.5, 54.6)	27.8 (5)	5.6 (1)
p-values		0.01	0.9	0.0003
Adult social class (2000)†				
I & II	2675 (40.9)	53.3 (52.4, 54.2)	6.7 (180)	1.4 (38)
III non-manual	1363 (20.9)	52.2 (50.9, 53.5)	7.9 (107)	1.1 (15)
III manual	1220 (18.7)	54.5 (53.1, 56.0)	6.1 (74)	2.0 (25)
IV & IV	1013 (15.5)	51.2 (49.8, 52.7)	8.0 (81)	1.0 (10)
Other	267 (4.1)	48.8 (45.9, 51.8)	12.7 (34)	1.1 (3)
p-values		0.007	0.02	0.8

* Values are *n* (%) or geometric mean. *P*-values from test for trend in linear or logistic regression where appropriate adjusted for season and sex. Unknown values excluded.

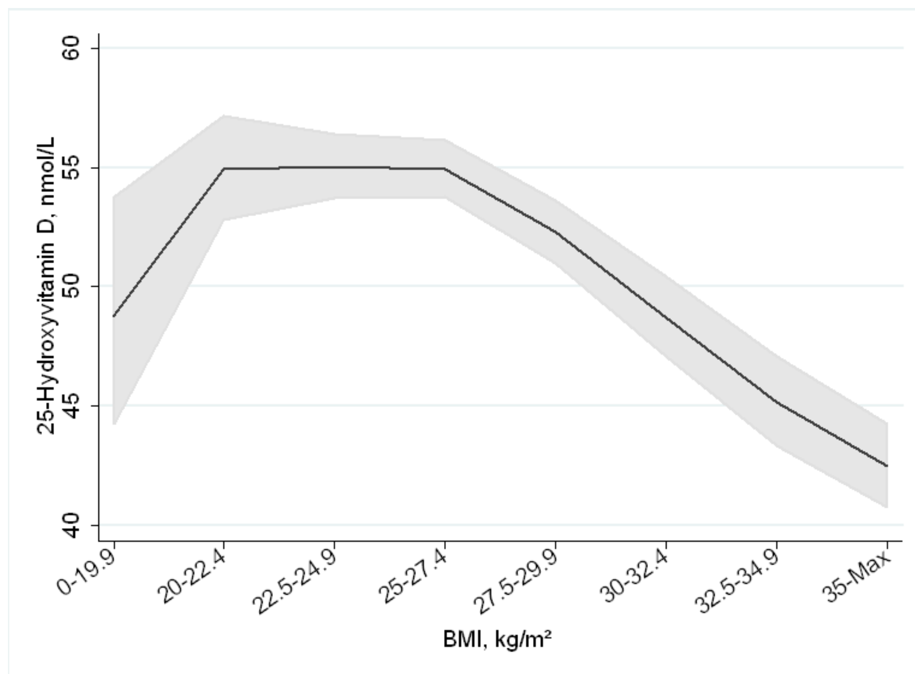
**Waist circumference quartiles: for men; 65.4-90.6, 90.7-96.7, 96.8-103.5, 103.6-151.2 cm; for women; 56.2-75.8, 75.9-82.6, 82.7-91.6, 91.7-138.3 cm.

† Classes I&II are managerial/professional, IV/V unskilled manual. “Other” includes cohort members who are institutionalised, retired, unemployed and other unclassifiable.

Participants with BMI below 25 kg/m^2 had higher levels of 25(OH)D, compared with obese participants (BMI > 30 kg/m^2) [(46.8 nmol/L (95% CI 45.7, 47.9) vs 55.1 nmol/L (95% CI 54.0, 56.2), *p*-value ≤0.0001)] (Table 3.3). Overweight participants (BMI 25-30 kg/m^2 or waist circumference in third quartile) had a lower prevalence of being 25(OH)D deficiency (<25 nmol/L), compared with those in the lighter and

heavier BMI categories or in the bottom and top quartiles of waist circumference. There was evidence that adiposity had a non-linear relationship with 25(OH)D (curved BMI *p*-value ≤ 0.0001, curved waist circumference *p*-value=0.04 adjusted for sex and month of blood sample). The curved relationship between 25(OH)D and measures of adiposity were visibly evident in Figure 3.6, where participants with BMI between 20-27.5 kg/m^2 had higher 25(OH)D levels, than underweight or obese participants.

A



B

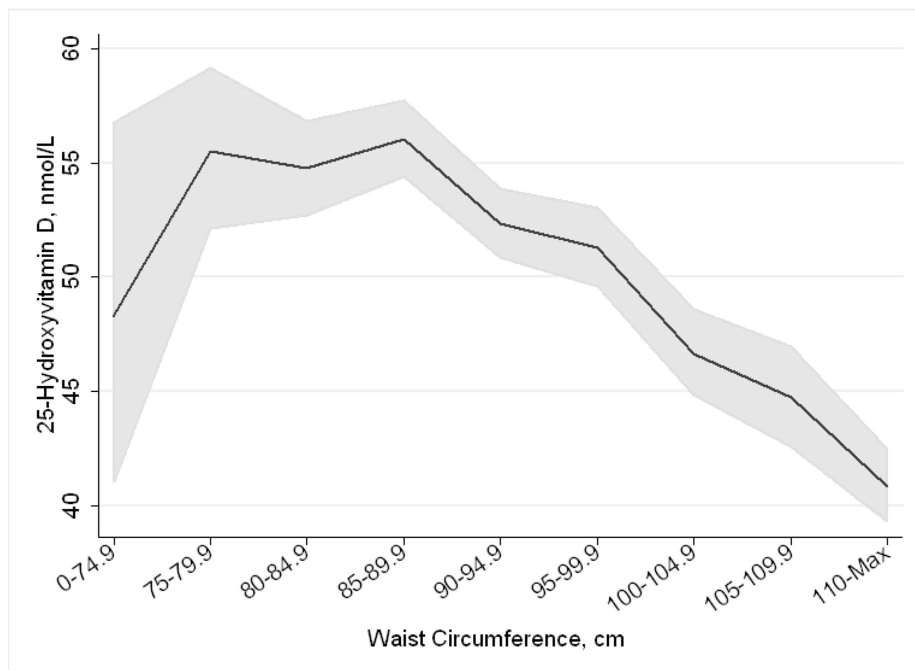
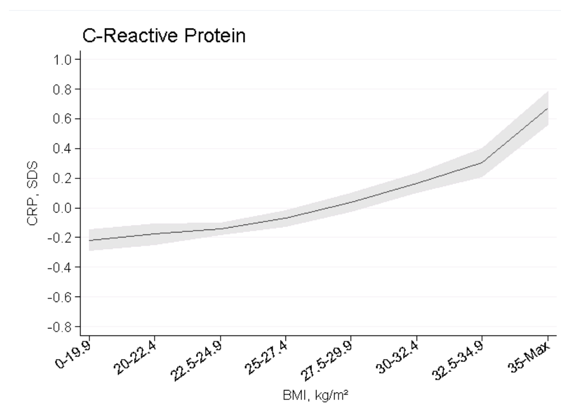


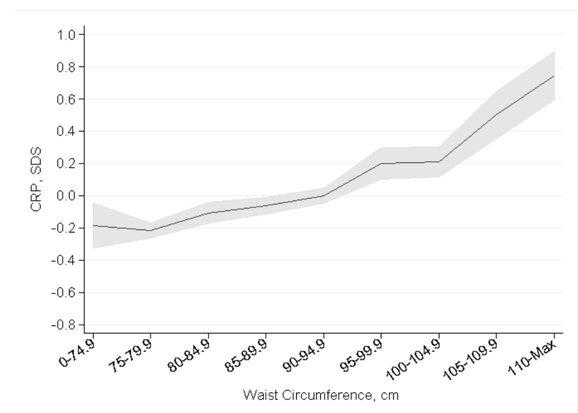
Figure 3.6 Variation in the mean 25(OH)D concentration by (A) BMI and (B) waist circumference. Values are geometric means (95% CI) standardized by sex and season in which measurements were taken.

Significant adverse relationships were observed between the measures of adiposity and the haemostatic and inflammatory biomarkers (Figure 3.7) (p-value <0.0001 for all outcomes adjusted for sex and month of blood sample). There was evidence of a non-linear BMI association with CRP and vWF (p-value ≤ 0.0001 for CRP and vWF adjusted for sex and month of blood sample), but not with the other biomarkers (p-value ≥0.23 with D-dimer, fibrinogen, tPA adjusted for sex and month of blood sample). Evidence of non-linear waist circumference association was found only with vWF (p-value < 0.0001 for vWF, p-value ≥0.20 for all other outcomes, adjusting for sex and month of blood sample). The association between 25(OH)D and the haemostatic/inflammatory biomarkers did not vary by levels of adiposity (BMI, waist circumference and 25(OH)D interactions p-value ≥0.13 for all comparisons, adjusting for sex and month of blood sample).

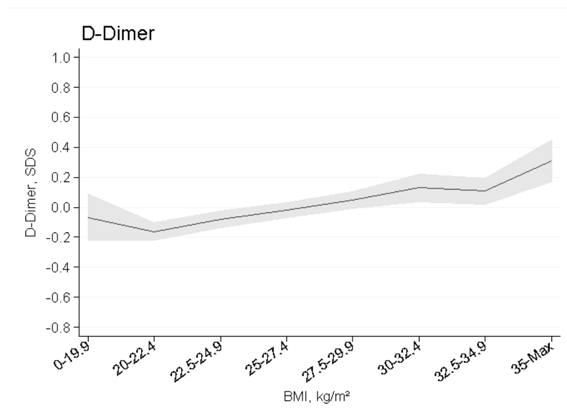
1.A



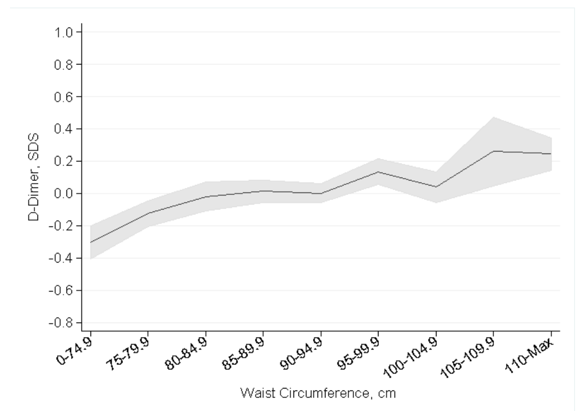
1.B



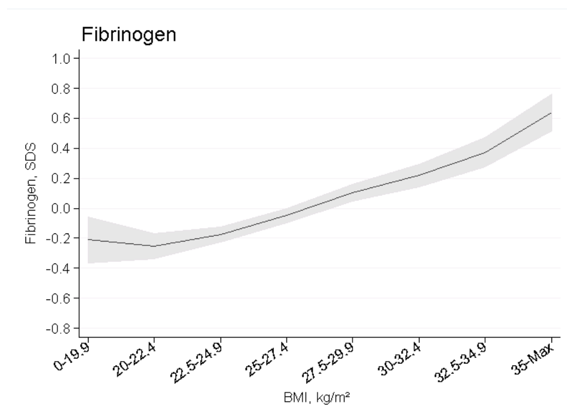
2.A



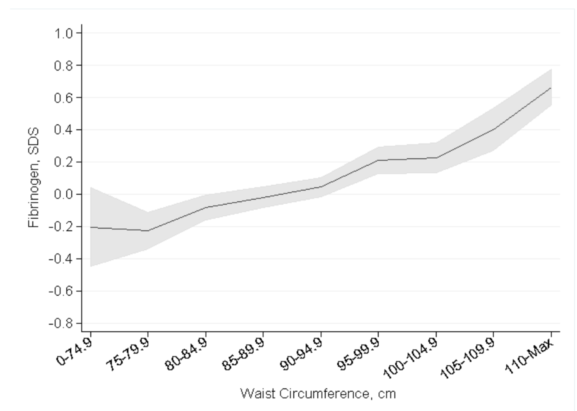
2.B



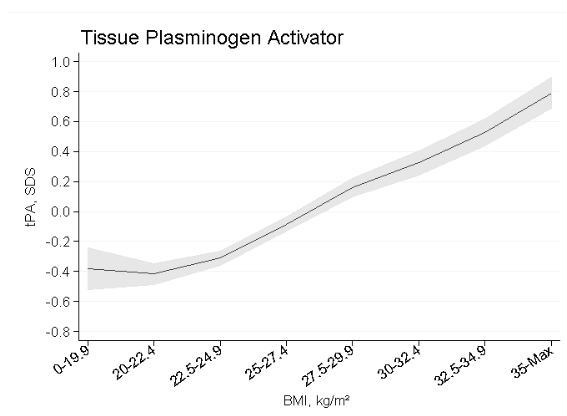
3.A



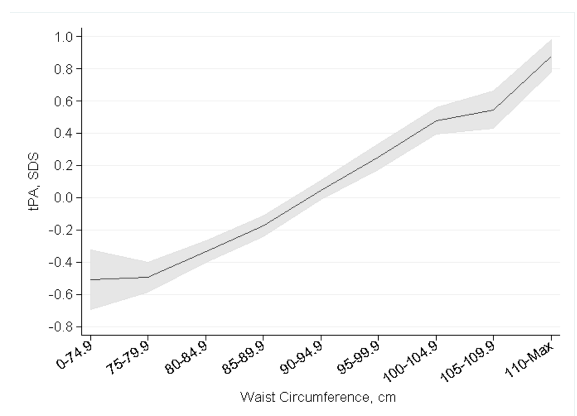
3.B



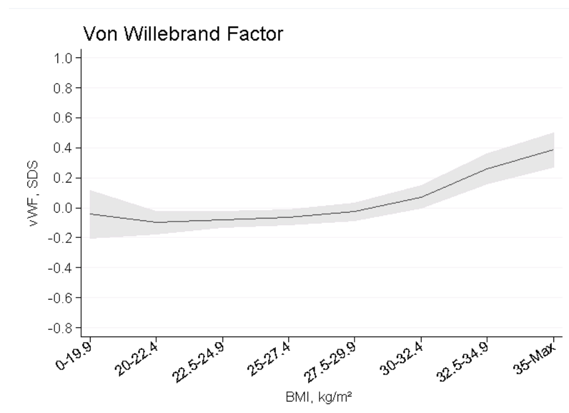
4.A



4.B



5.A



5.B

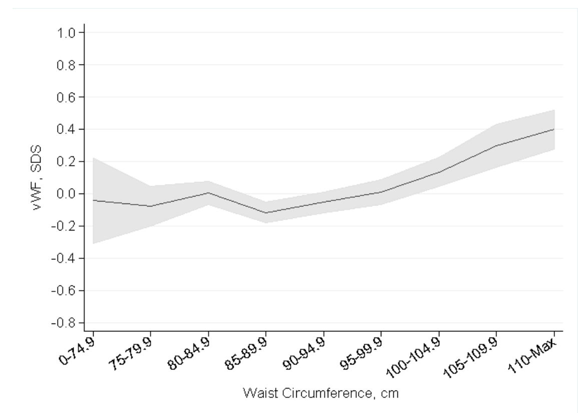
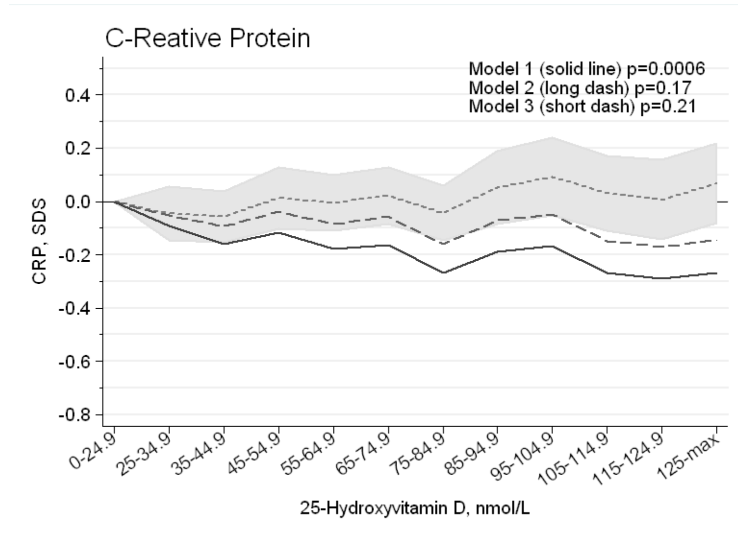


Figure 3.7 Variation in the mean sex specific standard deviation (SDS) concentrations of 1. CRP, 2. D-dimer, 3. fibrinogen, 4. tPA and 5. vWF by (A) body mass index and (B) waist circumference with 95% CI in the shaded area.

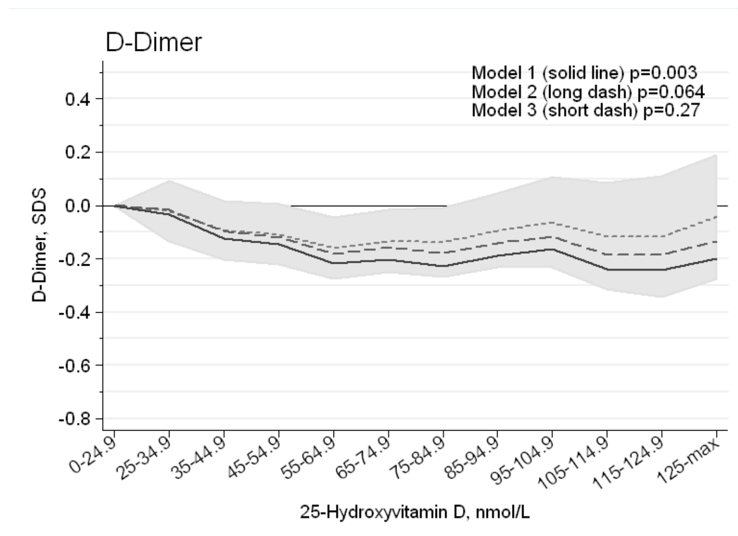
A protective 25(OH)D association was found with CRP, D-dimer, fibrinogen, tPA and vWF (p-value ≤0.015 for all comparisons, adjusting for sex and month of blood sample) (Figure 3.8) using multiple linear regression. After the additional adjustment for lifestyle and socio-economic covariates, 25(OH)D was no longer related to CRP and and vWF (p-value ≥0.17 for all comparisons). For tPA, the protective 25(OH)D relationship remained (p-value=0.015) after full adjustment (sex, month, lifestyle, socio-economic and adiposity covariates) although the size of the association was reduced with each model adjustment (Figure 3.8 D). Compared with participants with 25(OH)D < 25 nmol/L, those with 25(OH)D ≥ 75 nmol/L had 18.4% (95% CI 8.1, 28.8) lower tPA concentrations after full adjustment. After adjusting for sex, month, lifestyle, socio-economic and adiposity covariates, there was some evidence that 25(OH)D had a non-linear relationship with D-dimer (curved 25(OH)D; p-value=0.01), and to some extent with fibrinogen (curved 25(OH)D; p-value=0.06), although the non-linear association with fibrinogen was not statistically significant (Figure 3.8 B and C). Levels of fibrinogen were higher in

participants with 25(OH)D concentrations ≥ 125 nmol/L, compared with those with 25(OH)D < 125 nmol/L after full adjustment (Figure 3.8 C).

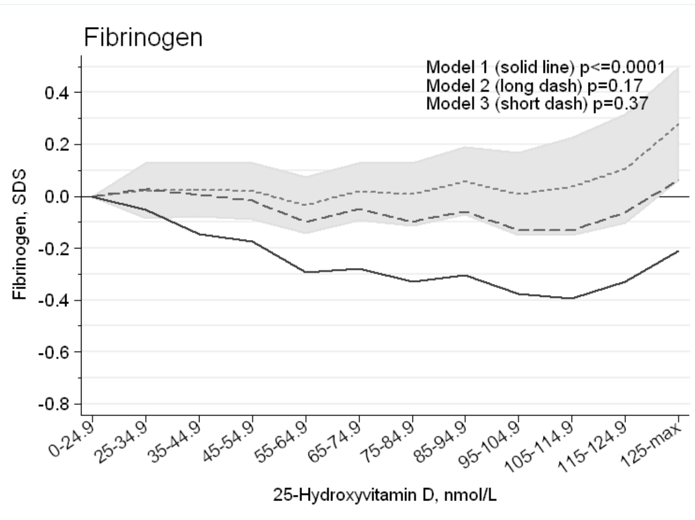
A



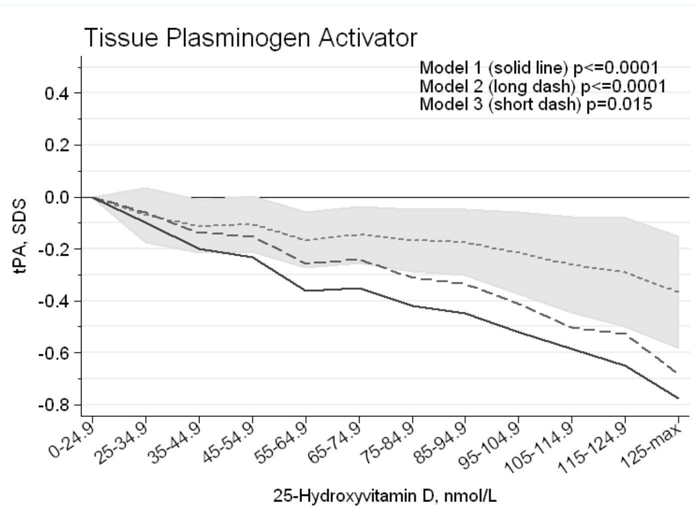
B



C



D



E

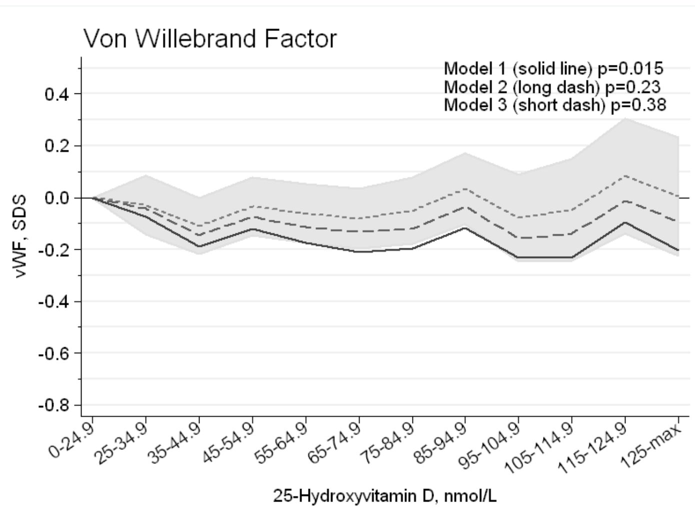


Figure 3.8 Variation in inflammatory and haemostatic markers by serum 25(OH)D concentration. Model 1 (solid line): adjusted for month of blood sample and sex. Model 2 (dashed, long): adjusted for lifestyle and socio-economic indicators

(physical activity, time spent watching TV/using PC, smoking, alcohol consumption and birth and adult social class), month of blood sample and sex. Model 3 (dashed, short): adjusted for adiposity (BMI and waist circumference), lifestyle/socio-economic indicators, month of blood sample, and sex. Values are coefficients from linear regression (reference <25nmol/l). 95% CI presented for Model 3 by the shaded area.

3.4.2 Mediation analysis results

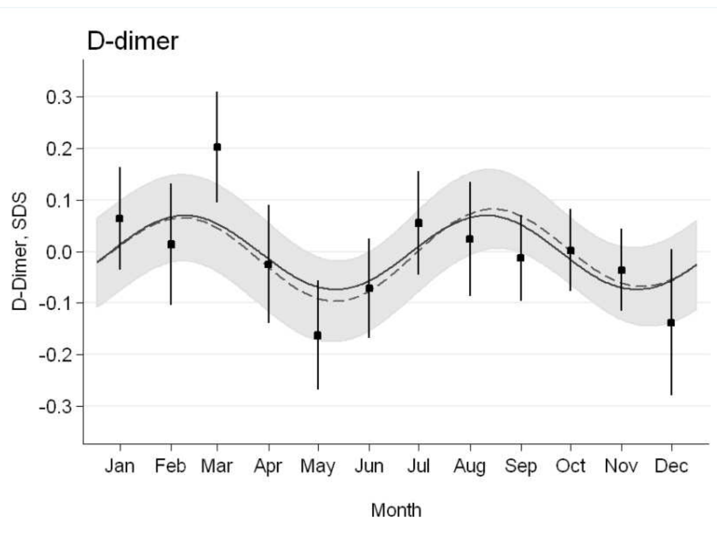
In the Frequentist setting, D-dimer, fibrinogen, tPA and vWF showed evidence of seasonal variation (*p*-value ≤0.016 adjusted for seasonal covariates; chest infection, alcohol consumption, PC/TC leisure time, recreation MET hours, and social class at adulthood) (Figure 3.9) (Appendix 2 Table 1). A yearly pattern was the best fit for fibrinogen, tPA and vWF over all other combinations of the frequencies in harmonic function. For D-dimer, it was the semestral pattern that had the best fit of the patterns. Whereas for CRP, none of the combinations of the yearly, semestral or quarterly patterns were significantly associated (*p*-value ≥0.77 adjusted for the seasonal covariates).

The same seasonal patterns for the haemostatic and inflammation biomarkers were found using Bayesian inference (Appendix 2 Table 2). These included the lack of seasonal variation of CRP, where all the harmonic function parameters' 95% credible intervals overlapped with zero. Therefore, CRP was excluded from any further mediation analysis.

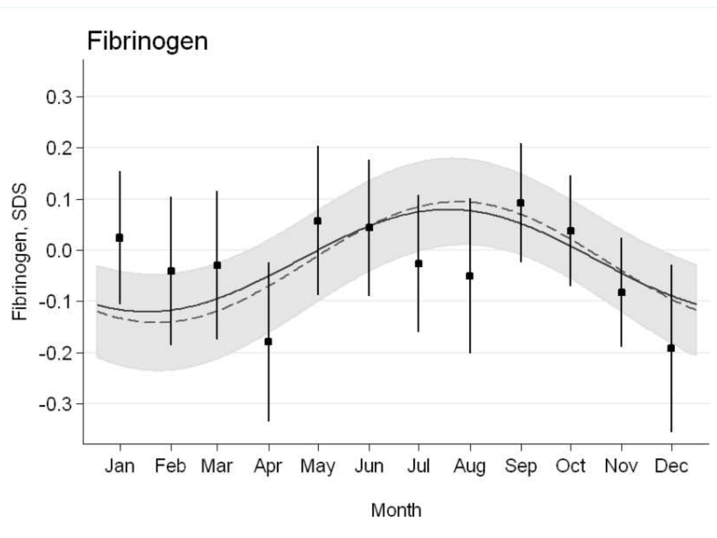
The frequencies of the seasonal components remained the same for the biomarkers after adjusting for 25(OH)D in the Bayesian and Frequentist settings (Appendix 2

Table 3 for Frequentist, and Table 4 for Bayesian). For tPA, fibrinogen, and D-dimer, the amplitude of the seasonal pattern slightly increased (Figure 3.9) after adjusting for 25(OH)D, compared with the models unadjusted for 25(OH)D.

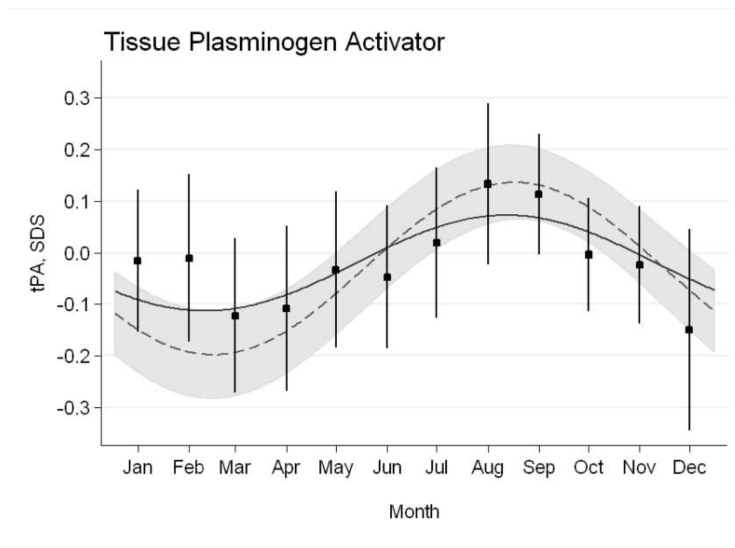
A



B



C



D

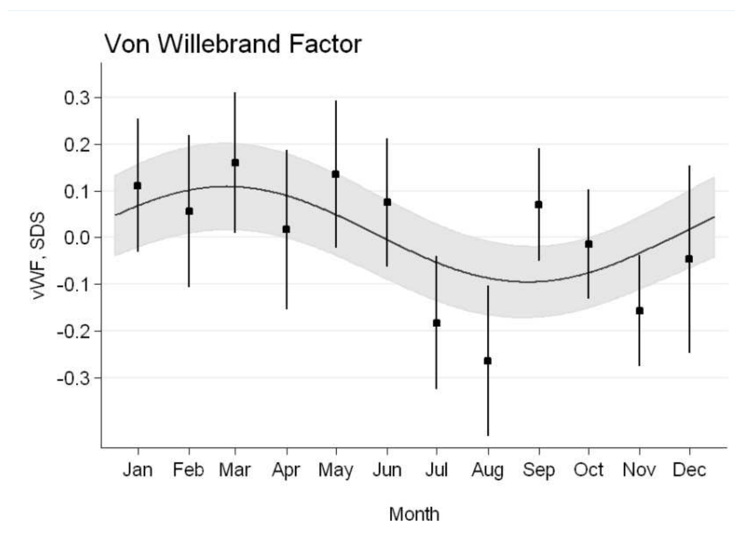


Figure 3.9 Seasonal patterns of A. D-Dimer; B. fibrinogen; C. tPA; and D. vWF. The solid line is from the model adjusted for chest infection, alcohol consumption, PC/TV leisure time, recreation MET hours, social class at adulthood. The dashed line is from a model that is in addition adjusted for 25(OH)D in the 1958BC (aged 45 years). Values are the predicted outcome from models and 95% prediction intervals are presented by the shaded area for the model adjusted for 25(OH)D. The squares are the SDS average concentrations and vertical bars the 95% CI.

For the mediation product of coefficient test in the Frequentist setting, the seasonal magnitude of 25(OH)D was used as a measure of seasonal variation of the mediator. Including all three frequencies in the harmonic function (yearly, semestral and quarterly) was the preferred fit for the seasonal variation of 25(OH)D in the Frequentist setting, and this combination was reconfirmed in the Bayesian setting (Appendix 2 Tables 1 for Frequentist setting, and Table 2 for Bayesian setting). In the Frequentist setting, the seasonal model predicted that over a year period a minimum in 25(OH)D concentrations of 42.0 nmol/L (95% Prediction Interval PI 40.5, 43.6) occurred during day 29 (end of Jan) and a maximum in concentrations during day 252 (end of Sept) of 77.3 nmol/L (95% PI 75.9, 78.7). The magnitude of the seasonal variation of 25(OH)D was 17.9 nmol/L (SE 0.592), as derived from the individual frequencies' amplitudes and the fraction of frequencies contribution at the extrema (Table 3.4). The magnitude as estimated by parametric bootstrap was the same as estimate derived from the fraction of the harmonic frequencies.

Table 3.4 The individual amplitude and magnitude of the seasonal variation of 25(OH)D. Estimated from a three frequency harmonic function model (adjusted for sex) used in the mediation test within the Frequentist setting.

Amplitude (SE) (nmol/l) for the individual and combined frequencies					
	Yearly	**Semestral**	**Quarterly**	**Magnitude** α_M	**LRT *p*-value**
25(OH)D amplitude	15.8 (0.425)	3.67 (0.428)	2.19 (0.421)	17.9 (0.592)	≤0.0001
Fraction*	0.903	0.414	0.984	-	-

* The fraction each of the harmonic frequencies contributed to the magnitude. The maximum and minimum of 25(OH)D seasonal variation within a year time frame occurs at time points t_{min}=29 days and t_{max}=252 days

In the Frequentist setting, the mediation product of coefficients test was the product of the maximum likelihood estimates for the parameters of the 25(OH)D association with the haemostatic/inflammation biomarker and the magnitude of the seasonal variation of 25(OH)D. A significant 25(OH)D association was found with tPA, fibrinogen, and D-dimer (p-value ≤ 0.004) after adjusting for seasonal covariates and seasonal pattern (Table 3.5). From the mediation test, it was inferred that 25(OH)D was associated with tPA, fibrinogen, and D-dimer, as there was evidence of 25(OH)D mediating the seasonal variations of tPA, fibrinogen and D-dimer (Table 3.5) (p-value ≤ 0.005). For vWF, there was no evidence of a 25(OH)D association in the model adjusted for seasonal covariates and yearly pattern (p-value=0.83) and, likewise, in the mediation test (p-value =0.99). The parametric bootstrap mediation test results also provided confirmation of the associations of 25(OH)D with the outcomes (Appendix 2 Table 5).

Table 3.5 The mediation product of coefficients test and 25(OH)D association in the Frequentist setting with the haemostatic markers adjusted for seasonal variation in mixed effects model

	Outcome adjusted for best fit harmonic terms, 25(OH)D and seasonal covariates*			
	D-Dimer, SDS adjusted for semestral pattern	**Fibrinogen, SDS adjusted for yearly pattern**	**tPA, SDS adjusted for yearly pattern**	**vWF, SDS adjusted for yearly pattern**
25(OH)D per 10 nmol/L coefficient (SE)	-0.019 (0.005)	-0.016 (0.006)	-0.047 (0.006)	-0.001 (0.006)
LRT *p*-value	<0.001	0.004	<0.001	0.83
Mediation test estimate	-0.33 (0.10)	-0.28 (0.10)	-0.85 (0.10)	-0.00 (0.010)
Mediation test *p*-value	<0.001	0.005	<0.001	0.99

* Seasonal covariates of chest infection, alcohol consumption, PC/TC leisure time, recreation MET hours, social class at adulthood adjusted for as fixed effects and laboratory batch included as a random effect.

In the Bayesian setting, the mediation product of coefficient test was the product of the parameters of 25(OH)D association with the haemostatic/inflammation biomarker and the harmonic parameters associated with 25(OH)D. From the Bayesian inference, a 25(OH)D association with D-dimer, tPA, fibrinogen was confirmed, as was the lack of association with vWF (Appendix 2 Table 6). For tPA, D-dimer and fibrinogen, it was inferred from the product of coefficients test that seasonal variation was mediated through 25(OH)D (Appendix 2 Table 7). The results of mediation tests in the Bayesian setting agreed with those from the tests in the Frequentist setting.

For tPA, the change in amplitude that was due to 25(OH)D was 0.079 SDS (SD 0.046), in comparison to the amplitude of the yearly seasonal pattern of 0.176 SDS (SD 0.034) (after adjusting for 25(OH)D and seasonal covariates) (Table 3.6). The change in the amplitude for fibrinogen was 0.028 SDS (SD 0.053), compared with 0.131 SDS (SD 0.038) for the fully adjusted model). The findings for tPA from the mediation analysis were in keeping with the results from the multiple regression models.

Table 3.6 The seasonal amplitude and change in amplitude associated with the mediator 25(OH)D in D-Dimer, Fibrinogen, tPA (Bayesian setting).

	D-dimer, SDS Amplitude mean and SD	**Fibrinogen, SDS Amplitude mean and SD**	**tPA, SDS Amplitude mean and SD**
Amplitude (model adjusted for 25(OH)D and seasonal confounders)	0.082 (0.025)	0.131 (0.038)	0.176 (0.034)
Change in amplitude due to 25(OH)D	0.009 (0.037)	0.028 (0.053)	0.079 (0.046)

3.5 Discussion

From the different analytical approaches presented in this chapter, it was found that 25(OH)D concentrations were associated with levels of tPA. There was some evidence from analysis using multiple linear regression models that vitamin D status was associated with D-dimer and fibrinogen, and the shape of the associations were non-linear. It was also inferred from the findings of the mediation analysis that 25(OH)D was associated with D-dimer and fibrinogen. In the Bayesian setting, based on the change in amplitude of the seasonal patterns it appeared that magnitude of the 25(OH)D association was largest with tPA, then fibrinogen and less so with D-dimer. No significant associations were observed between 25(OH)D and CRP, vWF, despite evidence of vWF showing a yearly seasonal rhythm.

3.5.1 Comparisons with other studies

The finding of a yearly pattern of vitamin D status, with peak levels occurring in late summer and low occurring in mid to late winter was expected, and has been observed in several studies in the UK, and studies in countries at similar latitudes (Ashwell *et al.* 2010; Reusch *et al.* 2009). The majority of vitamin D intake occurs from casual sunlight exposure that initiates the skin synthesis of pre-vitamin D_3 (Haddad & Hahn 1973). In the UK, as wintertime progresses towards the winter solstice, the sun becomes lower in the sky and ultraviolet B rays needed to initiate the synthesis fail to reach the surface, causing a drop in vitamin D status (Webb *et al.* 2010).

Coinciding with the low vitamin D status, a wintertime excess of coronary heart disease mortality has been observed (Pell & Cobbe 1999). However, although seasonal patterns are common in nature, the causes of the rhythms are complex and often not well understood. Seasonal weather triggers a cascade of interlinking factors that range from environment, such as temperature, humidity, rain, cloud cover and pollution, to social and societal behaviour to adjust to the change in the environment. Extremes in temperature reached in winter and summer have been associated with a greater risk of mortality in the elderly population in the UK (Hajat *et al.* 2007). However, a change in temperature alone is unlikely to be the cause of the seasonal variation in CVD mortality, as the excess wintertime mortality did not significantly decrease from 1964 to 1984, even though considerable improvements were made in heating homes over those decades (Keatinge *et al.* 1989). Whereas, underlying pre-existing conditions and seasonal illness, such as influenza, have been associated with an increased risk of cardiovascular mortality in winter (Wilkinson *et al.* 2004).

There have been several studies that have investigated seasonal patterns of the haemostatic/inflammatory biomarkers, in particular on CRP and fibrinogen. An earlier study of the 1958BC biomedical survey participants found evidence of a seasonal pattern for tPA and vWF, but no evidence of a pattern in CRP, D-dimer, and fibrinogen, (Rudnicka *et al.* 2007). However, the earlier study did not exclude any individuals unlike the work presented in this chapter, where subjects on CVD medication were excluded. Furthermore, the seasonal analyses presented in this chapter used random effect regression models to control for the nuisance variable of multiple assay batches of the haemostatic biomarkers. The assay batches of the

biomarkers were treated as random effects since they were correlated with the date of the assay, and therefore, season of when blood was collected. However, the interest was not in estimating changes in the outcome variables predicted by this correlation. Differences in the approaches between the earlier study and the work presented here also altered when the seasonal peaks occurred in tPA and vWF. A harmonic function was used to model the seasonal variations in the earlier study, but only a yearly pattern was investigated (Rudnicka *et al.* 2007).

Of the studies that have investigated seasonal variations in CRP, several studies in healthy individuals have been consistent with findings of this work (Frohlich *et al.* 2002; Rogowski *et al.* 2005; Rudez *et al.* 2009). Moreover, a longitudinal study also in healthy individuals found a lack of seasonal variation in CRP and that fibrinogen levels peaked in summer and had a low in late winter, which agrees with the findings presented here (Rudez *et al.* 2009). In contrast, studies in the elderly have found seasonal variations in CRP and different seasonal peaks in fibrinogen (Crawford *et al.* 2000; Woodhouse *et al.* 1994). Both studies had participants that were at least 65 years old that were repeatedly measured for a year, and also utilised a harmonic function to model seasonality. It was found that the seasonal patterns of CRP and fibrinogen levels peaked during mid-to-late winter. CRP and fibrinogen concentrations are known to increase with age, especially in the older years (Hutchinson *et al.* 2000; Lowe 1997). A study stratifying participants into age groups of 55 to 75 years old and above, found that the seasonal variation of fibrinogen was more pronounced in the older group, than younger (van der Bom *et al.* 1997).

Seasonal variation of fibrinogen is consistently observed in studies, although results conflict as to when the peak (and low) occurs during the year (Crawford *et al.* 2000; Crawford *et al.* 2003b; Frohlich *et al.* 1997; Rudez *et al.* 2009; Woodhouse *et al.* 1994). A large four year cross-sectional study found that each year fibrinogen had a yearly seasonal pattern, but the season of when the peaks occurred varied across the years (Steinvil *et al.* 2009). This study concluded that the timing of the peaks of fibrinogen levels had occurred shortly after episodes of respiratory illness, which may offer some explanation as to why there are differences in the timing of seasonal peaks found across studies.

Outside of the studies in the 1958BC participants, there have been no prior studies investigating seasonal rhythms in D-dimer. However, two studies' findings may have indirectly linked the intake of vitamin D with venous thromboembolism, which due to increased fibrin turnover can be diagnosed by elevated D-dimer levels (Bounameaux *et al.* 1991; Di Nisio M. *et al.* 2007). Venous thromboembolism has a peak in incidence during the wintertime (Dentali *et al.* 2011). Furthermore, a large prospective population based study of 29,317 Swedish women, found 30% reduced risk of venous thromboembolism in those women that used sunbeds or sunbathed, compared with those who did not (Lindqvist *et al.* 2009).

3.5.2 Explanations

There is evidence that vitamin D endocrine system influences cardiovascular health. Vitamin D receptors (VDR) that bind to and are activated by the active form of vitamin D have been found in cells from most tissues and organs throughout the body (Norman & Bouillon 2010). The activated VDR forms a heterodimer complex

with retinoid X receptor that associates with vitamin D response elements in the target gene (Pike & Shevde 2005). It is estimated that up to 3% of the human genome maybe regulated in some way by the vitamin D hormone (Bouillon *et al.* 2008). Evidence suggests that VDR are expressed in endothelial cells (Suzuki *et al.* 2009), vascular smooth muscle cells (Wu-Wong *et al.* 2007a) within human blood vessels and cardiac myocytes contained within human heart tissue (O'Connell & Simpson 1996). Cardiac myocytes and fibroblasts have also exhibited expression of 1α-hydroxylase, and therefore are capable of extra-renal synthesis of active vitamin D (Chen *et al.* 2008). Studies in animals have found that VDR knockout mice develop hypertrophied hearts (Simpson *et al.* 2007) and renin hypertension (Xiang *et al.* 2005), that can lead to systemic hypertension.

Vitamin D may also have a direct role in maintaining normal antithrombotic homeostasis. In the normal haemostatic response, the expression of tissue factor in the damaged tissue initiates the blood coagulation cascade, and thrombomodulin and antithrombin lead to the breakdown of thrombin to prevent thromobosis (Hoffbrand & Moss 2011a). VDR knockout mice have displayed thrombus formation in the liver and kidney after intervention, compared with wild type mice (Aihara *et al.* 2004). Compared with wild type mice, VDR knockout mice have also had a reduced gene expression of antithrombin in the liver and thrombomodulin in the arota, but increased tissue factor gene expression in the liver and kidney. A RCT in cancer patients that intervened with high dosage calcitriol analogue found a reduction in adverse thrombotic events in participants receiving treatment, compared with those receiving a placebo (Beer *et al.* 2006).

Vascular calcification is an indicator of CVD (van der Bom *et al.* 1997), and evidence suggests that vitamin D may have a dual role in its development (Rodríguez et al. 2011). In patients with chronic kidney disease (CKD), evidence has linked treatment with an active vitamin D analogue to developing vascular calcification (Briese *et al.* 2006; Goldsmith *et al.* 1997). Conversely, CKD patients treated with an active vitamin D analogue have shown a decreased risk of CVD mortality, compared with those not treated (Shoji *et al.* 2004). Furthermore, in a case series of paediatric CKD patients both high and low dosages of active vitamin D were associated with vascular calcification (Shroff *et al.* 2008). In the case series, there was some evidence of a curved association between fibrinogen and 25(OH)D concentrations above > 125 nmol/L, which also happens to correspond with these findings of a visibly non-linear relationship between fibrinogen and 25(OH)D at 25(OH)D levels > 125 nmol/L (Figure 3.8 C). Toxicity by vitamin D is rarely evident when 25(OH)D concentrations are below 250 nmol/L (Jones 2008), however recently the IOM has lowered the safe level of 25(OH)D to 125-150 nmol/L (Institute of Medicine 2011).

3.5.3 Methodological considerations

Several analytical approaches have been employed in this work to investigate the associations of 25(OH)D with the inflammation/haemostatic biomarkers. The first approach used multiple linear regression models and adjusted in successive models for lifestyle, socio-economic and adiposity variables. A benefit of using multiple linear regression is the ease at which the exposure's association with the outcome can be interpreted. Due to the cross-sectional design of this study, issues of confounding and temporality of when cause and effect occurred remain (Greenland

et al. 2008). It is feasible that individuals who have poor cardiovascular health and elevated tPA levels opt to spend less time outside due to their poor health, thereby reducing their vitamin D intake/status. In order to control confounding, the linear regression models were adjusted for possible confounders of lifestyle and social-economic factors and adiposity. However, confounders might be unknown and not included in the models, or inadequately measured therefore not be completely controlled for in the models (Glymour & Greenland 2008). Furthermore, each additional adjustment for the lifestyle, socio-economic and adiposity variables would reduce the power in the linear regression models to detect an effect between 25(OH)D and the haemostatic/inflammation biomarker (Cohen 1988a).

Seasonal patterns of the inflammation/haemostatic markers were investigated since it would be unexpected that these would be affected by adiposity. A harmonic function was used to model the seasonal variations of the biomarkers and the function has several desirable features over the alternatives. One alternative could have been to use a categorical variable for month of blood sample in the models. However, by doing so information would have been lost since day and month of when blood was taken would have reduced to a simple month category. Compared with a categorical month variable, the harmonic function describes continually the seasonal rhythm of biomarker in terms of frequency, phase and amplitude (Refinetti *et al.* 2007). Furthermore it is an efficient way of modelling seasonality, for instance, for a yearly pattern, the harmonic function requires just two parameters (derived from day of year) in the model, compared with a categorical month variable that would have required 11 contrasting parameters. Hence, statistical power would have been reduced in the seasonality analyses if the harmonic function had not been

used. The differences in power across models that adjusted for the harmonic function as opposed for the month variable, may partially explain the additional findings of the 25(OH)D associations with fibrinogen and D-dimer in the mediation analysis. The seasonal rhythm is modelled with smooth sinusoidal curves that may not precisely fit all fluctuations of the biomarker across the period. In this study, multiple frequencies of yearly, semestral and quaterly were investigated with the biomarkers that would have reduce the residual variance of the model and improved model fit (Refinetti *et al.* 2007).

Modelling a seasonal pattern of an inflammation/haemostatic biomarker that was similar to the seasonal pattern of 25(OH)D is not sufficient in itself to conclude a correlation between the two. The framework of mediation analysis was adapted to investigate whether the seasonal pattern of the inflammation/haemostatic biomarkers was mediated through the seasonal variation of 25(OH)D. In a cross-sectional study design, a seasonal variation of the inflammation/haemostatic biomarkers can be observed, but a seasonal effect on the inflammation/haemostatic outcomes as implied by the mediation framework cannot be inferred (Mackinnon *et al.* 2007). However, the direction of the path between season and biomarkers is certain, since it is nonsensical that a biomarker would cause seasonal changes in the environment. Furthermore, as the cause of the seasonal variation in 25(OH)D is well understood, it is more likely that the seasonal variation in the biomarkers is mediated by the seasonal variation in 25(OH)D rather than the other way around. Therefore, the adaption of the mediation analysis with the seasonal variations reduces the possibility of reverse causation between 25(OH)D and the haemostatic biomarkers.

However, the findings from the mediation analyses are still susceptible to confounding, as were the analyses that used multiple linear regression models.

The mediation test of product of coefficients was adapted to accommodate the seasonal patterns. The association of the mediator with the outcome (adjusted for the independent variable) largely determines the statistical significance of the mediation product of coefficient test. In mediation analysis, it is assumed that to be a valid mediator of the independent variable that a relationship between the two would exist. An alternative test to the product of coefficients is the direct test of mediation, i.e. the difference between the independent variable associations with the outcome adjusted and unadjusted for the mediator. In the situation of different seasonal frequencies being observed in the outcome when adjusted and unadjusted for 25(OH)D, the direct test of mediation could not be easily reduced down to one statistic in the Frequentist setting. Furthermore, the direct test can be affect by multi-collinearity when the independent variable is strongly associated with the mediator (Baron & Kenny 1986). In the Bayesian setting, attempts were made to quantify the magnitude of the association of 25(OH)D with tPA, fibrinogen and D-dimer. More work is required to develop the seasonal adaption of the mediation analysis in the Bayesian setting. Future work could investigate the use of the direct mediation test with seasonal patterns, whether there are differences in results from the two mediation tests, and whether the direction of the mediator's association can be inferred from the adapted mediation tests.

4 Genome-wide association study on vitamin D status (Paper II)

4.1 Introduction

Lifestyle, environment and personal factors are influential determinants of the nutritional status of vitamin D (as measured by 25(OH)D). The majority of intake of vitamin D is obtained through skin exposure to ultraviolet (UVB) rays from sunlight that initiates a cutaneous conversion of 7-DHC to pre-vitamin D_3 (Holick 1994; Holick 2002). Season, time of day, cloud cover and altitude play a part in the reach of the right type of UVB rays needed to initiate the vitamin D synthesis (Webb 2006). Likewise, time spent outside, use of sunscreen and clothing cover affect whether the UVB rays will reach the skin (Hyppönen & Power 2007; Springbett *et al.* 2010). Diet can play a role in providing vitamin D though it exists in narrow range of foods (Lamberg-Allardt 2006), and fortification of vitamin D in foods in the UK is scarce (Scientific Advisory Committee on Nutrition 2007). Furthermore, individual factors such as age, skin colour and weight can influence levels of 25(OH)D (Bell *et al.* 1985; Chen *et al.* 2007; MacLaughlin & Holick 1985; Snijder *et al.* 2005).

Despite these considerable determinants of vitamin D status and intake, genetic heritability estimates for 25(OH)D concentrations are large, ranging from 29 to 80% (Hunter *et al.* 2001; Shea *et al.* 2009; Wjst *et al.* 2007). So far, rare mutations in genes that are expressed in the metabolism pathway of vitamin D and its uptake have been found, that can cause functional vitamin D deficiency and diseases such as

rickets (Glorieux & St Arnaud 2005; Malloy *et al.* 2005). However, at the time of initiation of this thesis, it was unknown whether common genetic variants influence nutritional vitamin D status (as measured by 25(OH)D), and contribute to the risk of vitamin D deficiency/insufficiency.

In order, to investigate the association of common genetic variants with 25(OH)D concentrations, this work was done as part of a collaborative effort of cohorts that comes under the title of the SUNLIGHT consortium, Study of Underlying Genetic Determinants of Vitamin D and Highly Related Traits. A paper related to these findings has been published in a peer-reviewed journal (Wang *et al.* 2010b) and is shown in Appendix 1. The focus of this work was on analysing the 1958BC genetic data of the three genome-wide sub-samples labelled WTCCC, WTCCC2 and T1DGC (as outlined in Chapter 2), and the de-novo genotyped data of the remaining subjects. Work presented in this chapter has extended the work presented in the paper and includes detailed linkage disequilibrium (LD) plots, estimation of single nucleotide polymorphisms (SNP) effect size and allele score with 25(OH)D and the risk of vitamin D deficiency/insufficiency.

Several key developments led to the realisation of genome-wide association studies (GWAS)(Seng & Seng 2008). During the 1990s, it became apparent that genetic linkage studies were limited in their abilities to detect genetic variants with many types of common diseases (Risch & Merikangas 1996). Linkage studies rely on genetic data from related individuals, such as siblings or multi-generational families, and are ideal for finding rare genetic variations in Mendelian (monogenetic) diseases (Thomas 2004a). However, linkage studies have limited in power for finding

multiple genetic variants that have a small or modest effect on multi-factorial/complex diseases (Risch & Merikangas 1996). During the first half of the 2000s, candidate gene association studies on common diseases were popular, but there was little success when replicating the results in other studies. A 2002 review of candidate gene studies, found that of the 600 genetic variants associated with diseases that were reported, only six were consistently replicated in other studies (Hirschhorn *et al.* 2002).

The sequencing of the human genome was vital for GWAS, as was related technology to capture genetic variations. In the early 2000s, there was a move away from the design of tandem repeats that were typically used in linkage studies, to single nucleotide polymorphisms (SNPs) (Gray *et al.* 2000). High throughput genotyping of large SNP collections became commercially available and economically feasible during the last decade. Sequencing of the whole genome was started in 1990 and draft versions of the human genome were published some 11 years later in 2001 (Lander *et al.* 2001; Venter *et al.* 2001), and was finally completed in 2003. The sequence of three billion chemical base pairs that make up the human genome were identified by the Human Genome Project and included identification of some 20,000 genes (US Department of Energy Genome Programs 2011). The Human Genome Project in collaboration with the SNP consortium in 2001 identified 1.42 million SNPs throughout the genome (Sachidanandam *et al.* 2001). The SNP data were placed in the public domain, and today data on some 14 million SNPs identified in the human genome are freely available (National Center for Biotechnology Information & National Library of Medicine 2011).

Another significant project in the public domain essential for conducting GWAS is the International HapMap project (International HapMap Consortium 2003), initiated in 2002. The HapMap project set out to catalogue the genetic commonality and differences across and within populations. This was done by identifying sequences of SNPs that are inherited together within a population (haplotypes), which may vary across populations. Linkage disequilibrium (Thomas 2004b) is defined as any correlation beyond what is expected by chance, between groups of SNPs that typically reside nearby each other on a chromosome. Due to LD, not all SNPs need to be genotyped in a region since carefully selected SNPs (referred to as tag SNPs) will represent the region's genetic variation.

The first GWAS were published in 2005 (Seng & Seng 2008) and by 2007 several key GWAS emerged including "Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls" (Wellcome Trust Case Control Consortium 2007b) and "A common variant of HMGA2 is associated with adult and childhood height in the general population" (Weedon *et al.* 2007). The GWAS for height was a significant publication for several reasons. Human height is a model polygenic trait and heritably is as high as 90%, but prior to GWAS there had been limited success in identifying common genetic determinants. Furthermore, height is a quantitative trait rather than a disease, and case/control studies were the design used in most of the GWAS at that time (McCarthy *et al.* 2008). The early GWAS have undergone iterations where sample size has increased, and new loci related to the condition of interest have been found; the recent GWAS of 180,000 subjects found 180 loci associated with height (Lango *et al.* 2010).

GWAS are free from any underlying hypothesis, as polymorphisms across the genome are statistically tested for an association with a trait or disease regardless of prior knowledge (Hirschhorn & Daly 2005). Unlike candidate gene studies, the function of the polymorphism might be unknown or unrelated to the trait, and it is this open approach that has led to an exponential growth of consistent and reproducible SNP associations with common diseases. The design of a GWAS is critical for its success, and is typically done in four stages (Pearson & Manolio 2008). The first stage is the identification of a sufficiently large homogeneous study (or studies) that is representative of a certain population, and that can be genotyped. Consideration needs also to be given to the outcome of interest, whether it is a disease or a quantitative trait, as many GWAS are a collaborative effort and the outcome needs to be sufficiently comparable across studies. The second stage is genotyping, which is followed by rigorous DNA quality control procedures so that the genotypes reported are reliable and repeatable. Statistical imputation of genetic variants follows quality control, so that results of statistical analyses from multiple studies, that may have genotyped different SNPs, can be meta-analysed together for maximum power. The third stage is the discovery phase of the statistical analysis, where a common analysis plan is run across collaborating studies and results are then combined together through meta-analyses. A pre-defined significance threshold is applied to the discovery stage SNP associations and those smaller than the threshold are taken forward for replication. Replication is the final stage, where the analyses of the genetic variants are repeated in independent studies/samples to confirm findings (McCarthy *et al.* 2008; Moonesinghe *et al.* 2007). It is also possible to examine the function of the genetic variants to confirm their functional feasibility with the outcome, however statistical replication is essential.

The remainder of the chapter is divided into two parts. The first part describes the data used in the GWAS from the 1958BC including the number of participants, (which were covered with greater depth in Chapter 2). The earlier stages of study selection, DNA quality control and imputation of missing SNP data are dealt in greater depth in Chapter 2. The statistical methods used in the discovery and replication stages of the GWAS are also described in the Methods section. The second part presents and discusses the results of the analyses.

4.2 Methods

4.2.1 Participants

The discovery stage of the GWAS included the two genetic 1958BC sub-studies of WTCCC (*n*=1480) and T1DGC (*n*=2530). In the replication stage for the purposes of work presented in this chapter, the WTCCC2 data were analysed as *in-silico* replication (*n*=1386) exclusive of subjects in WTCCC. SNPs taken to the final stage of GWAS replication were de-novo genotyped in the participants of the 1958BC, and for this work it included additional samples (*n*=1807) that were not genotyped in the genome-wide sub-studies. The main analyses were restricted to those participants with full data on 25(OH)D concentrations and body mass index (BMI). The number of participants used in the analysis from the sub-samples of WTCCC, T1DGC, WTCCC2 and participants only de-novo genotyped, were *n*=1447, *n*=2390, *n*=1310 and *n*=1470 subjects respectively.

4.2.2 Procedures

The genome-wide data of the three sub-studies, WTCCC, WTCCC2 and TIDGC, used the gene-chips Affymetrix 500K, Affymetrix 6.0 and Illumina 550K Infinium respectively (Barrett *et al.* 2009; Sawcer *et al.* 2011; Wellcome Trust Case Control Consortium 2007b). De-novo genotyping was done by JDRF/WT Diabetes and Inflammation Laboratory, Cambridge, UK (Cooper *et al.* 2011) using the Taqman platform (Applied Biosystems, Warrington, UK).

Additional details of the genotyping methods including quality control procedures and imputation procedure, and laboratory assays used in the 1958BC are in Chapter 2.

4.2.3 Statistical methods

Power calculations

The power of the GWAS study was calculated with a technique used for simple linear regression (Hsieh *et al.* 1998). In a simple linear model, $y_i = \beta_0 + \beta_1 x_i + \varepsilon$, with individuals $i = 1,..,n$ and $\varepsilon \sim N(0,\sigma)$, the correlation coefficient ρ is related to the regression parameter by $\beta_1 = \rho\,\sigma_y / \sigma_x$, where σ_y and σ_x are the standard deviations of outcome *y* and exposure *x*. A SNP single exposure (*x*) with alleles *A* and *B* in HWE and has an allele frequency of *q*, then the genotype distribution (*AA, AB, BB*) has frequencies of (q^2, $2q(1-q)$, $(1-q)^2$). If the SNP penetrance is additive with the quantitative trait then the genotypes *AA, AB, BB* are coded as 0, 1, 2; and the exposure has mean $\overline{x} = 2-2q$ and variance $\sigma_x^2 = \dfrac{2nq(1-q)}{n-1}$ for a sample size of *n*. For testing $H_0 : \rho = 0$ against $H_A : \rho = r$, the sample size required to achieve power of $1-\beta$ with a significance level α is:

$$n = \frac{\left(Z_{1-\alpha/2} + Z_{1-\beta}\right)^2}{\left[1/2 \ln\left((1+r)/(1-r)\right)\right]^2} + 3$$
,

which includes the Fisher transformation for the correlation (Cohen 1988b). For the 1958BC discovery stage, with a total sample size *n*=3837 and significance threshold of $\alpha = 5\times10^{-8}$, there was 80% power to detect an additive SNP (with minor allele

frequency (MAF) of 0.01) association that explains 1.03% of the variation of 25(OH)D concentrations, and 90% power to detect a SNP association that explains 1.18% of 25(OH)D variation (Figure 4.1). In comparison, the SUNLIGHT consortium analysis at the discovery stage had greater power to detect smaller SNP associations with 25(OH)D concentrations, as it included five cohorts with an overall sample size of *n*=16125, so had 80% power to detect an additive SNP association that explained 0.28% variation of 25(OH)D concentrations (Wang *et al.* 2010b).

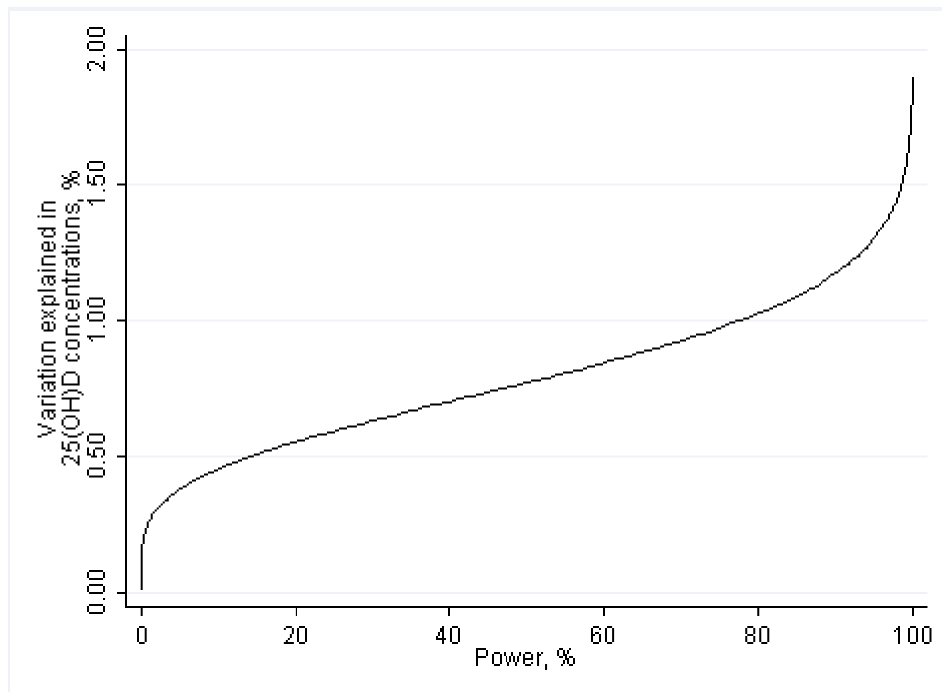


Figure 4.1 The variation explained in 25(OH)D concentration as a function of the power to detect an association of a SNP, which has a MAF 0.01, for the 1958BC discovery samples (WTCCC and T1DGC) with *n*=3,837 and significance threshold $\alpha = 5\times10^{-8}$.

Analysis plan

In the genome-wide analysis, the standardised residuals of 25(OH)D were used as the outcome in the model. The 25(OH)D concentrations were slightly left-skewed

and transforming with the natural log achieved a more normal distribution (as seen in Chapter 3). The standardised residuals came from a linear regression model with natural log transformed 25(OH)D concentrations that adjusted for sex, age, BMI and season. Season was coded as categorical variable with categories: summer (July-September); autumn (October-December); winter (January-March); and spring (April-June). The seasonal coding used in the SUNLIGHT analysis reflected the seasons of the studies across the Northern Hemisphere. The decision to use standardised residuals for the outcome was also part of the SUNLIGHT analysis, as at the time of the collaboration some of the studies were using genetic statistical software where it was not possible to adjust for covariates. The penetrance function of the SNPs in the analysis was assumed to be additive, that is, for each copy of the risk allele there is an equal increased effect on 25(OH)D concentrations. The SNP genotypes of *AA, AB, BB* are coded as 0, 1, 2 for an additive penetrance, which is equivalent to testing for trend across genotypes (van der Sluis & Posthuma 2008).

For the 1958BC genome-wide analysis, the statistical software used was `SNPtest v1.1.5` (Marchini *et al.* 2007), that complements the genotype probability distribution from the imputation program `Impute`. The analysis was run using a frequentist approach to account for the uncertainty in the imputed SNP's genotype probability distribution, so the log likelihood incorporated the degree of missingness with the standardised residuals of 25(OH)D concentrations. For each SNP association there is a measure of relative statistical information referred to as "proper info" with range 0-1, 1 being complete information and 0 no information. The SNPtest program is a UNIX -based command line application and scripts were written to run the analyses in batches by chromosome. Depending on the hardware

and operating system available, the chromosome batches can run at the same time to reduce computational time (which can be several hours for a large chromosome and days for a complete analysis).

The results of the 1958BC discovery samples were checked for discrepancies and errors that may have occurred during the genome-wide analysis, prior to meta-analysing results together. Specifically, the histogram and extreme values of SNP coefficients, standard errors and *p*-values were visually inspected. A cut-off value of 0.4 is applied to the statistical score "proper info" to remove the imputed SNPs with most uncertain genotype calls. Also excluded were SNPs whose MAF is less than 0.0,1 as the power to detect an association is limited and the estimates from the statistical software tend to be unreliable (Marchini *et al.* 2007; Sullivan & Purcell 2008). To reduce false-positive findings, the *p*-values are divided by the genomic control λ, which approximates the inflation that has occurred due to population stratification within a study (Evans 2008). The genomic control is estimated from the median of the test statistic *Y* of the SNPs ($i=1,...,s$) trend test, and squared is assumed to be distributed as χ_1^2. An estimate of λ is;

$$\hat{\lambda}=\left[\text{median}\left(Y_1^2,...,Y_s^2\right)/0.456\right],$$

where the divisor of 0.456 is the median of the χ_1^2 distribution, since the test statistics *Y*'s are assumed to be normally distributed.

Meta-analysis

To meta-analyse the results of some 2.3 million SNP associations, a weighted *Z*-score approach was taken as implemented in the genetic statistical software, `METAL`

(Abecasis 2009). Here, a two-sided *p*-value is transformed to a *Z*-score with sign of direction for the effect allele of the SNP. The *Z*-scores are weighted by the square root of the sample size in the study, so $w_i = \sqrt{n_i}$ for each study $i = 1,..,k$ and then aggregated. A statistic to test the null hypothesis of no effect, also known as Stouffer's *Z* score (Borenstein *et al.* 2009b), follows a standard normal distribution and is calculated as:

$$Z = \frac{\sum_{i=1}^{k} Z_i w_i}{\sqrt{\sum_{i=1}^{k} w_i^2}},$$

and generates corresponding two-sided *p*-value. The SNPs taken forward to replication are selected on the basis of their association's *p*-value with the standardised residuals of 25(OH)D and whether similar "hits" have occurred nearby. The threshold used for the genome-wide significance was *p*-value$<5\times10^{-8}$ (Dudbridge & Gusnanto 2008; Pe'er *et al.* 2006), which comes from a Bonferroni-type correction for the number of independent tests (or tests for genes) across the genome (Dudbridge & Gusnanto 2008).

Quantile-quantile (QQ) plots were used to assess how the observed *p*-values deviate from their expected values of their plot position. This is done by ranking the observed *p*-values and plotting $-\log(p)$ vs $-\log\left(i/(n+1)\right)$ where *n* is the number of observations. If the *p*-values in the top-right of the plot deviate strongly from their expected values it is taken as indication of a possible "hit", so evidence against the null hypothesis of no association (Balding 2006).

The degree of LD in the loci associated with 25(OH)D was assessed by LD plots that showed correlation amongst the SNPs as drawn by using Haploview v4.2 (Wang *et al.* 2005) with CEU population (v2 release 24).

The size of the relationship between SNP and 25(OH)D concentrations was assessed by using a linear regression model with ln 25(OH)D adjusted for age, sex, BMI and season. The coefficients from the models were reported as percentages, since the outcome was transformed by the natural logarithm and can be interpreted as a fractional difference on the original scale (Cole 2000). The variation explained in 25(OH)D concentrations by the SNP was calculated from the difference of the adjusted coefficient of determination (R^2). The overall mean variation explained by the SNP in 25(OH)D concentrations across studies was weighted by the sample size of the study. The coefficient and standard error of the SNP association with 25(OH)D were meta-analysed using the inverse-variance method (Borenstein *et al.* 2009a; Willer *et al.* 2010), which weights the coefficient by the inverse of its variance (v), so $w_i = 1/v_i$ for each study $i = 1,..,k$. The meta-analyzed coefficient calculated as is:

$$\hat{\beta} = \sum_{i=1}^{k} \beta_i w_i \bigg/ \sum_{i=1}^{k} w_i ,$$

and its standard error is:

$$se\left(\hat{\beta}\right) = \sqrt{1 \bigg/ \sum_{i=1}^{k} w_i} ,$$

can be used to give a Z score and corresponding two-sided p-value.

Allele score

To observe how the size of SNP association changed with thresholds of 25(OH)D concentrations, dichotomous outcomes were created at the deficient level of 25(OH)D < 25 nmol/l, the proposed sufficiency threshold of < 50 nmol/l, and the higher proposed sufficiency threshold of < 75 nmol/l. The 25(OH)D thresholds of 25 nmol/l, 50 nmol/l and 75 nmol/l were used as outcomes in a logistic regression models adjusted for age, sex and BMI. A weighted allele score was created to assess the influence of multiple common genetic variants with the risk of vitamin D deficiency/insufficiency. The allele score aggregated SNPs weighted by their effect size and rescaled to a relative score:

$$\text{Allele Score} = \frac{w_1 \times \text{SNP}_1 + w_2 \times \text{SNP}_2 + ... + w_s \times \text{SNP}_s}{2\sum_{i=1}^{s} w_i},$$

where *s* is the number of SNPs (as described in (Lin *et al.* 2009)). The allele score was used in the logistic regression models to assess the combined allele risk with thresholds of 25(OH)D from deficiency through to sufficiency.

Candidate gene analysis

Genetic variants within candidate genes were also investigated for an association with the standardised residuals of 25(OH)D, since it is feasible that some of the genes expressed in the metabolism pathway of vitamin D and regulation of its active form may influence its status. At the discovery stage, polymorphisms within 30 kb of the candidate genes were taken through to replication if the meta-analysed discovery results had *p*-value<0.001. The candidate genes were vitamin D receptor (*VDR*), retinoid X receptor α and β (*RXRA* and *RXRB*), 1α-hydoxylase (*CYP27B1*),

24-hydroxylase (*CYP24A1*), and 27-hydroxylase and 25-hydroxylase (*CYP27A1*). Regional LD plots for the candidate gene regions were drawn using the web-based SNAP program at the Broad Institute (Johnson *et al.* 2008), to elucidate the level of significance of the SNP associations with 25(OH)D in the 1958BC.

The results presented in this chapter are mostly from the analysis of the available 1958BC genetic sub-samples. The results of the SUNLIGHT consortium analysis have been included where necessary to clarify the SUNLIGHT consortium findings.

4.3 Results

The mean 25(OH)D concentrations across the 1958BC genetic sub-samples ranged from 50.7 nmol/L (95% CI 49.4, 52.0) to 54.9 nmol/L (95% CI 53.7, 56.1) (Table 4.1). The sexes were evenly distributed across the sub-samples. The proportion of participants who had blood taken in the winter (Jan-Mar) was different between the discovery sub-samples of WTCCC and T1DGC (8.4% vs 24.9%). However, a reasonable number of participants had blood taken in each season across the sub-samples. The BMI distribution of the participants appeared similar across the sub-samples, and 65.7% of participants had a BMI above 25 kg/m^2 (above normal), with 41.2% of the participants falling into BMI overweight category of 25-30 kg/m^2.

Table 4.1 Characteristics of 1958BC genetic sub-samples with 25(OH)D concentrations and covariates used in genome-wide analysis

	Discovery		Replication	
	WTCCC	T1DGC	WTCCC2 in-silico*	De-novo genotyping
n†	1447	2390	1310	1470
25(OH)D				
geometric mean nmol/l (95% CI)	54.9 (53.7, 56.1)	51.0 (50.0, 52.0)	50.7 (49.4, 52.0)	54.1 (52.8, 55.4)
< 25 nmol/l, %	4.5	9.3	8.3	7.0
< 50 nmol/l, %	35.6	44.0	44.6	39.3
< 75 nmol/l, %	76.4	78.3	79.8	73.8
Women, %	49.6	51.8	47.0	50.9
Age years, mean	44.9	45.3	45.2	45.3
Season				
Jan-Mar, %	8.4	24.9	26.8	20.3
Apr-Jun, %	22.8	23.6	23.4	15.8
Jul-Sep, %	32.5	23.7	24.4	29.1
Oct-Dec, %	36.4	27.7	25.3	34.8
BMI				
< 20 kg/m^2, %	2.8	2.4	2.6	3.3
20-24.9 kg/m^2, %	32.1	30.9	32.5	31.2
25-29.9 kg/m^2, %	39.9	41.8	42.5	42.6
≥30 kg/m^2, %	25.2	24.9	22.4	22.9

* The sub-sample of WTCCC2 excludes participants in the WTCCC sample

† Excludes missing data for 25(OH)D concentrations and BMI

4.3.1 Meta-analyses results

From the meta-analysis of the 1958BC discovery samples, the lowest observed *p*-values of the SNP associations (with standardised residuals of 25(OH)D) deviated strongly from their expected *p*-values in the QQ-plot (Figure 4.2). The genomic control for the WTCCC results was 0.990 and for the T1DGC results was 1.000, so no adjustment was made for controlling population stratification in either result set. Of the 1958BC samples analysed at the discovery stage, 11 polymorphisms had a *p*-value $< 5 \times 10^{-8}$ with all of the SNPs located in 4p12, as seen by peak of SNPs in that region (Figure 4.3). This region is near to the *GC* gene that encodes the vitamin D binding protein. The association of SNP rs2298850 had the lowest *p*-value $=2.05 \times 10^{-14}$ of all the SNPs with the standardised residuals of 25(OH)D in the 1958BC (Table 4.2).

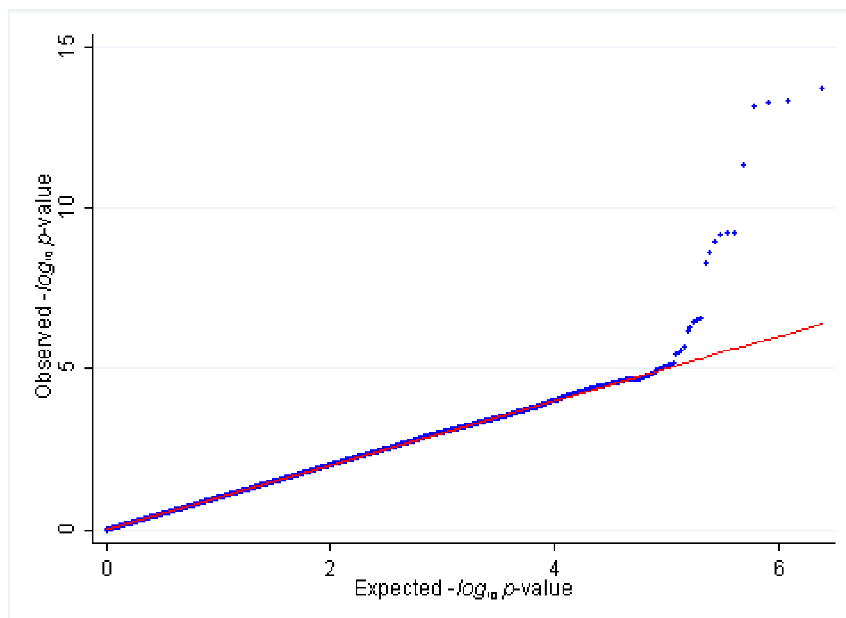


Figure 4.2 QQ Plot of the *Z*-score meta-analysed 1958BC sub-samples (WTCCC and T1DGC) SNP associations with standardised residuals of 25(OH)D. The blue dots are the observed log *p*-values and red line is their expected log values.

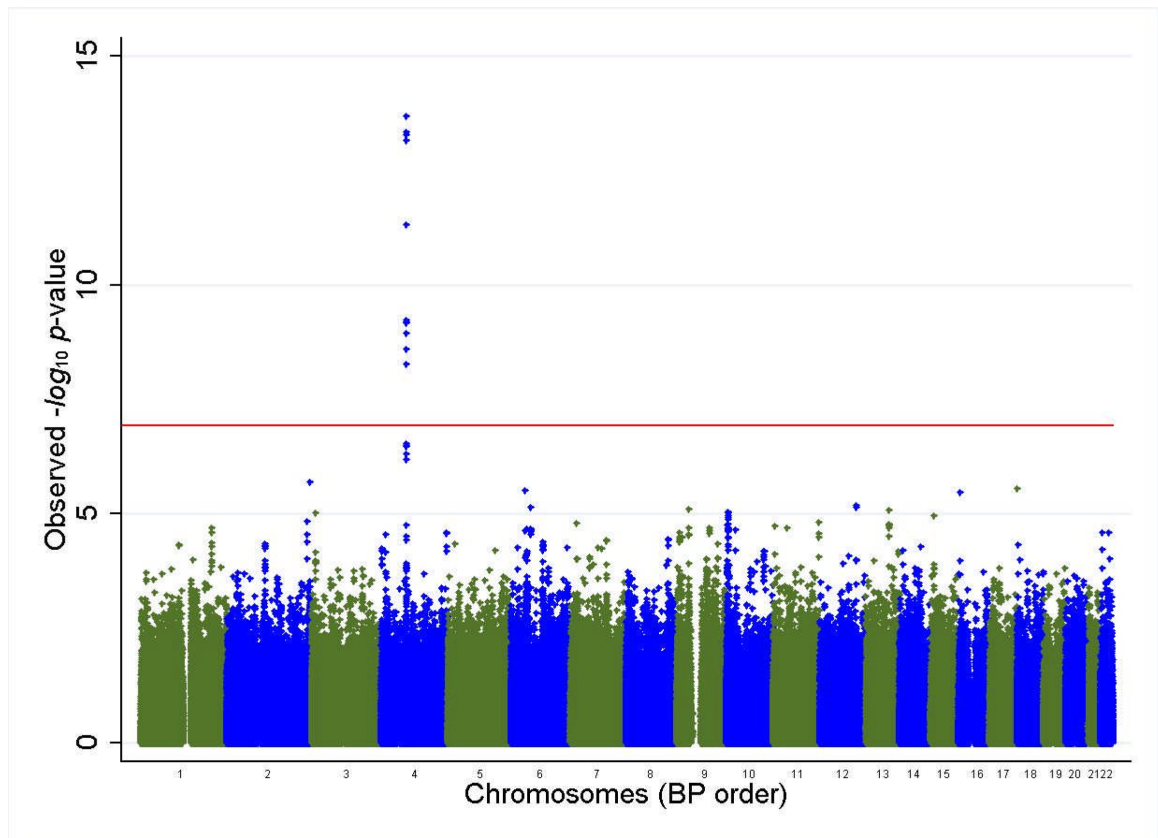


Figure 4.3 Manhattan plot of the observed *p*-values from *Z*-score meta-analysed 1958BC sub-samples (WTCCC and T1DGC) over chromosomes (alternating blue/green shading). The *y*-line in red is *p*-value= 5×10^{-8}.

Of the SUNLIGHT consortium results that included 12288 additional samples from another four cohorts, there was much greater deviation of the observed *p*-values from the expected *p*-values, than seen in the QQ-Plot of the 1958BC results (Figure 4.4). As with the 1958BC results, the same locus of *GC* the association of SNP rs2282679 had the smallest *p*-value=4.57×10^{-63} of all the SNPs. In the SUNLIGHT consortium meta-analysis, regions of 11q12 and 11p15 also had SNP associations reaching genome-wide significance at *p*-value<5×10^{-8}. In the region 11q12, the associated locus was near the genes *DHCR7* and *NADSYN1* that encode 7-dehydrocholesterol reductase and nicotinamide adenine dinucleotide synthetase, respectively. Near the associated locus in region 11p15, was gene *CYP2R1*, which is a member of the

cytochrome P450 superfamily of enzymes. Although the SNPs' associations of the 1958BC discovery samples did not reach GWA significance at *p*-value < 5×10^{-8}, all of the top six SNP associations in these loci were conventionally significant at *p*-value < 0.05 (Table 4.2). In the 1958BC replication analysis, the associations of the top six SNPs from the loci of *GC* and *DHCR7*/*NADSYN1* were successfully replicated with 25(OH)D at *p*-value < 0.05. For the polymorphisms near the *CYP2R1* locus, two SNPs had additional 1958BC samples from de-novo genotyping. The associations of the two *CYP2R1* SNPs (rs12794714 and rs10741657) with 25(OH)D were successfully confirmed at *p*-value< 0.05.

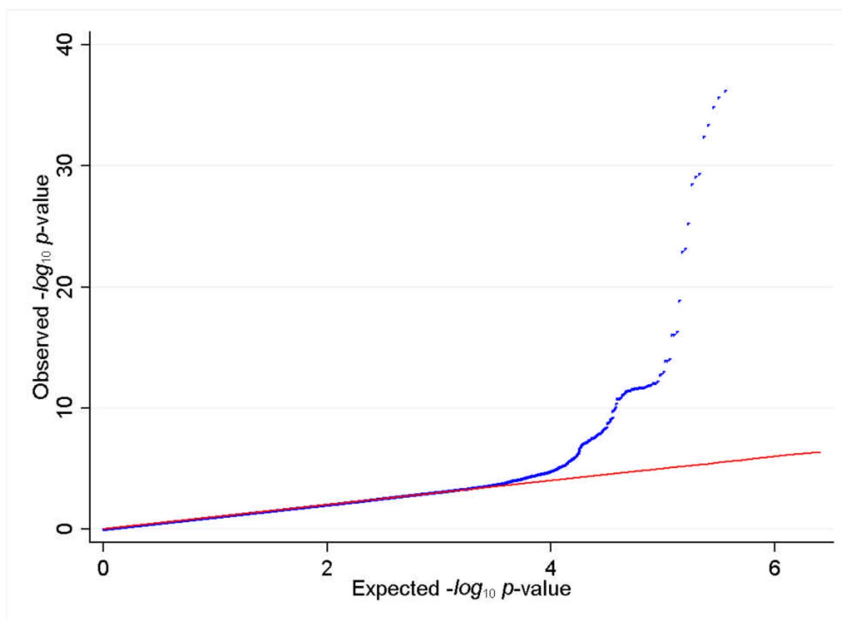


Figure 4.4 QQ Plot of the *Z*-score meta-analysed SUNLIGHT discovery samples SNP associations with 25(OH)D standardised residuals. The blue dots are the observed log *p*-values and the red line is their expected log values

Table 4.2 SNPs identified in the SUNLIGHT consortium GWAS with the results from the 1958BC

					1958BC results						
SNP*	Chr	Position	Nearest gene(s)	Hapmap MAF†	MAF	*p*-value for discovery	*p*-value for replication	Hetero-geneity *p*-value	*n*	Combined *p*-value	SUNLIGHT combined *p*-value
rs17467825	4	72824381	*GC*	0.30	0.29	5.15×10^{-14}	4.53×10^{-7}	0.64	5147	1.50×10^{-19}	6.75×10^{-74}
rs2282679**	4	72827247	*GC*	0.29	0.29	6.88×10^{-14}	9.10×10^{-15}	0.48	6445	9.14×10^{-27}	1.90×10^{-109}
rs3755967	4	72828262	*GC*	0.30	0.30	4.49×10^{-14}	5.61×10^{-7}	0.64	5146	1.57×10^{-19}	2.42×10^{-75}
rs2298850	4	72833131	*GC*	0.27	0.29	2.05×10^{-14}	3.36×10^{-7}	0.74	5147	4.44×10^{-20}	2.03×10^{-71}
rs7041	4	72837198	*GC*	0.42	0.44	6.35×10^{-10}	1.02×10^{-9}	0.30	6288	5.59×10^{-18}	6.31×10^{-59}
rs1155563	4	72862352	*GC*	0.32	0.31	4.71×10^{-12}	4.46×10^{-6}	0.89	5145	1.19×10^{-16}	2.37×10^{-73}
rs7944926	11	70843273	*DHCR7/ NADSYN1*	0.27	0.21	8.14×10^{-3}	7.61×10^{-3}	0.60	5147	2.84×10^{-4}	8.96×10^{-16}
rs12785878**	11	70845097	*DHCR7/ NADSYN1*	0.27	0.21	1.00×10^{-2}	2.89×10^{-5}	0.46	6589	3.03×10^{-6}	2.12×10^{-27}
rs4944957	11	70845683	*DHCR7/ NADSYN1*	0.27	0.21	9.92×10^{-3}	8.94×10^{-3}	0.58	5147	3.92×10^{-4}	8.70×10^{-15}
rs12800438	11	70848651	*DHCR7/ NADSYN1*	0.29	0.21	1.03×10^{-2}	7.71×10^{-3}	0.56	5147	3.72×10^{-4}	2.54×10^{-15}
rs3794060	11	70865327	*DHCR7/ NADSYN1*	0.27	0.21	7.93×10^{-3}	9.03×10^{-3}	0.62	5147	3.07×10^{-4}	3.38×10^{-15}

rs4945008	11	70898896	*DHCR7/ NADSYN1*	0.28	0.22	6.83×10^{-3}	1.98×10^{-2}	0.73	5147	4.47×10^{-4}	4.55×10^{-15}
rs7116978	11	14838347	*CYP2R1*	0.34	0.37	1.38×10^{-2}	0.20	0.77	5147	1.30×10^{-2}	4.99×10^{-9}
rs1993116	11	14866810	*CYP2R1*	0.37	0.40	5.28×10^{-3}	0.19	0.78	5147	2.19×10^{-3}	6.25×10^{-11}
rs10500804	11	14866849	*CYP2R1*	0.43	0.43	1.36×10^{-2}	0.19	0.92	5147	5.19×10^{-3}	2.67×10^{-9}
rs12794714	11	14870151	*CYP2R1*	0.40	0.43	1.38×10^{-2}	4.05×10^{-3}	0.64	6430	1.94×10^{-4}	1.84×10^{-9}
rs10741657**	11	14871454	*CYP2R1*	0.37	0.40	6.22×10^{-3}	4.97×10^{-2}	0.91	6430	7.82×10^{-4}	3.27×10^{-20}
rs2060793	11	14871886	*CYP2R1*	0.37	0.40	1.32×10^{-2}	0.19	0.68	5145	5.14×10^{-3}	1.73×10^{-11}

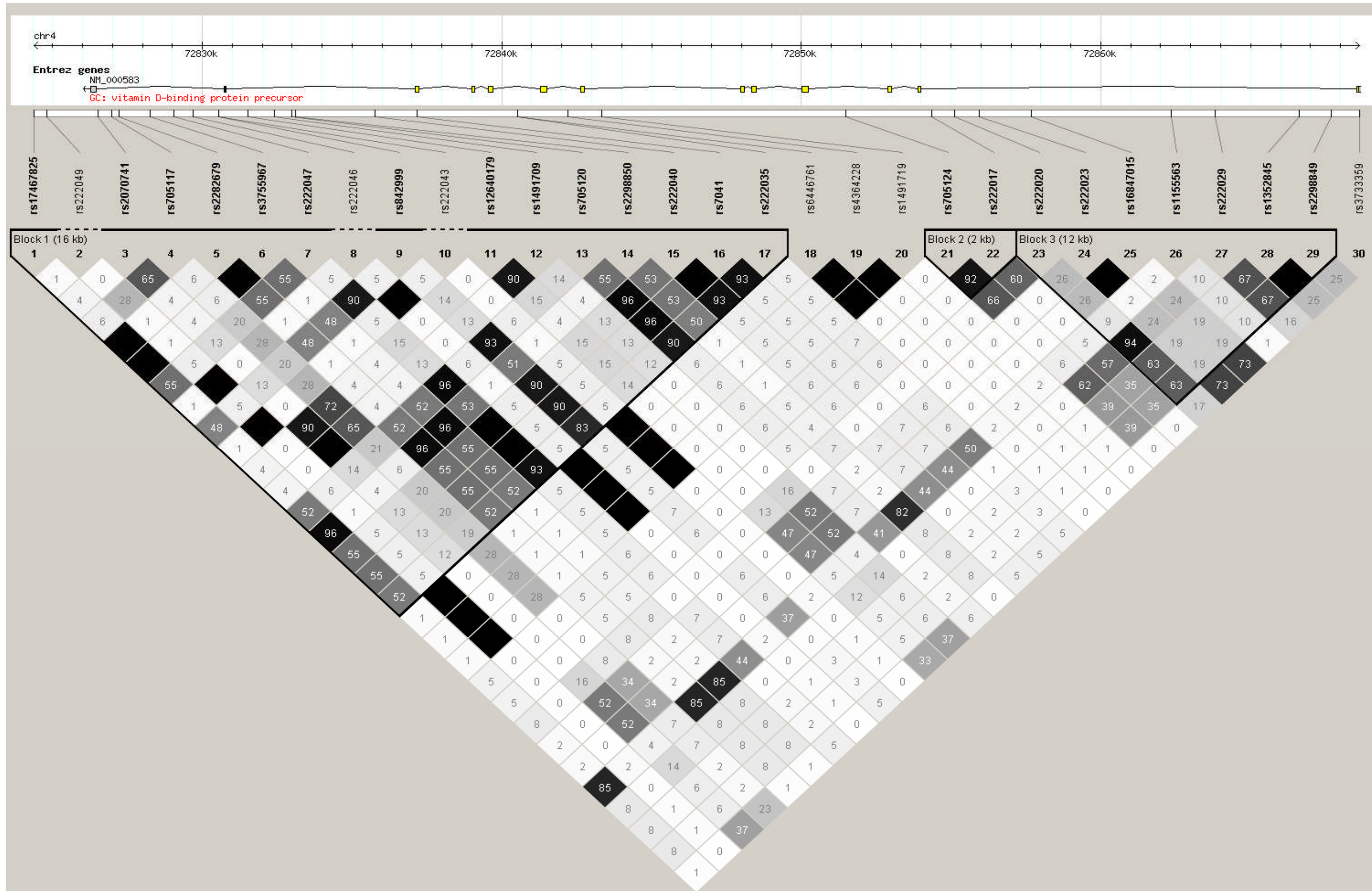
* The SNPs that were identified in the SUNLIGHT consortium GWAS that were successfully replicated (Wang *et al.* 2010b)

† Hapmap build 36, release 24/phase 11 for the CEU population

** SNPs that had the lowest *p*-value overall in the SUNLIGHT consortium GWAS

The LD structures of the regions near the genes *GC*, *CYP2R1* and *DHCR7/NADSYN1* were assessed using Haploview. Using the technique of (Gabriel *et al.* 2002), three LD blocks were estimated in the 4kb region of *GC*, with five out of the six top SUNLIGHT SNP hits found in the first block (Figure 4.5 A). The first LD block of *GC* contained the two non-synonymous SNPs, rs7041 and rs4588. The SNP rs4588 was not included in the Hapmap dataset at the time of imputation, but is in high LD with several of the top *GC* SNPs found. The top SUNLIGHT *GC* SNPs did include the other non-synonymous missense SNP rs7041, which alters the amino acid sequence from aspartic to glutamic. The five out of the six top *GC* SNPs from the SUNLIGHT results are located in the intron region of the gene, except for SNP rs17467825, which is downstream of the gene. In 38kb region for *DHCR7*/*NADSYN1* (Figure 4.5 B), all of the six top SNPs were located in the intron region of *NADSYN1*. However, one LD block was estimated for the region that included both genes *DHCR7* and *NADSYN1*, and contained the non-synonymous SNP, rs2276360 (amino acid sequence change from valine to leucine), that did not feature in the top six SUNLIGHT SNPs. The top SNP rs12785878 for *DHCR7* was in perfect LD (R^2=1.0) with the non-synonymous SNP rs2276360. For *CYP2R1* (Figure 4.5 C), in the 14kb region one LD block was estimated that had all of the top six SUNLIGHT SNPs contributing towards the haplotype. Five out of the six top SNPs for *CYP2R1* were within the intron region of the gene, with the exception of synonymous SNP rs12794714.

A



B

C

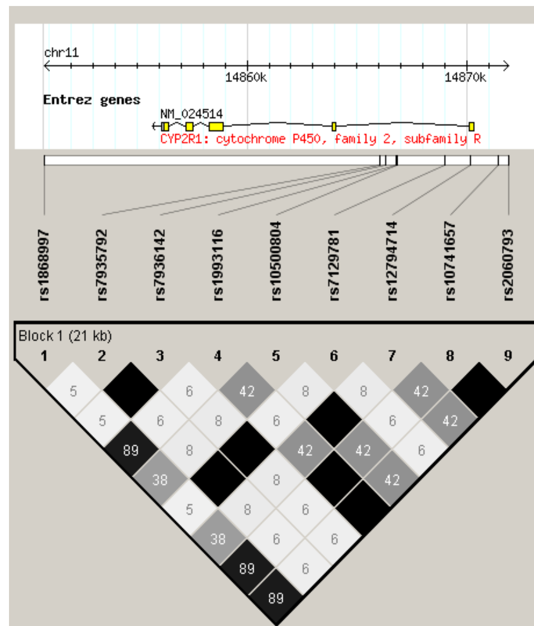


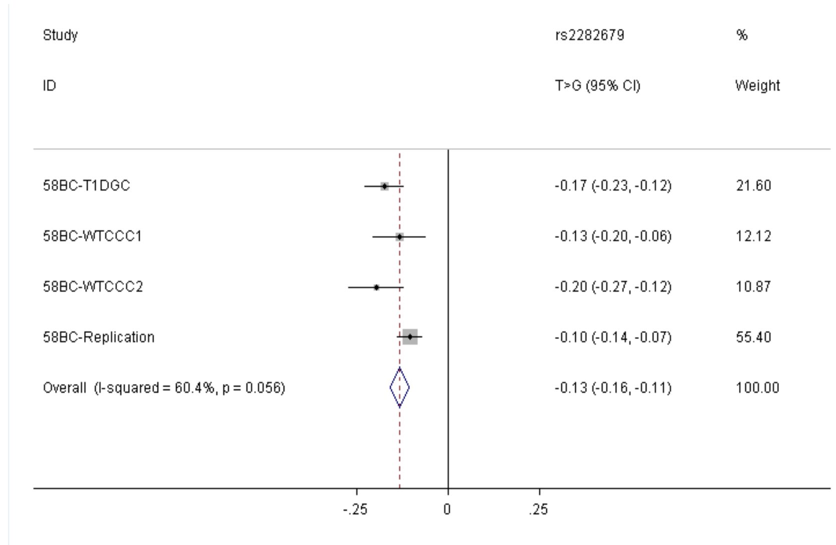
Figure 4.5 LD blocks of SNPs for regions A) region 4q12 near *GC*, B) region 11q12 near *DHCR7/NADSYN1* and C) region 11p15 near *CYP2R1*. The numbers within the blocks are the R^2 between SNPs, and the darkness of the block indicates the degree that the SNPs are LD, i.e. black is high LD and white no LD. The black lines grouping SNPs show the LD block. Dashed lines below the SNP indicate that the SNP is not necessary for the block's haplotype.

4.3.2 Associations of SNPs with 25(OH)D and 25(OH)D deficiency/insufficiency in the 1958BC

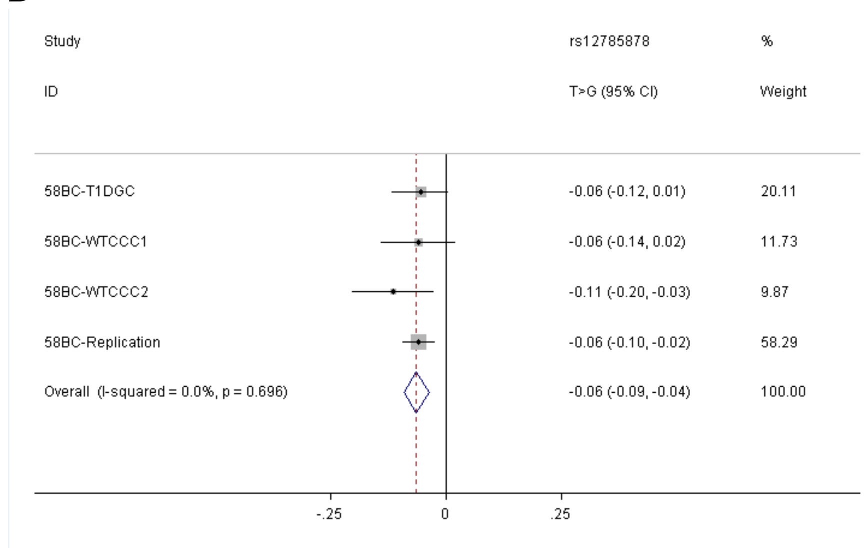
The *GC*, *CYP2R1* and *DHCR7* SNPs that had the lowest *p*-values of the meta-analysed 1958BC discovery and replication sub-samples, were re-analysed to assess their size of association with 25(OH)D after adjusting for age, sex, BMI and season. For *GC* SNP rs2298850 and *DHCR7* SNP rs12785878, the minor alleles were associated with lower 25(OH)D concentrations, compared with the major alleles (Figure 4.6 A,B). For *CYP2R1* SNP rs10741657, it was the major allele that was associated with lower 25(OH)D concentrations; participants carrying allele G had 4% (95% CI 2, 6, *p*-value=8.1×10^{-4}) lower 25(OH)D concentrations, compared with participants carrying the minor allele A (Figure 4.6 C). For the nearby non-synonymous (functional) *CYP2R1* SNP rs12794714, the effect direction of the minor allele was reversed, and it was participants carrying the minor allele A that had 5% (95% CI 2, 7, *p*-value =6.6×10^{-5}) lower 25(OH)D concentrations, compared with participants carrying the major allele G. The difference in effect direction of the two *CYP2R1* SNPs was somewhat surprising given that they were in moderate LD (R^2=0.42).

The risk allele of *GC* SNP rs2282679 had the largest association size with 25(OH)D at -13% (95% CI -16, -11 *p*-value=9.9×10^{-25}), compared with two SNPs from *DHCR7* and *CYP2R1* (Figure 4.6). The *GC* SNP rs2298850 explained the most variation in 25(OH)D concentrations of 2.00% out of the top SNPs found; SNPs rs12785878 and rs10741657 explained less of the variation at 0.71% and 0.65%, respectively.

A



B



C

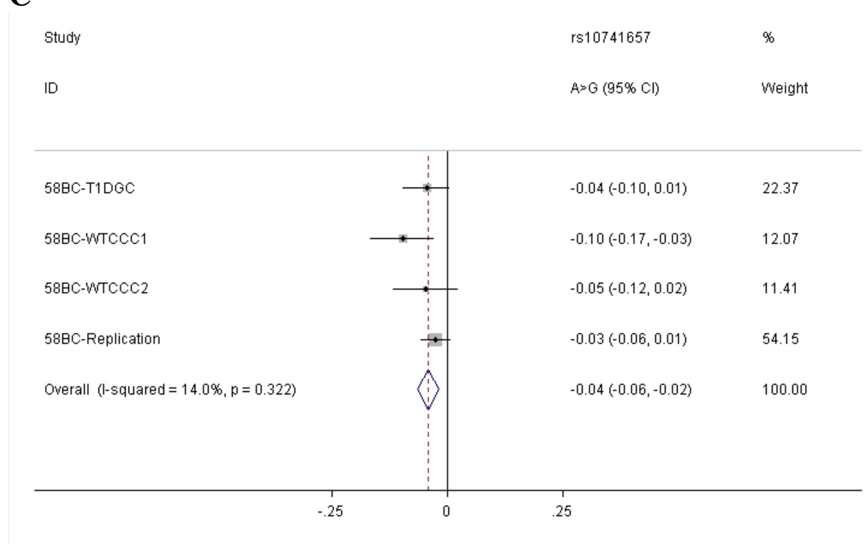


Figure 4.6 The top SUNLIGHT consortium SNPS for *GC*, *DHCR7/NADSYN1* and *CYP2R1* meta-analysis results of the 1958BC samples.

The prevalence risk of vitamin D deficiency/insufficiency was assessed in participants carrying the risk alleles of SNPs from the *GC*, *DHCR7/NADSYN1* and *CYP2R1* (Figure 4.7). For the risk of 25(OH)D deficiency (< 25 nmol/l), only carriers of the *GC* SNP (rs4588) risk allele had 47% (95% OR 1.26, 1.72, *p*-value = 1.27×10^{-6}) increased risk, compared with the other two variants for *CYP2R1* and *DHCR7*, where there was no evidence of risk (*p*-value >0.14). However, for the risk of 25(OH)D insufficiency (< 75 nmo/L), there was evidence of an increased risk per copy of any of risk alleles from the three variants of *GC*, *CYP2R1*, *DHCR7* (*p*-value$\leq3.07\times10^{-6}$ for all comparisons).

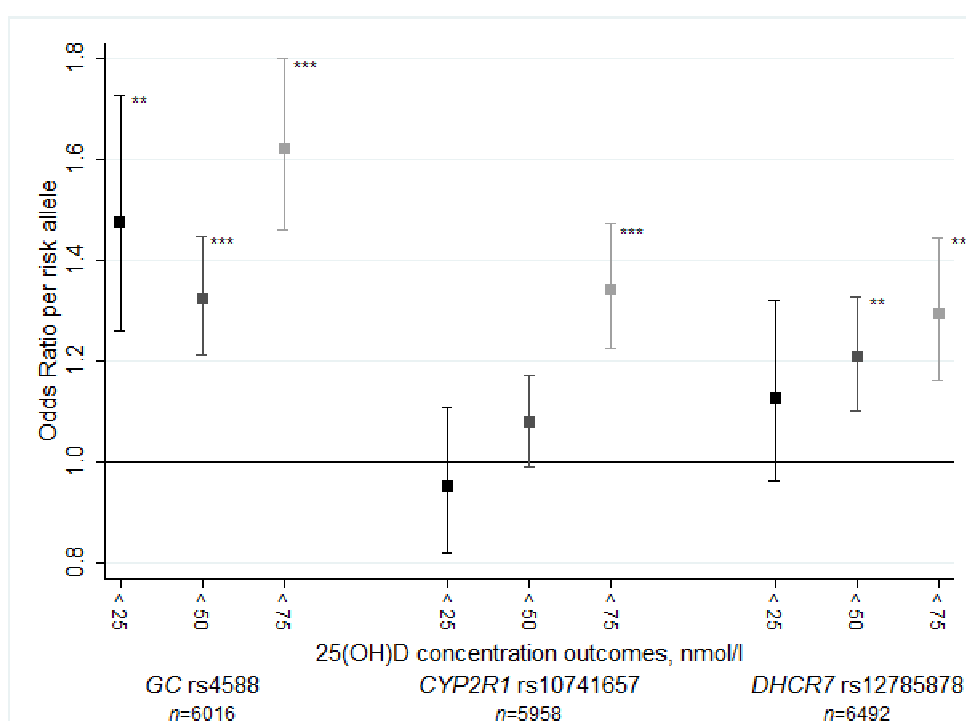


Figure 4.7 The *GC*, *CYPR21* and *DHCR7* SNP OR with thresholds of 25(OH)D at < 25 nmol/l, < 50 nmol/l and < 75 nmol/l (dark grey to light grey) in the 1958BC. All logistic regression models were adjusted for age, sex, BMI and season. The 95%

CI are indicated by the extended bars. * indicates the OR has p-value < 0.05 and > 5×10^{-8}, ** p-value $\leq 5\times10^{-8}$ and > 5×10^{-10}, *** p-value $\leq 5\times10^{-10}$.

There was an increased risk of 25(OH)D deficiency/insufficiency among participants at the top 25% of the genotype score combined from three variants of *GC*, *DHCR7/NADSYN1* and *CYP2R1*, compared with participants at the bottom 25% of the genotype score, after adjusting for age, sex, season and BMI (Figure 4.8). The risk of 25(OH)D insufficiency (< 50 nmol/L) was the same as for the risk of 25(OH)D deficiency (< 25 nmol/L) for participants in the top quarter of score, compared with participants in the bottom quarter. For the threshold of 25(OH)D insufficiency (< 75 nmol/l), the risk was nearly 3-fold higher among participants in the top quarter of the genotype score [OR 2.80 (95% CI 2.32, 3.38, trend p-value $=7.5\times10^{-27}$)], compared with participants in the bottom quarter. The risk of 25(OH)D insufficiency (< 75 nmol/L) was much larger than that for the lower thresholds of 25(OH)D at < 50 nmol/L and 25 nmol/L.

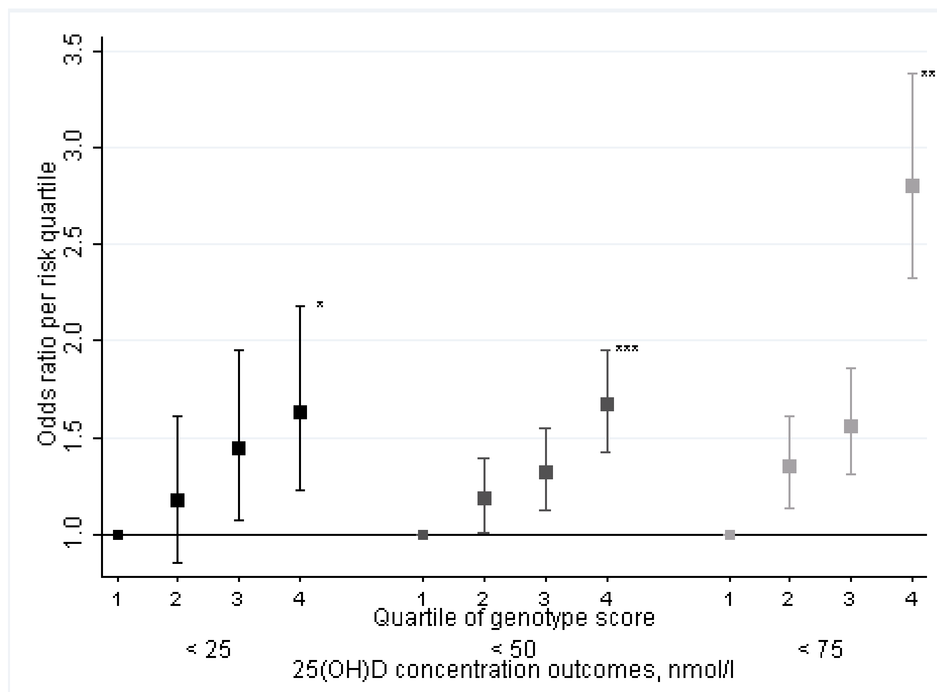


Figure 4.8 The genotype score odds ratio with thresholds of 25(OH)D at < 25 nmol/l, < 50 nmol/l and < 75 nmol/l (dark grey to light grey) in the 1958BC. All logistic regression models were adjusted for age, sex, BMI and season. The 95% CI are indicated by the extended bars. * indicates the OR has p-value < 0.05 and > 5×10^{-8}, ** p-value ≤ 5×10^{-8} and > 5×10^{-10}, *** p-value ≤ 5×10^{-10}

4.3.3 Candidate gene results

Of the candidate gene analysis of the 1958BC discovery samples, none of the SNPs associations, except for those near *GC*, reached the pre-specified threshold at p-value< 1×10^{-3}. The associations of polymorphisms near genes *CYP24A1*, *CYP27B1* and *VDR* showed conventional significance of p-value < 0.05 (Appendix 3) with the standardised residuals of 25(OH)D. In the SUNLIGHT consortium results, SNP rs6013897 near the *CYP24A1* locus, was associated with standardised residuals of 25(OH)D (p-value =6.0×10^{-10}) after meta-analysing the discovery and replication samples together. The SNP rs6013897 was de-novo genotyped in 6522 samples of the 1958BC, and participants carrying the minor allele A had 2% (95% CI 0, 4, p-value=0.023) lower concentrations of 25(OH)D, compared with participants carrying the major allele T.

4.4 Discussion

Common polymorphisms from the vitamin D binding protein gene (*GC*) were found to be associated with 25(OH)D and the risk of vitamin D insufficiency in the 1958BC. The top *GC* SNP rs2282679 of the GWAS is common in the Europeans, with 30% of the population carrying the risk allele. The common loci found in the SUNLIGHT consortium GWAS could assist in identifying groups in the population susceptible to vitamin D insufficiency. These findings may also lead to the identification of novel causal mechanisms in vitamin D deficiency and elucidate its role in other diseases, such as cardiovascular disease.

A comprehensive review of the genetic variants found to be associated with 25(OH)D concentrations from the SUNLIGHT GWAS and other recent genetic studies is in Chapter 5. The following discussion will focus primarily on methodological considerations of GWAS.

The success of GWAS has come from its hypothesis free approach, that is testing genome-wide genetic variants regardless of function or prior knowledge with the trait (Hirschhorn & Daly 2005). This approach incorporates a multi-stage design and includes rigorous DNA quality control, genetic imputation, and discovery and replication statistical analysis. The identification of common polymorphisms by GWAS has been prolific, some 4300 polymorphisms have been identified in 800 GWAS since 2008 (Hindoff *et al.* 2011). Polymorphisms found in the first waves of GWAS were common (MAF ≥ 5%), had small to modest effect size with complex traits (Manolio *et al.* 2008; Wellcome Trust Case Control Consortium 2007b), and followed the “common variant to common disease” model. The model assumes

multiple common variants have an additive and/or multiplicative effect with the disease, as suggested in a meta-analysis of genetic associations prior to the advent of GWAS (Lohmueller *et al.* 2003). The *GC* variants found in this work had the largest association and explained the most variation in 25(OH)D concentrations, compared with the other loci found. However, the *GC* association with 25(OH)D is still modest compared with the familial heritability estimates of 28% and upwards (Hunter *et al.* 2001; Shea *et al.* 2009; Wjst *et al.* 2007).

Typically in GWAS, variants found to be associated with the trait are unlikely to be the underlying causal variant, but are in LD with the causal variant (McCarthy *et al.* 2008). SNP rs2282679 in the *GC* locus is in high LD (R^2=0.973) with the non-synonymous SNP rs4588, which was not covered by the gene-chips (or in the HapMap 2 CEU population, that was used for the reference population in the imputation analysis). The other non-synonymous SNP rs7041 in the *GC* locus, that is commonly studied with rs4588, was also highlighted in the results as its association achieved low *p*-values with 25(OH)D. The allele combinations of rs4588 and rs7041 form the *GC* haplotypes of *GC1F*, *GC1S* and *GC2* (Braun *et al.* 1992; Fang *et al.* 2009); where each respective haplotype varies by a single amino acid (Speeckaert *et al.* 2006). It has been argued that non-synonymous polymorphisms are likely to be causal variants, due to the potential impact an amino acid substitution has on gene expression (Botstein & Risch 2003; Jegga *et al.* 2007). However, as findings have accumulated from GWAS it has been discovered that variants other than non-synonymous polymorphisms can influence a common disease (Hirschhorn & Gajdos 2011). In a single locus, there might be more than one underlying causal variant that has an independent risk with the disease, a combination of common variants with

weak effects and rare variants with large effects, or a different type of mutation (i.e. deletion) (Altshuler *et al.* 2008). In this study, LD blocks were estimated for the identified loci to demonstrate the genetic variation that occurs within inherited blocks, which may expand across multiple genes - as seen with *DHCR7* and *NADSYN1*. The start and end location of the blocks estimated were based on the location of the genes (Gabriel *et al.* 2002), but can expand over a larger area of the chromosome than shown here (Price *et al.* 2008). Nevertheless, it is biologically feasible that the genes found were associated with 25(OH)D concentrations (as expanded in Chapter 5). However, as demonstrated by the two polymorphisms near the *CYP2R1* locus, the minor alleles associated with 25(OH)D were in opposite directions, despite moderate LD of the SNPs. Therefore, much can still be learnt from further investigation of the found regions, for instance, through fine mapping and functional translation of the variants.

In the analysis of the allele risk score that combined the top *GC*, *CYP2R1* and *CYP24A1* variants, there was little difference in the odds ratios of the 25(OH)D thresholds at 25 nmol/l and 50 nmol/l of participants who carried the most risk compared with those who carried the least. This was somewhat surprising as an increasing trend of risk by increasing thresholds of 25(OH)D concentrations was expected. Since each allele of the SNPs used in the score indicated an increased risk with increasing thresholds of 25(OH)D concentrations, with the exception of *GC*. There are several possible reasons for the static risk over vitamin D status thresholds. It might be due to rarity of the prevalence of severe vitamin D deficiency, overall in the 1958BC sub-samples 7.5% had a vitamin D status of less than 25 nmol/l. As noted, the polymorphism for *GC* had a higher allele risk associated with 25(OH)D <

25 nmol/l than that with 25(OH)D < 50 nmol/l, which may have also changed the risk profile of the allele score over the thresholds of vitamin D status. As part of the SUNLIGHT consortium analysis, it was found that the *GC* polymorphism rs2282679 risk allele for lower 25(OH)D concentrations, was also associated with lower vitamin D binding protein concentrations in the Twins UK cohort (Wang *et al.* 2010b). The *GC* haplotypes in the order of most to least common in European populations are *GC1S, GC2* and *GC1F* (Fang *et al.* 2009). The haplotype *GC1S* contains the combination of the major alleles from the two SNPs. Vitamin D binding protein has been demonstrated to effect the bio-availability of 25(OH)D to monocyctes (Chun *et al.* 2010). From the same study, it was also found that human serum with the *GC2* haplotypes show a higher induction of cathelicidin by 25(OH)D compared with the serum with GC1F haplotype (Chun *et al.* 2010). The differences in response to the *GC* genetic variation might be a form of genetic canalisation (Debat & David 2001), that dampens the risk allele's associated risk with 25(OH)D at certain thresholds, where biological mechanisms have been developed to attempt to keep vitamin D homoeostasis.

Although the analyses in this work were done solely with the 1958BC data, population stratification may still bias the results. In this study, genomic control was used to control for the over-inflation of results due to underlying population stratification, which was a null adjustment for both sub-studies. The use of genomic control does have issues (Price *et al.* 2006; Tian *et al.* 2008). An uniform inflation statistic adjusts all SNP associations as oppose to adjusting individual variants according to their allele frequency differences, hence some variants maybe under adjusted whilst other might be over adjusted (Price *et al.* 2006; Tian *et al.* 2008). As

part of the WTCCC "Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls" study, the WTCCC 1958BC sub-sample (as used in this analysis) was compared against the other control sample (UK Blood service) to investigate population sub-structure differences (Wellcome Trust Case Control Consortium 2007b). For the control samples in the WTCCC study, it was concluded that "population structure has at most a small confounding effect" and did not adjust for population structure. Given the decision made in the WTCCC publication, the use of the genomic control statistic should suffice to control over-inflation of the results presented in this chapter.

In the broader context of the SUNLIGHT consortium, the results from the studies were meta-analysed together at the discovery and replication stages with a *Z*-score weighted approach. Heterogeneity of results across studies can be an issue, which may be caused by differences in the definition of the trait (Zeggini & Ioannidis 2009). Techniques for measuring 25(OH)D concentrations are known to produce substantial differences (Carter 2009). To address the reliability of 25(OH)D measures, the International Vitamin D External Quality Assessment Scheme (DEQAS) was established in 1989 with the aim of "monitor(ing) the performance of individual laboratories" (Carter *et al.* 2010). Furthermore, variability from study-specific 25(OH)D measurement techniques are kept to a minimum as the *Z*-score meta-analysis approach is not scale-dependent.

In the SUNLIGHT consortium results, two additional loci from the *CYP2R1* and *DHCR7*/*NADSYN1* regions were identified using the genome-wide approach, and *CYP24A1* locus was identified using a candidate gene approach. The additional

samples confirmed that sample size is critical to GWAS for identifying true associations (Wang *et al.* 2005). The loci of *CYP2R1*, *DHCR7* and *CYP24A1* were conventionally significant (p-value<0.05) in the discovery analysis of the 1958BC, however so were numerous other loci. The level of significance used in this study was $\alpha<5\times10^{-8}$, which has come from the Bonferroni correction to adjust for a million independent tests (Risch & Merikangas 1996). This threshold is strict compared with other thresholds used in GWAS (Wang *et al.* 2005; Wellcome Trust Case Control Consortium 2007b). Alternative procedures for ranking p-values are Bayesian approaches, False Discovery Rate or permutation methods, but these provide equally stringent significance thresholds as those used here (Dudbridge & Gusnanto 2008) However, the consensus is that only low p-values are evidence of an association and should be followed-up by replication in additional samples (McCarthy *et al.* 2008; Moonesinghe *et al.* 2007). The sample size of the discovery and replication analyses presented in this work was relatively low for GWAS and came from the same birth cohort. Therefore, the results might suffer from the "winners' curse", where the results are upwardly biased due to an overestimate of the significance of the associated loci (Garner 2007). However, it is unlikely that the association of the *GC* variants is upwardly biased, since the main GWAS took place within the collaboration of SUNLIGHT that confirmed this association.

Further study is underway to increase the sample size of the GWAS with nutritional vitamin D status under the SUNLIGHT consortium with the primary aim of discovering novel loci. The analysis strategy has changed to reflect the enhancements of the genetic statistical software, and requires the studies to explicitly adjust for the covariates in the model, rather than the use residuals as the outcome.

Furthermore, in the analysis plan the covariates used to account for seasonal differences in 25(OH)D concentrations is month of measure, which is a more precise adjustment of season than what was used in this work. It also gives the opportunity to study the association sizes of the loci with 25(OH)D concentrations. For the analysis of 1958BC data, the larger genome-wide dataset of the WTCCC2 will be included in the discovery stage. Correction for potential population stratification in the larger genetic dataset will also include using components from multi-dimensional scaling. Further study has also been proposed within the consortium to investigate gene-environment interactions. Initial environmental factors that are being considered for modifying genetic effects are the vitamin D determinants of diet (including vitamin D supplementation) and season of when blood was taken for vitamin D assay.

5 A Mendelian randomisation study on 25-hydroxyvitamin D and inflammation/haemostatic biomarkers in the British1958 birth cohort (Papers III & IV)

5.1 Introduction

In observational studies, the relationship between exposure and disease is commonly affected by a mutual risk factor, which is termed a confounder (Hennekens & Buring 1987a). Undoubtedly, the hypothesised association between 25(OH)D (a nutritional marker for vitamin D status) and sub-clinical cardiovascular disease (CVD) (as measured by the inflammation/haemostatic biomarkers) in participants of the British 1958 birth cohort (1958BC) will be affected by confounding. This occurs despite the comprehensive background information that exists on the 1958BC, as there will always be some uncertainty as to whether the confounders are known, adequately measured and fully taken into consideration (Rothman & Greenland 1998). In Mendelian randomisation (MR) studies, genetic variants that influence the exposure are used as its proxy measure (or instrument) in estimating the relationship between the exposure and outcome (Davey Smith & Ebrahim 2003). Due to the random nature of how organisms are assigned deoxyribonucleic acid (DNA) from both parents via meiosis, it has been proposed that using genetic instruments to mimic the exposure avoids confounding and removes the possibility of reverse causation in cross-sectional studies (Davey Smith & Ebrahim 2005).

The term "Mendelian randomisation" has been coined from the laws first proposed by Gregor Mendel in 1865 after studying inheritance of traits in pea plants (Thomas 2004b). His first law states that a single polygenetic locus has two copies of an allele with one copy inherited from each parent. Furthermore, the inherited alleles are randomly selected from a parent's two alleles. The second law is that variants are independently inherited from one another. The random allocation of genetic variation within the population has meant that MR studies have been described as "Nature's randomised trial" (Hingorani & Humphries 2005) (Figure 5.1). The random allocation of the treatment in a trial is comparable with the allocation of genotypes that influence the level of the exposure, and as such, the confounders are evenly distributed amongst the groups (Davey Smith & Ebrahim 2005). A disease that is yet to develop is unlikely to affect the allocation of genotypes that takes place prior to inception, and implies that the reduced exposure levels precedes the disease in a cross-sectional setting. The lack of confounding that typically affects observational studies, and inferred temporality of the effects has meant that MR can strengthen the evidence for causality (Lawlor *et al.* 2008).

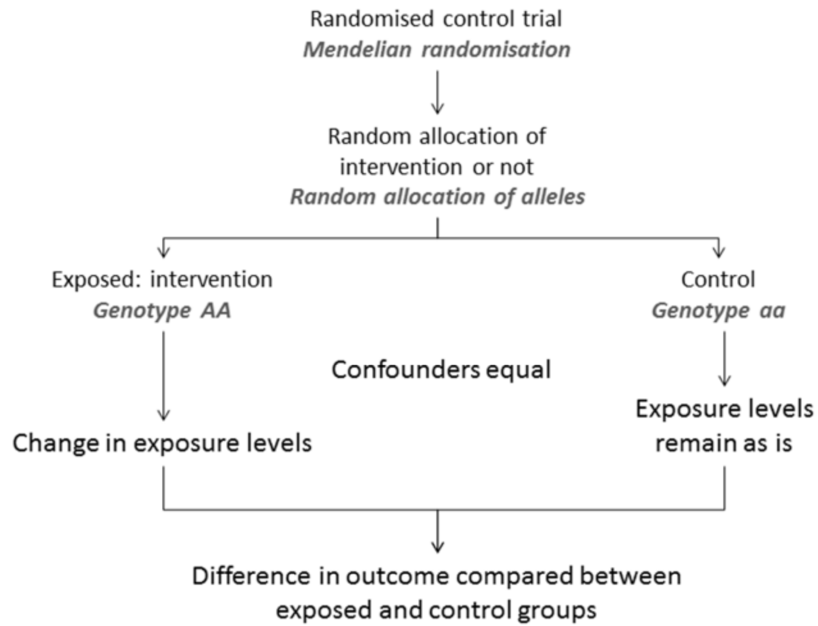


Figure 5.1 The similarities between a randomised control trial and a MR study. Taken from (Davey Smith & Ebrahim 2005).

In the framework of MR, a genetic variant *Z* is associated with the modifiable environmental exposure *X*, and the exposure *X* associated with the outcome ***Y*.** The confounders *U* are associated with both the exposure *X* and the outcome *Y*, but not with the genetic variant Z (Figure 5.2)**.**

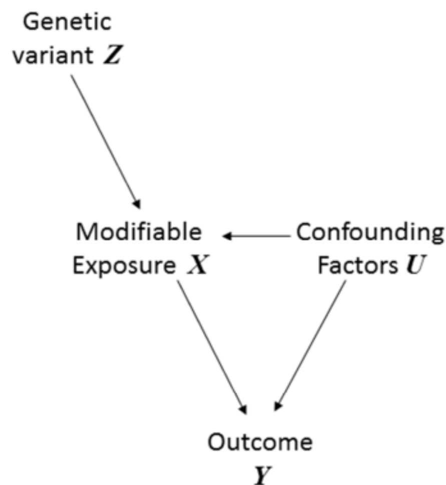


Figure 5.2 The pathways of the MR framework. The genetic variant *Z* is causally associated with exposure *X* and used as *X*'s proxy/instrument. *Z* is not associated with confounders *U* which affect the relationship between outcome *Y* and exposure *X*. Exposure *X* is assumed to be causally associated with outcome *Y*. The figure is adapted from reference (Didelez & Sheehan 2007)

There are three main assumptions in MR studies (Lawlor *et al.* 2008):

1) The genetic variant *Z* is causal in changing levels of the exposure *X*;
2) Exposure *X* is causal of the outcome *Y*. There might be unknown confounders *U* that are causal of both *X* and *Y* affecting their relationship, but the genetic variant *Z* is independent of those confounders *U*;
3) Genetic variant *Z* is only associated with the outcome *Y* through its association with the exposure *X*.

The assumption of null association between the genetic variant *Z* and exposure *X* can be formally tested; however the second and third assumptions cannot (Sheehan *et al.* 2008). A fourth assumption is made when inferring causality:

4) The relationships between the genetic variant *Z*, exposure *X* and outcome *Y* are linear, and so have a dose-response trend or biological gradient (Lawlor *et al.* 2008).

The identification of genetic proxies (instruments) for the exposure is critical to the success of a MR study. Ideally the function of a genetic variant's that alters the levels of the exposure is well known, reliable and singular, and will only affect the outcome through the mechanism of the exposure (Lawlor *et al.* 2008). MR studies can also be affected by ethnic differences within the study (population stratification), but there are methods to identify and control for these differences (Evans 2008). Genetic confounding may affect MR studies as the independent inheritance of genetic variants in close proximity may not hold due to linkage disequilibrium (LD). However, this will only be an issue when LD masks an underlying genetic variant associated with the outcome outside of the relationship with the exposure (Lawlor *et al.* 2008).

The remainder of the chapter is divided into three sections. The first section is a review of genetic determinants of 25-hydroxyvitamin D and lists potential genetic candidates from genome-wide association studies (GWAS) on hair and skin colour and tanning ability. Genetic variants associated with variations of hair, skin colour and ability to tan may also affect vitamin D status due to its dependence on ultraviolet B (UVB) rays from sun exposure. A paper related to the review of genetic determinants of 25(OH)D has been published in a peer-reviewed journal (Berry & Hyppönen 2011) and is shown in Appendix 1. The second section covers the methods and data used, including number of participants used from the 1958BC,

and gives an overview of the biomarkers and lifestyle and socio-economic covariates considered (which are discussed in greater detail in Chapter 2). The statistical methods used to determine the suitability of genetic variants for instruments of 25(OH)D concentrations and the statistical techniques used for the MR study are also described in this section. The third section presents and discusses the results of the analyses and the MR approach. A paper related to the work of evaluating genetic variants for use in MR studies is under review in a journal and is shown in Appendix 1.

Work presented in this Chapter has been expanded on and differs slightly from the work presented in the papers. Single nucleotide polymorphisms (SNP) from the candidate gene loci were tested for an association with 25(OH)D concentrations, since when discussing the plausibility of a variant with a trait, it is the genes that are considered with trait rather than the variant. The protein isoforms of *GC* were also investigated to determine their suitability as instruments for 25(OH)D concentrations, since variants from *GC* were the top hits in the SUNLIGHT consortium GWAS. To control for multiple testing for the SNP association with 25(OH)D concentrations, a False Discovery Rate (FDR) procedure that was adapted for groups was used, since there were SNPs grouped by genes and number of SNPs for each gene varied greatly. This work also included a simulation MR study on the effects of a pleiotropic SNP to gain an understanding of how results from instrumental variable (IV) regression might be distorted. Finally, presented in this chapter is a MR study to investigate the relationship between 25(OH)D concentrations and inflammation/haemostatic biomarkers using the evaluated genetic instruments.

5.2 Genetic determinants of 25-hydroxyvitamin D for instruments in Mendelian randomisation

Recent genetic studies with nutritional vitamin D status were identified by searching in PubMed for articles published within the two previous years (from November 2010) with the following search criteria:

```
("vitamin D" or ergocalciferols or calcifediol or calcidiol or
"25-hydroxyvitamin D" or "25-hydroxyvitamin D" or "25-
hydroxyvitamin D2" or "25-hydroxyvitamin D3") AND
(single or polymorphism or nucleotide or snp or gene or
"genome-wide" or genome or loci or gwas or haplotype )
```

The search identified 1364 articles and, after scrutinising the abstracts, three independent GWAS studies were identified, including the SUNLIGHT consortium publication, 14 candidate gene studies (Table 5.1) and one linkage study. The rest of the articles were not considered as the studies had not evaluated associations between genetic variants and 25(OH)D concentrations. The linkage study of some 1700 related individuals of European ancestry reconfirmed that 25-hydroxyvitamin D was highly heritable at 29% (Shea *et al.* 2009), with earlier estimates from studies being 43% and 80% (Hunter *et al.* 2001; Wjst *et al.* 2007).

Table 5.1 25-hydroxyvitamin D candidate gene studies since 2008

Study description					Results		
Country	**Ethnicity**	**Sex**	**Design**	**Sample Size**	**Gene tested**	**Genetic variant associated with 25(OH)D**	**Author**
USA	Caucasians non-Hispanic	Men	Prostate cancer case-control	749 cases 781 controls	CYP27A1, GC, CYP27B1,VDR, RXRA, RXRB, PPAR, NCOA1, NCOA2, NCO3, SMAD3 (212 snps)	GC rs12512631 rs2282679 rs7041 rs1155563	(Ahn *et al.* 2009)
USA and Sweden	Not stated	Women	Epithelial ovarian cancer case-control	123 controls from NY 189 controls from Sweden	VDR (SNPs and haplotype)	None reached significance	(Arslan *et al.* 2009)
Spain	Not stated	Mixed	Bariatric surgery with secondary hyper-parathyroidism case-control	26 cases 31 controls	VDR (4 polymorphisms into 4 haplotypes)	None reached significance	(Balsa *et al.* 2010)
USA	Caucasians	Mixed	Population based study	156 discovery 340 replication (women only)	ALPL, CYP24A1, CYP27A1, CYP27B1, CYP2R1, CYP3A4,	CYP2R1 rs12794714, rs10741657 rs1562902 rs10766197 GC	(Bu *et al.* 2010)

Study description					Results		
Country	Ethnicity	Sex	Design	Sample Size	Gene tested	Genetic variant associated with 25(OH)D	Author
					GC, VDR, PTH (49 SNPs)	rs222020, rs2298849	
Netherlands	Dutch Caucasians	Mixed	The Rotterdam study, a prospective population-based cohort study aged > 55 years	1317	GC (2 SNPs)	GC rs4588 and rs7041 (Haplotypes)	(Fang *et al.* 2009)
Canada	Not stated	Mostly women (91%)	Crossover trial for vitamin D supplementation	98	GC (2 SNPs)	GC rs4588 rs7041	(Fu *et al.* 2009)
Netherlands	Not stated	Mixed	Case series, subjects had cured differentiated thyroid carcinoma	154	D2 polymorphism Thr92AIa	None reached significance	(Heemstra *et al.* 2010)
USA	Caucasians	Mixed	Urso-deoxycholic (UDCA) randomised control trial	415	VDR (35 SNPs) RXRA (23 SNPs)	None reached significance	(Hibler *et al.* 2010)
UK	European	Mixed	The 1958BC, a prospective birth cohort	6429	CYP27B1 (1 SNP)	CYP27B1 rs10877012	(Hyppönen *et al.* 2009)

Study description					Results		
Country	**Ethnicity**	**Sex**	**Design**	**Sample Size**	**Gene tested**	**Genetic variant associated with 25(OH)D**	**Author**
	Not stated	Mixed	COPD case-control	262 cases 152 control	GC (2 SNPs)	GC rs4588, rs7041	(Janssens *et al.* 2010)
Netherlands	Dutch Caucasians	Mixed	The Rotterdam Study and Longitudinal Aging Study Amsterdam (LASA), both population based cohorts	6367 Rotterdam study 844 LASA	LPH (1 SNP)	None reached significance	(Koek *et al.* 2010)
UK	Gujarati Asians	Mixed	Case series, subjects had TB contacts	36	GC (haplotypes – 2 genotypes)	None reached significance	(Martineau *et al.* 2010)
Germany	German ethnicity	Mixed	Type 1 diabetes case-control	100 cases 100 controls	Cubilin (5 SNPs)	None reached significance	(Ramos-López *et al.* 2010)
Canada	Caucasians (mostly French ancestry)	Women	Cross-sectional premenopausal women	741	GC (2 SNPs)	GC rs4588, rs7041	(Sinotte *et al.* 2009)
Netherlands	Not stated	Mixed	Multiple sclerosis case-control	212 cases 289 controls	VDR (1 SNP)	Fok-I rs10735810	(Smolders *et al.* 2009)

In addition to the SUNLIGHT consortium GWAS, a similar independent GWAS was published on participants of European ancestry (Ahn *et al.* 2010). This consisted of nine cohorts, mainly from the USA, with 4501 participants in the discovery phase and an additional 2221 in the replication analysis. As with the SUNLIGHT meta-analysis, the study identified three SNPs (*DHCR7, CYP2R1* and *GC*), which were all in or near genes encoding steps in the vitamin D metabolism or associated with its transport. The SUNLIGHT meta-analysis also confirmed a variant that is a catalyst for the clearance of 25(OH)D (*CYP24A1*) (Wang *et al.* 2010b). A much smaller third GWAS on a US Hispanic population was published in 2010. This study was based on a discovery sample of 229 and found no associations at the GWAS significance level (Engelman *et al.* 2010). These findings, alongside candidate gene studies, have provided evidence on the following genetic determinants of nutritional vitamin D status.

DHCR7 [7-dehydrocholesterol reductase] expresses a reductase catalysing the conversion of 7-DHC to cholesterol. The *DHCR7* reductase removes the pre-cholesterol from the vitamin D pathway and reduces the substrate of 7-DHC available for conversion to 25(OH)D. In the synthesis of vitamin D, 7-DHC is converted to a pre-cursor of vitamin D3 after the skin is exposed to UVB rays. Mutations in *DHCR7* are known to lead to the rare Smith-Lemli-Opitz syndrome, where impaired activity of the gene leads to accumulation of 7-DHC and a deficiency of cholesterol along with congenital abnormalities and intellectual disabilities (Tint *et al.* 1994). Limited information exists on individuals with the syndrome and whether their vitamin D status is affected (Rossi *et al.* 2005).

However, in earlier animal studies, the use of *DHCR7* inhibitors have led to increased 7-DHC and 25(OH)D concentrations (Bonjour *et al.* 1987; Morris 1999).

CYP2R1 locus was one of the top hits in both GWAS in participants of European ancestry (Ahn *et al.* 2010; Wang *et al.* 2010b) and there was prior evidence for an association with 25(OH)D concentrations from a candidate gene study (Bu *et al.* 2010). *CYP2R1* is a member of the cytochrome P450 superfamily of enzymes, which are the catalysts of numerous reactions in the synthesis of cholesterol, lipids and steroids (Schuster 2010). GWAS' findings confirmed the role of *CYP2R1* as the enzyme primarily responsible for the hepatic hydroxylation of vitamin D by the microsomal enzyme of 25-hydroxylase. Furthermore, a genetic mutation in *CYP2R1* is known to lead to vitamin D deficiency (Cheng *et al.* 2004). It has been demonstrated in animal studies that vitamin D does not accumulate in the liver, and the 25-hydroxylation is primarily regulated by the quantity of available circulating vitamin D (Gascon-Barré 2005).

Four cytochrome P450 forms: *CYP2R1*; *CYP27A1*; *CYP2J;* and *CYP3A4I;* had been considered candidates for the catalyst of hydroxylation in the liver (Gascon-Barré 2005; Schuster 2010). Of these, *CYP2R1* has the highest affinity and specificity to vitamin D, however other CYP genes may contribute to a lesser extent to the 25-hydroxylation (Schuster 2010).

Variants from the gene region ***GC*** **(group-specific component or Gc globulin)**, which encodes the vitamin D binding protein (VDBP), have consistently been associated with 25(OH)D concentrations in candidate gene studies (Ahn *et al.* 2009;

Bu *et al.* 2010; Fang *et al.* 2009; Fu *et al.* 2009; Janssens *et al.* 2010; Sinotte *et al.* 2009) and were the top hits in the two GWAS (Ahn *et al.* 2010; Wang *et al.* 2010b). Most of the circulating 25(OH)D is bound by the protein (83-85%; high affinity), less by albumin (12-15%; low affinity) and 0.04% of 25(OH)D is free (Bikle *et al.* 1986a). In an *in vitro* study, protein isoforms of *GC* were associated with 25(OH)D bioavailability to target cells, that of monocytes, and the cells' subsequent response (Chun *et al.* 2010).

No individuals have been found to be deplete in VDBP, but *GC* null type animals are viable (Safadi *et al.* 1999; Zella *et al.* 2008). Compared with the wild type, the *GC* null animals are more likely to develop secondary hyperparathyroidism and hypocalcaemia when subjected to a vitamin D deficient diet, but they are also likely to be more resistant to hypercalcemia and vitamin D toxicity after being overloaded with vitamin D (Safadi *et al.* 1999). Protein bound vitamin D metabolites have a longer half-life in circulation as they are less susceptible to hydroxylation and degradation (Speeckaert *et al.* 2006). This suggests that an important function for VDBP is to stabilise and maintain circulating concentrations of 25(OH)D and other vitamin D metabolites (Safadi *et al.* 1999). There are three common isoforms in the VDBP, namely *GC1F, GC1S*, and *GC2*. These protein isoforms can be based on the combination of alleles from non-synonymous SNPs rs4588 and rs7041 (Fu *et al.* 2009; Speeckaert *et al.* 2006). The common variants differ by amino acid substitutions and/or by the degree of glycosylation (Braun *et al.* 1992). In populations of European ancestry, either for SNPs rs4588 (or rs2282679, a close proxy with R^2>0.99) and/or rs7041, the minor allele is consistently associated with lower 25(OH)D concentrations (Ahn *et al.* 2009; Fu *et al.* 2009; Janssens *et al.* 2010;

Sinotte *et al.* 2009). Concentrations of the circulating VDBP tend to be lower in carriers with *GC2*, compared to those with *GC1* (Lauridsen *et al.* 2001). It has been proposed that carriers of *GC2* might also be related to a faster metabolism of 25(OH)D (Kawakami *et al.* 1981; Lauridsen *et al.* 2005). There is evidence for ethnic and geographical differences in the *GC* variants, suggesting that the *GC1F* type is a more common isoform in people with dark skin compared with those with pale skin, whilst *GC1S* and *GC2* are more common in those with pale skin compared with dark (Kamboh & Ferrell 1986). This can be confirmed by the genetic information available for the Hapmap populations (International HapMap Consortium 2003). This suggests that, compared with *GC2*, VDBP in people carrying *GC1* may have a higher affinity to vitamin D metabolites (Chun *et al.* 2010; Speeckaert *et al.* 2006), and potentially, that there may be variations in the bioavailability of 25(OH)D between the different ethnicities.

A locus near the ***CYP24A1*** achieved the genome-wide significance threshold (*p*-value$<5\times10^{-8}$) with 25(OH)D concentrations after combining the discovery and replication studies in the SUNLIGHT meta-analysis (Wang *et al.* 2010b). As with *CYP2R1* and *CYP27B1* (next gene), *CYP241* is a member of the cytochrome P450 superfamily. The 24-hydroxylation is crucial in the regulation of the concentrations of the active 1,25$(OH)_2$D hormone in renal and in extra-renal sites. It is also known to degrade 25(OH)D into an inactive compound that can be excreted from the system (Omdahl & May 2005). Within the kidney, the primary regulators of the 24-hydroxylation are the parathyroid hormone (PTH) and the active metabolite of 1,25$(OH)_2$D; in a normal state increased PTH concentrations will block expression of *CYP24A1*, whilst increased 1,25$(OH)_2$D will activate it (Horst *et al.* 2005).

Calcium deficient animals have elevated 1,25$(OH)_2$D but very limited *CYP24A1* activity (Horst *et al.* 2005), and it is suspected that PTH is the primary mediator in a calcium deficient state (Reinhardt & Horst 1990). The expression of *CYP24A1* has been detected in most of the active vitamin D target cells and tissues (Schuster 2010). *CYP24A1* is considered as a drug target, and inhibitors for the enzyme are under development (Schuster 2010). In the future, these might be used on chronic kidney disease patients who are susceptible to vitamin D deficiency by blocking the 24-hydroxylation breakdown (Schuster 2010).

Other vitamin D related genetic variations have been considered in candidate gene studies. A large population based cohort with 6288 participants (Hyppönen *et al.* 2009) has reported an association between the gene coding 1α-hydroxylase ***CYP27B1*** and 25(OH)D. Additional evidence of an association has been seen in a smaller case-control study for gestational diabetes mellitus for the same polymorphism (Ramos-López *et al.* 2008) and in a twin multiple sclerosis study for a different polymorphism (Orton *et al.* 2008). However, there was no genome-wide significant *CYP27B1* locus association in the SUNLIGHT candidate analysis (Wang *et al.* 2010b) or in two other studies (Ahn *et al.* 2009; Bu *et al.* 2010), where it was analyzed with 25(OH)D. It is biologically plausible for *CYP27B1* to influence 25(OH)D concentrations as it converts 25(OH)D to its active hormone, 1,25$(OH)_2$D. However, given the tight regulation of circulating 1,25$(OH)_2$D concentrations and 1α-hydroxylation, it is unsurprising that its association with 25(OH)D is fairly inconsistent. The principal renal activator of 1α-hydroxylase is PTH, acting in its regulatory role for calcium homeostasis (DeLuca 2004). Rare mutations in *CYP27B1* prevent the renal hydroxylation of 25(OH)D, causing a deficiency in

$1{,}25(OH)_2D$, a disorder known as vitamin D-dependent rickets type 1 (Glorieux & St Arnaud 2005).

The **vitamin D receptor (*VDR*)** mediates the hormonal actions of $1{,}25(OH)_2D$ and is the classical gene for vitamin D metabolism; until recently, most vitamin D relevant genetic studies were focused on *VDR*. There has been interest in assessing *VDR* – disease associations and recent meta-analyses suggest links between selected *VDR* variants and the risk of fractures (Ji *et al.* 2010), rheumatoid arthritis (Lee *et al.* 2011), tuberculosis (Gao *et al.* 2010) and various types of cancer (Chen *et al.* 2009; Kostner *et al.* 2009; Raimondi *et al.* 2009; Randerson-Moor *et al.* 2009; Tang *et al.* 2009; Yin *et al.* 2009). Given the distance of *VDR* from 25(OH)D in the metabolic pathway, strong associations between them would not be expected. Evidence for *VDR* association with 25(OH)D concentrations is weak (Arslan *et al.* 2009; Balsa *et al.* 2010; Bu *et al.* 2010; Hibler *et al.* 2010), although some evidence has been found in small multiple sclerosis studies. In a case-control study based in the Netherlands, the VDR polymorphism Fok-1 was associated with 25(OH)D in season, case and control stratified analysis of 289 controls and 212 cases (Smolders *et al.* 2009). Similarly, in a Canadian twin-based study of 99 twin pairs, again a Fok-1 polymorphism association persisted after controlling for multiple testing of genetic markers (Orton *et al.* 2008). Additionally, no association has been found for the heterodimer partner of *VDR*, *RXR* (*RXRA* and *RXRB*), with 25(OH)D (Ahn *et al.* 2009; Hibler *et al.* 2010).

5.2.1 Candidate genes from hair and skin colour, tanning GWAS for instruments of 25(OH)D

The majority of vitamin D is sourced through sun exposure (Holick 1994), and pigmentation of skin and hair colour can affect the level of tolerable UVB exposure (Andreassi *et al.* 1987; Cummings *et al.* 1997). It has been hypothesised that human pigmentation has partially evolved to optimise levels of vitamin D (Jablonski & Chaplin 2010). Findings from three GWAS on hair and skin colour and tanning ability have found associations in seven loci (Table 5.2).

Table 5.2 Genetic loci associated with skin, hair colour and tanning ability in GWAS

Gene	Region	SNP	Phenotype	Author
MATP/ SLC45A2	5p13.2	rs16891982	Black vs blonde hair Tanning ability Skin colour	(Han *et al.* 2008)
		rs28777	Black vs blonde hair Black vs red hair	(Han *et al.* 2008)
IRF4 / EXOC2	6p25-p23	rs1540771	Freckles	(Sulem *et al.* 2007)
		rs12203592	Black vs blonde hair Skin colour Tanning ability	(Han *et al.* 2008)
		rs12210050	Tanning ability	(Nan *et al.* 2009)
MC1R	16q24.3	rs1805008	Red vs non red hair Blond vs brown hair Skin sensitivity to sun Freckles	(Sulem *et al.* 2007)
		rs258322	Black vs. red hair	(Han *et al.* 2008)
		rs11648785	Tanning ability	(Nan *et al.* 2009)
		rs464349	Tanning ability	(Nan *et al.* 2009)
TYR	11q14-q21	rs1042602	Freckles	(Sulem *et al.* 2007)
		rs1393350	Skin sensitivity to sun	(Sulem *et al.* 2007)
		rs1393350	Tanning ability	(Nan *et al.* 2009)
KITLG	12q22	rs12821256	Blond vs brown hair	(Sulem *et al.* 2007)
SLC24A4	14q32.12	rs12896399	Blond vs brown hair	(Sulem *et al.* 2007)
		rs12896399	Black vs blond hair	(Han *et al.* 2008)
OCA2/ HERC2	15q12-q13	rs1667394	Blond vs brown hair	(Sulem *et al.* 2007)
		rs11855019	Black vs blond hair Black vs red hair	(Han *et al.* 2008)
		rs12913832	Hair colour Tanning ability	(Han *et al.* 2008)

5.2.2 Summary of genetic variants considered as instruments for 25(OH)D in MR

The genetic variants highlighted in the two reviews were taken forward as potential instruments for vitamin D status in a MR study. For the purpose of learning in this thesis, the strategy to select SNPs to take forward as potential instruments differed slightly from paper IV. Excluding SNPs highlighted in GWAS on 25(OH)D, all SNPs from candidate genes were tested for an association with 25(OH)D. Multiple testing of the candidate gene SNPs with 25(OH)D was corrected for by a false discovery rate procedure. Haplotypes from *GC* were estimated from the two functional *GC* SNPs, and the association of the haplotypes with 25(OH)D was assessed. The genetic variants underwent statistical analysis to assess their suitability as instruments. Formal MR analysis was applied by IV regression using the suitable genetic instruments for 25(OH)D with the outcomes of the inflammatory/haemostatic biomarkers.

5.3 *Methods*

5.3.1 Participants

The number of participants from the 1958BC Biomedical survey analysed to assess the suitability of the instruments were: WTCCC2 (*n*=2703), T1DGC (*n*=2530) and the additional de novo genotyped (*n*=1807). These analyses were restricted to participants with full data on 25(OH)D concentrations, and reduced the number in the sub-studies to *n*=2598 in WTCCC2, *n*=2390 in T1DGC and *n*=1470 in the de-novo genotyped samples.

The formal MR analyses were done using IV regression for the outcomes of C-reactive protein (CRP), D-dimer, fibrinogen, tissue plasminogen activator (tPA) and von Willebrand factor (vWF). The analyses were further restricted to participants with any data on the outcomes and full data on the covariates of body mass index (BMI) at age 45 years and geographical region of residence (*n*=5898). Since the focus was on individuals free from CVD, subjects who had reported taking hypertensive medication and/or lipid regulating medication (*n*=401) were excluded from the IV regression analyses. The final analyses were performed on at most 5497 participants who had information on at least one SNP or more.

5.3.2 Procedures

A full description on the laboratory techniques used for the serum biomarkers, procedures to measure lung function, height, waist circumference and weight from the biomedical survey, when the cohort members were aged 44-45 years, are in Chapter 2. Genome-data was used from the two sub-studies of WTCCC2 and

TIDGC, genotyped on chips Affymetrix 6.0 and Illumina 550K Infinium and respectively (Barrett *et al.* 2009; Sawcer *et al.* 2011). De-novo genotyping was done by JDRF/WT Diabetes and Inflammation Laboratory, Cambridge, UK using the Taqman platform (Applied Biosystems, Warrington, UK).

The serum biomarkers used as outcomes or to assess the suitability of the instruments were 25(OH)D, CRP, D-dimer, fibrinogen, tPA, vWF, IGF-1, Glycosylated haemoglobin (HbA1c), triglycerides, total and HDL cholesterol and IgE. BMI was calculated from height and weight measurements at age 45 years. Other measurements used in this chapter were forced expiratory volume 1 (FEV_1) and blood pressure.

5.3.3 Lifestyle and socio-economic covariates

A detailed description of the lifestyle and socio-economic covariates used in these analyses are in Chapter 2. The lifestyle and socio-economic variables were used to assess whether the instruments (SNPs) associated with 25(OH)D were mediated through them, or whether the instrument was directly associated with them. The following covariates used in these analyses recorded at age 44-45 years were: physical activity as coded by recreation metabolic equivalent of task (MET) hours; alcohol consumption; time spent watching a television and/or using a PC; time spent outside; use of sun cover; oily fish consumption; use of vitamin D supplements; and geographical region. Variables used in these analyses taken from earlier 1958BC sweeps were socio-economic class and smoking at age 42 years.

The variables were dichotomised when used as outcomes by collapsing into the following categories: time spent watching a television/using a PC (coded as <1 h vs ≥ 1 h); time spent outside (coded as <1 h vs ≥ 1 h); protecting skin when in the sun (often/sometimes vs rarely/never); oily fish consumption (≥3 days/ week vs. < 3 days/ week); use of vitamin D supplement (yes/no); and season of blood drawn (winter/spring vs. summer/autumn).

5.3.4 Selection of SNPs, candidate genes and genotyping procedures

Five SNPs were de-novo genotyped from four genes found as recent hits in the SUNLIGHT GWAS (Wang *et al.* 2010b). For the candidate gene analysis, in total 11 loci were identified from literature that have a potential role in skin and hair pigmentation and tanning ability (Han *et al.* 2008; Nan *et al.* 2009; Sulem *et al.* 2007), or have a role in the hydroxylation of vitamin D or 25(OH)D (Gascon-Barré 2005; Horst *et al.* 2005; Schuster 2010) (Figure 5.3). The *GC* SNP rs2282679 found to be the top hit in the SUNLIGHT consortium GWAS was in high LD with the candidate non-synonymous SNP rs4588 (R^2= 0.98), and rs4588 was chosen due to its functional role. The functional SNP rs7041 was also selected to allow the investigation of the association of GC haplotypes with 25(OH)D. The analysis of the candidate genes was run using the data from the genome-wide sub-studies of WTCCC2 and T1DGC and supplemented with two additional SNPs that were genotyped in the 1958BC. The 1447 SNPs genotyped by the gene-chips selected for further assessment as potential instruments were imported into Stata and a threshold of 0.8 was applied to the genotype certainty, via the program GTOOL (Freeman & Marchini 2011).

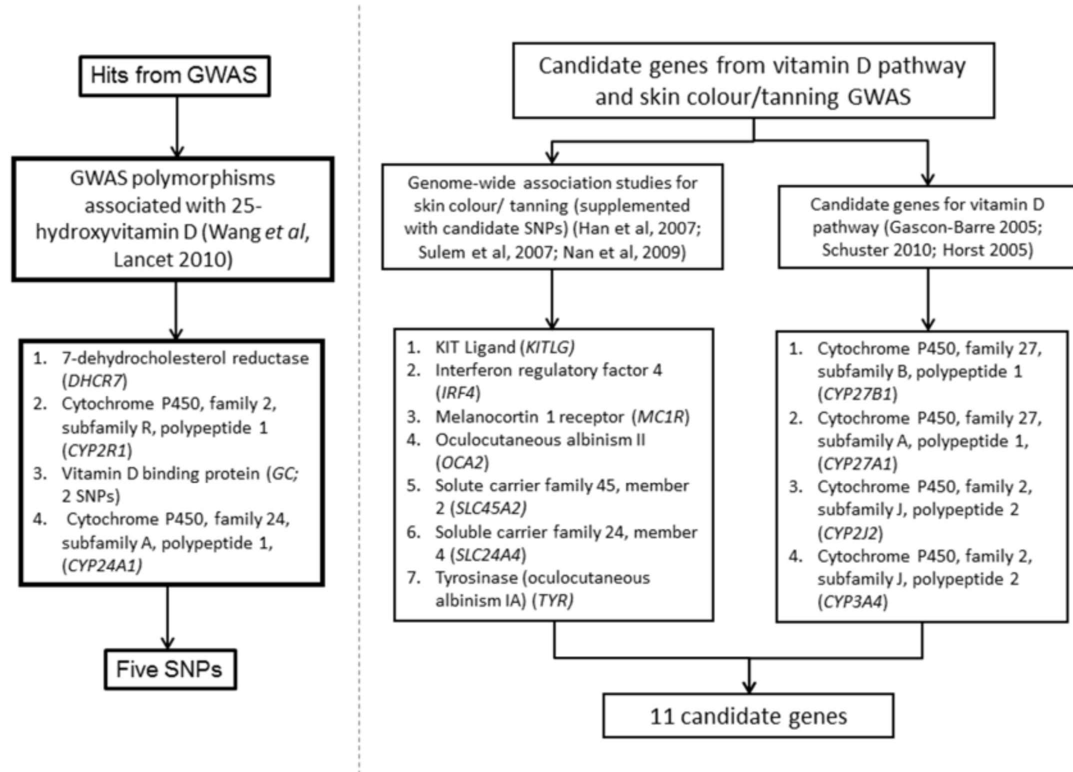


Figure 5.3 Polymorphisms and candidate regions considered as instruments for 25(OH)D

5.3.5 Statistical Methods

The hypothesised relationships between the genetic markers, exposure and outcome in the MR study for 25(OH)D on the inflammation/haemostatic markers are characterised in Figure 5.4.

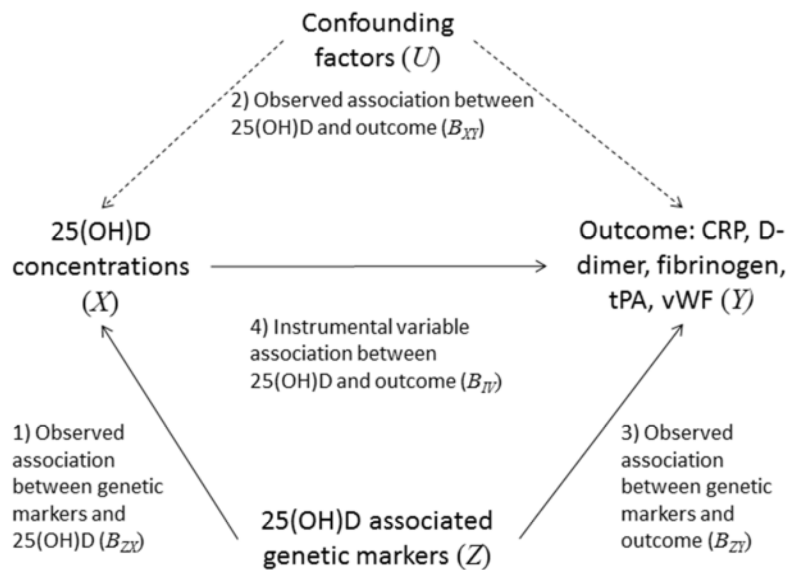


Figure 5.4 The relationships between the genetic markers, exposure, and outcome in MR.

Methods used to assess and evaluate genetic variants as instruments for 25(OH)D

Testing $H_0 : \beta_{ZX} = 0$

The SNPs identified as possible instruments for 25(OH)D were tested for their associations with 25(OH)D in the 1958BC. The 25(OH)D concentrations were left skewed in the 1958BC and were transformed by natural logarithm (ln) to achieve a normal distribution, when used as an outcome in the linear regression models. The models used to assess the SNP associations were adjusted for sex, and a *p*-value threshold was applied depending on how the SNP was initially identified as a possible instrument. The SNPs identified through GWAS on 25(OH)D had a significance threshold of p-value < 0.05 applied to reflect the prior evidence. For candidate gene polymorphisms, in total 1449 SNPs from within the 11 loci were

tested for an association with 25(OH)D and adjusted for sex. As all but two of the SNPs were from the genetic sub-samples of WTCCC2 and T1DGC genotyped on the gene-chips, the analysis was run using the program `SNPtest` (Marchini *et al.* 2007), and then meta-analysed using the inverse-variance method with the program METAL (Abecasis 2009). Prior to the meta-analysis, the SNP results of the WTCCC2 and T1DGC were filtered out by minor allele frequency (MAF) of less than 0.01, and lack of genotype certainty in the model as measured by the score statistic (i.e. proper info) with a cut-off of less than 0.4, as typically used for GWAS (Freathy *et al.* 2010; Marchini *et al.* 2007).

A false discovery rate (FDR) procedure was applied to the candidate SNP associations with 25(OH)D to control for multiple testing (Benjamini & Hochberg 1995). The FDR procedure has been adapted to grouped results (Hu *et al.* 2010), which lent itself this situation of 11 genetic loci and number of SNPs varying in each locus from 7 to 440. A FDR procedure was used as opposed to a simple Bonferroni correction, since a Bonferroni correction may have been too stringent and led to rejecting true findings.

The FDR procedure (Benjamini & Hochberg 1995) ranks the *p*-values $P_1, P_2, ..., P_m$ from *m* number of tests, $P_{(1)} \le P_{(2)} \le, ..., \le P_{(m)}$, where P_i corresponds to test *i*, with a significance level of α'. Then *k* is the largest *P* ranked i^{th} in the list that satisfies $P_{(i)} \le \frac{i}{m}\alpha'$, and the null hypotheses are rejected according to their associated order of $P_{(i)} i = 1, 2, ..., k$. A two-stage FDR has been adapted to a grouped data structure (Hu *et al.* 2010), and each group (i.e. loci as in this work) is weighted according to

the proportion of true null hypotheses in the group $\eta_{g,0}$, $g = 1, \ldots G$. The true value of $\eta_{g,0}$ is unknown for each of the G loci, so an adaptive two-stage approach of FDR was used with an initial significance level of $\alpha = 0.05$, that was adjusted to $\alpha' = \alpha/(1+\alpha)$. The first stage of the FDR procedure was applied to the SNPs grouped by their regions at level α'. In each group, the number of tests that did not reject the null hypothesis is used to estimate the proportion of true hypotheses $\eta_{g,0}$ and the proportion of alternative hypotheses. The *p*-values that have passed the first stage (i.e. rejected null hypotheses) are weighted by the ratio of the proportion of true hypotheses over alternative hypotheses, to give $P^{w}_{g,i}$. The overall proportion of null hypotheses is estimated from the average number of null hypotheses and tests in each group for the second stage significance level α^{w}. The FDR procedure is then re-run on the weighted *p*-values of $P^{w}_{g,i}$ with the second stage significance level α^{w}. Polymorphisms passing the two-stage FDR control were taken forward in the analysis to assess suitability as possible instruments.

<u>Relationship of SNPs with social, lifestyle and dietary measures and disease-relevant biomarkers</u>

In a MR study, the second assumption that the genetic variant does not have risk factors in common with the exposure or outcome is primarily justified from Mendel's laws (Lawlor *et al.* 2008). Although the lack of confounding between the SNP and exposure cannot be proven, several multiple linear regression models were run with 25(OH)D as the outcome to provide as much possible evidence. To assess the strength and sensitivity of the SNP association with 25(OH)D, the models were

adjusted for the social, lifestyle and dietary covariates and then separately for disease-relevant biomarkers. The social, lifestyle and dietary covariates adjusted for were: time spent outside; sun cover; oily fish consumption; vitamin D supplements; season; smoking; alcohol consumption; personal computer (PC)/television (TV) time; recreational metabolic equivalent task hours; adult social-economic class; geographical region of residence; BMI and waist circumference; alongside sex. The adjustment for the disease-relevant biomarkers, also assessed whether the SNP association with 25(OH)D was due to other health factors. The biomarkers adjusted for in the model together were: vWF; tPA; D-dimer; fibrinogen; CRP; IgE; triglycerides; low-density lipoproteins (LDL); high-density lipoproteins (HDL); total cholesterol; forced expiratory volume in 1 second (FEV); diastolic and systolic blood pressure; IGF1; HbA1c; BMI and waist circumference.

The association of SNP with the social, lifestyle and dietary variables and biomarkers was next considered, as if pleiotropy did exist the association should be strong and not affected by 25(OH)D adjustment. The social, lifestyle and dietary variables were dichotomised and used as outcomes in a logistic regression model. The SNP associations with the social, lifestyle and dietary variables were corrected for multiple testing, 14 in total, and with a threshold of p-value $< 0.004 = 0.05/14$ that was applied. Interactions between the social/lifestyle/dietary variables and SNPs with 25(OH)D concentrations were also investigated and corrected for multiple testing of the 14 variables. For each of the biomarkers as outcomes (ln transformed as appropriate to normalise left skew distributions), linear regression models were used and adjusted for 25(OH)D and geographical region of residence. Again, the

significance threshold was corrected for multiple testing with the biomarkers, and a threshold of p-value $\leq 0.003 = 0.05/17$ was applied.

The final step before combining the polymorphisms into single allele counts, scores and haplotypes was testing for interaction between pairs of SNPs with 25(OH)D adjusted for sex to investigate whether a non-linear relationship existed between the genetic proxy and exposure.

Haplotypes, allele scores and counts

The SNPs that passed evaluation were taken forward as instruments and combined as an allele count or a weighted allele score. Genetic variants where the related gene expressions took place prior to or as part of the 25-hydroxylation of vitamin D were referred to as the "synthesis" polymorphisms. Variants that were expressed post the 25-hydroxylation were referred to as the "metabolism" polymorphisms. A distinction was made in this work between the SNPs found in GWAS (labelled "GWAS") and the SNPs found in both GWAS and candidate gene analysis (labelled "non-GWAS") when creating the allele scores and counts.

The *GC* non-synonymous SNPs were also investigated separately as an instrument by estimating their *GC* haplotypes, since these polymorphisms alone explained a relatively sizable proportion of variation in 25(OH)D compared with the other variants from different loci. The haplotypes were estimated from rs4588 and rs7041 using SimHap, and haplotypes with a posterior probability < 0.95 were removed (Carter *et al.* 2008). After considering various forms of categorisation of the haplotypes (Appendix 4), a variable categorising participants by the number of *Gc2*

haplotype copies carried was taken forward as a potential instrument for 25(OH)D concentrations.

For creating the allele count and the weighted allele score, the SNPs genotypes were coded as (0-2) with their risk allele for lowering 25(OH)D concentrations as the effect allele (i.e. homozygous genotype with both risk alleles is coded as 2). The allele count for the synthesis and metabolism SNPs were then summed on the basis of the risk alleles. When the tails of the counts were carried by only a few participants, those carrying the maximum number of risk alleles were collapsed into the next count category. For the allele score, the technique is as outlined in the Statistical methods section of Chapter 4 and created separately for the synthesis and metabolism SNPs. The weights used were taken from the SNP association with 25(OH)D from linear regression models adjusted for sex in the 1958BC. For scale display purposes in the figures, the scale of the score was put on a similar scale to the count by multiplying by a quarter. Each score and count was created solely with the SNPs found in the GWAS studies and recreated again with the SNPs combined from the GWAS studies and the candidate gene analysis.

The proportion of variation explained in 25(OH)D concentrations by the counts, scores and haplotypes was estimated from the adjusted R^2 for the number of parameters in the linear regression models with 25(OH)D as the outcome. The *F*-statistic for the allele counts, scores and haplotype was also considered from the linear regression models with 25(OH)D, since it is used as indicator of instrument strength in IV analysis. The *F*-statistic should ideally be greater than 10 in order for an instrument to be considered strong enough to use (Staiger & Stock 1997). The

proportion of variance explained (R^2) and the F-statistic of an instrument(s) with the exposure are tightly related since

$$F = \frac{R^2 (n-1-k)}{(1-R^2)k},$$

where k is the number of instruments (or variables) and n is the sample size (Rice 1995). An increase in sample size will increase the value of F even when an instrument only explains a small proportion of the variation in an exposure. However, biasing will increase in IV analysis compared with ordinary least squares (OLS) (linear regression) when multiple weak instruments are used together (Palmer *et al.* 2011a). The relative bias of the IV analysis to linear regression is defined as k/nR^2. The final allele score or count from each group of the synthesis, metabolism and *GC* SNPs/haplotypes was assessed in the same way as the individual SNPs with the social, lifestyle and dietary variables and biomarkers.

Methods used for the IV analyses between 25(OH)D and the haemostatic/ inflammation outcomes with genetic instruments

Testing $H_0 : \beta_{XY} = 0$, $H_0 : \beta_{ZY} = 0$, $H_0 : \beta_{IV} = 0$

Linear regression models were used with each of the haemostatic/inflammation biomarkers ln transformed as outcome adjusted for sex, BMI, month of measure, geographical region of residence, and ln transformed 25(OH)D. The normal probability plot of the studentised residuals of each model was used to assess the assumption of normality (Appendix 4). The untransformed curvature term of 25(OH)D was significantly associated with each of the haemostatic/inflammation outcomes after adjustment for sex, BMI, month of measure and geographical region.

Plots of model residuals and 25(OH)D (ln and raw) were visually assessed for a trend.

The covariates adjusted for in IV regression were sex, month of blood sample, BMI and geographical region. In IV analysis each (exogenous) covariate is treated as its own instrument (Baum 2006). Therefore, co-linearity amongst the exogenous variables should be avoided to minimise the standard error of the IV estimator (Wooldridge 2009), therefore additional lifestyle covariates were not considered for this reason.

To establish the association size of ln 25(OH)D with the ln outcomes (i.e. to test for $H_0 : \beta_{XY} = 0$), linear regression models were run adjusting for sex, month of blood sample, BMI and geographical region of residence. The linear regression models were repeated with the same adjustments, to investigate whether the individual SNPs and allele counts or scores had an association with the ln outcomes (i.e. to test for $H_0 : \beta_{ZY} = 0$). For a significant relationship between 25(OH)D and any of the haemostatic/inflammation outcomes to exist, it would be expected that there would be some evidence against both null hypotheses. The final step was to run IV regression between 25(OH)D and the haemostatic/inflammation outcomes using the genetic instruments to test the null hypothesis of $H_0 : \beta_{IV} = 0$.

The first model form of IV regression can be written as (Baum 2006; Wooldridge 2009)

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 z_{1i} + \varepsilon_{1i}$$

where $i = 1, ..., n$ are for the individuals in the model, and $\varepsilon_{1i} \sim N\left(0, \sigma^2\right)$, independently. The parameters of interest are β_0, β_1, β_2, and, unlike OLS regression, the exogenous variables are distinguished from the endogenous. The exogenous variables are assumed to be uncorrelated with the error term, unlike the endogenous variables. In the above equation, the main exposure x is correlated with the error term ε_1 making it an endogenous variable, but z_1 is uncorrelated with ε_1, making it an exogenous variable. Furthermore, the outcome y will also be correlated with its error term ε_1 making it endogenous. An instrument must be correlated with the exposure, but exogenous (thus uncorrelated with ε_1). The second model form in IV regression defines the relationship between the exposure and the exogenous variables as:

$$x_i = \pi_0 + \pi_1 z_{1i} + \pi_2 z_{2i} + \varepsilon_{2i},$$

where the assumptions are as before for the distribution of the error term and number of individuals. To be a valid instrument of x, it is assumed that the defining exogenous variables are $\pi_1, \pi_2 \neq 0$. It is also worth noting that variable z_1 is in both equations and is its own instrument, but that the endogenous variable x must have at least one associated exogenous (instrument) variable (i.e. z_2).

For each of the haemostatic/inflammation biomarkers, a two-stage least squares (2SLS) regression was run with the genetic variants as exogenous variables for ln 25(OH)D and adjusted for sex, BMI, month of sample and geographical region of residence. To test that the errors of the 2SLS model were identical and independently distributed, i.e. homoscedastic, the Pagan-Hall statistic was used (Baum *et al.* 2007). This statistic is distributed as a χ_p^2 under the null hypothesis of

homoscedasticity in 2SLS regression and is also valid when heteroscedasticity is present in the regression model outside the IV (Baum *et al.* 2003). When evidence of heteroscedasticity was present in the 2SLS regression model, a generalised method of moments (GMM) estimator with a robust variance estimate was used instead as it is heteroscedastic consistent (Baum 2006). Hansen's *J* statistic was used to assess whether any of the SNPs as individual instruments were over-identified in the IV regression (Baum *et al.* 2007), therefore whether there unnecessary instruments for the exposure in the model. Over-identification of instruments relies on there being sufficient correlation between the instruments and exposure, however in large sample sizes over-identified endogenous variables can be more efficient (i.e. have smaller standard errors) at estimating (Baum 2006).

The use of IV regression as opposed to OLS regression does come at a cost of efficiency (Baum 2006). The asymptotic variance of an IV estimator will always be larger than that of an estimator from OLS (Baum 2006). The Durbin-Wu-Hausan test for endogeneity determines whether IV regression is appropriate to use over OLS regression, given the loss of efficiency of using an IV estimator against using a potentially biased and inconsistent OLS estimator. Under a null hypothesis of endogeneity, OLS is an appropriate method and evidence against this null hypothesis suggests that the OLS estimator is inconsistent. The analysis was run in Stata 11 (StataCorp 2010) and the IV regression were fitted using the `ivreg2` command (Baum 2006).

The interpretation of the IV estimate – elasticity

In the IV regression the ln of 25(OH)D (x) was regressed on the ln of the biomarkers (y) of CRP, D-dimer, fibrinogen, tPA and vWF. A ln coefficient with ln outcome in econometrics is often referred to as an elasticity (Wooldridge 2009). An elastic effect size is can be considered as the relative change in y with respect to the relative change in x;

$$\beta_1 = \frac{\%\,\Delta y}{\%\,\Delta x}$$

Power calculations

For 2SLS regression there is no standard approach for estimating power and sample size, however one way is through repetitive simulation of a dataset. This was done by simulating a dataset of a fixed sample size (n) by imposing constraints on the known relationships, running IV regression and then testing the parameter of interest at a fixed significant level α. The dataset was re-simulated a 1000 number times for the same sample size (n) and the proportion of times the parameter of interest was less than the significant level α is an estimate of the power yielded by the sample size (Feiveson 2002).

The model based on the MR assumptions was used for the simulated dataset, and defined as:

$$\begin{aligned} x_i &= \alpha_0 + \alpha_1 g_i + \varepsilon_{xi} \\ y_i &= \beta_0 + \beta_1 x_i + \varepsilon_{yi} \end{aligned},$$

where i is the number of individuals, x_i, g_i, y_i are the exposure, genetic variant, and outcome with $\varepsilon_{xi} \sim N\left(0, \sigma_1^2\right)$ and $\varepsilon_{yi} \sim N\left(0, \sigma_2^2\right)$. The genetic variant g_i takes the

values for the single marker $\{0,1,2\}$ and joint allele count $\{0,1,2,3\}$. The parameters were set to the effect sizes as found in the 1958BC from the relationship between ln 25(OH)D and ln tPA. tPA was used as the example outcome in the simulation since a relationship was observed between 25(OH)D and tPA in Chapter 3. The SNP and allele count genotypes in the simulated dataset were set to the observed proportions in the 1958BC.

Simulation of pleiotropic SNP in a MR study

Simulation was also used to understand the effect pleiotropy might have on the reliability of 2SLS regression estimates, when the assumptions of MR are invalidated by one SNP that is subsequently combined with other markers. The datasets created had pathways from the genetic variant Z_p going to different intermediate exposures X and P (Figure 5.5 Scenario A). The exposures X and P were created in such way so that both were independently associated with the outcome Y and correlated to each other. The correlation between exposures X and P were set to 0.57 (based on the reported correlation between 25(OH)D and VDBP (Bikle *et al.* 1986b)). The initial parameters were based on the ln 25(OH)D (as exposure X) and ln tPA (as outcome Y). The effect size of exposure P regressed on Y was set at three values based on the effect size of X on Y; 50% smaller ("small"), the same size ("moderate") and 50% larger ("large"). The offending pleiotropic SNP was combined into an allele count Z_{C1} with a SNP Z_1, which did meet the assumptions of MR in its own right. The allele count Z_{C1} was also used together with another allele count Z_{C2} (which did met the assumptions of MR) as instruments for the main exposure X (Figure 5.5

Scenarios B-C). As before, the simulation was repeated 1000 times for each scenario.

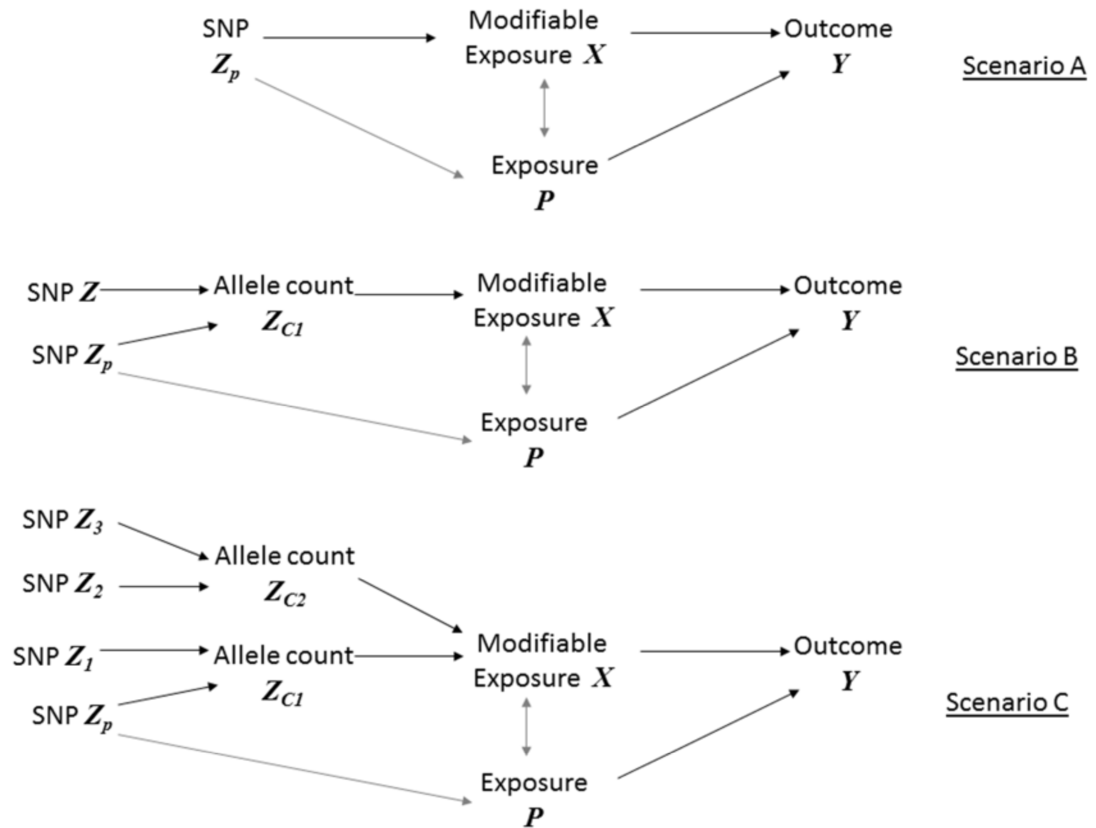


Figure 5.5 Proposed scenarios of pleiotropic SNP in a MR study. SNP Z_p through exposure P to outcome Y outside of the pathway of the exposure of interest X (scenario A). The offending pleiotropic SNP Z_p is then used in an allele count Z_{C1} along with another SNP Z_1, that meets the assumptions of MR (scenario B). The allele count Z_{C1} is used together with another allele count Z_{C2}, which also meets the assumptions of MR, as instruments for exposure X (scenario C).

5.4 Results

5.4.1 Evaluation of the genetic instruments

All six polymorphisms selected from GWAS results on 25(OH)D had significant associations with 25(OH)D (*p*-value<0.02 for all), after adjustment for sex (Table 5.3). The non-synonymous SNP rs4588 from the *GC* locus explained the largest proportion of variation in 25(OH)D at 1.18% and had the largest *F* statistic, 73.5, compared with the other SNPs (Table 5.3). Of the 11 candidate genes tested for an association with 25(OH)D, eight SNPs passed an adapted two-stage FDR control. Seven of those SNPs were from the *CYP27B1* gene and the one other was from *SLC45A2* (Appendix 5 Figure 1). The association of SNP rs10877012 with 25(OH)D (adjusted for sex) had the lowest *p*-value of the *CYP27B1* SNPs, and was one of the de-novo genotyped SNPs in the 1958BC. The other candidate gene SNP was rs16892096 from the gene *SLC45A2*, which had a risk MAF of 1.5% and larger effect size with 25(OH)D than the other SNPs found (Table 5.3). Participants who carried the T allele of rs16892096 had 12% lower 25(OH)D concentrations (95% CI 4, 20, *p*-value=0.0007), than non-carriers. In total, seven SNPs from six genes (two SNPs were from *GC*) were taken forward for further assessment as potential instruments for 25(OH)D concentrations in IV analyses. Of the seven SNPs taken forward, SNPs from *CYP24A1* and *CYP27B1* were considered to be weak instruments on the basis of the $F < 10$.

Table 5.3 The SNP associations with ln 25(OH)D concentrations adjusting for sex

Gene	SNPs	*n*	MAF	Alleles*	β	Std error	*p*-value	% of variance explained	*F*-statistic
GC	rs4588	6027	0.30	G > T	-0.08	0.009	1.5×10^{-17}	1.18	73.5
GC	rs7041	5412	0.45	G > T	-0.06	0.009	9.5×10^{-11}	0.75	42.5
DHCR7/NADSYN1	rs12785878	6504	0.22	T > G	-0.05	0.009	1.2×10^{-6}	0.35	22.4
CYP2R1	rs10741657	5968	0.40	G > A	0.03	0.009	0.0003	0.21	13.3
CYP24A1	rs6013897	6534	0.20	T > A	-0.03	0.010	0.016	0.07	5.4
SLC45A2†	rs16892096	4923	0.02	C > T	-0.12	0.039	0.0007	0.19	10.3
CYP27B1†‡	rs10877012	6877	0.33	C > A	-0.02	0.008	0.008	0.09	7.2

* Major allele as reference > Minor allele as effect.

† Genes *SLC45A2* and *CYP27B1* were the top hits from the candidate gene analysis that passed an adapted two-stage FDR

‡ SNP rs10877012 was the top hit in the *CYP27B1* locus.

The consistency of the SNP associations with 25(OH)D were considered next and whether the SNP-25(OH)D associations were sensitive to potential lifestyle confounders (region, dietary, lifestyle and adiposity variables), or to confounding from different pathways (biomarkers). All of the SNPs remained significantly associated with 25(OH)D after adjusting for social, lifestyle, region and dietary covariates, and after adjusting for biomarkers (Figure 5.6). The *SLC45A2* SNP association with 25(OH)D did attenuate somewhat after adjusting for the social, lifestyle, region and dietary covariates, but still remained significant (*p*-value=0.010).

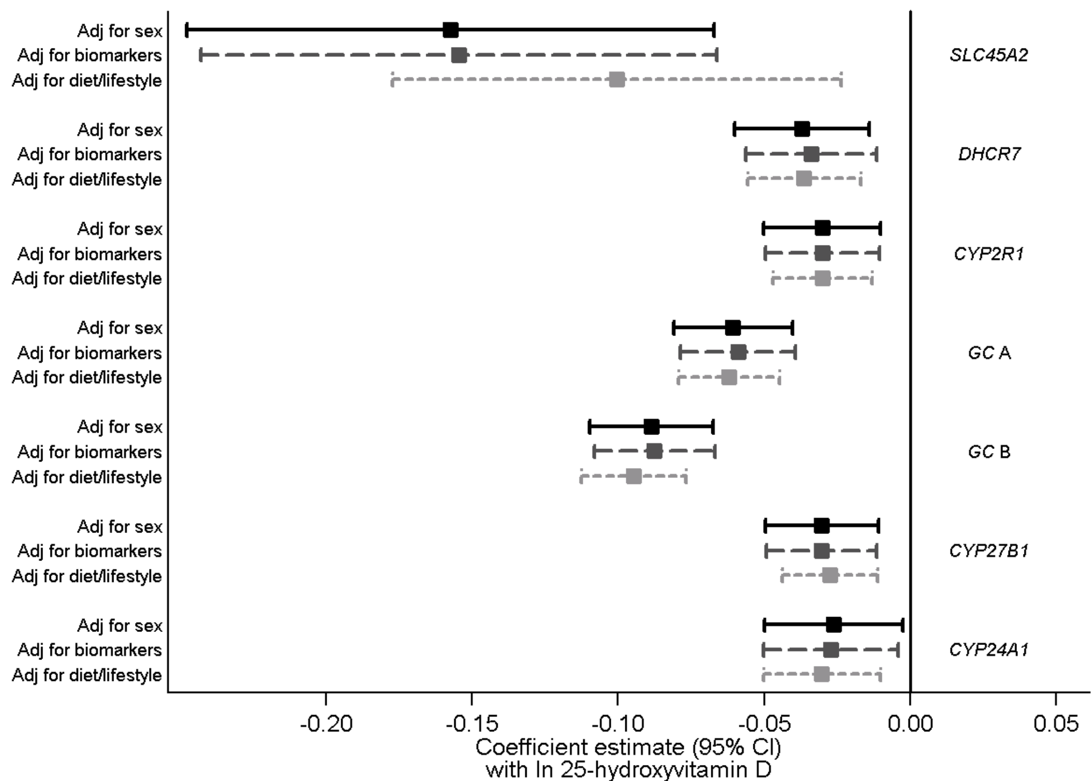


Figure 5.6 SNP associations with 25(OH)D after three model adjustments for; 1) adjusted for sex (black solid line); 2) adjusted for biomarkers (dark grey long dashed line); 3) adjusted for social, dietary, lifestyle and region covariates (light grey short

dashed line). The coefficient of the SNP association is denoted by the solid square and 95% CI by the horizontal bars.

The SNPs were individually regressed on the dichotomous variables representing time spent outside; sun cover; oily fish consumption; vitamin D supplements; season; smoking; alcohol consumption; PC/TV time; recreational metabolic equivalent task hours; adult social-economic class; geographical region; BMI and waist circumference. After correcting the *p*-values for multiple testing, some relationships between SNPs and lifestyle variables were observed (Figure 5.7). Both of the *GC* SNPs, rs7041 and rs4588, were associated with oily fish consumption (*p*-value=0.028, *p*-value=0.027 respectively). The *DHCR7* SNP was associated with region (*p*-value=0.0004), and the *CYP2R1* SNP was associated with adulthood socio-economic position (*p*-value=0.040). The *DHCR7* SNP rs12785878 MAF in the 1958BC is 22%, compared with Hapmap CEU population at 27% (International HapMap Consortium 2003), which is higher. There was no evidence of an interaction between region, social, dietary, lifestyle, and adiposity variables and SNPs with 25(OH)D (*p*-value=0.083 corrected for multiple testing).

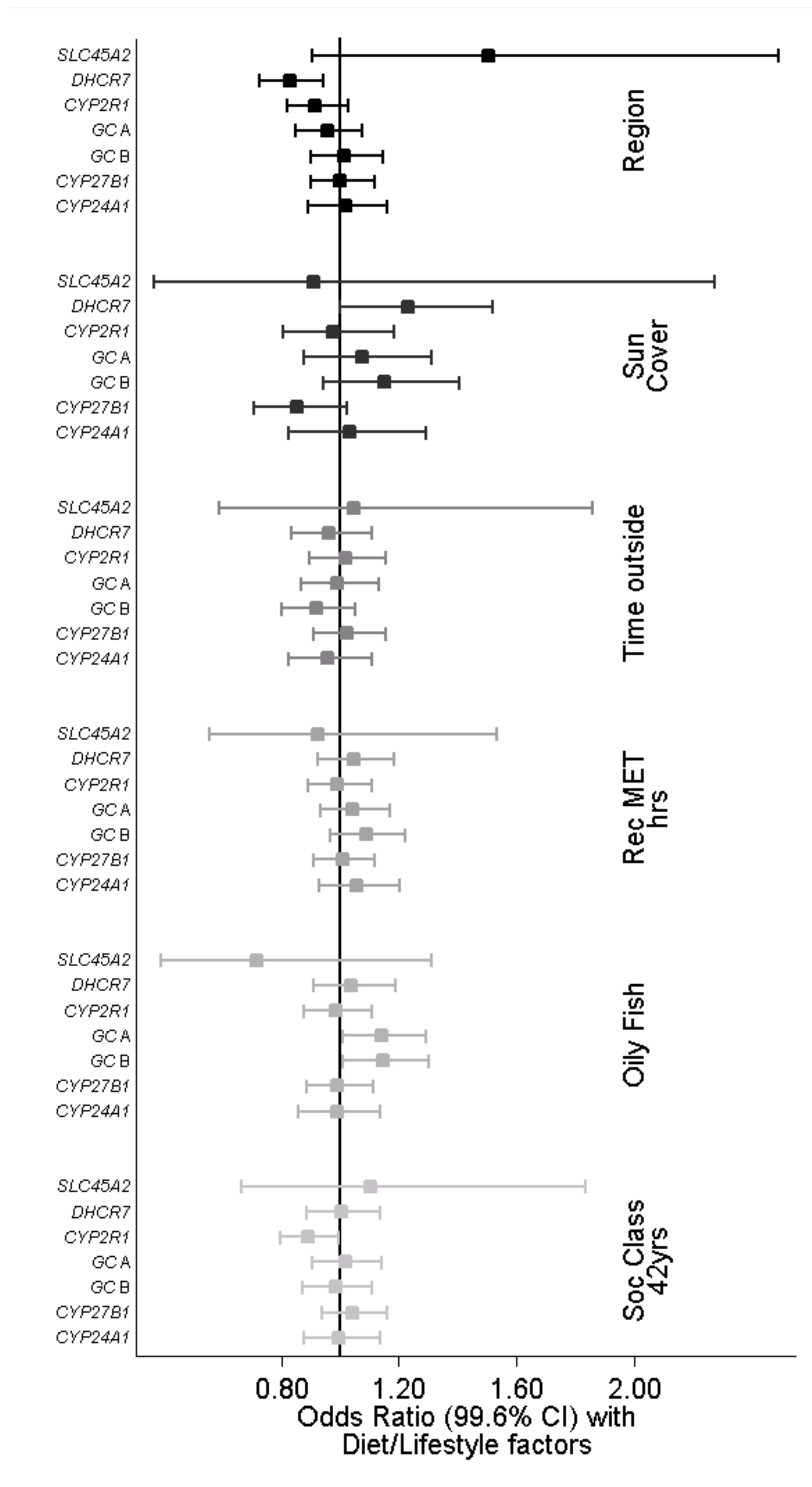


Figure 5.7 SNP odds ratios with selected dichotomous dietary, lifestyle and region variables (light to dark shading for each outcome), and all models were adjusted for sex except region, where the model was also adjusted for season. The odds ratio of the SNP association is denoted by the solid square and 99.6% CI (corrected for multiple testing) by the horizontal bars.

The final test of possible SNP pleiotropy with the biomarkers, was assessing the SNP association with the biomarkers after adjusting for 25(OH)D, sex and region (Appendix 6 Table 1), and correcting for multiple testing. The SNP from *SLC45A2* was associated with LDL, cholesterol and lung function (p-value=0.019 and p-value=0.005 respectively). After these associations with LDL and FEV_1, the *SLC45A2* SNP was removed from any further analysis with the allele counts and scores.

A distinction was made between the SNPs from genes expressed prior to and as part of the 25-hydroxylation (labelled "synthesis"), genes post 25-hydroxylation of vitamin D (labelled "metabolism"), SNPs found in GWAS on 25(OH)D (labelled "GWAS"), and SNPs found from the GWAS and candidate gene analysis (labelled "non-GWAS"), an un-weighted allele count (labelled "count"), weighted score (labelled "score") (Table 5.4). The metabolism scores that were weighted in favour of the *GC* rs4588 association with 25(OH)D, explained the greatest proportion of variance in 25(OH)D and had the largest F-statistics, compared with all other combinations of SNPs. For the synthesis SNPs, it was the synthesis count that explained slightly more variation in 25(OH)D concentrations, compared with the synthesis score. All combinations of SNPs had $F > 10$, which implied that none of the potential instruments were "weak".

Table 5.4 Allele counts/scores and haplotype association with ln 25(OH)D adjusted for sex.

	SNP from genes	*n*	ln 25(OH)D Coefficient (95% CI)	*p*-value	R^2, %	*F*-statistic	Rel. bias, %
Synthesis allele count	*DHCR7, CYP2R1*	5856	-0.040 (-0.053, -0.026)	5.99×10^{-9}	0.56	33.4	3.0
Synthesis allele score	*DHCR7, CYP2R1*	5856	-0.154 (-0.206, -0.101)	9.04×10^{-9}	0.55	33.4	3.1
Metabolism GWAS allele count	*GC**, *CYP24A1*	5936	-0.057 (-0.071, -0.043)	1.81×10^{-15}	1.04	63.1	1.6
Metabolism GWAS allele score	*GC**, *CYP24A1*	5936	-0.196 (-0.240, -0.152)	3.76×10^{-18}	1.24	75.8	1.4
Gc2 haplotype	*GC*	5301	-0.077 (-0.096, -0.058)	2.94×10^{-15}	1.14	62.6	1.7
Metabolism non-GWAS allele count	*GC**, *CYP24A1, CYP27B1*	5623	-0.046 (-0.057, -0.035)	8.60×10^{-16}	1.12	64.6	1.6
Metabolism non-GWAS allele score	*GC**, *CYP24A1, CYP27B1*	5623	-0.238 (-0.289, -0.186)	2.01×10^{-19}	1.41	81.7	1.3

* GC SNP rs4588

The strength of the association of the synthesis count, metabolism scores, and *Gc2* haplotype with 25(OH)D was assessed against potential lifestyle confounders and the health biomarkers. The allele count/score associations showed minimal attenuation after the model adjustments for both groups of covariates (Figure 5.8). Furthermore,

the allele counts/scores were no longer directly associated with some of the social, dietary, lifestyle and region variables that the individual SNPs had been associated with (Figure 5.7 vs Figure 5.9). Unlike the *DHCR7* SNP, the synthesis count was not associated with adulthood socio-economic class (*p*-value=0.57 corrected for multiple testing). The metabolism count (non-GWAS) was not associated with oily fish consumption (*p*-value=0.10 corrected for multiple testing), unlike the individual *GC* SNPs. However, the synthesis score was associated with region (*p*-value=0.0002 corrected for multiple testing), as had been observed for the *DHCR7* SNP. None of the allele scores/counts were significantly associated with the health biomarkers (*p*-value ≥0.16 for all allele scores/counts and corrected for multiple testing) and showed no direct evidence of pleiotropy (Appendix 6 Table 2).

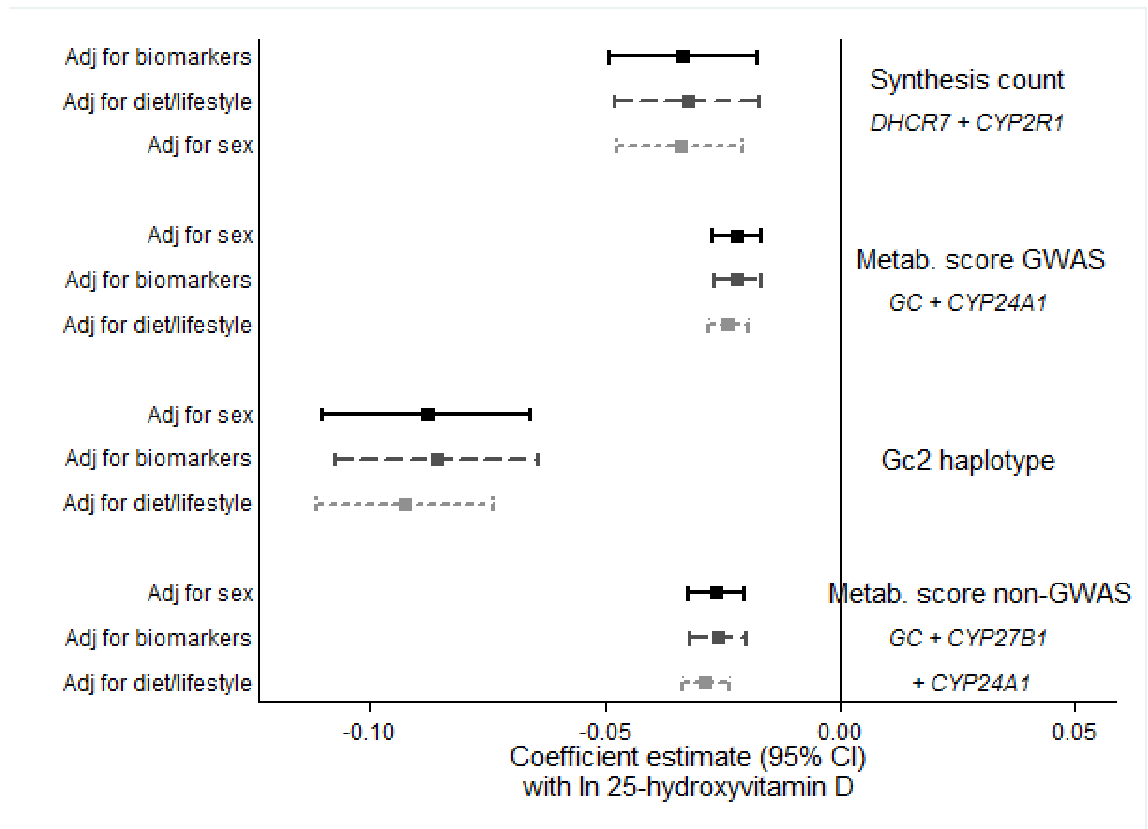


Figure 5.8 Allele score, counts and haplotype associations with 25(OH)D after three model adjustments; 1) adjusted for sex (black solid line); 2) adjusted for biomarkers (dark grey long dashed line); 3) adjusted for diet/lifestyle and region covariates (light grey short dashed line). The coefficient of the count/score association is denoted by the solid square and 95% CI by the horizontal bars.

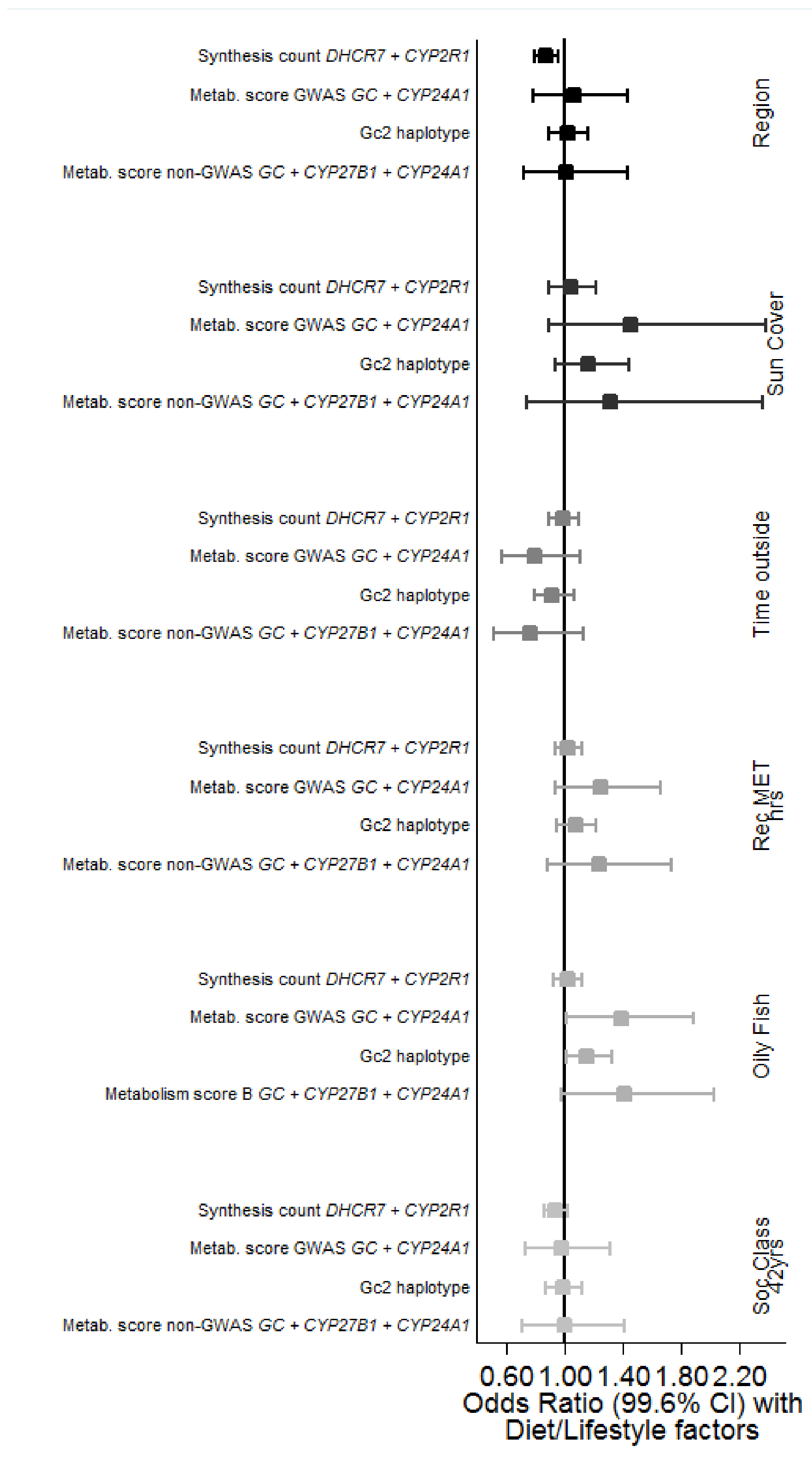


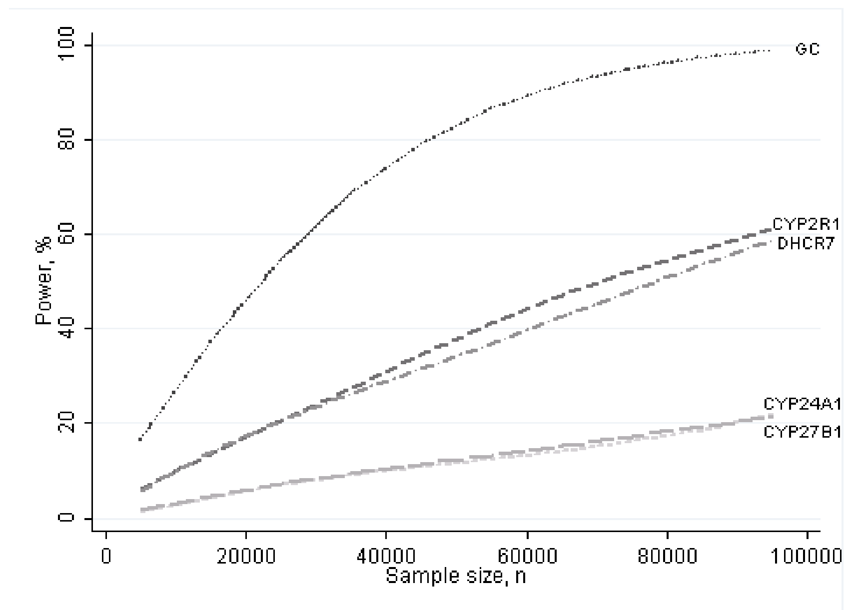
Figure 5.9 Allele score, counts and haplotype odds ratios with selected dichotomous dietary, lifestyle and region variables (light to dark shading for each outcome), and all models were adjusted for sex except for region, where the model was also

adjusted for season. The odds ratio of the count/score association is denoted by the solid square and 99.6% CI (corrected for multiple testing) by the horizontal bars.

5.4.2 Power calculation and MR simulation study

The effect size of the 25(OH)D association with tPA was used as the basis for the power calculations to detect a true effect by IV regression (Figure 5.10). The combination of synthesis and metabolism (non-GWAS) scores together as instruments in IV regression had the most power to detect an effect, compared with the power when using the other genetic instruments. There was little difference in power between the metabolism scores (GWAS vs non-GWAS). Of the individual SNPs, the *GC* SNP (rs4588) as an instrument in IV regression had the largest power to detect a true effect. The power in IV regression to detect an effect with the *GC* SNP as an instrument was similar to the power when using the metabolism scores as instruments.

A



B

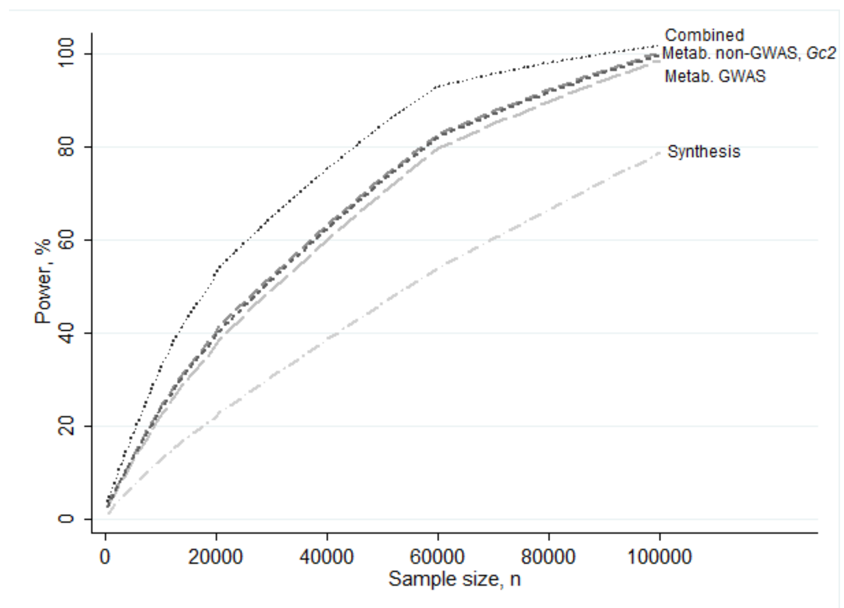


Figure 5.10 Power calculations for IV analysis with the single SNPs as instruments (A) and allele scores/counts/haplotypes as instruments (B) based on the ln 25(OH)D association with ln tPA.

The simulation study of a SNP with varying degrees of pleiotropy (none, small, moderate, large) had the following instrument scenarios in 2SLS regression: the SNP alone as a single instrument; combined into an allele count with another SNP (that

met the assumptions of MR as a single instrument); two instruments with an allele count that included the pleiotropic SNP and another allele count (that met the assumptions of MR) (as outlined in Figure 5.5). The simulation revealed that the size of the association between the exposure and outcome became larger as the strength of the SNP's pleiotropy increased from small to large (Appendix 5 Figure 2). The inflated estimate of the association between the exposure and outcome occurred regardless of the instruments used in the IV regression model. Furthermore, increasing the sample size had little effect on the IV estimate when the SNP had some degree of pleiotropy, as the exposure's effect size on the outcome remained inflated. Compared with a single SNP as the instrument, the allele count did not give a more accurate estimate of the true effect size of the exposure-outcome relationship. Of all three types of instruments (single SNP as instrument, allele count as instrument, and two allele counts as instruments) used in the IV regression, the two allele counts as instruments gave the least inflated IV estimate of the association between the exposure and outcome. It was the two allele counts used as two instruments in 2SLS regression that also had the greatest power to detect an effect, compared with the other combinations of SNPs.

5.4.3 IV regression of 25(OH)D with haemostatic/inflammation markers using genetic instruments

Concentrations of 25(OH)D were not associated with the haemostatic/inflammatory markers in IV regression that used the synthesis count or SNPs in anyway as instruments (p-value≥0.29) (Table 5.5), even when combined with the metabolism score or SNPs (p-value≥0.29) (Appendix 6 Table 5). In comparison with the IV regression results, 25(OH)D was negatively associated with CRP, D-dimer,

fibrinogen and tPA (*p*-value ≤0.004) in the multiple linear regression analysis, after adjusting for sex, BMI , month of measure and geographical region, except for vWF (*p*-value=0.37) (Table 5.5). However based on the Durbin-Wu-Hausman test, the associations between 25(OH)D and the haemostatic/inflammatory biomarkers did not differ when estimated by IV regression, when using the synthesis count or SNPs as instrument(s), and compared with the estimates from OLS (linear regression) (*p*-value≥0.19) (Appendix 6 Table 4).

There was no evidence that the synthesis count was directly associated with the haemostatic/inflammatory markers. However, there was some evidence that genetic markers that included *GC* SNPs were directly associated with CRP and tPA. Each unit increase of the metabolism (GWAS) score was associated with 10.4% (95% CI -21.3, 0.5 *p*-value= 0.061) lower CRP levels (Appendix 6 Table 3). For tPA, each copy of *Gc2* increased levels of tPA by 2.1% (95% CI -0.1, 4.2, *p*-value= 0.057) (Appendix 6 Table 3).

Initially the IV regression was run using 2SLS estimation, however in most instances there was evidence of heteroscedasticity in the model (Appendix 6 Table 4). The IV regression was re-run using GMM estimation and there was very little difference between the results estimated with 2SLS or GMM (Appendix 6 Table 5). For all outcomes, there no evidence that any of SNPs, allele scores and counts were redundant as instruments in the 2SLS regression (*p*-value≥0.09) (Appendix 6 Table 4), and in the GMM regression (*p*-value≥0.09) (Appendix 6 Table 5).

Increasing 25(OH)D concentrations were tentatively associated with higher CRP levels in the IV regression results, when using the metabolism allele (GWAS) score as an instrument [IV coefficient: 0.50 (95% CI -0.03, 1.02, *p*-value=0.065)] (Table 5.5). The association of 25(OH)D with CRP strengthen slightly when using *GC* (rs4588) and *CYP24A1* SNPs together as two instruments [IV coefficient: 0.53 (95% CI 0.00, 1.05, *p*-value= 0.049) (Appendix 6 Table 5). The direction of the 25(OH)D association with CRP observed in IV regression was opposite from that found in OLS (linear) regression. Furthermore, there was evidence that the 25(OH)D association with CRP substantially differed when estimated by IV regression, using metabolism scores as instruments, compared with the OLS estimate of the association (Durbin-Wu-Hausman test (p-value≤0.041) (Appendix 6 Table 4). As noted before, no association was observed between 25(OH)D and CRP in the IV regression results when using the synthesis count or SNPs, including in combination with the metabolism score (p-value≥0.34) (Table 5.5, Appendix 6 Table 5).

From IV regression, the observed inverse association between ln 25(OH)D with ln tPA fell slightly short of statistical significance when using *Gc2* as an instrument [IV coefficient: -0.24 (95% CI -0.49, 0.00, (*p*-value=0.054); which translates into a 6% decrease in tPA with every 25% increase of 25(OH)D concentrations] (Table 5.5). However, no association between 25(OH)D and tPA was observed when using any other instrument (*p*-value=0.21) (Table 5.5, Appendix 6 Table 5) in the IV regression, despite other instruments having the larger *F* values and more statistical power to detect an effect.

Table 5.5 25-hydroxyvitamin D association with the inflammation/haemostatic biomarkers and the 25(OH)D association using the genetic proxies as instruments for inflammation/haemostatic biomarkers adjusted for sex, month of measure, geographical region and BMI

n	ln 25(OHD) β (95% CI)	ln 25(OH)D *p*-value	Instrument(s) for ln 25(OH)D	ln 25(OH)D IV Beta (95% CI)	IV *p*-value
Outcome : ln CRP					
5592	-0.105 (-0.177, -0.032)	0.004	Synthesis allele count (*DHCR7*, *CYP2R1*)	-0.406 (-1.247, 0.435)	0.34
			Metabolism allele score A *(GC, CYP24A1)*	0.495 (-0.030, 1.021)	0.065
-	-	-	Gc2 haplotype	0.434 (-0.130, 0.997)	0.13
			Metabolism allele score B *(GC, CYP27B1, CYP24A1)*	0.427 (-0.082, 0.936)	0.10
Outcome : ln D-Dimer					
5576	-0.058 (-0.094, -0.023)	0.001	Synthesis allele count (*DHCR7*, *CYP2R1*)	-0.187 (-0.599, 0.225)	0.37
			Metabolism allele score A *(GC, CYP24A1)*	-0.101 (-0.360, 0.158)	0.45
-	-	-	Gc2 haplotype	-0.094 (-0.371, 0.184)	0.51
			Metabolism allele score B *(GC, CYP27B1, CYP24A1)*	-0.097 (-0.353, 0.158)	0.46
Outcome: ln fibrinogen					
5588	-0.031	<0.001	Synthesis allele count	-0.065	0.40

	n	ln 25(OHD) β (95% CI)	ln 25(OH)D *p*-value	Instrument(s) for ln 25(OH)D	ln 25(OH)D IV Beta (95% CI)	IV *p*-value
		(-0.044, -0.018)		(*DHCR7*, *CYP2R1*)	(-0.215, 0.085)	
	-	-	-	Metabolism allele score A *(GC, CYP24A1)*	0.025 (-0.068, 0.117)	0.60
	-	-	-	Gc2 haplotype	0.012 (-0.088, 0.111)	0.82
	-	-	-	Metabolism allele score B *(GC, CYP27B1, CYP24A1)*	0.028 (-0.064, 0.119)	0.55
Outcome: ln tPA						
	5592	-0.138 (-0.170, -0.106)	<0.001	Synthesis allele count (*DHCR7*, *CYP2R1*)	0.062 (-0.318, 0.441)	0.75
	-	-	-	Metabolism allele score A *(GC, CYP24A1)*	-0.132 (-0.360, 0.097)	0.26
	-	-	-	Gc2 haplotype	-0.240 (-0.485, 0.004)	0.054
	-	-	-	Metabolism allele score B *(GC, CYP27B1, CYP24A1)*	-0.140 (-0.361, 0.080)	0.21
Outcome: ln vWF						
	5593	-0.010 (-0.032, 0.012)	0.37	Synthesis allele count (*DHCR7*, *CYP2R1*)	-0.136 (-0.391, 0.118)	0.29
	-	-	-	Metabolism allele score A *(GC, CYP24A1)*	0.029 (-0.130, 0.187)	0.72

	n	ln 25(OHD) β (95% CI)	ln 25(OH)D *p*-value	Instrument(s) for ln 25(OH)D	ln 25(OH)D IV Beta (95% CI)	IV *p*-value
	-	-	-	Gc2 haplotype	0.023 (-0.145, 0.191)	0.79
	-	-	-	Metabolism allele score B *(GC, CYP27B1, CYP24A1)*	-0.013 (-0.166, 0.141)	0.87

5.5 Discussion

5.5.1 Methodological considerations

Choosing the right instruments

The choice of genetic variant as an instrument for the exposure is crucial to MR studies (Lawlor *et al.* 2008). Early studies of MR typically had a protein based exposure expressed by a singular gene, making the choice of gene obvious (Davey Smith *et al.* 2004; Timpson *et al.* 2005). However, vitamin D is a pro-hormone synthesised after UVB exposure and metabolised into an active form. Nevertheless, there are several key genes in the vitamin D synthesis and metabolism pathway that have been found to influence 25(OH)D concentrations. On the basis of their influence, these genes seem obvious candidates to be used as instruments in MR studies on 25(OH)D. Compared with single SNPs as instruments in IV regression, SNPs combined into allele scores/counts had greater power, explained a greater proportion of variance in 25(OH)D, and also in some cases the associations were less sensitive to changes after adjusting for lifestyle covariates.

In an MR study, the genetic variants should have a reliable association with the exposure that is not confounded, and is unrelated with the outcome except through the mechanism of the exposure (Lawlor *et al.* 2008). LD and population stratification can affect and introduce confounding in the relationship of the genetic variant with the exposure. To some extent, social behaviour may also mask the genetic marker-exposure relationship. Based solely on the SNPs associations with 25(OH)D, *DHCR7*, *CYP2R1*, *GC*, *CYP24A*, *CYP27B1* and *SLC45A2* were found to be suitable instruments, and thus can conclude the formal testing of MR (Sheehan *et*

al. 2008). To further address the issue of suitability, the assumptions of MR analysis were quantified where possible by assessing the potential effects of lifestyle, diet, socio-economic variables and health biomarkers.

Ethnic differences in the population as reflected in the differences in allele frequencies may confound the results in IV analysis (Lawlor *et al.* 2008). It was found that the SNP for *DHCR7* was associated with geographical region of residence, and its MAF in the 1958BC appeared to be different from the CEU population frequency in Hapmap (International HapMap Consortium 2003). There are also known ethnic differences in the distribution of *GC* SNPs (Kamboh & Ferrell 1986). Geographical region of residence was controlled for in the models to minimise the effects of population stratification in the IV analysis. However, further work will need to be done to fully understand the extent to which population stratification affects the 25(OH)D genetic determinants.

The second assumption, that the polymorphism is not associated to the confounding variables that affect the 25(OH)D -outcome's relationship, is inferred through Mendel's laws (Davey Smith & Ebrahim 2003). To assess this as much as possible, the second assumption was investigated by testing the SNP associations with lifestyle, dietary and socio-economic variables. Although some associations were evident between genetic variants and diet, lifestyle variables, combining the SNPs into allele scores/counts did reduce confounding (Palmer *et al.* 2011a).

The third assumption of MR, that the genetic variant *Z* is associated only with the outcome *Y* through *X*, cannot be formally tested (Didelez & Sheehan 2007). In this

work, the direct association between the genetic variants and the biomarkers were tested after adjustment for 25(OH)D, since if a relationship had been observed this would be taken as evidence of possible pleiotropy. Based on the analysis, there was no evidence of an association between *Gc2* and tPA outside of the pathway of 25(OH)D, and this was likewise with the relationship between metabolism score (GWAS) and CRP. However, the evaluation was not without merit as the *SLC45A2* SNP was deemed to be an unsuitable as instrument for 25(OH)D, since relationships were observed with some of the biomarkers. Furthermore, when testing the strength of the *SLC45A2* SNP association with 25(OH)D, after adjustment for the biomarkers this relationship did attenuate more than any of the other individual SNPs.

Instruments in IV regression

In the IV analysis on the five haemostatic/inflammation biomarkers, 25(OH)D was not significantly associated with any of the biomarkers when the synthesis count or SNPs were used as instruments. However, there was some evidence that 25(OH)D had an association with CRP and tPA, when using the variants from *GC* as instruments in the IV regression.

For the allele scores/counts, consideration was given to the position in the pathway where the gene was expressed by creating two distinct combined markers, i.e. "synthesis" and "metabolism". The genes *DHCR7* and *CYP2R1* are expressed prior to, or as part of, the 25-hydroxylation to 25(OH)D. The evidence for the gene *DHCR7* suggests that its chosen SNP is associated with lower concentrations of 25(OH)D, and it can be conceptualised that the SNP's underlying causal variant is more efficient at removing *DHC-7* from the vitamin D pathway, therein by reducing

the quantities of *DHC-7* from being synthesised to vitamin D. Likewise with *CYP2R1*, where again the conceptualisation can be that the causal variant is a more efficient catalyst of the 25-hydroxylation of vitamin D, thus increasing quantities of 25(OH)D. However using this conceptualisation, the quantification of the genes expected effects on 25(OH)D expressed past the 25-hydroxylation of vitamin D becomes less certain. As reviewed, evidence suggests that the VDBP affects the bio-availability of 25(OH)D (Chun *et al.* 2010) and so the conceptualisation of what direction *Gc* influences 25(OH)D becomes complex. Similarly, both *CYP27B1* and *CYP24A1* are regulated by other hormones (Horst *et al.* 2005), and the extent of the regulation can be dependent on levels of 25(OH)D, so it is unclear what direction the polymorphisms may truly influence 25(OH)D concentrations.

From the IV analysis, 25(OH)D had a tentative relationship with CRP and tPA, but significance of this relationship depended on the instrument used. The combination of both synthesis and metabolism (non-GWAS) scores together as instruments had slightly more power to detect an effect than other instruments, and it would be expected that an association using these two as instruments would be evident first. However, the borderline significant associations in IV regression were observed when weaker instruments were used than of the metabolism and synthesis instruments together. There would be much more confidence in the IV analysis results, if the relationship of 25(OH)D with CRP and tPA had been observed when the synthesis score was used as an instrument. Since conceptually the roles of the genes behind the synthesis score are more straightforward with regards to 25(OH)D levels, than the genes behind the metabolism scores. However the power calculations revealed that whilst it was unlikely that a relationship between 25(OH)D

and tPA would be observed in IV regression, there was even less chance that it would be seen when using the synthesis score as instrument in a study the size of the 1958BC.

The tests of endogeneity for CRP did suggest that the IV estimate differed from the OLS estimate when using the metabolism scores as instruments. The test is often used in econometrics as a justification for IV regression (Baum 2006). However, the large positive association of 25(OH)D with CRP by IV regression was in keeping with results of the simulation study. In the simulation, when a pleiotropic SNP was used as an instrument, the association size by IV regression was much larger than the true effect size. Therefore, suggesting that instruments using variants from *GC* were particularly not suitable for CRP.

The allele scores were based on coefficients from the same sample, and this can also bias IV results (Pierce *et al.* 2011). The SNPs behind the allele scores were used as single instruments to formally test whether any of the SNPs were unnecessary as instruments, and whether the IV estimates were biased by the weighted scores. In the IV regression, individual SNPs used as instruments gave a comparable result to their associated genetic score, without assuming the effect size or direction of the association. In situations in which individual data are available, it has been shown that single SNPs vs allele scores have similar power as instruments in MR analysis (Palmer *et al.* 2011a; Pierce *et al.* 2011).

The final assumption of MR analysis is that the relationships between genetic variants and exposure, and exposure and outcome, are linear and are unaffected by

statistical interactions (Lawlor *et al.* 2008). No interactions were evident between SNPs and lifestyle variables with 25(OH)D. However, a curved 25(OH)D relationship was observed for the inflammation and haemostatic outcomes, so 25(OH)D was transformed to reduce curvature. There has been ongoing debate about whether a threshold for 25(OH)D concentrations exists for diseases (Bischoff-Ferrari *et al.* 2006; Dawson-Hughes *et al.* 2005; Heaney 2005; Vieth & El-Hajj 2005), which may affect the linearity assumption and the ability to estimate the magnitude of the effect (Didelez & Sheehan 2007; Sheehan *et al.* 2008).

5.5.2 Explanations

The relationship between 25(OH)D concentrations, VDBP, and tPA

In this work, a negative 25(OH)D association was found with tPA. This association was observed in the IV analysis when the instrument was the functionally known *Gc2* haplotype of VDBP. Although 25(OH)D is associated with tPA, an association between VBDP and tPA could also exist, implying that VDBP has a pleiotropic effect on tPA. A possible mechanism between VDBP and tPA, might be via actin filaments, which are released from the vascular smooth muscle cells after their death caused by tissue damage (Janmey *et al.* 1992). Elevated concentrations of actin in the blood can increase blood viscosity to dangerous levels of resistance and can cause clot formation by aggregating platelets (Bogaerts *et al.* 2005). The majority of 25(OH)D is carried by VDBP, however only 5% of the protein is occupied by vitamin D sterols (Laing & Cooke 2005). The binding protein can act as an actin-sequestering agent and lead to its clearance from circulation (Bogaerts *et al.* 2005). The release of tPA occurs in fibrinolysis that forms plasmin from the fibrin-bound

plasminogen (Hoffbrand & Moss 2011a). Some evidence suggests that actin can interact with tPA and other fibrinolytic proteins to affect clot structure and lysis (Janmey *et al.* 1992).

There also is plausible evidence of a 25(OH)D-tPA pathway. In a cross-sectional study, a similar negative 25(OH)D association with tPA remained after adjusting for age and BMI (Jorde *et al.* 2007). However, in a subsequent trial no change was observed in tPA levels after overweight and obese participants took high weekly dosages of cholecalciferol for a year compared with the placebo group, although their 25(OH)D levels had increased (Jorde *et al.* 2010d). Evidence that vitamin D may have a direct influence on fibrinolysis has been observed in-vitro studies, where the expression of vitamin D receptors has been found in vascular smooth muscle cells (Wu-Wong *et al.* 2006; Wu-Wong *et al.* 2007a). In smooth muscle cells from the aorta, expression of PAI-1 mRNA was down-regulated to a maximum of 60% after treatment with active vitamin D analogues, whereas expression of thrombomodulin mRNA was up-regulated (Wu-Wong *et al.* 2007a). Thrombomodulin binds to thrombin on the endothelial cell surface and leads to the expression of tPA, and subsequent breakdown of fibrin by plasmin after the inhibition of PAI-1 (Hoffbrand & Moss 2011a). The expression of tPA was directly observed in rat microvascular cells, which mostly consisted of endothelial cells, after treating with active vitamin D (Puri *et al.* 2000). However, not all in-vitro studies have consistently seen expression of VDR agonists in endothelial cells (Koyama *et al.* 1998; Wu-Wong *et al.* 2006; Wu-Wong *et al.* 2007a). Further evidence of VDR enhancing fibrinolysis and potentially having antithrombotic effects has come from an in-vivo study. In the animal study, VDR knockout mice had higher platelet

aggregation and reduced expression of thrombomodulin mRNA in the aorta, liver and kidney compared with wild type mice (Aihara *et al.* 2004).

The role of tPA is not limited to the fibrinolysis cascade, and its association found in this work might represent another process. tPA is produced by other cells, such as osteoblasts and osteoclasts, that are essential for bone remodelling (Allan & Martin 1995). In osteoblasts cells, tPA was secreted after administration of active vitamin D, and this was confirmed by evidence of the expression of the tPA mRNA (Merchiers *et al.* 1999). Two vitamin D response elements (VDRE) were discovered upstream of the tPA gene, and the mutated VDRE showed a decrease of active vitamin D induction of tPA (Merchiers *et al.* 1999). Treatment of active vitamin D has also up-regulated thrombomodulin mRNA and down-regulated tissue factor in monocyctic leukaemia cells found in marrow (Koyama *et al.* 1998).

The relationship between 25(OH)D concentrations, VDBP, and CRP

The pleiotropic nature of VDBP was also potentially observed with CRP. A detrimental association 25(OH)D with CRP was observed when using the metabolism score (non-GWAS) as instrument, which is heavily weighted in favour of a functional SNP from *GC*. Furthermore, the direction of the effect of vitamin D lowering alleles for the SNPs in the metabolism score were opposite to the relationship observed in the IV regression between 25(OH)D and CRP.

Recent observational evidence does support a relationship between CRP and VDBP. In a study of 6,720 subjects from the Twins UK registry, serum CRP was significantly correlated with serum VDBP (Arora *et al.* 2011). More interestingly, a

SNP from the CRP gene locus that was positively associated with CRP concentrations was also positively associated with serum VDBP. In an animal experiment it was found that in rats who had had a biomaterial patch placed on the gastrointestinal tract, CRP and VDBP RNA were present locally, which was not evident in the control rats (Lobler *et al.* 2002). Less convincing evidence of a relationship between CRP and VDBP was presented in a study following 33 patients who had undergone elective knee surgery (Reid *et al.* 2011). Patients' serum CRP levels were elevated in the days immediately after surgery, whereas VDBP did decrease the same day as the surgery and then started to rise, although the levels were not significantly different to those prior to the surgery. The lack of change in VDBP levels was in contrast to the decrease in 25(OH)D and free 25(OH)D concentrations in the patients after surgery.

The role of VDBP in the inflammatory process is multifaceted and has been established in several areas (Laing & Cooke 2005): the prevention of thrombotic events by binding to actin; the stimulation of chemotaxis by neutrophil; and the activation of macrophages. Neutrophils are the dominant white cell and are recruited by the chemotactic activity of complement activation peptides as part of the acute phase of inflammatory response (Roitt & Rabson 2000). VDBP can greatly enhance the effect of complement peptide C5a and its derivate, and subsequent chemotaxis of neutrophils (Kew & Webster 1988). Macrophages are the key cells in the immunity response (Roitt & Rabson 2000) and a de-glycosylated VDBP can act as their activating factor (Yamamoto & Homma 1991). Macrophages activated by the de-glycosylated VDBP have shown anti-tumour behaviour and can inhibit the

proliferation of endothelial cells from several different tissues, including the aorta (Kanda *et al.* 2002).

Evidence suggests that the active form of vitamin D also has a role in immunity outside VDBP (Mora *et al.* 2008). However, in this chapter the inferred causal positive relationship of 25(OH)D on CRP is in contrast to findings from several trials. So far, trials investigating the inflammatory response after supplementing participants with vitamin D metabolites have observed either no change or a decrease in CRP concentrations (Barnes *et al.* 2011; Jorde *et al.* 2010d; Matias *et al.* 2010; Pittas *et al.* 2007; Wu *et al.* 2011). Trials reporting a decrease in CRP levels have been in patients of hemodialysis after supplementing with analogues of active vitamin D (Matias *et al.* 2010; Wu *et al.* 2011). Randomised control trials (RCT) based on healthy subjects have observed no changes in CRP concentrations after supplementing with cholecalciferol. The study reported by (Barnes *et al.* 2011) had two age groups from the ages of 20-40 years and ≥ 65 years, supplemented with three dosages of cholecalciferol over winter. In the younger group, the 25(OH)D concentrations were lower in all but the highest dosage group of cholecalciferol, and for the highest supplemented group the 25(OH)D concentrations had not changed. In a three-year RCT with participants aged ≥ 65 years, analysis was stratified by baseline glucose tolerance and neither group had a change in CRP levels prior to and after treatment of cholecalciferol and calcium citrate, although 25(OH)D levels had increased for those treated (Pittas *et al.* 2007). Similarly, no change was observed in CRP levels after overweight and obese participants took 40,000 IU of cholecalciferol weekly for a year, compared with the placebo group, although their 25(OH)D levels had increased (Jorde *et al.* 2010d).

There has been one population based observational study that reported a positive association between 25(OH)D and CRP (Sacheck *et al.* 2011), whilst similar studies have reported no association (Ganji *et al.* 2011; Jorde *et al.* 2007; Michos *et al.* 2009) or an inverse association (Ngo *et al.* 2010). The US study reporting a positive 25(OH)D relationship with CRP was based on 263 ethnically diverse children aged 9-14 years and remained significant after adjusting for BMI, sex and puberty (Sacheck *et al.* 2011). However, the analysis of the much larger group of 5867 children aged 12-19 years, surveyed in NHANES 2001-2006, observed no difference in CRP concentrations across the in-study tertiles of 25(OH)D (Ganji *et al.* 2011).

5.5.3 Conclusions

There are several ways in which assumptions in a MR (or IV) study maybe violated by poor instrument choice. Thorough statistical testing can identify unsuitable genes as instruments for an exposure, despite the inability to formally test all possible violations. The final choice of gene variant as an instrument and how it is used with other variants in MR analyses ultimately relies on knowledge of its function. The combination of genetic markers as instruments for 25(OH)D were defined by the position of the gene expression in the vitamin D pathway, and quantification of how the gene may influence 25(OH)D levels. On this basis, the synthesis count was considered a more reliable instrument for 25(OH)D than the metabolism scores or *GC* variants. No evidence of a 25(OH)D association with tPA was observed in the IV analysis, when the synthesis count or SNPs were used as instruments. Nevertheless, findings in this investigation did tentatively suggest that 25(OH)D has a protective relationship with tPA, despite uncertainty over the use of *GC* variants as

an instrument. Due to uncertainty, there is sufficient doubt on whether this is truly evidence of causal relationship between vitamin D and CVD, or if it is evidence of pleiotropic gene or a chance finding. Therefore, more evidence is required from larger MR studies using the synthesis count or SNPs as instruments for 25(OH)D and from RCTs.

6 Discussion

6.1 Conclusions

The aim of this thesis was to investigate the associations of nutritional vitamin D status (25(OH)D) with novel biomarkers of sub-clinical cardiovascular disease (CVD) that pertain to inflammation and haemostasis, in participants of the British 1958 birth cohort (1958BC). Vitamin D insufficiency is widespread in the UK and the deficiency is common and unnecessary, as is death caused by CVD (Lanham-New *et al.* 2011; Scarborough *et al.* 2010). However, deficient levels of vitamin D are easily preventable by raising vitamin D intake throughout the year, as the cost of vitamin D supplementation is relatively small. Evidence of vitamin D having a beneficial influence on inflammation and haemostasis would have broad public health implications, in terms of contributing to the prevention of a common disease that blights the general population. Despite accumulating evidence from observational studies, further evidence is required before the role of vitamin D in CVD is conclusive. Evidence of a causal role remains to found by randomised control trials (RCTs), and to some extent by large Mendelian Randomisation (MR) studies that use reliable genetic proxies as instruments for 25(OH)D levels in the analysis.

In this thesis, different methodologies using cross-sectional data were applied to fulfil the objectives. This included a MR study that can be interpreted to infer causality, if it is felt that all of the assumptions made in the MR analysis are true. Some evidence for a protective association of 25(OH)D with tPA was seen in the results of all the approaches, tentatively suggesting a role in CVD through the

mechanism of fibrinolysis. For the other haemostatic and inflammatory biomarkers, inconsistent findings were observed in the results of the different methodologies applied. A cautious interpretation has been given in the context of the criteria used to assess causality as outlined below.

6.1.1 Judging the evidence

In order, to judge the existence of a cause-effect relationship between vitamin D status and the biomarkers of inflammation and haemostasis, judgement was made on whether the associations were due to chance, bias and/or confounding, and whether the methods used infers temporality of the effects (Hennekens & Buring 1987c). Following these considerations, the evidence was judged in terms of shape and strength of the association, consistency with other investigations and biological credibility.

Chance finding

A relationship was observed between 25(OH)D and tissue plasminogen activator (tPA) from the three analytical approaches used in this investigation of multiple linear regression, mediation analysis, and MR analysis using instrumental variable (IV) regression. A protective association of 25(OH)D with tPA was found when using multiple linear regression and MR analysis, and a relationship was inferred in the mediation analysis, suggesting that the significant association was not by chance alone. In contrast, conflicting results were observed for C-reactive protein (CRP), D-dimer and fibrinogen, which implies that the significance of these associations were more likely to be due to chance. However, in the mediation analysis it was

expected that seasonal variations of the biomarkers were not affected by confounding from adiposity so the adjustment for number of confounders was less, compared with the multiple linear regression models. Therefore, the power to detect an effect was probably greater in the mediation analysis, compared with the multiple linear regression models, and may explain in part why associations with D-dimer and fibrinogen were observed. The power to detect an effect by IV regression was very limited in the 1958BC, and large samples are required before causality can be inferred.

Bias

The main strength of this work lays in the data resource of the 1958BC, and therefore, the ability to apply multiple methodologies to the data. The 1958BC is a large sample with extensive information on the cohort members' lifestyle, economic situation, and health that were relevant to this thesis. This considerable resource meant that many factors could be taken into account in the statistical models, and where relevant exclusions could be made (i.e. excluding participants on cardiovascular medication). Furthermore, the survey took place over an 18 month period that made analytical work on seasonal variations possible. In addition, many of the variables in the 1958BC biomedical survey underwent thorough quality control prior to and during this thesis, so that results found here can be compared with other studies. The 25(OH)D concentrations in the 1958BC were standardized to the mean of the Vitamin D External Quality Assessment Scheme (Hyppönen et al. 2007). The DNA were subjected to specific quality control procedures and imputed as outlined in Chapter 2.

The 1958BC does have a few limitations that need mentioning, which may have introduced bias to the study design. It has suffered sample attrition since its inception, with disadvantaged groups and ethnic minorities are under-represented in biomedical survey (Atherton et al. 2008). However, it is still considered largely representative of the original cohort and broadly comparable with adults aged 45-49 years old surveyed in the 2001 census (Atherton et al. 2008). The lack of ethnic minorities has also worked to the advantage of 1958BC as it is desirable to have ethnicity homogenous populations for genetic studies. The sample homogeneity meant that the 1958BC participants were genotyped by large consortiums; data that this thesis has utilised. Nevertheless, it means that the findings here may not be generalised to other populations without further work. The data were restricted to individuals of European ancestry and further work is required to investigate whether in individuals of non-European ancestry the relationships found persist, and whether the same genes found to influence vitamin D status are as influential. Furthermore, the analyses were restricted to individuals free of CVD, so it uncertain whether the relationships observed in this work can be generalised to older age groups or individuals with CVD.

Confounding

The different methods used in this thesis had different approaches for controlling confounding. For the analysis presented in Chapter 3, the multiple linear regression models were adjusted for relevant confounders. However, it may be possible that low vitamin D status and elevated tPA levels were caused by a mutual unknown factor not accounted for or the association was affected by residual confounding. A known confounder of vitamin D and cardiovascular health is obesity. Adiposity was

accounted for in the models with measures of body mass index (BMI) and waist circumference, however this may not have adequately controlled for obesity and resulted in residual confounding (Glymour & Greenland 2008). Adiposity was not controlled for in the models used to explore seasonal patterns of the outcomes, as it would be unexpected that adiposity would vary by season or affect the seasonal variations of the biomarkers. However, it is not possible to eliminate confounding from the associations due to the cross-sectional design of study and the framework of the analysis. Therefore the seasonal variations observed in 25(OH)D and tPA maybe confounded. To my knowledge, mediation analysis has not been adapted for seasonal variations before, and more work is required to better understand its strengths and limitations in this context.

Genetic confounding may affect the findings from the genome-wide association study (GWAS) presented in Chapter 4. Furthermore, as the genetic variants found in the GWAS were used as instruments in IV regression, genetic confounding may have also affected the findings from the MR study presented in Chapter 5. Genetic variants found to be associated with a trait in GWAS are unlikely to be the underlying causal variant of the different levels in a trait, but in linkage disequilibrium with the causal variant (McCarthy et al. 2008). Nevertheless, GWAS do identify loci that influence the outcome that further studies, such as MR studies, can build on. The genetic variants found to be associated with 25(OH)D were all near or within genes that were expressed during the vitamin D pathway, which strengthen the plausibility of the GWAS findings. Underlying population differences within the study and across the studies may cause false findings and

techniques were used in the GWAS to reduce the possibility of this occurring (Evans 2008).

More weight is given to the evidence presented in Chapter 5, than to the evidence presented in Chapter 3. The analysis of MR has been equated to a RCT since DNA is randomly allocated, much like an intervention in a trial. Due to the random allocation of DNA this means that confounders should be equally distributed across individuals (Hingorani & Humphries 2005). When all of the assumptions are met, the findings from a MR study are the closest to a RCT in an observational study setting without the expense and complications of running a RCT (Nitsch et al. 2006). Moreover, extensive research is underway in the application of MR and has been applied in numerous studies (Burgess *et al.* 2010; Burgess *et al.* 2011; Palmer *et al.* 2011b; Pfister *et al.* 2011; van Durme *et al.* 2011).

Temporality of the effects

The data used in this work was cross-sectional and, so the order of when the changes in 25(OH)D and the biomarkers occurred are unknown for associations found from the analyses presented in Chapter 3. The approach of using seasonal variations in mediation analysis to some extent addresses the unknown order of effects. The external changes in the environment that are the causes of the seasonal patterns must logically precede the change in 25(OH)D and tPA concentrations. The cause of the seasonal variations in 25(OH)D are well understood, therefore it is more likely that the seasonal variations in 25(OH)D would influence and precede variation in tPA. In the MR study, the association of the exposure and outcome is not affected by reverse

causation since genetic variants are used as a proxy for the exposure, which have been allocated prior to inception and subsequent development of disease.

Shape and strength of association

A consistent dose-response relationship between 25(OH)D concentrations and tPA was observed after successive adjustments in multiple linear regression models for lifestyle and socio-economic variables and measures of adiposity, as presented in Chapter 3. In contrast, there was evidence of a curved relationship between 25(OH)D and D-dimer and fibrinogen. Observing a dose-response relationship in itself does not confirm a causal relationship; however it does add credibility to some degree that one may exist (Hennekens & Buring 1987c). Thresholds of 25(OH)D concentrations are required for health and the level of insufficiency is vigorously debated, yet conversely vitamin D toxicity can lead to death, as reviewed in Chapter 1 (Heaney & Holick 2011; Henry *et al.* 2010; Institute of Medicine 2011; Jones 2008). The existence of thresholds may suggest that a non-linear relationships between 25(OH)D and outcomes are likely, since it is fairly common that an exposure below a certain threshold will have a very different shape relationship with a disease than to what it is above that threshold (Rothman *et al.* 2008).

It has been argued that the relationship between a cause and effect should be strong in terms of magnitude of risk, whereas a weak relationship might be the result of confounding or bias (Rothman *et al.* 2008). Comparing with the other biomarkers, the strength of the association between 25(OH)D and tPA was strongest, which again does add more credibility to this relationship than the others. However, as with observing a dose-response relationship, a strong relationship between exposure and

outcome does not conclude causality, furthermore, a weak relationship does not eliminate the possibility of causality either (Rothman *et al.* 2008).

Consistency with other studies and biological credibility

There is observational evidence that vitamin D deficiency is a risk factor with CVD and subsequent mortality, as discussed in Chapter 1. In the progression of CVD, mechanisms that form a stable plug after vascular damage can also lead to thrombi, causing intermitted blood flow, or a single large thrombus that blocks an artery (Ross 1993). An essential factor in the feedback of the haemostatic response is tPA, as it prevents thrombus from forming after vascular damage (Hoffbrand & Moss 2011a). Elevated tPA levels can be taken as an indication of impaired fibrinolysis (Lowe 1997), and are associated with an increased incidence risk of coronary heart disease (Lowe *et al.* 2004).

It is biologically plausible that vitamin D does have a role in haemostasis, although very few epidemiology studies have investigated this relationship. In a small observational study of 206 subjects, a protective association of vitamin D status with tPA was observed after adjustment for age, gender, BMI and smoking status (Jorde *et al.* 2007). A RCT in obese and overweight subjects, observed no difference in tPA levels after supplementing subjects for a year with high dosage cholecalciferol, compared with those not supplemented with cholecalciferol (Jorde et al. 2010c). There is in-vitro evidence that vitamin D receptors have been found to exist in vascular smooth muscle cells (Wu-Wong *et al.* 2006; Wu-Wong *et al.* 2007b), and that the inhibitor of tPA is down regulated in vascular smooth muscle cells after treatment with active vitamin D analogues (Wu-Wong *et al.* 2007a). In animal

studies, after treatment with a calcitriol analogue the expression of tPA was observed in microvascular cells (Puri *et al.* 2000). Evidence also suggests that VDR knockout mice have an impaired fibrinolytic system (Aihara *et al.* 2004).

Further methodological considerations

Consideration must also be given to whether all of the assumptions made in the MR analysis were true in the context of 25(OH)D and the inflammation/haemostatic biomarkers. One assumption that appears particularly vulnerable, is the gene used as proxy for the exposure is only related to the outcome through the mechanism of the exposure (Lawlor *et al.* 2008). No association was seen between 25(OH)D and tPA when the synthesis count was used as an instrument in the MR analysis, which conceptually is a more robust instrument for 25(OH)D than the metabolism score or variants from *GC*. The protective association of 25(OH)D with tPA was only found when variants from *GC* were used as a proxy for 25(OH)D concentrations. It was not statistically proven that *GC* was associated with tPA outside of its relationship with 25(OH)D. In contrast, the evidence of *GC*'s pleiotropic nature is clearer for CRP and 25(OH)D. Concentrations of 25(OH)D were detrimentally associated with CRP when variants from *GC* formed part of the proxy for 25(OH)D. However, the significance of this association was contrary to the null findings presented in Chapter 3, and findings from RCTs (Barnes *et al.* 2011; Jorde *et al.* 2010d; Pittas *et al.* 2007). There is a clear biological case of vitamin D binding protein (VBDP) role in inflammation, but a case can also be made that it has a role in haemostasis, as discussed in Chapter 5.

6.1.2 Future work

Further evidence of the role vitamin D plays in cardiovascular health may be drawn from several sources. There is a need for high quality RCTs to investigate the mechanism and the role of vitamin D in cardiovascular health. RCTs are the gold standard for evidence as they "provide a degree of assurance about the validity of a result that is simply not possible with any observational design option" (Hennekens & Buring 1987b). A recent systematic review found that the majority of RCTs in this area were too heterogeneous to be meta-analysed (Pittas et al. 2010a).

MR studies can in part fill the evidence gap for vitamin D's role. The next round of GWAS on vitamin D status is underway and should provide novel genetic variants that can be used as instruments for 25(OH)D in MR studies. More power is gained when SNPs are combined into a single instrument, as demonstrated by the power calculations presented in Chapter 5. So far the genetic variants found to influence 25(OH)D have a relatively small effect, and MR studies will require a collaborative effort to have sufficient power to detect an association. Consideration will also need to be given as to what genetic variants should be used as instruments in MR studies on vitamin D, since evidence suggests that the gene for VDBP is pleiotropic and has many overlapping functions with vitamin D. Furthermore, the other variants used in the metabolism score influence on 25(OH)D were not as conceptually straightforward as the synthesis variants. However, there is less chance that a single pleiotropic gene will dominate an instrument when using multiple genetic variants as instruments, as it is unlikely that the combined genetic score will exhibit pleiotropy in the same manor (Davey Smith 2011).

Opportunities will also arise as the 1958BC cohort members grow older to do prospective studies on vitamin D status and the incidence of CVD. In future, it is hoped that another biomedical survey will take place providing invaluable opportunities to investigate the relationship of vitamin D status and the progression of CVD over time. It is also hoped that there will be opportunities to measure additional biomarkers from the stored blood of the 1958BC participants. So far 1958BC biomedical survey serum measures were restricted to common biomarkers. For this work it meant that the vitamin D exposure was solely limited to total 25-hydroxyvitamin D, however if there had been serum data on parathyroid hormone then thresholds of optimal vitamin D status with the inflammation/haemostatic biomarkers could have been explored. Furthermore, serum measures of VDBP could have been used to directly investigate its relationship with the haemostatic/inflammation biomarkers. Depending on whether a relationship was found between VDBP and biomarkers, this would have helped determine the suitability of the *GC* variants as proxies for 25(OH)D in the MR study.

References

Aaronson, P.I. & Ward, J.P.T. 2007. *The cardiovascular system at a glance*, Third ed. Blackwell Publishing.

Abecasis, G. R.: METAL Package. Website: www.sph.umich.edu/csg/abecasis/metal/. Last date accessed: 2009

Adam, S.S., Key, N.S., & Greenberg, C.S. 2009. D-dimer antigen: current concepts and future prospects. *Blood*, 113, (13) 2878-2887

Ahn, J., Albanes, D., Berndt, S.I., Peters, U., Chatterjee, N., Freedman, N.D., Abnet, C.C., Huang, W.Y., Kibel, A.S., Crawford, E.D., Weinstein, S.J., Chanock, S.J., Schatzkin, A., & Hayes, R.B. 2009. Vitamin D-related Genes, Serum Vitamin D Concentrations, and Prostate Cancer Risk. *Carcinogenesis*, 30, (5) 769-776

Ahn, J., Yu, K., Stolzenberg-Solomon, R., Simon, K.C., McCullough, M.L., Gallicchio, L., Jacobs, E.J., Ascherio, A., Helzlsouer, K., Jacobs, K.B., Li, Q., Weinstein, S.J., Purdue, M., Virtamo, J., Horst, R., Wheeler, W., Chanock, S., Hunter, D.J., Hayes, R.B., Kraft, P., & Albanes, D. 2010. Genome-wide association study of circulating vitamin D levels. *Hum.Mol.Genet.*, 19, (13) 2739-2745

Aihara, K., Azuma, H., Akaike, M., Ikeda, Y., Yamashita, M., Sudo, T., Hayashi, H., Yamada, Y., Endoh, F., Fujimura, M., Yoshida, T., Yamaguchi, H., Hashizume, S., Kato, M., Yoshimura, K., Yamamoto, Y., Kato, S., & Matsumoto, T. 2004. Disruption of nuclear vitamin D receptor gene causes enhanced thrombogenicity in mice. *J.Biol.Chem.*, 279, (34) 35798-35802

Albers, G.W., Bates, V.E., Clark, W.M., Bell, R., Verro, P., & Hamilton, S.A. 2000. Intravenous tissue-type plasminogen activator for treatment of acute stroke: the Standard Treatment with Alteplase to Reverse Stroke (STARS) study. *JAMA*, 283, (9) 1145-1150

Allan, E.H. & Martin, T.J. 1995. The plasminogen activator inhibitor system in bone cell function. *Clin.Orthop.Relat Res.* (313) 54-63

Allison, P.D. 2002. *Missing data* Thousand Oaks, Calif, Sage.

Altshuler, D., Daly, M.J., & Lander, E.S. 2008. Genetic mapping in human disease. *Science*, 322, (5903) 881-888

American Thoracic Society 1987. Standardization of spirometry--1987 update. *Am.Rev.Respir.Dis.*, 136, (5) 1285-1298

Andreassi, L., Simoni, S., Fiorini, P., & Fimiani, M. 1987. Phenotypic characters related to skin type and minimal erythemal dose. *Photodermatol.*, 4, (1) 43-46

Armas, L.A., Hollis, B.W., & Heaney, R.P. 2004. Vitamin D2 is much less effective than vitamin D3 in humans. *J.Clin.Endocrinol.Metab*, 89, (11) 5387-5391

Armitage, P. & Berry, G. 1994, "Data Editing," *In Statistical Methods in Medical Research*, Third ed. Osney Mead, Oxford, UK: Blackwell Science Ltd, pp. 386-401.

Arora, P., Garcia-Bailo, B., Dastani, Z., Brenner, D., Villegas, A., Malek, S., Spector, T.D., Richards, B., El-Sohemy, A., Karmali, M., & Badawi, A. 2011. Genetic polymorphisms of innate immunity-related inflammatory pathways and their association with factors related to type 2 diabetes. *BMC.Med.Genet.*, 12, (1) 95

Arslan, A.A., Clendenen, T.V., Koenig, K.L., Hultdin, J., Enquist, K., Agren, A., Lukanova, A., Sjodin, H., Zeleniuch-Jacquotte, A., Shore, R.E., Hallmans, G., Toniolo, P., & Lundin, E. 2009. Circulating vitamin d and risk of epithelial ovarian cancer. *J.Oncol.*, 2009,

Ashwell, M., Stone, E.M., Stolte, H., Cashman, K.D., Macdonald, H., Lanham-New, S., Hiom, S., Webb, A., & Fraser, D. 2010. UK Food Standards Agency Workshop Report: an investigation of the relative contributions of diet and sunlight to vitamin D status. *Br.J.Nutr.*, 104, (4) 603-611

Atherton, K., Berry, D.J., Parsons, T., Macfarlane, G.J., Power, C., & Hyppönen, E. 2009. Vitamin D and chronic widespread pain in a white middle-aged British population: evidence from a cross-sectional population survey. *Ann.Rheum.Dis.*, 68, (6) 817-822

Atherton, K., Fuller, E., Shepherd, P., Strachan, D.P., & Power, C. 2008. Loss and representativeness in a biomedical survey at age 45 years: 1958 British birth cohort. *J.Epidemiol.Community Health*, 62, (3) 216-223

Bailey, R., Cooper, J.D., Zeitels, L., Smyth, D.J., Yang, J.H., Walker, N.M., Hyppönen, E., Dunger, D.B., Ramos-Lopez, E., Badenhoop, K., Nejentsev, S., & Todd, J.A. 2007. Association of the vitamin D metabolism gene CYP27B1 with type 1 diabetes. *Diabetes*, 56, (10) 2616-2621

Balding, D.J. 2006. A tutorial on statistical methods for population association studies. *Nat.Rev.Genet.*, 7, (10) 781-791

Balsa, J.A., Iglesias, B., Peromingo, R., Conde, S., Vazquez, C., San-Millan, J.L., & Botella-Carretero, J.I. 2010. Vitamin D receptor polymorphisms in secondary hyperparathyroidism after scopinaro's biliopancreatic diversion. *Obes.Surg.*, 20, (10) 1415-1421

Barnes, M.S., Horigan, G., Cashman, K.D., Hill, T.R., Forsythe, L.K., Lucey, A.J., McSorley, E.M., Kiely, M., Bonham, M.P., Magee, P.J., Strain, J.J., & Wallace, J.M. 2011. Maintenance of wintertime vitamin D status with cholecalciferol supplementation is not associated with alterations in serum cytokine concentrations among apparently healthy younger or older adults. *J.Nutr.*, 141, (3) 476-481

Baron, R.M. & Kenny, D.A. 1986. The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *J.Pers.Soc.Psychol.*, 51, (6) 1173-1182

Baroncelli, G.I., Bereket, A., El, K.M., Audi, L., Cesur, Y., Ozkan, B., Rashad, M., Fernández-Cancio, M., Weisman, Y., Saggese, G., & Hochberg, Z. 2008. Rickets in

the Middle East: role of environment and genetic predisposition. *J.Clin.Endocrinol.Metab*, 93, (5) 1743-1750

Barrett, J.C., Clayton, D.G., Concannon, P., Akolkar, B., Cooper, J.D., Erlich, H.A., Julier, C., Morahan, G., Nerup, J., Nierras, C., Plagnol, V., Pociot, F., Schuilenburg, H., Smyth, D.J., Stevens, H., Todd, J.A., Walker, N.M., & Rich, S.S. 2009. Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. *Nat.Genet.*

Baum, C. F. 2006, "Instrumental-variables estimators," *In An Introduction to Modern Econometrics using Stata*, College Station, Texas, US: Stata Press, pp. 185-219.

Baum, C.F., Schaffer, M.E., & Stillman, S. 2003. Instrumental variables and GMM: Estimation and testing. *Stata Journal*, 3, (1) 1-31

Baum, C.F., Schaffer, M.E., & Stillman, S. 2007. Enhanced routines for instrumental variables/generalized method of moments estimation and testing. *Stata Journal*, 7, (4) 465-506

Beer, T.M., Venner, P.M., Ryan, C.W., Petrylak, D.P., Chatta, G., Dean, R.J., Chi, K.N., Curd, J.G., & DeLoughery, T.G. 2006. High dose calcitriol may reduce thrombosis in cancer patients. *Br.J.Haematol.*, 135, (3) 392-394

Bell, N.H., Epstein, S., Greene, A., Shary, J., Oexmann, M.J., & Shaw, S. 1985. Evidence for alteration of the vitamin D-endocrine system in obese subjects. *J.Clin.Invest*, 76, (1) 370-373

Benjamini, Y. & Hochberg, Y. 1995. Controlling the False Discovery Rate - A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B-Methodological*, 57, (1) 289-300

Bergwitz, C. & Juppner, H. 2010. Regulation of phosphate homeostasis by PTH, vitamin D, and FGF23. *Annu.Rev.Med.*, 61, 91-104

Berry, D. & Hyppönen, E. 2011. Determinants of vitamin D status: focus on genetic variations. *Curr.Opin.Nephrol.Hypertens.*, 20, (4) 331-336

Biancuzzo, R.M., Young, A., Bibuld, D., Cai, M.H., Winter, M.R., Klein, E.K., Ameri, A., Reitz, R., Salameh, W., Chen, T.C., & Holick, M.F. 2010. Fortification of orange juice with vitamin D(2) or vitamin D(3) is as effective as an oral supplement in maintaining vitamin D status in adults. *Am.J.Clin.Nutr.*, 91, (6) 1621-1626

Bikle, D.D., Gee, E., Halloran, B., Kowalski, M.A., Ryzen, E., & Haddad, J.G. 1986a. Assessment of the free fraction of 25-hydroxyvitamin D in serum and its regulation by albumin and the vitamin D-binding protein. *J.Clin.Endocrinol.Metab*, 63, (4) 954-959

Bikle, D.D., Halloran, B.P., Gee, E., Ryzen, E., & Haddad, J.G. 1986b. Free 25-hydroxyvitamin D levels are normal in subjects with liver disease and reduced total 25-hydroxyvitamin D levels. *J.Clin.Invest*, 78, (3) 748-752

Bingham, C., Arbogast, B., Guillaume, G.C., Lee, J.K., & Halberg, F. 1982. Inferential Statistical-Methods for Estimating and Comparing Cosinor Parameters. *Chronobiologia*, 9, (4) 397-439

Bischoff-Ferrari, H.A., Giovannucci, E., Willett, W.C., Dietrich, T., & Dawson-Hughes, B. 2006. Estimation of optimal serum concentrations of 25-hydroxyvitamin D for multiple health outcomes. *Am.J.Clin.Nutr.*, 84, (1) 18-28

Blann, A.D. 2006. Plasma von Willebrand factor, thrombosis, and the endothelium: the first 30 years. *Thromb.Haemost.*, 95, (1) 49-55

Bogaerts, I., Verboven, C., Van Baelen, H., & Bouillon, R. 2005, "New Aspects of DBP," *In Vitamin D*, 2nd ed ed. D. Feldman, F. H. Glorieux, & J. W. Pike, eds., Amsterdam: Elsevier Academic Press, pp. 135-152.

Bonjour, J.P., Trechsel, U., Granzer, E., Klopffer, G., Muller, K., & Scholler, D. 1987. The increase in skin 7-dehydrocholesterol induced by an hypocholesterolemic agent is associated with elevated 25-hydroxyvitamin D3 plasma level. *Pflugers Arch.*, 410, (1-2) 165-168

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. 2009a, "Fixed-Effect Model," *In Introduction to meta-analysis*, Chichester, West Sussex, UK: John Wiley & Sons, pp. 63-67.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. 2009b, "Meta-analysis methods based on direction and p-values," *In Introduction to meta-analysis*, Chichester, West Sussex, UK: John Wiley & Sons, pp. 325-330.

Botstein, D. & Risch, N. 2003. Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. *Nat.Genet.*, 33 Suppl, 228-237

Bouillon, R., Carmeliet, G., Verlinden, L., van, E.E., Verstuyf, A., Luderer, H.F., Lieben, L., Mathieu, C., & Demay, M. 2008. Vitamin D and human health: lessons from vitamin D receptor null mice. *Endocr.Rev*, 29, (6) 726-776

Bounameaux, H., Cirafici, P., de, M.P., Schneider, P.A., Slosman, D., Reber, G., & Unger, P.F. 1991. Measurement of D-dimer in plasma as diagnostic aid in suspected pulmonary embolism. *Lancet*, 337, (8735) 196-200

Bransby, E.R., Berry, W.T., & Taylor, D.M. 1964. Study of the vitamin-D intakes of infants in 1960. *Br.Med.J.*, 1, (5399) 1661-1663

Braun, A., Bichlmaier, R., & Cleve, H. 1992. Molecular analysis of the gene for the human vitamin-D-binding protein (group-specific component): allelic differences of the common genetic GC types. *Hum.Genet.*, 89, (4) 401-406

Brien, S.E., Ronksley, P.E., Turner, B.J., Mukamal, K.J., & Ghali, W.A. 2011. Effect of alcohol consumption on biological markers associated with risk of coronary heart disease: systematic review and meta-analysis of interventional studies. *BMJ*, 342, d636

Briese, S., Wiesner, S., Will, J.C., Lembcke, A., Opgen-Rhein, B., Nissel, R., Wernecke, K.D., Andreae, J., Haffner, D., & Querfeld, U. 2006. Arterial and cardiac disease in young adults with childhood-onset end-stage renal disease-impact of calcium and vitamin D therapy. *Nephrol.Dial.Transplant.*, 21, (7) 1906-1914

British Paediatric Association 1956. Hypercalcaemia in Infants and Vitamin D. *BMJ*, 2, (4985) 149

Brock, K.E., Huang, W.Y., Fraser, D.R., Ke, L., Tseng, M., Mason, R.S., Stolzenberg-Solomon, R.Z., Freedman, D.M., Ahn, J., Peters, U., McCarty, C., Hollis, B.W., Ziegler, R.G., Purdue, M.P., & Graubard, B.I. 2011. Diabetes prevalence is associated with serum 25-hydroxyvitamin D and 1,25-dihydroxyvitamin D in US middle-aged Caucasian men and women: a cross-sectional analysis within the Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial. *Br.J.Nutr.*, 106, (3) 339-344

Brown, W.V., Fujioka, K., Wilson, P.W., & Woodworth, K.A. 2009. Obesity: why be concerned? *Am.J.Med.*, 122, (4 Suppl 1) S4-11

Bu, F.X., Armas, L., Lappe, J., Zhou, Y., Gao, G., Wang, H.W., Recker, R., & Zhao, L.J. 2010. Comprehensive association analysis of nine candidate genes with serum 25-hydroxy vitamin D levels among healthy Caucasian subjects. *Hum.Genet.*, 128, (5) 549-556

Buckley, D.I., Fu, R., Freeman, M., Rogers, K., & Helfand, M. 2009. C-reactive protein as a risk factor for coronary heart disease: a systematic review and meta-analyses for the U.S. Preventive Services Task Force. *Ann.Intern.Med.*, 151, (7) 483-495

Bulter, N.R. & Bonham, D.G. 1963. *Perinatal Mortality* E & S Livingstone Ltd.

Burgaz, A., Byberg, L., Rautiainen, S., Orsini, N., Hakansson, N., Arnlov, J., Sundstrom, J., Lind, L., Melhus, H., Michaelsson, K., & Wolk, A. 2011. Confirmed hypertension and plasma 25(OH)D concentrations amongst elderly men. *J.Intern.Med.*, 269, (2) 211-218

Burgess, S., Seaman, S., Lawlor, D.A., Casas, J.P., & Thompson, S.G. 2011. Missing data methods in mendelian randomization studies with multiple instruments. *Am.J.Epidemiol.*, 174, (9) 1069-1076

Burgess, S., Thompson, S.G., Andrews, G., Samani, N.J., Hall, A., Whincup, P., Morris, R., Lawlor, D.A., Davey, S.G., Timpson, N., Ebrahim, S., Ben-Shlomo, Y., Davey, S.G., Timpson, N., Brown, M., Ricketts, S., Sandhu, M., Reiner, A., Psaty, B., Lange, L., Cushman, M., Hung, J., Thompson, P., Beilby, J., Warrington, N., Palmer, L.J., Nordestgaard, B.G., Tybjaerg-Hansen, A., Zacho, J., Wu, C., Lowe, G., Tzoulaki, I., Kumari, M., Sandhu, M., Yamamoto, J.F., Chiodini, B., Franzosi, M., Hankey, G.J., Jamrozik, K., Palmer, L., Rimm, E., Pai, J., Psaty, B., Heckbert, S., Bis, J., Anand, S., Engert, J., Collins, R., Clarke, R., Melander, O., Berglund, G., Ladenvall, P., Johansson, L., Jansson, J.H., Hallmans, G., Hingorani, A., Humphries, S., Rimm, E., Manson, J., Pai, J., Watkins, H., Clarke, R., Hopewell, J., Saleheen, D., Frossard, R., Danesh, J., Sattar, N., Robertson, M., Shepherd, J., Schaefer, E.,

Hofman, A., Witteman, J.C., Kardys, I., Ben-Shlomo, Y., Davey, S.G., Timpson, N., de, F.U., Bennet, A., Sattar, N., Ford, I., Packard, C., Kumari, M., Manson, J., Lawlor, D.A., Davey, S.G., Anand, S., Collins, R., Casas, J.P., Danesh, J., Davey, S.G., Franzosi, M., Hingorani, A., Lawlor, D.A., Manson, J., Nordestgaard, B.G., Samani, N.J., Sandhu, M., Smeeth, L., Wensley, F., Anand, S., Bowden, J., Burgess, S., Casas, J.P., Di, A.E., Engert, J., Gao, P., Shah, T., Smeeth, L., Thompson, S.G., Verzilli, C., Walker, M., Whittaker, J., Hingorani, A., & Danesh, J. 2010. Bayesian methods for meta-analysis of causal relationships estimated using genetic instrumental variables. *Stat.Med.*, 29, (12) 1298-1311

Carter, G.D. 2009. 25-Hydroxyvitamin D assays: the quest for accuracy. *Clin.Chem.*, 55, (7) 1300-1302

Carter, G.D., Berry, J.L., Gunter, E., Jones, G., Jones, J.C., Makin, H.L., Sufi, S., & Wheeler, M.J. 2010. Proficiency Testing of 25-Hydroxyvitamin D (25-OHD) Assays. *J.Steroid Biochem.Mol.Biol.*

Carter, K.W., McCaskie, P.A., & Palmer, L.J. 2008. SimHap GUI: an intuitive graphical user interface for genetic association analysis. *BMC Bioinformatics*, 9, 557

Casas, J.P., Shah, T., Hingorani, A.D., Danesh, J., & Pepys, M.B. 2008. C-reactive protein and coronary heart disease: a critical review. *J.Intern.Med.*, 264, (4) 295-314

Casella, G. & Berger, R. L. 1990, "Point Estimation," *In Statistical inference*, Pacific Grove, Calif: Brooks/Cole Pub. Co, pp. 284-344.

Cawthon, P.M., Parimi, N., Barrett-Connor, E., Laughlin, G.A., Ensrud, K.E., Hoffman, A.R., Shikany, J.M., Cauley, J.A., Lane, N.E., Bauer, D.C., Orwoll, E.S., & Cummings, S.R. 2010. Serum 25-hydroxyvitamin D, parathyroid hormone, and mortality in older men. *J.Clin.Endocrinol.Metab*, 95, (10) 4625-4634

Centre for Longitudinal Studies: National Child Development Study. Website: www.cls.ioe.ac.uk. Last date accessed: 2009

Chapuy, M.C., Preziosi, P., Maamer, M., Arnaud, S., Galan, P., Hercberg, S., & Meunier, P.J. 1997. Prevalence of vitamin D insufficiency in an adult normal population. *Osteoporos.Int.*, 7, (5) 439-443

Chen, L., Davey Smith G., Evans, D.M., Cox, A., Lawlor, D.A., Donovan, J., Yuan, W., Day, I.N., Martin, R.M., Lane, A., Rodriguez, S., Davis, M., Zuccolo, L., Collin, S.M., Hamdy, F., Neal, D., & Lewis, S.J. 2009. Genetic variants in the vitamin d receptor are associated with advanced prostate cancer at diagnosis: findings from the prostate testing for cancer and treatment study and a systematic review. *Cancer Epidemiol.Biomarkers Prev.*, 18, (11) 2874-2881

Chen, S., Glenn, D.J., Ni, W., Grigsby, C.L., Olsen, K., Nishimoto, M., Law, C.S., & Gardner, D.G. 2008. Expression of the vitamin d receptor is increased in the hypertrophic heart. *Hypertension*, 52, (6) 1106-1112

Chen, T.C., Chimeh, F., Lu, Z., Mathieu, J., Person, K.S., Zhang, A., Kohn, N., Martinello, S., Berkowitz, R., & Holick, M.F. 2007. Factors that influence the

cutaneous synthesis and dietary sources of vitamin D. *Arch.Biochem.Biophys.*, 460, (2) 213-217

Cheng, J.B., Levine, M.A., Bell, N.H., Mangelsdorf, D.J., & Russell, D.W. 2004. Genetic evidence that the human CYP2R1 enzyme is a key vitamin D 25-hydroxylase. *Proc.Natl.Acad.Sci.U.S.A*, 101, (20) 7711-7715

Cheng, S., Massaro, J.M., Fox, C.S., Larson, M.G., Keyes, M.J., McCabe, E.L., Robins, S.J., O'Donnell, C.J., Hoffmann, U., Jacques, P.F., Booth, S.L., Vasan, R.S., Wolf, M., & Wang, T.J. 2010. Adiposity, cardiometabolic risk, and vitamin D status: the Framingham Heart Study. *Diabetes*, 59, (1) 242-248

Chick, D. H., Palzell, E. J., & Hume, E. M. 1923, *Study of rickets in Vienna 1919-1922* 77.

Chun, R.F., Lauridsen, A.L., Suon, L., Zella, L.A., Pike, J.W., Modlin, R.L., Martineau, A.R., Wilkinson, R.J., Adams, J., & Hewison, M. 2010. Vitamin D-binding protein directs monocyte responses to 25-hydroxy- and 1,25-dihydroxyvitamin D. *J.Clin.Endocrinol.Metab*, 95, (7) 3368-3376

Cleeves, M. 1999. Hardy-Weinberg equilibrium test and allele frequency estimation. *Stata Technical Bulletin*, 48, 34-38

Cohen, H.J., Harris, T., & Pieper, C.F. 2003. Coagulation and activation of inflammatory pathways in the development of functional decline and mortality in the elderly. *Am.J.Med.*, 114, (3) 180-187

Cohen, J. 1988a, "Multiple Regression and Correlation Analysis," *In Statistical power analysis for the behavioral sciences*, 2nd ed. Hillsdale, N.J: L. Erlbaum Associates, pp. 407-444.

Cohen, J. 1988b, "The Significance of a Product Moment r," *In Statistical power analysis for the behavioral sciences*, 2nd ed. Hillsdale, N.J: L. Erlbaum Associates, pp. 75-105.

Cole, T.J. 2000. Sympercents: symmetric percentage differences on the 100 log(e) scale simplify the presentation of log transformed data. *Statistics in Medicine*, 19, (22) 3109-3125

Cooper, J.D., Smyth, D.J., Walker, N.M., Stevens, H., Burren, O.S., Wallace, C., Greissl, C., Ramos-Lopez, E., Hypponen, E., Dunger, D.B., Spector, T.D., Ouwehand, W.H., Wang, T.J., Badenhoop, K., & Todd, J.A. 2011. Inherited variation in vitamin D genes is associated with predisposition to autoimmune disease type 1 diabetes. *Diabetes*, 60, (5) 1624-1631

Craig, C.C. 1936. On the Frequency Function of xy. *The Annals of Mathematical Statistics*, 7, (1) 1-15

Crawford, V.L., McCann, M., & Stout, R.W. 2003a. Changes in seasonal deaths from myocardial infarction. *QJM.*, 96, (1) 45-52

Crawford, V.L., McNerlan, S.E., & Stout, R.W. 2003b. Seasonal changes in platelets, fibrinogen and factor VII in elderly people. *Age Ageing*, 32, (6) 661-665

Crawford, V.L., Sweeney, O., Coyle, P.V., Halliday, I.M., & Stout, R.W. 2000. The relationship between elevated fibrinogen and markers of infection: a comparison of seasonal cycles. *QJM.*, 93, (11) 745-750

Cummings, S.R., Tripp, M.K., & Herrmann, N.B. 1997. Approaches to the prevention and control of skin cancer. *Cancer Metastasis Rev.*, 16, (3-4) 309-327

Daiger, S.P., Schanfield, M.S., & Cavalli-Sforza, L.L. 1975. Group-specific component (Gc) proteins bind vitamin D and 25-hydroxyvitamin D. *Proc.Natl.Acad.Sci.U.S.A*, 72, (6) 2076-2080

Danesh, J., Lewington, S., Thompson, S.G., Lowe, G.D., Collins, R., Kostis, J.B., Wilson, A.C., Folsom, A.R., Wu, K., Benderly, M., Goldbourt, U., Willeit, J., Kiechl, S., Yarnell, J.W., Sweetnam, P.M., Elwood, P.C., Cushman, M., Psaty, B.M., Tracy, R.P., Tybjaerg-Hansen, A., Haverkate, F., de Maat, M.P., Fowkes, F.G., Lee, A.J., Smith, F.B., Salomaa, V., Harald, K., Rasi, R., Vahtera, E., Jousilahti, P., Pekkanen, J., D'Agostino, R., Kannel, W.B., Wilson, P.W., Tofler, G., Arocha-Pinango, C.L., Rodriguez-Larralde, A., Nagy, E., Mijares, M., Espinosa, R., Rodriquez-Roa, E., Ryder, E., Diez-Ewald, M.P., Campos, G., Fernandez, V., Torres, E., Marchioli, R., Valagussa, F., Rosengren, A., Wilhelmsen, L., Lappas, G., Eriksson, H., Cremer, P., Nagel, D., Curb, J.D., Rodriguez, B., Yano, K., Salonen, J.T., Nyyssonen, K., Tuomainen, T.P., Hedblad, B., Lind, P., Loewel, H., Koenig, W., Meade, T.W., Cooper, J.A., De, S.B., Knottenbelt, C., Miller, G.J., Cooper, J.A., Bauer, K.A., Rosenberg, R.D., Sato, S., Kitamura, A., Naito, Y., Palosuo, T., Ducimetiere, P., Amouyel, P., Arveiler, D., Evans, A.E., Ferrieres, J., Juhan-Vague, I., Bingham, A., Schulte, H., Assmann, G., Cantin, B., Lamarche, B., Despres, J.P., Dagenais, G.R., Tunstall-Pedoe, H., Woodward, M., Ben-Shlomo, Y., Davey, S.G., Palmieri, V., Yeh, J.L., Rudnicka, A., Ridker, P., Rodeghiero, F., Tosetto, A., Shepherd, J., Ford, I., Robertson, M., Brunner, E., Shipley, M., Feskens, E.J., Kromhout, D., Dickinson, A., Ireland, B., Juzwishin, K., Kaptoge, S., Lewington, S., Memon, A., Sarwar, N., Walker, M., Wheeler, J., White, I., & Wood, A. 2005. Plasma fibrinogen level and the risk of major cardiovascular diseases and nonvascular mortality: an individual participant meta-analysis. *JAMA*, 294, (14) 1799-1809

Danesh, J., Whincup, P., Walker, M., Lennon, L., Thomson, A., Appleby, P., Rumley, A., & Lowe, G.D. 2001. Fibrin D-dimer and coronary heart disease: prospective study and meta-analysis. *Circulation*, 103, (19) 2323-2327

Das, G., Crocombe, S., McGrath, M., Berry, J.L., & Mughal, M.Z. 2006. Hypovitaminosis D among healthy adolescent girls attending an inner city school. *Arch.Dis.Child*, 91, (7) 569-572

Davey Smith, G. 2011. Random allocation in observational data: how small but robust effects could facilitate hypothesis-free causal inference. *Epidemiology*, 22, (4) 460-463

Davey Smith, G. & Ebrahim, S. 2003. Mendelian randomization: can genetic epidemiology contribute to understanding environmental determinants of disease? *Int.J.Epidemiol.*, 32, (1) 1-22

Davey Smith, G. & Ebrahim, S. 2005. What can mendelian randomisation tell us about modifiable behavioural and environmental exposures? *BMJ*, 330, (7499) 1076-1079

Davey Smith, G., Harbord, R., & Ebrahim, S. 2004. Fibrinogen, C-reactive protein and coronary heart disease: does Mendelian randomization suggest the associations are non-causal? *QJM.*, 97, (3) 163-166

Dawson-Hughes, B., Heaney, R.P., Holick, M.F., Lips, P., Meunier, P.J., & Vieth, R. 2005. Estimates of optimal vitamin D status. *Osteoporos.Int.*, 16, (7) 713-716

Debat, V. & David, P. 2001. Mapping phenotypes: canalization, plasticity and developmental stability. *Trends in Ecology & Evolution*, 16, (10) 555-561

DeLuca, H. F. 2005, "Historical Perspective," *In Vitamin D*, 2nd ed ed. D. Feldman, F. H. Glorieux, & J. W. Pike, eds., Amsterdam: Elsevier Academic Press, pp. 3-12.

DeLuca, H.F. 2004. Overview of general physiologic features and functions of vitamin D. *Am.J.Clin.Nutr.*, 80, (6 Suppl) 1689S-1696S

DeLuca, H.F. 2008. Evolution of our understanding of vitamin D. *Nutr.Rev.*, 66, (10 Suppl 2) S73-S87

DeLuca, H.F., Prahl, J.M., & Plum, L.A. 2011. 1,25-Dihydroxyvitamin D is not responsible for toxicity caused by vitamin D or 25-hydroxyvitamin D. *Arch.Biochem.Biophys.*, 505, (2) 226-230

Dentali, F., Ageno, W., Rancan, E., Donati, A.V., Galli, L., Squizzato, A., Venco, A., Mannucci, P.M., & Manfredini, R. 2011. Seasonal and monthly variability in the incidence of venous thromboembolism. A systematic review and a meta-analysis of the literature. *Thromb.Haemost.*, 106, (3) 439-447

Department of Health 2002, *Reform of the Welfare Food Scheme*, Crown.

Department of Health. Vitamin D. 20 Jan. 2010. 7-10-0011.
Ref Type: Pamphlet

Department of Health. Healthy Start. 7 Oct. 2011.
Ref Type: Online Source

Di Nisio M., Squizzato, A., Rutjes, A.W., Buller, H.R., Zwinderman, A.H., & Bossuyt, P.M. 2007. Diagnostic accuracy of D-dimer test for exclusion of venous thromboembolism: a systematic review. *J.Thromb.Haemost.*, 5, (2) 296-304

Didelez, V. & Sheehan, N. 2007. Mendelian randomization as an instrumental variable approach to causal inference. *Stat.Methods Med.Res.*, 16, (4) 309-330

Donin, A.S., Nightingale, C.M., Owen, C.G., Rudnicka, A.R., McNamara, M.C., Prynne, C.J., Stephen, A.M., Cook, D.G., & Whincup, P.H. 2010. Nutritional composition of the diets of South Asian, black African-Caribbean and white European children in the United Kingdom: the Child Heart and Health Study in England (CHASE). *Br.J.Nutr.*, 104, (2) 276-285

Dudbridge, F. & Gusnanto, A. 2008. Estimation of significance thresholds for genomewide association scans. *Genet.Epidemiol.*, 32, (3) 227-234

Earthman, C.P., Beckman, L.M., Masodkar, K., & Sibley, S.D. 2011. The link between obesity and low circulating 25-hydroxyvitamin D concentrations: considerations and implications. *Int.J.Obes.(Lond)*

Empana, J.P., Canoui-Poitrine, F., Luc, G., Juhan-Vague, I., Morange, P., Arveiler, D., Ferrieres, J., Amouyel, P., Bingham, A., Montaye, M., Ruidavets, J.B., Haas, B., Evans, A., & Ducimetiere, P. 2008. Contribution of novel biomarkers to incident stable angina and acute coronary syndrome: the PRIME Study. *Eur.Heart J.*, 29, (16) 1966-1974

Engelman, C.D., Meyers, K.J., Ziegler, J.T., Taylor, K.D., Palmer, N.D., Haffner, S.M., Fingerlin, T.E., Wagenknecht, L.E., Rotter, J.I., Bowden, D.W., Langefeld, C.D., & Norris, J.M. 2010. Genome-wide association study of vitamin D concentrations in Hispanic Americans: the IRAS family study. *J.Steroid Biochem.Mol.Biol.*, 122, (4) 186-192

Engelsen, O., Brustad, M., Aksnes, L., & Lund, E. 2005. Daily duration of vitamin D synthesis in human skin with relation to latitude, total ozone, altitude, ground cover, aerosols and cloud thickness. *Photochem.Photobiol.*, 81, (6) 1287-1290

Evans, D. M. 2008, "Factors affecting power and type one error in association," *In Statistical Genetics: Gene Mapping through Linkage and Association*, B. M. Neale et al., eds., Taylor & Francis, pp. 487-533.

Everitt, B. S. 1987, "Optimization in multivariate analysis," *In Introduction to optimization methods and their application in statistics*, London: Chapman and Hall, pp. 59-77.

Fabsitz, R. & Feinleib, M. 1980. Geographic patterns in county mortality rates from cardiovascular diseases. *Am.J.Epidemiol.*, 111, (3) 315-328

Fang, Y., van Meurs, J.B., Arp, P., van Leeuwen, J.P., Hofman, A., Pols, H.A., & Uitterlinden, A.G. 2009. Vitamin D binding protein genotype and osteoporosis. *Calcif.Tissue Int*, 85, (2) 85-93

Feiveson, A.H. 2002. Power by simulation. *Stata Journal*, 2, (2) 107-124

Folsom, A.R. 1995. Epidemiology of fibrinogen. *Eur.Heart J.*, 16 Suppl A, 21-23

Forman, J.P., Curhan, G.C., & Taylor, E.N. 2008. Plasma 25-hydroxyvitamin D levels and risk of incident hypertension among young women. *Hypertension*, 52, (5) 828-832

Forman, J.P., Giovannucci, E., Holmes, M.D., Bischoff-Ferrari, H.A., Tworoger, S.S., Willett, W.C., & Curhan, G.C. 2007. Plasma 25-hydroxyvitamin D levels and risk of incident hypertension. *Hypertension*, 49, (5) 1063-1069

Forrest, K.Y. & Stuhldreher, W.L. 2011. Prevalence and correlates of vitamin D deficiency in US adults. *Nutr.Res.*, 31, (1) 48-54

Fraser, D. 1967. The relation between infantile hypercalcemia and vitamin D--public health implications in North America. *Pediatrics*, 40, (6) 1050-1061

Freathy, R.M., Mook-Kanamori, D.O., Sovio, U., Prokopenko, I., Timpson, N.J., Berry, D.J., Warrington, N.M., Widen, E., Hottenga, J.J., Kaakinen, M., Lange, L.A., Bradfield, J.P., Kerkhof, M., Marsh, J.A., Magi, R., Chen, C.M., Lyon, H.N., Kirin, M., Adair, L.S., Aulchenko, Y.S., Bennett, A.J., Borja, J.B., Bouatia-Naji, N., Charoen, P., Coin, L.J., Cousminer, D.L., de Geus, E.J., Deloukas, P., Elliott, P., Evans, D.M., Froguel, P., Glaser, B., Groves, C.J., Hartikainen, A.L., Hassanali, N., Hirschhorn, J.N., Hofman, A., Holly, J.M., Hyppönen, E., Kanoni, S., Knight, B.A., Laitinen, J., Lindgren, C.M., McArdle, W.L., O'Reilly, P.F., Pennell, C.E., Postma, D.S., Pouta, A., Ramasamy, A., Rayner, N.W., Ring, S.M., Rivadeneira, F., Shields, B.M., Strachan, D.P., Surakka, I., Taanila, A., Tiesler, C., Uitterlinden, A.G., van Duijn, C.M., Wijga, A.H., Willemsen, G., Zhang, H., Zhao, J., Wilson, J.F., Steegers, E.A., Hattersley, A.T., Eriksson, J.G., Peltonen, L., Mohlke, K.L., Grant, S.F., Hakonarson, H., Koppelman, G.H., Dedoussis, G.V., Heinrich, J., Gillman, M.W., Palmer, L.J., Frayling, T.M., Boomsma, D.I., Davey, S.G., Power, C., Jaddoe, V.W., Jarvelin, M.R., & McCarthy, M.I. 2010. Variants in ADCY5 and near CCNL1 are associated with fetal growth and birth weight. *Nat Genet.*, 42, (5) 430-435

Freeman, C and Marchini, J.: GTOOL. Website: http://www.well.ox.ac.uk/~cfreeman/software/gwas/gtool.html. Last date accessed: 2011

Frohlich, M., Sund, M., Russ, S., Hoffmeister, A., Fischer, H.G., Hombach, V., & Koenig, W. 1997. Seasonal variations of rheological and hemostatic parameters and acute-phase reactants in young, healthy subjects. *Arterioscler.Thromb.Vasc.Biol.*, 17, (11) 2692-2697

Frohlich, M., Sund, M., Thorand, B., Hutchinson, W.L., Pepys, M.B., & Koenig, W. 2002. Lack of seasonal variation in C-reactive protein. *Clin.Chem.*, 48, (3) 575-577

Fu, L., Yun, F., Oczak, M., Wong, B.Y., Vieth, R., & Cole, D.E. 2009. Common genetic variants of the vitamin D binding protein (DBP) predict differences in response of serum 25-hydroxyvitamin D [25(OH)D] to vitamin D supplementation. *Clin.Biochem.*, 42, (10-11) 1174-1177

Fuller, E., Power, C., Shepherd, P., & Strachan, D. P. 2006, *Technical report on the National Child Development Study biomedical survey 2002-2004*, National Centre for Social Research.

Gabriel, S.B., Schaffner, S.F., Nguyen, H., Moore, J.M., Roy, J., Blumenstiel, B., Higgins, J., Defelice, M., Lochner, A., Faggart, M., Liu-Cordero, S.N., Rotimi, C., Adeyemo, A., Cooper, R., Ward, R., Lander, E.S., Daly, M.J., & Altshuler, D. 2002.

The structure of haplotype blocks in the human genome. *Science*, 296, (5576) 2225-2229

Gagnon, C., Lu, Z.X., Magliano, D.J., Dunstan, D.W., Shaw, J.E., Zimmet, P.Z., Sikaris, K., Grantham, N., Ebeling, P.R., & Daly, R.M. 2011. Serum 25-hydroxyvitamin D, calcium intake, and risk of type 2 diabetes after 5 years: results from a national, population-based prospective study (the Australian Diabetes, Obesity and Lifestyle study). *Diabetes Care*, 34, (5) 1133-1138

Galitzer, H., Ben-Dov, I., Lavi-Moshayoff, V., Naveh-Many, T., & Silver, J. 2008. Fibroblast growth factor 23 acts on the parathyroid to decrease parathyroid hormone secretion. *Curr.Opin.Nephrol.Hypertens.*, 17, (4) 363-367

Galle, C., Papazyan, J.P., Miron, M.J., Slosman, D., Bounameaux, H., & Perrier, A. 2001. Prediction of pulmonary embolism extent by clinical findings, D-dimer level and deep vein thrombosis shown by ultrasound. *Thromb.Haemost.*, 86, (5) 1156-1160

Ganji, V., Zhang, X., Shaikh, N., & Tangpricha, V. 2011. Serum 25-hydroxyvitamin D concentrations are associated with prevalence of metabolic syndrome and various cardiometabolic risk factors in US children and adolescents based on assay-adjusted serum 25-hydroxyvitamin D data from NHANES 2001-2006. *Am.J.Clin.Nutr.*, 94, (1) 225-233

Gannage-Yared, M.H., Chemali, R., Yaacoub, N., & Halaby, G. 2000. Hypovitaminosis D in a sunny country: relation to lifestyle and bone markers. *J.Bone Miner.Res.*, 15, (9) 1856-1862

Gao, L., Tao, Y., Zhang, L., & Jin, Q. 2010. Vitamin D receptor genetic polymorphisms and tuberculosis: updated systematic review and meta-analysis. *Int J.Tuberc.Lung Dis.*, 14, (1) 15-23

Garner, C. 2007. Upward bias in odds ratio estimates from genome-wide association studies. *Genet.Epidemiol.*, 31, (4) 288-295

Gascon-Barré, M. 2005, "The Vitamin D 25-Hydroxylase," *In Vitamin D*, 2nd ed ed. D. Feldman, F. H. Glorieux, & J. W. Pike, eds., Amsterdam: Elsevier Academic Press, pp. 47-68.

Ginde, A.A., Liu, M.C., & Camargo, C.A., Jr. 2009a. Demographic differences and trends of vitamin D insufficiency in the US population, 1988-2004. *Arch.Intern.Med.*, 169, (6) 626-632

Ginde, A.A., Scragg, R., Schwartz, R.S., & Camargo, C.A., Jr. 2009b. Prospective study of serum 25-hydroxyvitamin D level, cardiovascular disease mortality, and all-cause mortality in older U.S. adults. *J.Am.Geriatr.Soc.*, 57, (9) 1595-1603

Ginde, A.A., Wolfe, P., Camargo, C.A., Jr., & Schwartz, R.S. 2011. Defining Vitamin D Status by Secondary Hyperparathyroidism in the U.S. Population. *J.Endocrinol.Invest*

Giovannucci, E., Liu, Y., Hollis, B.W., & Rimm, E.B. 2008. 25-hydroxyvitamin D and risk of myocardial infarction in men: a prospective study. *Arch.Intern.Med.*, 168, (11) 1174-1180

Glagov, S., Weisenberg, E., Zarins, C.K., Stankunavicius, R., & Kolettis, G.J. 1987. Compensatory enlargement of human atherosclerotic coronary arteries. *N.Engl.J.Med.*, 316, (22) 1371-1375

Glorieux, F. H. & St Arnaud, R. 2005, "Vitamin D Pseudodeficiency," *In Vitamin D*, 2nd ed ed. D. Feldman, F. H. Glorieux, & J. W. Pike, eds., Amsterdam: Elsevier Academic Press, pp. 1197-1198.

Glymour, M. A. & Greenland, S. 2008, "Causal Diagrams," *In Modern Epidemiology*, Third ed. K. J. Rothman, S. Greenland, & T. L. Lash, eds., Philadelphia: Lippincott Williams & Wilkins, pp. 183-209.

Goldsmith, D.J., Covic, A., Sambrook, P.A., & Ackrill, P. 1997. Vascular calcification in long-term haemodialysis patients in a single unit: a retrospective analysis. *Nephron*, 77, (1) 37-43

Goldstein, H: Review of Winbugs. Website: http://www.cmm.bristol.ac.uk/learning-training/multilevel-m-software/winbugs.shtml. Last date accessed: 30 Nov. 2010

Goodnight, S. H. & Hathaway, W. E. 2001, "Mechanisms of hemostasis and thrombosis," *In Disorders of hemostasis and thrombosis: a clinical guide*, 2nd ed. McGraw Hill Medical, pp. 3-19.

Gray, I.C., Campbell, D.A., & Spurr, N.K. 2000. Single nucleotide polymorphisms as tools in human genetics. *Hum.Mol.Genet.*, 9, (16) 2403-2408

Greenland, S., Pearl, J., & Robins, J.M. 1999. Causal diagrams for epidemiologic research. *Epidemiology*, 10, (1) 37-48

Greenland, S., Rothman, K. J., & Lash, T. L. 2008, "Measures of effect and measures of association," *In Modern Epidemiology*, Third ed. Lippincott Williams & Wilkins, pp. 51-70.

Grimnes, G., Emaus, N., Joakimsen, R.M., Figenschau, Y., Jenssen, T., Njolstad, I., Schirmer, H., & Jorde, R. 2010. Baseline serum 25-hydroxyvitamin D concentrations in the Tromso Study 1994-95 and risk of developing type 2 diabetes mellitus during 11 years of follow-up. *Diabet.Med.*, 27, (10) 1107-1115

Hacke, W., Kaste, M., Bluhmki, E., Brozman, M., Davalos, A., Guidetti, D., Larrue, V., Lees, K.R., Medeghri, Z., Machnig, T., Schneider, D., von, K.R., Wahlgren, N., & Toni, D. 2008. Thrombolysis with alteplase 3 to 4.5 hours after acute ischemic stroke. *N.Engl.J.Med.*, 359, (13) 1317-1329

Haddad, J.G.Jr. & Hahn, T.J. 1973. Natural and synthetic sources of circulating 25-hydroxyvitamin D in man. *Nature*, 244, (5417) 515-517

Hajat, S., Kovats, R.S., & Lachowycz, K. 2007. Heat-related and cold-related deaths in England and Wales: who is at risk? *Occup.Environ.Med.*, 64, (2) 93-100

Halberg, F., Tong, Y. L., & Johnson, A. E. Circadian System Phase - An Aspect of Temporal Morphology; Procedures and Illustrative Examples, *In Cellular Aspects of Biorhythms*, H. von Mayersbach, ed., New York: Springer, pp. 20-48.

Han, J., Kraft, P., Nan, H., Guo, Q., Chen, C., Qureshi, A., Hankinson, S.E., Hu, F.B., Duffy, D.L., Zhao, Z.Z., Martin, N.G., Montgomery, G.W., Hayward, N.K., Thomas, G., Hoover, R.N., Chanock, S., & Hunter, D.J. 2008. A genome-wide association study identifies novel alleles associated with hair color and skin pigmentation. *PLoS.Genet.*, 4, (5) e1000074

Hansson, G.K. 2005. Inflammation, atherosclerosis, and coronary artery disease. *N.Engl.J.Med.*, 352, (16) 1685-1695

Heaney, R.P. 2005. Serum 25-hydroxyvitamin D and parathyroid hormone exhibit threshold behavior. *J.Endocrinol.Invest*, 28, (2) 180-182

Heaney, R.P. & Holick, M.F. 2011. Why the IOM recommendations for vitamin D are deficient. *J.Bone Miner.Res.*

Heaney, R.P., Recker, R.R., Grote, J., Horst, R.L., & Armas, L.A. 2011. Vitamin d3 is more potent than vitamin d2 in humans. *J.Clin.Endocrinol.Metab*, 96, (3) E447-E452

Heemstra, K.A., Hoftijzer, H., van der Deure, W.M., Peeters, R.P., Hamdy, N.A., Pereira, A., Corssmit, E.P., Romijn, J.A., Visser, T.J., & Smit, J.W. 2010. The type 2 deiodinase Thr92Ala polymorphism is associated with increased bone turnover and decreased femoral neck bone mineral density. *J.Bone Miner.Res.*, 25, (6) 1385-1391

Hennekens, C.H. & Buring, J.E. 1987a. *Epidemiology in Medicine* Little, Brown and Company.

Hennekens, C. H. & Buring, J. E. 1987b, "Intervention Studies," *In Epidemiology in Medicine*, 1st ed. S. L. Mayrent, ed., Little, Brown and Company, pp. 178-212.

Hennekens, C. H. & Buring, J. E. 1987c, "Statistical Association and Cause-Effect Relationships," *In Epidemiology in Medicine*, 1st ed. S. L. Mayrent, ed., Little, Brown and Company, pp. 30-53.

Henry, H.L., Bouillon, R., Norman, A.W., Gallagher, J.C., Lips, P., Heaney, R.P., Vieth, R., Pettifor, J.M., Dawson-Hughes, B., Lamberg-Allardt, C.J., & Ebeling, P.R. 2010. 14th Vitamin D Workshop consensus on vitamin D nutritional guidelines. *J.Steroid Biochem.Mol.Biol.*, 121, (1-2) 4-6

Hess, A.F. 1922. The Prevention and Cure of Rickets by Sunlight. *Am.J.Public Health (N.Y.)*, 12, (2) 104-107

Hess, A.F. & Weinstock, M. 1924. Antirachitic properties imparted to lettuce and to growing wheat by ultraviolet irradiation. *Proc Soc Exp Biol Med*, 22, (1) 5-6

Hibler, E.A., Jurutka, P.W., Egan, J.B., Hu, C., LeRoy, E.C., Martinez, M.E., Thompson, P.A., & Jacobs, E.T. 2010. Association between polymorphic variation

in VDR and RXRA and circulating levels of vitamin D metabolites. *J.Steroid Biochem.Mol.Biol.*, 121, (1-2) 438-441

Hindoff, L. A, Junkins, H. A, Hall, P. N, Mehta, J. P, and Manolio, T. A.: A Catalog of Published Genome-Wide Association Studies. Website: www.genome.gov/gwastudies. Last date accessed: 24 May 2011

Hingorani, A. & Humphries, S. 2005. Nature's randomised trials. *Lancet*, 366, (9501) 1906-1908

Hirani, V., Tull, K., Ali, A., & Mindell, J. 2010. Urgent action needed to improve vitamin D status among older people in England! *Age Ageing*, 39, (1) 62-68

Hirschhorn, J.N. & Daly, M.J. 2005. Genome-wide association studies for common diseases and complex traits. *Nat.Rev.Genet.*, 6, (2) 95-108

Hirschhorn, J.N. & Gajdos, Z.K. 2011. Genome-wide association studies: results from the first few years and potential implications for clinical medicine. *Annu.Rev.Med.*, 62, 11-24

Hirschhorn, J.N., Lohmueller, K., Byrne, E., & Hirschhorn, K. 2002. A comprehensive review of genetic association studies. *Genet.Med.*, 4, (2) 45-61

Hoffbrand, A. V. & Moss, P. A. H. 2011a, "Platelets, blood coagulation and haemostasis," *In Essential Haemoatology*, 6th ed. Wiley-Blackwell, pp. 314-329.

Hoffbrand, A. V. & Moss, P. A. H. 2011b, "Thrombosis and antithrombotic therapy," *In Essential Haemoatology*, 6th ed. Wiley-Blackwell, pp. 362-380.

Holden, J.M., Lemar, L.E., & Exler, J. 2008. Vitamin D in foods: development of the US Department of Agriculture database. *Am.J.Clin.Nutr.*, 87, (4) 1092S-1096S

Holick, M.F. 1994. McCollum Award Lecture, 1994: vitamin D--new horizons for the 21st century. *Am.J.Clin.Nutr.*, 60, (4) 619-630

Holick, M.F. 1995. Environmental factors that influence the cutaneous production of vitamin D. *Am.J.Clin.Nutr.*, 61, (3 Suppl) 638S-645S

Holick, M.F. 2002. Vitamin D: the underappreciated D-lightful hormone that is important for skeletal and cellular health. *Curr.Opin.Endocrinol.Diabetes*, 9, (1) 87-98

Holick, M.F. 2006. Resurrection of vitamin D deficiency and rickets. *J.Clin.Invest*, 116, (8) 2062-2072

Holick, M.F. 2008. Vitamin D: a D-Lightful health perspective. *Nutr.Rev.*, 66, (10 Suppl 2) S182-S194

Holick, M.F., Biancuzzo, R.M., Chen, T.C., Klein, E.K., Young, A., Bibuld, D., Reitz, R., Salameh, W., Ameri, A., & Tannenbaum, A.D. 2008. Vitamin D2 is as effective as vitamin D3 in maintaining circulating concentrations of 25-hydroxyvitamin D. *J.Clin.Endocrinol.Metab*, 93, (3) 677-681

Holick, M.F., Chen, T.C., Lu, Z., & Sauter, E. 2007. Vitamin D and skin physiology: a D-lightful story. *J.Bone Miner.Res.*, 22 Suppl 2, V28-V33

Holick, M.F., MacLaughlin, J.A., & Doppelt, S.H. 1981. Regulation of cutaneous previtamin D3 photosynthesis in man: skin pigment is not an essential regulator. *Science*, 211, (4482) 590-593

Hollis, B.W. 2005. Circulating 25-hydroxyvitamin D levels indicative of vitamin D sufficiency: implications for establishing a new effective dietary intake recommendation for vitamin D. *J.Nutr.*, 135, (2) 317-322

Horst, R., Reinhardt, T., & Satyanarayana Reddy, G. 2005, "Vitamin D Metabolism," *In Vitamin D*, 2nd ed ed. D. Feldman, F. H. Glorieux, & J. W. Pike, eds., Amsterdam: Elsevier Academic Press, pp. 15-36.

Hsieh, F.Y., Bloch, D.A., & Larsen, M.D. 1998. A simple method of sample size calculation for linear and logistic regression. *Stat.Med.*, 17, (14) 1623-1634

Hu, J.X., Zhao, H.Y., & Zhou, H.H. 2010. False Discovery Rate Control With Groups. *Journal of the American Statistical Association*, 105, (491) 1215-1227

Hunter, D., De Lange, M., Snieder, H., MacGregor, A.J., Swaminathan, R., Thakker, R.V., & Spector, T.D. 2001. Genetic contribution to bone metabolism, calcium excretion, and vitamin D and parathyroid hormone regulation. *J.Bone Miner.Res.*, 16, (2) 371-378

Hutchinson, W.L., Koenig, W., Frohlich, M., Sund, M., Lowe, G.D., & Pepys, M.B. 2000. Immunoradiometric assay of circulating C-reactive protein: age-related values in the adult general population. *Clin.Chem.*, 46, (7) 934-938

Hyppönen, E., Berry, D., Cortina-Borja, M., & Power, C. 2010. 25-Hydroxyvitamin D and pre-clinical alterations in inflammatory and hemostatic markers: a cross sectional analysis in the 1958 British Birth Cohort. *PLoS One.*, 5, (5) e10801

Hyppönen, E., Berry, D.J., Wjst, M., & Power, C. 2009. Serum 25-hydroxyvitamin D and IgE - a significant but nonlinear relationship. *Allergy*, 64, (4) 613-620

Hyppönen, E., Boucher, B.J., Berry, D.J., & Power, C. 2008. 25-hydroxyvitamin D, IGF-1, and metabolic syndrome at 45 years of age: a cross-sectional study in the 1958 British Birth Cohort. *Diabetes*, 57, (2) 298-305

Hyppönen, E. & Power, C. 2007. Hypovitaminosis D in British adults at age 45 y: nationwide cohort study of dietary and lifestyle predictors. *Am.J.Clin.Nutr.*, 85, (3) 860-868

Hyppönen, E., Turner, S., Cumberland, P., Power, C., & Gibb, I. 2007. Serum 25-hydroxyvitamin D measurement in a large population survey with statistical harmonization of assay variation to an international standard. *J.Clin.Endocrinol.Metab*, 92, (12) 4615-4622

Ingles, S. 2004, "Basic Concepts of Molecular Genetics," *In Statistical Methods in Genetic Epidemiology*, D. C. Thomas, ed., Oxford University Press, pp. 25-44.

Institute of Medicine 2011, *Dietary Reference Intakes for Calcium and Vitamin D*, The National Academies Press, Washington, DC.

International HapMap Consortium 2003. The International HapMap Project. *Nature*, 426, (6968) 789-796

Jablonski, N.G. & Chaplin, G. 2010. Colloquium paper: human skin pigmentation as an adaptation to UV radiation. *Proc.Natl.Acad.Sci.U.S.A*, 107 Suppl 2, 8962-8968

Janmey, P.A., Lamb, J.A., Ezzell, R.M., Hvidt, S., & Lind, S.E. 1992. Effects of actin filaments on fibrin clot structure and lysis. *Blood*, 80, (4) 928-936

Janssens, W., Bouillon, R., Claes, B., Carremans, C., Lehouck, A., Buysschaert, I., Coolen, J., Mathieu, C., Decramer, M., & Lambrechts, D. 2010. Vitamin D deficiency is highly prevalent in COPD and correlates with variants in the vitamin D-binding gene. *Thorax*, 65, (3) 215-220

Jassal, S.K., Chonchol, M., von, M.D., Smits, G., & Barrett-Connor, E. 2010. Vitamin d, parathyroid hormone, and cardiovascular mortality in older adults: the Rancho Bernardo study. *Am.J.Med.*, 123, (12) 1114-1120

Jegga, A.G., Gowrisankar, S., Chen, J., & Aronow, B.J. 2007. PolyDoms: a whole genome database for the identification of non-synonymous coding SNPs with the potential to impact disease. *Nucleic Acids Res.*, 35, (Database issue) D700-D706

Ji, G.R., Yao, M., Sun, C.Y., Li, Z.H., & Han, Z. 2010. BsmI, TaqI, ApaI and FokI polymorphisms in the vitamin D receptor (VDR) gene and risk of fracture in Caucasians: a meta-analysis. *Bone*, 47, (3) 681-686

Johnson, A.D., Handsaker, R.E., Pulit, S.L., Nizzari, M.M., O'Donnell, C.J., & de Bakker, P.I. 2008. SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics.*, 24, (24) 2938-2939

Johnson, J. D-Dimer. Hypponen, E. 9 Aug. 2007.

Jones, G. 2000, "Vitamin D," *In The Cambridge World History of Food*, K. F. Kiple & K. Conee Ornelas, eds., Cambridge, United Kingdom: Cambridge University Press.

Jones, G. 2008. Pharmacokinetics of vitamin D toxicity. *Am.J.Clin.Nutr.*, 88, (2) 582S-586S

Jorde, R., Figenschau, Y., Emaus, N., Hutchinson, M., & Grimnes, G. 2010a. Serum 25-hydroxyvitamin D levels are strongly related to systolic blood pressure but do not predict future hypertension. *Hypertension*, 55, (3) 792-798

Jorde, R., Haug, E., Figenschau, Y., & Hansen, J.B. 2007. Serum levels of vitamin D and haemostatic factors in healthy subjects: the Tromso study. *Acta Haematol.*, 117, (2) 91-97

Jorde, R., Sneve, M., Emaus, N., Figenschau, Y., & Grimnes, G. 2010b. Cross-sectional and longitudinal relation between serum 25-hydroxyvitamin D and body mass index: the Tromso study. *Eur.J.Nutr.*, 49, (7) 401-407

Jorde, R., Sneve, M., Torjesen, P., Figenschau, Y., & Hansen, J.B. 2010c. Parameters of the thrombogram are associated with serum 25-hydroxyvitamin D levels at baseline, but not affected during supplementation with vitamin D. *Thromb.Res.*, 125, (5) e210-e213

Jorde, R., Sneve, M., Torjesen, P.A., Figenschau, Y., Goransson, L.G., & Omdal, R. 2010d. No effect of supplementation with cholecalciferol on cytokines and markers of inflammation in overweight and obese subjects. *Cytokine*, 50, (2) 175-180

Kamboh, M.I. & Ferrell, R.E. 1986. Ethnic variation in vitamin D-binding protein (GC): a review of isoelectric focusing studies in human populations. *Hum.Genet.*, 72, (4) 281-293

Kanda, S., Mochizuki, Y., Miyata, Y., Kanetake, H., & Yamamoto, N. 2002. Effects of vitamin D(3)-binding protein-derived macrophage activating factor (GcMAF) on angiogenesis. *J.Natl.Cancer Inst.*, 94, (17) 1311-1319

Kanis, J. A. 2002, "Vitamin D Deficiency," *In Encyclopedia of life sciences*, London: Nature Publishing Group.

Kaptoge, S., Di, A.E., Lowe, G., Pepys, M.B., Thompson, S.G., Collins, R., & Danesh, J. 2010. C-reactive protein concentration and risk of coronary heart disease, stroke, and mortality: an individual participant meta-analysis. *Lancet*, 375, (9709) 132-140

Kawakami, M., Blum, C.B., Ramakrishnan, R., Dell, R.B., & Goodman, D.S. 1981. Turnover of the plasma binding protein for vitamin D and its metabolites in normal human subjects. *J.Clin.Endocrinol.Metab*, 53, (6) 1110-1116

Keatinge, W.R., Coleshaw, S.R., & Holmes, J. 1989. Changes in seasonal mortalities with improvement in home heating in England and Wales from 1964 to 1984. *Int.J.Biometeorol.*, 33, (2) 71-76

Kendrick, J., Targher, G., Smits, G., & Chonchol, M. 2009. 25-Hydroxyvitamin D deficiency is independently associated with cardiovascular disease in the Third National Health and Nutrition Examination Survey. *Atherosclerosis*, 205, (1) 255-260

Kenward, M.G. & Carpenter, J. 2007. Multiple imputation: current perspectives. *Stat.Methods Med.Res.*, 16, (3) 199-218

Kew, R.R. & Webster, R.O. 1988. Gc-globulin (vitamin D-binding protein) enhances the neutrophil chemotactic activity of C5a and C5a des Arg. *J.Clin.Invest*, 82, (1) 364-369

Kilkkinen, A., Knekt, P., Aro, A., Rissanen, H., Marniemi, J., Heliovaara, M., Impivaara, O., & Reunanen, A. 2009. Vitamin D status and the risk of cardiovascular disease death. *Am.J.Epidemiol.*, 170, (8) 1032-1039

Kimlin, M.G. 2004. The climatology of Vitamin D producing ultraviolet radiation over the United States. *J.Steroid Biochem.Mol.Biol.*, 89-90, (1-5) 479-483

Knekt, P., Laaksonen, M., Mattila, C., Harkanen, T., Marniemi, J., Heliovaara, M., Rissanen, H., Montonen, J., & Reunanen, A. 2008. Serum vitamin D and subsequent occurrence of type 2 diabetes. *Epidemiology*, 19, (5) 666-671

Koek, W.N., van Meurs, J.B., van der Eerden, B.C., Rivadeneira, F., Zillikens, M.C., Hofman, A., Obermayer-Pietsch, B., Lips, P., Pols, H.A., Uitterlinden, A.G., & van Leeuwen, J.P. 2010. The T-13910C polymorphism in the lactase phlorizin hydrolase gene is associated with differences in serum calcium levels and calcium intake. *J.Bone Miner.Res.*, 25, (9) 1980-1987

Koenig, W. & Ernst, E. 1992. The possible role of hemorheology in atherothrombogenesis. *Atherosclerosis*, 94, (2-3) 93-107

Kostner, K., Denzer, N., Muller, C.S., Klein, R., Tilgen, W., & Reichrath, J. 2009. The relevance of vitamin D receptor (VDR) gene polymorphisms for cancer: a review of the literature. *Anticancer Res.*, 29, (9) 3511-3536

Koyama, T., Shibakura, M., Ohsawa, M., Kamiyama, R., & Hirosawa, S. 1998. Anticoagulant effects of 1alpha,25-dihydroxyvitamin D3 on human myelogenous leukemia cells and monocytes. *Blood*, 92, (1) 160-167

Kuha, J. 2004. AIC and BIC - Comparisons of assumptions and performance. *Sociological Methods & Research*, 33, (2) 188-229

Laing, C. J. & Cooke, N. K. 2005, "Vitamin D binding protein," *In Vitamin D*, 2nd ed ed. D. Feldman, F. H. Glorieux, & J. W. Pike, eds., Amsterdam: Elsevier Academic Press, pp. 117-134.

Lamberg-Allardt, C. 2006. Vitamin D in foods and as supplements. *Prog.Biophys.Mol.Biol.*, 92, (1) 33-38

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J.P., Miranda, C., Morris, W., Naylor, J., Raymond, C., Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J.C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R.H., Wilson, R.K., Hillier, L.W., McPherson, J.D., Marra, M.A., Mardis, E.R., Fulton, L.A., Chinwalla, A.T., Pepin, K.H., Gish, W.R., Chissoe, S.L., Wendl, M.C., Delehaunty, K.D., Miner, T.L., Delehaunty, A., Kramer, J.B., Cook, L.L., Fulton, R.S., Johnson, D.L., Minx, P.J., Clifton, S.W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J.F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., Gibbs, R.A., Muzny, D.M., Scherer, S.E., Bouck, J.B., Sodergren, E.J.,

Worley, K.C., Rives, C.M., Gorrell, J.H., Metzker, M.L., Naylor, S.L., Kucherlapati, R.S., Nelson, D.L., Weinstock, G.M., Sakaki, Y., Fujiyama, A., Hattori, M., Yada, T., Toyoda, A., Itoh, T., Kawagoe, C., Watanabe, H., Totoki, Y., Taylor, T., Weissenbach, J., Heilig, R., Saurin, W., Artiguenave, F., Brottier, P., Bruls, T., Pelletier, E., Robert, C., Wincker, P., Smith, D.R., Doucette-Stamm, L., Rubenfield, M., Weinstock, K., Lee, H.M., Dubois, J., Rosenthal, A., Platzer, M., Nyakatura, G., Taudien, S., Rump, A., Yang, H., Yu, J., Wang, J., Huang, G., Gu, J., Hood, L., Rowen, L., Madan, A., Qin, S., Davis, R.W., Federspiel, N.A., Abola, A.P., Proctor, M.J., Myers, R.M., Schmutz, J., Dickson, M., Grimwood, J., Cox, D.R., Olson, M.V., Kaul, R., Raymond, C., Shimizu, N., Kawasaki, K., Minoshima, S., Evans, G.A., Athanasiou, M., Schultz, R., Roe, B.A., Chen, F., Pan, H., Ramser, J., Lehrach, H., Reinhardt, R., McCombie, W.R., de la Bastide, M., Dedhia, N., Blocker, H., Hornischer, K., Nordsiek, G., Agarwala, R., Aravind, L., Bailey, J.A., Bateman, A., Batzoglou, S., Birney, E., Bork, P., Brown, D.G., Burge, C.B., Cerutti, L., Chen, H.C., Church, D., Clamp, M., Copley, R.R., Doerks, T., Eddy, S.R., Eichler, E.E., Furey, T.S., Galagan, J., Gilbert, J.G., Harmon, C., Hayashizaki, Y., Haussler, D., Hermjakob, H., Hokamp, K., Jang, W., Johnson, L.S., Jones, T.A., Kasif, S., Kaspryzk, A., Kennedy, S., Kent, W.J., Kitts, P., Koonin, E.V., Korf, I., Kulp, D., Lancet, D., Lowe, T.M., McLysaght, A., Mikkelsen, T., Moran, J.V., Mulder, N., Pollara, V.J., Ponting, C.P., Schuler, G., Schultz, J., Slater, G., Smit, A.F., Stupka, E., Szustakowski, J., Thierry-Mieg, D., Thierry-Mieg, J., Wagner, L., Wallis, J., Wheeler, R., Williams, A., Wolf, Y.I., Wolfe, K.H., Yang, S.P., Yeh, R.F., Collins, F., Guyer, M.S., Peterson, J., Felsenfeld, A., Wetterstrand, K.A., Patrinos, A., Morgan, M.J., de, J.P., Catanese, J.J., Osoegawa, K., Shizuya, H., Choi, S., & Chen, Y.J. 2001. Initial sequencing and analysis of the human genome. *Nature*, 409, (6822) 860-921

Lango, A.H., Estrada, K., Lettre, G., Berndt, S.I., Weedon, M.N., Rivadeneira, F., Willer, C.J., Jackson, A.U., Vedantam, S., Raychaudhuri, S., Ferreira, T., Wood, A.R., Weyant, R.J., Segre, A.V., Speliotes, E.K., Wheeler, E., Soranzo, N., Park, J.H., Yang, J., Gudbjartsson, D., Heard-Costa, N.L., Randall, J.C., Qi, L., Vernon, S.A., Magi, R., Pastinen, T., Liang, L., Heid, I.M., Luan, J., Thorleifsson, G., Winkler, T.W., Goddard, M.E., Sin, L.K., Palmer, C., Workalemahu, T., Aulchenko, Y.S., Johansson, A., Zillikens, M.C., Feitosa, M.F., Esko, T., Johnson, T., Ketkar, S., Kraft, P., Mangino, M., Prokopenko, I., Absher, D., Albrecht, E., Ernst, F., Glazer, N.L., Hayward, C., Hottenga, J.J., Jacobs, K.B., Knowles, J.W., Kutalik, Z., Monda, K.L., Polasek, O., Preuss, M., Rayner, N.W., Robertson, N.R., Steinthorsdottir, V., Tyrer, J.P., Voight, B.F., Wiklund, F., Xu, J., Zhao, J.H., Nyholt, D.R., Pellikka, N., Perola, M., Perry, J.R., Surakka, I., Tammesoo, M.L., Altmaier, E.L., Amin, N., Aspelund, T., Bhangale, T., Boucher, G., Chasman, D.I., Chen, C., Coin, L., Cooper, M.N., Dixon, A.L., Gibson, Q., Grundberg, E., Hao, K., Juhani, J.M., Kaplan, L.M., Kettunen, J., Konig, I.R., Kwan, T., Lawrence, R.W., Levinson, D.F., Lorentzon, M., McKnight, B., Morris, A.P., Muller, M., Suh, N.J., Purcell, S., Rafelt, S., Salem, R.M., Salvi, E., Sanna, S., Shi, J., Sovio, U., Thompson, J.R., Turchin, M.C., Vandenput, L., Verlaan, D.J., Vitart, V., White, C.C., Ziegler, A., Almgren, P., Balmforth, A.J., Campbell, H., Citterio, L., De, G.A., Dominiczak, A., Duan, J., Elliott, P., Elosua, R., Eriksson, J.G., Freimer, N.B., Geus, E.J., Glorioso, N., Haiqing, S., Hartikainen, A.L., Havulinna, A.S., Hicks, A.A., Hui, J., Igl, W., Illig, T., Jula, A., Kajantie, E., Kilpelainen, T.O., Koiranen, M., Kolcic, I., Koskinen, S., Kovacs, P., Laitinen, J., Liu, J., Lokki, M.L., Marusic, A., Maschio, A., Meitinger,

T., Mulas, A., Pare, G., Parker, A.N., Peden, J.F., Petersmann, A., Pichler, I., Pietilainen, K.H., Pouta, A., Ridderstrale, M., Rotter, J.I., Sambrook, J.G., Sanders, A.R., Schmidt, C.O., Sinisalo, J., Smit, J.H., Stringham, H.M., Bragi, W.G., Widen, E., Wild, S.H., Willemsen, G., Zagato, L., Zgaga, L., Zitting, P., Alavere, H., Farrall, M., McArdle, W.L., Nelis, M., Peters, M.J., Ripatti, S., van Meurs, J.B., Aben, K.K., Ardlie, K.G., Beckmann, J.S., Beilby, J.P., Bergman, R.N., Bergmann, S., Collins, F.S., Cusi, D., den, H.M., Eiriksdottir, G., Gejman, P.V., Hall, A.S., Hamsten, A., Huikuri, H.V., Iribarren, C., Kahonen, M., Kaprio, J., Kathiresan, S., Kiemeney, L., Kocher, T., Launer, L.J., Lehtimaki, T., Melander, O., Mosley, T.H., Jr., Musk, A.W., Nieminen, M.S., O'Donnell, C.J., Ohlsson, C., Oostra, B., Palmer, L.J., Raitakari, O., Ridker, P.M., Rioux, J.D., Rissanen, A., Rivolta, C., Schunkert, H., Shuldiner, A.R., Siscovick, D.S., Stumvoll, M., Tonjes, A., Tuomilehto, J., van Ommen, G.J., Viikari, J., Heath, A.C., Martin, N.G., Montgomery, G.W., Province, M.A., Kayser, M., Arnold, A.M., Atwood, L.D., Boerwinkle, E., Chanock, S.J., Deloukas, P., Gieger, C., Gronberg, H., Hall, P., Hattersley, A.T., Hengstenberg, C., Hoffman, W., Lathrop, G.M., Salomaa, V., Schreiber, S., Uda, M., Waterworth, D., Wright, A.F., Assimes, T.L., Barroso, I., Hofman, A., Mohlke, K.L., Boomsma, D.I., Caulfield, M.J., Cupples, L.A., Erdmann, J., Fox, C.S., Gudnason, V., Gyllensten, U., Harris, T.B., Hayes, R.B., Jarvelin, M.R., Mooser, V., Munroe, P.B., & Ouwehand, W.H. 2010. Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature*, 467, (7317) 832-838

Lanham-New, S.A., Buttriss, J.L., Miles, L.M., Ashwell, M., Berry, J.L., Boucher, B.J., Cashman, K.D., Cooper, C., Darling, A.L., Francis, R.M., Fraser, W.D., de Groot, C.P., Hyppönen, E., Kiely, M., Lamberg-Allardt, C., Macdonald, H.M., Martineau, A.R., Masud, T., Mavroeidi, A., Nowson, C., Prentice, A., Stone, E.M., Reddy, S., Vieth, R., & Williams, C.M. 2011. Proceedings of the Rank Forum on Vitamin D. *Br.J.Nutr.*, 105, (1) 144-156

Lauridsen, A.L., Vestergaard, P., Hermann, A.P., Brot, C., Heickendorff, L., Mosekilde, L., & Nexo, E. 2005. Plasma concentrations of 25-hydroxy-vitamin D and 1,25-dihydroxy-vitamin D are related to the phenotype of Gc (vitamin D-binding protein): a cross-sectional study on 595 early postmenopausal women. *Calcif.Tissue Int*, 77, (1) 15-22

Lauridsen, A.L., Vestergaard, P., & Nexo, E. 2001. Mean serum concentration of vitamin D-binding protein (Gc globulin) is related to the Gc phenotype in women. *Clin.Chem.*, 47, (4) 753-756

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

Lee, Y.H., Bae, S.C., Choi, S.J., Ji, J.D., & Song, G.G. 2011. Associations between vitamin D receptor polymorphisms and susceptibility to rheumatoid arthritis and systemic lupus erythematosus: a meta-analysis. *Mol.Biol.Rep.*, 38, (6) 3643-51

Lensmeyer, G., Wiebe, D.A., Binkley, N., & Drezner, M.K. 2006. Letters to the Editor. *Clin.Chem.*, 52, (12) 2305-2306

Lin, X., Song, K., Lim, N., Yuan, X., Johnson, T., Abderrahmani, A., Vollenweider, P., Stirnadel, H., Sundseth, S.S., Lai, E., Burns, D.K., Middleton, L.T., Roses, A.D., Matthews, P.M., Waeber, G., Cardon, L., Waterworth, D.M., & Mooser, V. 2009. Risk prediction of prevalent diabetes in a Swiss population using a weighted genetic score--the CoLaus Study. *Diabetologia*, 52, (4) 600-608

Lindqvist, P.G., Epstein, E., & Olsson, H. 2009. Does an active sun exposure habit lower the risk of venous thrombotic events? A D-lightful hypothesis. *J.Thromb.Haemost.*, 7, (4) 605-610

Lips, P. 2005, "How to define normal values for serum concentrations of 25-hydroxyvitamin D? An overview," *In Vitamin D*, 2nd ed ed. D. Feldman, F. H. Glorieux, & J. W. Pike, eds., Amsterdam: Elsevier Academic Press, pp. 1019-1028.

Liu, E., Meigs, J.B., Pittas, A.G., Economos, C.D., McKeown, N.M., Booth, S.L., & Jacques, P.F. 2010. Predicted 25-hydroxyvitamin D score and incident type 2 diabetes in the Framingham Offspring Study. *Am.J.Clin.Nutr.*, 91, (6) 1627-1633

Lobler, M., Sass, M., Kunze, C., Schmitz, K.P., & Hopt, U.T. 2002. Biomaterial implants induce the inflammation marker CRP at the site of implantation. *J.Biomed.Mater.Res.*, 61, (1) 165-167

Lohmueller, K.E., Pearce, C.L., Pike, M., Lander, E.S., & Hirschhorn, J.N. 2003. Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. *Nat.Genet.*, 33, (2) 177-182

Looker, A.C., Dawson-Hughes, B., Calvo, M.S., Gunter, E.W., & Sahyoun, N.R. 2002. Serum 25-hydroxyvitamin D status of adolescents and adults in two seasonal subpopulations from NHANES III. *Bone*, 30, (5) 771-777

Lowe, G. D. 1997, "Haemostatic risk factors for arterial and venous thrombosis," *In Recent advances in blood coagulation*, 7th ed. L. Poller & C. A. Ludham, eds., Churchill Livingstone, pp. 69-96.

Lowe, G.D., Danesh, J., Lewington, S., Walker, M., Lennon, L., Thomson, A., Rumley, A., & Whincup, P.H. 2004. Tissue plasminogen activator antigen and coronary heart disease. Prospective study and meta-analysis. *Eur.Heart J.*, 25, (3) 252-259

Lowe, N.M., Mitra, S.R., Foster, P.C., Bhojani, I., & McCann, J.F. 2010. Vitamin D status and markers of bone turnover in Caucasian and South Asian postmenopausal women living in the UK. *Br.J.Nutr.*, 103, (12) 1706-1710

Lunn, D.J., Thomas, A., Best, N., & Spiegelhalter, D. 2000. WinBUGS - A Bayesian modelling framework: Concepts, structure, and extensibility. *Statistics and Computing*, 10, (4) 325-337

Macdonald, H.M., Mavroeidi, A., Fraser, W.D., Darling, A.L., Black, A.J., Aucott, L., O'Neill, F., Hart, K., Berry, J.L., Lanham-New, S.A., & Reid, D.M. 2011. Sunlight and dietary contributions to the seasonal vitamin D status of cohorts of healthy postmenopausal women living at northerly latitudes: a major cause for concern? *Osteoporos.Int.*, 22, 2461-2472

Macfarlane, R.G. 1977. Haemostasis: Introduction. *Br.Med.Bull*, 33, (3) 1-5

Mackinnon, D.P., Fairchild, A.J., & Fritz, M.S. 2007. Mediation analysis. *Annu.Rev.Psychol.*, 58, 593-614

Mackinnon, D.P. & Fritz, M.S. 2007. Distribution of the product confidence limits for the indirect effect: Program PRODCLIN. *Behavior Research Methods*, 39, (3) 384-389

Mackinnon, D.P., Lockwood, C.M., Hoffman, J.M., West, S.G., & Sheets, V. 2002. A comparison of methods to test mediation and other intervening variable effects. *Psychol.Methods*, 7, (1) 83-104

Mackinnon, D.P., Warsi, G., & Dwyer, J.H. 1995. A Simulation Study of Mediated Effect Measures. *Multivariate Behavioral Research*, 30, (1) 41-62

MacLaughlin, J. & Holick, M.F. 1985. Aging decreases the capacity of human skin to produce vitamin D3. *J.Clin.Invest*, 76, (4) 1536-1538

Malloy, P. L., Pike, J. W., & Feldman, D. 2005, "Hereditary 1,25-Dihydroxyvitamin D-Resistant Rickets," *In Vitamin D*, 2nd ed ed. D. Feldman, F. H. Glorieux, & J. W. Pike, eds., Amsterdam: Elsevier Academic Press, pp. 1207-1237.

Mannucci, P.M. 1998. von Willebrand factor: a marker of endothelial damage? *Arterioscler.Thromb.Vasc.Biol.*, 18, (9) 1359-1362

Manolio, T.A., Brooks, L.D., & Collins, F.S. 2008. A HapMap harvest of insights into the genetics of common disease. *J.Clin.Invest*, 118, (5) 1590-1605

Marchini, J., Howie, B., Myers, S., McVean, G., & Donnelly, P. 2007. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat.Genet.*, 39, (7) 906-913

Martineau, A.R., Leandro, A.C., Anderson, S.T., Newton, S.M., Wilkinson, K.A., Nicol, M.P., Pienaar, S.M., Skolimowska, K.H., Rocha, M.A., Rolla, V.C., Levin, M., Davidson, R.N., Bremner, S.A., Griffiths, C.J., Eley, B.S., Bonecini-Almeida, M.G., & Wilkinson, R.J. 2010. Association between Gc genotype and susceptibility to TB is dependent on vitamin D status. *Eur.Respir.J.*, 35, (5) 1106-1112

Martins, D., Wolf, M., Pan, D., Zadshir, A., Tareen, N., Thadhani, R., Felsenfeld, A., Levine, B., Mehrotra, R., & Norris, K. 2007. Prevalence of cardiovascular risk factors and the serum levels of 25-hydroxyvitamin D in the United States: data from the Third National Health and Nutrition Examination Survey. *Arch.Intern.Med.*, 167, (11) 1159-1165

Matias, P.J., Jorge, C., Ferreira, C., Borges, M., Aires, I., Amaral, T., Gil, C., Cortez, J., & Ferreira, A. 2010. Cholecalciferol supplementation in hemodialysis patients: effects on mineral metabolism, inflammation, and cardiac dimension parameters. *Clin.J.Am.Soc.Nephrol.*, 5, (5) 905-911

May, M., Lawlor, D.A., Patel, R., Rumley, A., Lowe, G., & Ebrahim, S. 2007. Associations of von Willebrand factor, fibrin D-dimer and tissue plasminogen

activator with incident coronary heart disease: British Women's Heart and Health cohort study. *Eur.J.Cardiovasc.Prev.Rehabil.*, 14, (5) 638-645

McCarthy, M.I., Abecasis, G.R., Cardon, L.R., Goldstein, D.B., Little, J., Ioannidis, J.P., & Hirschhorn, J.N. 2008. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat.Rev.Genet.*, 9, (5) 356-369

Melamed, M.L., Michos, E.D., Post, W., & Astor, B. 2008. 25-hydroxyvitamin D levels and the risk of mortality in the general population. *Arch.Intern.Med.*, 168, (15) 1629-1637

Mellanby, E. 1919. An experimental investigation on rickets. *The Lancet*, 193, (4985) 407-448

Merchiers, P., Bulens, F., Stockmans, I., De, V.A., Convents, R., Bouillon, R., Collen, D., Belayew, A., & Carmeliet, G. 1999. 1,25-Dihydroxyvitamin D(3) induction of the tissue-type plasminogen activator gene is mediated through its multihormone-responsive enhancer. *FEBS Lett.*, 460, (2) 289-296

Michaelsson, K., Baron, J.A., Snellman, G., Gedeborg, R., Byberg, L., Sundstrom, J., Berglund, L., Arnlov, J., Hellman, P., Blomhoff, R., Wolk, A., Garmo, H., Holmberg, L., & Melhus, H. 2010. Plasma vitamin D and mortality in older men: a community-based prospective cohort study. *Am.J.Clin.Nutr.*, 92, (4) 841-848

Michos, E.D., Streeten, E.A., Ryan, K.A., Rampersaud, E., Peyser, P.A., Bielak, L.F., Shuldiner, A.R., Mitchell, B.D., & Post, W. 2009. Serum 25-hydroxyvitamin d levels are not associated with subclinical vascular disease or C-reactive protein in the old order amish. *Calcif.Tissue Int.*, 84, (3) 195-202

Ministry of Agriculture, F. a. F. The infant formula and follow-on formula regulations 1995. 77. 1995a. Statutory Instruments.
Ref Type: Bill/Resolution

Ministry of Agriculture, F. a. F. The spreadable fats (marketing standards) regulations 1995. 3116. 1995b. Statutory Instruments.
Ref Type: Bill/Resolution

Ministry of Health 1957, *Report of the joint subcommittee on welfare foods*, H.M.S.O, London.

Mithal, A., Wahl, D.A., Bonjour, J.P., Burckhardt, P., Dawson-Hughes, B., Eisman, J.A., El-Hajj, F.G., Josse, R.G., Lips, P., Morales-Torres, J., & IOF Committee of Scientific Advisors (CSA) Nutrition Working Group. 2009. Global vitamin D status and determinants of hypovitaminosis D. *Osteoporos.Int*, 20, (11) 1807-1820

Moonesinghe, R., Khoury, M.J., & Janssens, A.C. 2007. Most published research findings are false-but a little replication goes a long way. *PLoS.Med.*, 4, (2) e28

Mora, J.R., Iwata, M., & von Andrian, U.H. 2008. Vitamin effects on the immune system: vitamins A and D take centre stage. *Nat Rev Immunol*, 8, (9) 685-698

Morris, J.G. 1999. Ineffective vitamin D synthesis in cats is reversed by an inhibitor of 7-dehydrocholestrol-delta7-reductase. *J.Nutr.*, 129, (4) 903-908

Mortimer, E.A., Jr., Monson, R.R., & MacMahon, B. 1977. Reduction in mortality from coronary heart disease in men residing at high altitude. *N.Engl.J.Med.*, 296, (11) 581-585

Nan, H., Kraft, P., Qureshi, A.A., Guo, Q., Chen, C., Hankinson, S.E., Hu, F.B., Thomas, G., Hoover, R.N., Chanock, S., Hunter, D.J., & Han, J. 2009. Genome-wide association study of tanning phenotype in a population of European ancestry. *J.Invest Dermatol.*, 129, (9) 2250-2257

National Center for Biotechnology Information and National Library of Medicine: Database of Single Nucleotide Polymorphisms (dbSNP). Website: http://www.ncbi.nlm.nih.gov/SNP/. Last date accessed: 24 Apr. 2011

Ngo, D.T., Sverdlov, A.L., McNeil, J.J., & Horowitz, J.D. 2010. Does vitamin D modulate asymmetric dimethylarginine and C-reactive protein concentrations? *Am.J.Med.*, 123, (4) 335-341

Ni, M.C., Rodgers, A., Pan, W.H., Gu, D.F., & Woodward, M. 2004. Body mass index and cardiovascular disease in the Asia-Pacific Region: an overview of 33 cohorts involving 310 000 participants. *Int J.Epidemiol.*, 33, (4) 751-758

Nitsch, D., Molokhia, M., Smeeth, L., DeStavola, B.L., Whittaker, J.C., & Leon, D.A. 2006. Limits to causal inference based on Mendelian randomization: a comparison with randomized controlled trials. *Am.J.Epidemiol.*, 163, (5) 397-403

Norman, A.W. & Bouillon, R. 2010. Vitamin D nutritional policy needs a vision for the future. *Exp.Biol.Med.(Maywood.)*, 235, (9) 1034-1045

O'Connell, T.D. & Simpson, R.U. 1996. Immunochemical identification of the 1,25-dihydroxyvitamin D3 receptor protein in human heart. *Cell Biol.Int.*, 20, (9) 621-624

Ockene, I.S., Matthews, C.E., Rifai, N., Ridker, P.M., Reed, G., & Stanek, E. 2001. Variability and classification accuracy of serial high-sensitivity C-reactive protein measurements in healthy adults. *Clin.Chem.*, 47, (3) 444-450

Omdahl, J. L. & May, B. 2005, "The 25-Hydroxyvitamin D 24-Hydroxylase," *In Vitamin D*, 2nd ed ed. D. Feldman, F. H. Glorieux, & J. W. Pike, eds., Amsterdam: Elsevier Academic Press, pp. 85-104.

Orton, S.M., Morris, A.P., Herrera, B.M., Ramagopalan, S.V., Lincoln, M.R., Chao, M.J., Vieth, R., Sadovnick, A.D., & Ebers, G.C. 2008. Evidence for genetic regulation of vitamin D status in twins with multiple sclerosis. *Am.J.Clin.Nutr.*, 88, (2) 441-447

Ozkan, B. 2010. Nutritional rickets. *J.Clin.Res.Pediatr.Endocrinol.*, 2, (4) 137-143

Palmer, T.M., Lawlor, D.A., Harbord, R.M., Sheehan, N.A., Tobias, J.H., Timpson, N.J., Smith, G.D., & Sterne, J.A. 2011a. Using multiple genetic variants as instrumental variables for modifiable risk factors. *Stat.Methods Med.Res.*

Palmer, T.M., Sterne, J.A., Harbord, R.M., Lawlor, D.A., Sheehan, N.A., Meng, S., Granell, R., Smith, G.D., & Didelez, V. 2011b. Instrumental variable estimation of causal risk ratios and causal odds ratios in Mendelian randomization analyses. *Am.J.Epidemiol.*, 173, (12) 1392-1403

Parsons, T.J., Thomas, C., & Power, C. 2009. Estimated activity patterns in British 45 year olds: cross-sectional findings from the 1958 British birth cohort. *Eur.J.Clin.Nutr.*, 63, (8) 978-985

Pe'er, I., de Bakker, P.I., Maller, J., Yelensky, R., Altshuler, D., & Daly, M.J. 2006. Evaluating and improving power in whole-genome association studies using fixed marker sets. *Nat.Genet.*, 38, (6) 663-667

Pearson, T.A. & Manolio, T.A. 2008. How to interpret a genome-wide association study. *JAMA*, 299, (11) 1335-1344

Pell, J.P. & Cobbe, S.M. 1999. Seasonal variations in coronary heart disease. *QJM.*, 92, (12) 689-696

Pepys, M.B. & Hirschfield, G.M. 2003. C-reactive protein: a critical update. *J.Clin.Invest*, 111, (12) 1805-1812

Pfister, R., Sharp, S., Luben, R., Welsh, P., Barroso, I., Salomaa, V., Meirhaeghe, A., Khaw, K.T., Sattar, N., Langenberg, C., & Wareham, N.J. 2011. Mendelian randomization study of B-type natriuretic Peptide and type 2 diabetes: evidence of causal association from population studies. *PLoS.Med.*, 8, (10) e1001112

Pickle, L.W. & Gillum, R.F. 1999. Geographic variation in cardiovascular disease mortality in US blacks and whites. *J.Natl.Med.Assoc.*, 91, (10) 545-556

Pierce, B.L., Ahsan, H., & Vanderweele, T.J. 2011. Power and instrument strength requirements for Mendelian randomization studies using multiple genetic variants. *Int.J.Epidemiol.*, 40, (3) 740-752

Pike, J. W. & Shevde, N. K. 2005, "The Vitamin D Receptor," *In Vitamin D*, 2nd ed ed. D. Feldman, F. H. Glorieux, & J. W. Pike, eds., Amsterdam: Elsevier Academic Press, pp. 167-192.

Pilz, S., Dobnig, H., Nijpels, G., Heine, R.J., Stehouwer, C.D., Snijder, M.B., van Dam, R.M., & Dekker, J.M. 2009. Vitamin D and mortality in older men and women. *Clin.Endocrinol.(Oxf)*, 71, (5) 666-672

Pilz, S., Marz, W., Wellnitz, B., Seelhorst, U., Fahrleitner-Pammer, A., Dimai, H.P., Boehm, B.O., & Dobnig, H. 2008. Association of vitamin D deficiency with heart failure and sudden cardiac death in a large cross-sectional study of patients referred for coronary angiography. *J.Clin.Endocrinol.Metab*, 93, (10) 3927-3935

Pittas, A.G., Chung, M., Trikalinos, T., Mitri, J., Brendel, M., Patel, K., Lichtenstein, A.H., Lau, J., & Balk, E.M. 2010a. Systematic review: Vitamin D and cardiometabolic outcomes. *Ann.Intern.Med.*, 152, (5) 307-314

Pittas, A.G., Harris, S.S., Stark, P.C., & Dawson-Hughes, B. 2007. The effects of calcium and vitamin D supplementation on blood glucose and markers of inflammation in nondiabetic adults. *Diabetes Care*, 30, (4) 980-986

Pittas, A.G., Sun, Q., Manson, J.E., Dawson-Hughes, B., & Hu, F.B. 2010b. Plasma 25-hydroxyvitamin D concentration and risk of incident type 2 diabetes in women. *Diabetes Care*, 33, (9) 2021-2023

Power, C. & Elliott, J. 2006. Cohort profile: 1958 British birth cohort (National Child Development Study). *Int.J.Epidemiol.*, 35, (1) 34-41

Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A., & Reich, D. 2006. Principal components analysis corrects for stratification in genome-wide association studies. *Nat.Genet.*, 38, (8) 904-909

Price, A.L., Weale, M.E., Patterson, N., Myers, S.R., Need, A.C., Shianna, K.V., Ge, D., Rotter, J.I., Torres, E., Taylor, K.D., Goldstein, D.B., & Reich, D. 2008. Long-range LD can confound genome scans in admixed populations. *Am.J.Hum.Genet.*, 83, (1) 132-135

Purcell, S.: PLINK (v1.05). Website: pngu.mgh.harvard.edu/purcell/plink. Last date accessed: 23 Apr. 2009

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J., & Sham, P.C. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am.J.Hum.Genet.*, 81, (3) 559-575

Puri, S., Bansal, D.D., Uskokovic, M.R., & MacGregor, R.R. 2000. Induction of tissue plasminogen activator secretion from rat heart microvascular cells by fM 1,25(OH)(2)D(3). *Am.J.Physiol Endocrinol.Metab*, 278, (2) E293-E301

R Development Core Team. R: A Language and Environment for Statistical Computing. 2008. R Foundation for Statistical Computing.
Ref Type: Computer Program

Rader, D.J. & Daugherty, A. 2008. Translating molecular discoveries into new therapies for atherosclerosis. *Nature*, 451, (7181) 904-913

Raimondi, S., Johansson, H., Maisonneuve, P., & Gandini, S. 2009. Review and meta-analysis on vitamin D receptor polymorphisms and cancer risk. *Carcinogenesis*, 30, (7) 1170-1180

Ramagopalan, S.V., Heger, A., Berlanga, A.J., Maugeri, N.J., Lincoln, M.R., Burrell, A., Handunnetthi, L., Handel, A.E., Disanto, G., Orton, S.M., Watson, C.T., Morahan, J.M., Giovannoni, G., Ponting, C.P., Ebers, G.C., & Knight, J.C. 2010. A ChIP-seq defined genome-wide map of vitamin D receptor binding: associations with disease and evolution. *Genome Res.*, 20, (10) 1352-1360

Ramos-López, E., Bruck, P., Jansen, T., Herwig, J., & Badenhoop, K. 2007. CYP2R1 (vitamin D 25-hydroxylase) gene is associated with susceptibility to type 1 diabetes and vitamin D levels in Germans. *Diabetes Metab Res.Rev.*, 23, (8) 631-636

Ramos-López, E., Kahles, H., Weber, S., Kukic, A., Penna-Martinez, M., Badenhoop, K., & Louwen, F. 2008. Gestational diabetes mellitus and vitamin D deficiency: genetic contribution of CYP27B1 and CYP2R1 polymorphisms. *Diabetes Obes.Metab*, 10, (8) 683-685

Ramos-López, E., Lange, B., Penna-Martinez, M., Bruck, P., Swiech, K., Mauf, S., Kahles, H., & Badenhoop, K. 2010. The role of cubilin gene polymorphisms and their influence on 25(OH)D3 and 1,25(OH)2D3 plasma levels in type 1 diabetes patients. *J.Steroid Biochem.Mol.Biol.*, 121, (1-2) 442-444

Randerson-Moor, J.A., Taylor, J.C., Elliott, F., Chang, Y.M., Beswick, S., Kukalizch, K., Affleck, P., Leake, S., Haynes, S., Karpavicius, B., Marsden, J., Gerry, E., Bale, L., Bertram, C., Field, H., Barth, J.H., Silva, I.S., Swerdlow, A., Kanetsky, P.A., Barrett, J.H., Bishop, D.T., & Bishop, J.A. 2009. Vitamin D receptor gene polymorphisms, serum 25-hydroxyvitamin D levels, and melanoma: UK case-control comparisons and a meta-analysis of published VDR data. *Eur.J.Cancer*, 45, (18) 3271-3281

Refinetti, R., Cornelissen, G., & Halberg, F. 2007. Procedures for numerical analysis of circadian rhythms. *Biological Rhythm Research*, 38, (4) 275-325

Reid, D., Toole, B.J., Knox, S., Talwar, D., Harten, J., O'Reilly, D.S., Blackwell, S., Kinsella, J., McMillan, D.C., & Wallace, A.M. 2011. The relation between acute changes in the systemic inflammatory response and plasma 25-hydroxyvitamin D concentrations after elective knee arthroplasty. *Am.J.Clin.Nutr.*, 93, (5) 1006-1011

Reinhardt, T.A. & Horst, R.L. 1990. Parathyroid hormone down-regulates 1,25-dihydroxyvitamin D receptors (VDR) and VDR messenger ribonucleic acid in vitro and blocks homologous up-regulation of VDR in vivo. *Endocrinology*, 127, (2) 942-948

Reusch, J., Ackermann, H., & Badenhoop, K. 2009. Cyclic changes of vitamin D and PTH are primarily regulated by solar radiation: 5-year analysis of a German (50 degrees N) population. *Horm.Metab Res.*, 41, (5) 402-407

Rice, J.A. 1995. *Mathematical Statistics and Data Analysis*, 2nd ed. International Thomson Publishing.

Ridker, P.M. 2003. Clinical application of C-reactive protein for cardiovascular disease detection and prevention. *Circulation*, 107, (3) 363-369

Risch, N. & Merikangas, K. 1996. The future of genetic studies of complex human diseases. *Science*, 273, (5281) 1516-1517

Robinson, J.G., Manson, J.E., Larson, J., Liu, S., Song, Y., Howard, B.V., Phillips, L., Shikany, J.M., Allison, M., Curb, J.D., Johnson, K.C., & Watts, N. 2011. Lack of association between 25(OH)D levels and incident type 2 diabetes in older women. *Diabetes Care*, 34, (3) 628-634

Rockell, J.E., Green, T.J., Skeaff, C.M., Whiting, S.J., Taylor, R.W., Williams, S.M., Parnell, W.R., Scragg, R., Wilson, N., Schaaf, D., Fitzgerald, E.D., & Wohlers,

M.W. 2005. Season and ethnicity are determinants of serum 25-hydroxyvitamin D concentrations in New Zealand children aged 5-14 y. *J.Nutr.*, 135, (11) 2602-2608

Rodríguez, M., Martínez-Moreno, J.M., Rodríguez-Ortiz, M.E., Muñoz-Castañeda, J.R., & Almaden, Y. 2011. Vitamin D and vascular calcification in chronic kidney disease. *Kidney Blood Press Res.*, 34, (4) 261-268

Rogowski, O., Toker, S., Shapira, I., Melamed, S., Shirom, A., Zeltser, D., & Berliner, S. 2005. Values of high-sensitivity C-reactive protein in each month of the year in apparently healthy individuals. *Am.J.Cardiol.*, 95, (1) 152-155

Roitt, I. & Rabson, A. 2000. *Really essential medical immunology* Blackwell Science Ltd.

Ross, R. 1993. The pathogenesis of atherosclerosis: a perspective for the 1990s. *Nature*, 362, (6423) 801-809

Rossi, M., Federico, G., Corso, G., Parenti, G., Battagliese, A., Frascogna, A.R., Della, C.R., Dello, R.A., Strisciuglio, P., Saggese, G., & Andria, G. 2005. Vitamin D status in patients affected by Smith-Lemli-Opitz syndrome. *J.Inherit.Metab Dis.*, 28, (1) 69-80

Rothman, K. J., Greenland, S., Poole, C., & Lash, T. L. 2008, "Causation and Causal Inference," *In Modern Epidemiology*, Third ed. K. J. Rothman, S. Greenland, & T. L. Lash, eds., Lippincott Williams & Wilkins, pp. 5-31.

Rothman, K.J. & Greenland, S. 1998. *Modern Epidemiology*, Second ed. Lippincott Williams & Wilkins.

Royston, P. 2004. Multiple imputation of missing values. *Stata Journal*, 4, (3) 227-241

Royston, P., Carlin, J.B., & White, I.R. 2009. Multiple imputation of missing values: New features for mim. *Stata Journal*, 9, (2) 252-264

Rudez, G., Meijer, P., Spronk, H.M., Leebeek, F.W., ten, C.H., Kluft, C., & de Maat, M.P. 2009. Biological variation in inflammatory and hemostatic markers. *J.Thromb.Haemost.*, 7, (8) 1247-1255

Rudnicka, A.R., Rumley, A., Lowe, G.D., & Strachan, D.P. 2007. Diurnal, seasonal, and blood-processing patterns in levels of circulating fibrinogen, fibrin D-dimer, C-reactive protein, tissue plasminogen activator, and von Willebrand factor in a 45-year-old population. *Circulation*, 115, (8) 996-1003

Ruggeri, Z.M. & Ware, J. 1992. The structure and function of von Willebrand factor. *Thromb.Haemost.*, 67, (6) 594-599

Rumley, A., Emberson, J.R., Wannamethee, S.G., Lennon, L., Whincup, P.H., & Lowe, G.D. 2006. Effects of older age on fibrin D-dimer, C-reactive protein, and other hemostatic and inflammatory variables in men aged 60-79 years. *J.Thromb.Haemost.*, 4, (5) 982-987

Ruston, D., Office for National Statistics, Medical Research Council, & Human Nutrition Research 2004. *The National Diet & Nutrition Survey adults aged 19 to 64 years* London, TSO.

Sacheck, J., Goodman, E., Chui, K., Chomitz, V., Must, A., & Economos, C. 2011. Vitamin D Deficiency, Adiposity, and Cardiometabolic Risk in Urban Schoolchildren. *J.Pediatr.*

Sachidanandam, R., Weissman, D., Schmidt, S.C., Kakol, J.M., Stein, L.D., Marth, G., Sherry, S., Mullikin, J.C., Mortimore, B.J., Willey, D.L., Hunt, S.E., Cole, C.G., Coggill, P.C., Rice, C.M., Ning, Z., Rogers, J., Bentley, D.R., Kwok, P.Y., Mardis, E.R., Yeh, R.T., Schultz, B., Cook, L., Davenport, R., Dante, M., Fulton, L., Hillier, L., Waterston, R.H., McPherson, J.D., Gilman, B., Schaffner, S., Van Etten, W.J., Reich, D., Higgins, J., Daly, M.J., Blumenstiel, B., Baldwin, J., Stange-Thomann, N., Zody, M.C., Linton, L., Lander, E.S., & Altshuler, D. 2001. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature*, 409, (6822) 928-933

Safadi, F.F., Thornton, P., Magiera, H., Hollis, B.W., Gentile, M., Haddad, J.G., Liebhaber, S.A., & Cooke, N.E. 1999. Osteopathy and resistance to vitamin D toxicity in mice null for vitamin D binding protein. *J.Clin.Invest*, 103, (2) 239-251

Sawcer, S., Hellenthal, G., Pirinen, M., Spencer, C.C., Patsopoulos, N.A., Moutsianas, L., Dilthey, A., Su, Z., Freeman, C., Hunt, S.E., Edkins, S., Gray, E., Booth, D.R., Potter, S.C., Goris, A., Band, G., Oturai, A.B., Strange, A., Saarela, J., Bellenguez, C., Fontaine, B., Gillman, M., Hemmer, B., Gwilliam, R., Zipp, F., Jayakumar, A., Martin, R., Leslie, S., Hawkins, S., Giannoulatou, E., D'alfonso, S., Blackburn, H., Boneschi, F.M., Liddle, J., Harbo, H.F., Perez, M.L., Spurkland, A., Waller, M.J., Mycko, M.P., Ricketts, M., Comabella, M., Hammond, N., Kockum, I., McCann, O.T., Ban, M., Whittaker, P., Kemppinen, A., Weston, P., Hawkins, C., Widaa, S., Zajicek, J., Dronov, S., Robertson, N., Bumpstead, S.J., Barcellos, L.F., Ravindrarajah, R., Abraham, R., Alfredsson, L., Ardlie, K., Aubin, C., Baker, A., Baker, K., Baranzini, S.E., Bergamaschi, L., Bergamaschi, R., Bernstein, A., Berthele, A., Boggild, M., Bradfield, J.P., Brassat, D., Broadley, S.A., Buck, D., Butzkueven, H., Capra, R., Carroll, W.M., Cavalla, P., Celius, E.G., Cepok, S., Chiavacci, R., Clerget-Darpoux, F., Clysters, K., Comi, G., Cossburn, M., Cournu-Rebeix, I., Cox, M.B., Cozen, W., Cree, B.A., Cross, A.H., Cusi, D., Daly, M.J., Davis, E., de Bakker, P.I., Debouverie, M., D'hooghe, M.B., Dixon, K., Dobosi, R., Dubois, B., Ellinghaus, D., Elovaara, I., Esposito, F., Fontenille, C., Foote, S., Franke, A., Galimberti, D., Ghezzi, A., Glessner, J., Gomez, R., Gout, O., Graham, C., Grant, S.F., Guerini, F.R., Hakonarson, H., Hall, P., Hamsten, A., Hartung, H.P., Heard, R.N., Heath, S., Hobart, J., Hoshi, M., Infante-Duarte, C., Ingram, G., Ingram, W., Islam, T., Jagodic, M., Kabesch, M., Kermode, A.G., Kilpatrick, T.J., Kim, C., Klopp, N., Koivisto, K., Larsson, M., Lathrop, M., Lechner-Scott, J.S., Leone, M.A., Leppa, V., Liljedahl, U., Bomfim, I.L., Lincoln, R.R., Link, J., Liu, J., Lorentzen, A.R., Lupoli, S., Macciardi, F., Mack, T., Marriott, M., Martinelli, V., Mason, D., McCauley, J.L., Mentch, F., Mero, I.L., Mihalova, T., Montalban, X., Mottershead, J., Myhr, K.M., Naldi, P., Ollier, W., Page, A., Palotie, A., Pelletier, J., Piccio, L., Pickersgill, T., Piehl, F., Pobywajlo, S., Quach, H.L., Ramsay, P.P., Reunanen, M., Reynolds, R., Rioux, J.D., Rodegher, M., Roesner, S., Rubio, J.P., Ruckert, I.M., Salvetti, M., Salvi, E., Santaniello, A., Schaefer, C.A., Schreiber, S.,

Schulze, C., Scott, R.J., Sellebjerg, F., Selmaj, K.W., Sexton, D., Shen, L., Simms-Acuna, B., Skidmore, S., Sleiman, P.M., Smestad, C., Sorensen, P.S., Sondergaard, H.B., Stankovich, J., Strange, R.C., Sulonen, A.M., Sundqvist, E., Syvanen, A.C., Taddeo, F., Taylor, B., Blackwell, J.M., Tienari, P., Bramon, E., Tourbah, A., Brown, M.A., Tronczynska, E., Casas, J.P., Tubridy, N., Corvin, A., Vickery, J., Jankowski, J., Villoslada, P., Markus, H.S., Wang, K., Mathew, C.G., Wason, J., Palmer, C.N., Wichmann, H.E., Plomin, R., Willoughby, E., Rautanen, A., Winkelmann, J., Wittig, M., Trembath, R.C., Yaouanq, J., Viswanathan, A.C., Zhang, H., Wood, N.W., Zuvich, R., Deloukas, P., Langford, C., Duncanson, A., Oksenberg, J.R., Pericak-Vance, M.A., Haines, J.L., Olsson, T., Hillert, J., Ivinson, A.J., De Jager, P.L., Peltonen, L., Stewart, G.J., Hafler, D.A., Hauser, S.L., McVean, G., Donnelly, P., & Compston, A. 2011. Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature*, 476, (7359) 214-219

Sayre, R.M. & Dowdy, J.C. 2007. Darkness at noon: sunscreens and vitamin D3. *Photochem.Photobiol.*, 83, (2) 459-463

Scarborough, P, Bhatnagar, P, Wickramasinghe, K, Smolina, K, Mitchel, C, and Rayner, M: Mortality. Website: http://www.heartstats.org/datapage.asp?id=9075. Last date accessed: 2010

Schuster, I. 2010. Cytochromes P450 are essential players in the vitamin D signaling system. *Biochim.Biophys.Acta*, 1814, (1) 186-199

Scientific Advisory Committee on Nutrition 2007. *Update on vitamin D position statement* London, Stationery Office.

Scragg, R. 1981. Seasonality of cardiovascular disease mortality and the possible protective effect of ultra-violet radiation. *Int J.Epidemiol.*, 10, (4) 337-341

Scragg, R., Jackson, R., Holdaway, I.M., Lim, T., & Beaglehole, R. 1990. Myocardial infarction is inversely associated with plasma 25-hydroxyvitamin D3 levels: a community-based study. *Int J.Epidemiol.*, 19, (3) 559-563

Selkirk, T.K., Greenebaum, J.V., & Mitchell, A.G. 1928. Studies in Rickets: III Prevention by means of ultraviolet irradiation. *JAMA*, 91, (26) 2035-2088

Semba, R.D., Houston, D.K., Bandinelli, S., Sun, K., Cherubini, A., Cappola, A.R., Guralnik, J.M., & Ferrucci, L. 2010. Relationship of 25-hydroxyvitamin D with all-cause and cardiovascular disease mortality in older community-dwelling adults. *Eur.J.Clin.Nutr.*, 64, (2) 203-209

Seng, K.C. & Seng, C.K. 2008. The success of the genome-wide association approach: a brief story of a long struggle. *Eur.J.Hum.Genet.*, 16, (5) 554-564

Shea, M.K., Benjamin, E.J., Dupuis, J., Massaro, J.M., Jacques, P.F., D'Agostino, R.B., Sr., Ordovas, J.M., O'Donnell, C.J., Dawson-Hughes, B., Vasan, R.S., & Booth, S.L. 2009. Genetic and non-genetic correlates of vitamins K and D. *Eur.J.Clin.Nutr.*, 63, (4) 458-464

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

Shephard, R.M. & DeLuca, H.F. 1980. Plasma concentrations of vitamin D3 and its metabolites in the rat as influenced by vitamin D3 or 25-hydroxyvitamin D3 intakes. *Arch.Biochem.Biophys.*, 202, (1) 43-53

Shepherd, P. 1995, *The National Child Development Study (NCDS): An Introduction to the Origins of the Study and the Methods of Data Collection*, Centre for Longitudinal Studies.

Shoji, T., Shinohara, K., Kimoto, E., Emoto, M., Tahara, H., Koyama, H., Inaba, M., Fukumoto, S., Ishimura, E., Miki, T., Tabata, T., & Nishizawa, Y. 2004. Lower risk for cardiovascular mortality in oral 1alpha-hydroxy vitamin D3 users in a haemodialysis population. *Nephrol.Dial.Transplant.*, 19, (1) 179-184

Shroff, R., Egerton, M., Bridel, M., Shah, V., Donald, A.E., Cole, T.J., Hiorns, M.P., Deanfield, J.E., & Rees, L. 2008. A bimodal association of vitamin D levels and vascular disease in children on dialysis. *J.Am.Soc.Nephrol.*, 19, (6) 1239-1246

Shultz, T.D., Fox, J., Heath, H., III, & Kumar, R. 1983. Do tissues other than the kidney produce 1,25-dihydroxyvitamin D3 in vivo? A reexamination. *Proc.Natl.Acad.Sci.U.S.A*, 80, (6) 1746-1750

Simpson, R.U., Hershey, S.H., & Nibbelink, K.A. 2007. Characterization of heart size and blood pressure in the vitamin D receptor knockout mouse. *J.Steroid Biochem.Mol.Biol.*, 103, (3-5) 521-524

Sinha, S., Luben, R.N., Welch, A., Bingham, S., Wareham, N.J., Day, N.E., & Khaw, K.T. 2005. Fibrinogen and cigarette smoking in men and women in the European Prospective Investigation into Cancer in Norfolk (EPIC-Norfolk) population. *Eur.J.Cardiovasc.Prev.Rehabil.*, 12, (2) 144-150

Sinotte, M., Diorio, C., Berube, S., Pollak, M., & Brisson, J. 2009. Genetic polymorphisms of the vitamin D binding protein and plasma concentrations of 25-hydroxyvitamin D in premenopausal women. *Am.J.Clin.Nutr.*, 89, (2) 634-640

Smith, A., Patterson, C., Yarnell, J., Rumley, A., Ben-Shlomo, Y., & Lowe, G. 2005. Which hemostatic markers add to the predictive value of conventional risk factors for coronary heart disease and ischemic stroke? The Caerphilly Study. *Circulation*, 112, (20) 3080-3087

Smith, W.C. & Tunstall-Pedoe, H. 1984. European regional variation in cardiovascular mortality. *Br.Med.Bull*, 40, (4) 374-379

Smolders, J., Damoiseaux, J., Menheere, P., Tervaert, J.W., & Hupperts, R. 2009. Fok-I vitamin D receptor gene polymorphism (rs10735810) and vitamin D metabolism in multiple sclerosis. *J.Neuroimmunol.*, 207, (1-2) 117-121

Snellman, G., Melhus, H., Gedeborg, R., Olofsson, S., Wolk, A., Pedersen, N.L., & Michaelsson, K. 2009. Seasonal genetic influence on serum 25-hydroxyvitamin D levels: a twin study. *PLoS One.*, 4, (11) e7747

Snijder, M.B., van Dam, R.M., Visser, M., Deeg, D.J., Dekker, J.M., Bouter, L.M., Seidell, J.C., & Lips, P. 2005. Adiposity in relation to vitamin D status and parathyroid hormone levels: a population-based study in older men and women. *J.Clin.Endocrinol.Metab*, 90, (7) 4119-4123

Speeckaert, M., Huang, G., Delanghe, J.R., & Taes, Y.E. 2006. Biological and clinical aspects of the vitamin D binding protein (Gc-globulin) and its polymorphism. *Clin.Chim.Acta*, 372, (1-2) 33-42

Spiegelhalter, D. & Lunn, D. J. 2009, *Introduction to Bayesian Analysis Using Winbugs*, MRC Biostatistics Unit, Institute of Public Health.

Spiegelhalter, D.J., Best, N.G., Carlin, B.R., & van der Linde, A. 2002. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society Series B-Statistical Methodology*, 64, 583-616

Springbett, P., Buglass, S., & Young, A.R. 2010. Photoprotection and vitamin D status. *J.Photochem.Photobiol.B*, 101, (2) 160-168

Staiger, D. & Stock, J.H. 1997. Instrumental Variables Regression with Weak Instruments. *Econometrica*, 65, (3) 557

StataCorp. Stata Statistical Software: Release 10. 2010. StataCorp LP.
Ref Type: Computer Program

Steenbock, H. 1924. The induction growth promoting and calcifying properties in a ration by exposure to light. *Science*, 60, (1549) 224-225

Steinvil, A., Berliner, S., Bromberg, M., Cohen, M., Shalev, V., Shapira, I., & Rogowski, O. 2009. Micro-inflammatory changes in asymptomatic healthy adults during bouts of respiratory tract infections in the community: potential triggers for atherothrombotic events. *Atherosclerosis*, 206, (1) 270-275

Sulem, P., Gudbjartsson, D.F., Stacey, S.N., Helgason, A., Rafnar, T., Magnusson, K.P., Manolescu, A., Karason, A., Palsson, A., Thorleifsson, G., Jakobsdottir, M., Steinberg, S., Palsson, S., Jonasson, F., Sigurgeirsson, B., Thorisdottir, K., Ragnarsson, R., Benediktsdottir, K.R., Aben, K.K., Kiemeney, L.A., Olafsson, J.H., Gulcher, J., Kong, A., Thorsteinsdottir, U., & Stefansson, K. 2007. Genetic determinants of hair, eye and skin pigmentation in Europeans. *Nat.Genet.*, 39, (12) 1443-1452

Sullivan, P. F. & Purcell, S. 2008, "Analyzing genome-wide association study data: a tutorial using PLINK," *In Statistical Genetics: Gene Mapping through Linkage and Association*, B. M. Neale et al., eds., Taylor & Francis, pp. 355-394.

Suzuki, Y., Ichiyama, T., Ohsaki, A., Hasegawa, S., Shiraishi, M., & Furukawa, S. 2009. Anti-inflammatory effect of 1alpha,25-dihydroxyvitamin D(3) in human

coronary arterial endothelial cells: Implication for the treatment of Kawasaki disease. *J.Steroid Biochem.Mol.Biol.*, 113, (1-2) 134-138

Swales, H.H. & Wang, T.J. 2010. Vitamin D and cardiovascular disease risk: emerging evidence. *Curr.Opin.Cardiol.*, 25, (5) 513-517

Tang, C., Chen, N., Wu, M., Yuan, H., & Du, Y. 2009. Fok1 polymorphism of vitamin D receptor gene contributes to breast cancer susceptibility: a meta-analysis. *Breast Cancer Res.Treat.*, 117, (2) 391-399

Teo, Y.Y., Inouye, M., Small, K.S., Gwilliam, R., Deloukas, P., Kwiatkowski, D.P., & Clark, T.G. 2007. A genotype calling algorithm for the Illumina BeadArray platform. *Bioinformatics.*, 23, (20) 2741-2746

Thomas, D. C. 2004a, "Linkage Analysis," *In Statistical Methods in Genetic Epidemiology*, Oxford University Press, pp. 167-226.

Thomas, D. C. 2004b, "Overview of Genetic Epidemiology," *In Statistical Methods in Genetic Epidemiology*, Oxford University Press, pp. 3-24.

Thomas, D.C. 2004c. *Statistical Methods in Genetic Epidemiology* Oxford University Press.

Tian, C., Gregersen, P.K., & Seldin, M.F. 2008. Accounting for ancestry: population substructure and genome-wide association studies. *Hum.Mol.Genet.*, 17, (R2) R143-R150

Timpson, N.J., Lawlor, D.A., Harbord, R.M., Gaunt, T.R., Day, I.N., Palmer, L.J., Hattersley, A.T., Ebrahim, S., Lowe, G.D., Rumley, A., & Davey, S.G. 2005. C-reactive protein and its role in metabolic syndrome: mendelian randomisation study. *Lancet*, 366, (9501) 1954-1959

Tint, G.S., Irons, M., Elias, E.R., Batta, A.K., Frieden, R., Chen, T.S., & Salen, G. 1994. Defective cholesterol biosynthesis associated with the Smith-Lemli-Opitz syndrome. *N.Engl.J.Med.*, 330, (2) 107-113

Tofler, G.H., Massaro, J., Levy, D., Mittleman, M., Sutherland, P., Lipinska, I., Muller, J.E., & D'Agostino, R.B. 2005. Relation of the prothrombotic state to increasing age (from the Framingham Offspring Study). *Am.J.Cardiol.*, 96, (9) 1280-1283

Tong, Y.L. 1976. Parameter Estimation in Studying Circadian-Rhythms. *Biometrics*, 32, (1) 85-94

Trang, H.M., Cole, D.E., Rubin, L.A., Pierratos, A., Siu, S., & Vieth, R. 1998. Evidence that vitamin D3 increases serum 25-hydroxyvitamin D more efficiently than does vitamin D2. *Am.J.Clin.Nutr.*, 68, (4) 854-858

Unal, B., Critchley, J.A., & Capewell, S. 2004. Explaining the decline in coronary heart disease mortality in England and Wales between 1981 and 2000. *Circulation*, 109, (9) 1101-1107

US Department of Energy Genome Programs: Human Genome Project Information. Website: http://genomics.energy.gov. Last date accessed: 24 Apr. 2011

van der Bom, J.G., de Maat, M.P., Bots, M.L., Hofman, A., Kluft, C., & Grobbee, D.E. 1997. Seasonal variation in fibrinogen in the Rotterdam Study. *Thromb.Haemost.*, 78, (3) 1059-1062

van der Sluis, S. & Posthuma, D. 2008, "Single-locus association models," *In Statistical Genetics: Gene Mapping through Linkage and Association*, B. M. Neale et al., eds., Taylor & Francis, pp. 323-354.

van Durme, Y.M., Lahousse, L., Verhamme, K.M., Stolk, L., Eijgelsheim, M., Loth, D.W., Uitterlinden, A.G., Breteler, M.M., Joos, G.F., Hofman, A., Stricker, B.H., & Brusselle, G.G. 2011. Mendelian Randomization Study of Interleukin-6 in Chronic Obstructive Pulmonary Disease. *Respiration*

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., Gocayne, J.D., Amanatides, P., Ballew, R.M., Huson, D.H., Wortman, J.R., Zhang, Q., Kodira, C.D., Zheng, X.H., Chen, L., Skupski, M., Subramanian, G., Thomas, P.D., Zhang, J., Gabor Miklos, G.L., Nelson, C., Broder, S., Clark, A.G., Nadeau, J., McKusick, V.A., Zinder, N., Levine, A.J., Roberts, R.J., Simon, M., Slayman, C., Hunkapiller, M., Bolanos, R., Delcher, A., Dew, I., Fasulo, D., Flanigan, M., Florea, L., Halpern, A., Hannenhalli, S., Kravitz, S., Levy, S., Mobarry, C., Reinert, K., Remington, K., Abu-Threideh, J., Beasley, E., Biddick, K., Bonazzi, V., Brandon, R., Cargill, M., Chandramouliswaran, I., Charlab, R., Chaturvedi, K., Deng, Z., Di, F., V, Dunn, P., Eilbeck, K., Evangelista, C., Gabrielian, A.E., Gan, W., Ge, W., Gong, F., Gu, Z., Guan, P., Heiman, T.J., Higgins, M.E., Ji, R.R., Ke, Z., Ketchum, K.A., Lai, Z., Lei, Y., Li, Z., Li, J., Liang, Y., Lin, X., Lu, F., Merkulov, G.V., Milshina, N., Moore, H.M., Naik, A.K., Narayan, V.A., Neelam, B., Nusskern, D., Rusch, D.B., Salzberg, S., Shao, W., Shue, B., Sun, J., Wang, Z., Wang, A., Wang, X., Wang, J., Wei, M., Wides, R., Xiao, C., Yan, C., Yao, A., Ye, J., Zhan, M., Zhang, W., Zhang, H., Zhao, Q., Zheng, L., Zhong, F., Zhong, W., Zhu, S., Zhao, S., Gilbert, D., Baumhueter, S., Spier, G., Carter, C., Cravchik, A., Woodage, T., Ali, F., An, H., Awe, A., Baldwin, D., Baden, H., Barnstead, M., Barrow, I., Beeson, K., Busam, D., Carver, A., Center, A., Cheng, M.L., Curry, L., Danaher, S., Davenport, L., Desilets, R., Dietz, S., Dodson, K., Doup, L., Ferriera, S., Garg, N., Gluecksmann, A., Hart, B., Haynes, J., Haynes, C., Heiner, C., Hladun, S., Hostin, D., Houck, J., Howland, T., Ibegwam, C., Johnson, J., Kalush, F., Kline, L., Koduru, S., Love, A., Mann, F., May, D., McCawley, S., McIntosh, T., McMullen, I., Moy, M., Moy, L., Murphy, B., Nelson, K., Pfannkoch, C., Pratts, E., Puri, V., Qureshi, H., Reardon, M., Rodriguez, R., Rogers, Y.H., Romblad, D., Ruhfel, B., Scott, R., Sitter, C., Smallwood, M., Stewart, E., Strong, R., Suh, E., Thomas, R., Tint, N.N., Tse, S., Vech, C., Wang, G., Wetter, J., Williams, S., Williams, M., Windsor, S., Winn-Deen, E., Wolfe, K., Zaveri, J., Zaveri, K., Abril, J.F., Guigo, R., Campbell, M.J., Sjolander, K.V., Karlak, B., Kejariwal, A., Mi, H., Lazareva, B., Hatton, T., Narechania, A., Diemer, K., Muruganujan, A., Guo, N., Sato, S., Bafna, V., Istrail, S., Lippert, R., Schwartz, R., Walenz, B., Yooseph, S., Allen, D., Basu, A., Baxendale, J., Blick, L., Caminha, M., Carnes-Stine, J., Caulk, P., Chiang, Y.H., Coyne, M., Dahlke, C., Mays, A., Dombroski, M., Donnelly, M., Ely, D., Esparham, S., Fosler, C., Gire, H., Glanowski, S., Glasser, K., Glodek, A., Gorokhov, M., Graham, K., Gropman, B.,

Harris, M., Heil, J., Henderson, S., Hoover, J., Jennings, D., Jordan, C., Jordan, J., Kasha, J., Kagan, L., Kraft, C., Levitsky, A., Lewis, M., Liu, X., Lopez, J., Ma, D., Majoros, W., McDaniel, J., Murphy, S., Newman, M., Nguyen, T., Nguyen, N., & Nodell, M. 2001. The sequence of the human genome. *Science*, 291, (5507) 1304-1351

Vieth, R. 1999. Vitamin D supplementation, 25-hydroxyvitamin D concentrations, and safety. *Am.J.Clin.Nutr.*, 69, (5) 842-856

Vieth, R. & El-Hajj, F.G. 2005. There is no lower threshold level for parathyroid hormone as 25-hydroxyvitamin D concentrations increase. *J.Endocrinol.Invest*, 28, (2) 183-186

Vigushin, D.M., Pepys, M.B., & Hawkins, P.N. 1993. Metabolic and scintigraphic studies of radioiodinated human C-reactive protein in health and disease. *J.Clin.Invest*, 91, (4) 1351-1357

Virtanen, J.K., Nurmi, T., Voutilainen, S., Mursu, J., & Tuomainen, T.P. 2010. Association of serum 25-hydroxyvitamin D with the risk of death in a general older population in Finland. *Eur.J.Nutr.*

Vischer, U.M. 2006. von Willebrand factor, endothelial dysfunction, and cardiovascular disease. *J.Thromb.Haemost.*, 4, (6) 1186-1193

Vonesh, E.F. & Chinchilli, V.M. 1997. *Linear and Nonlinear Models for the Analysis of Repeated Measurements* Marcel Dekker.

Wallace, A.M., Gibson, S., de la Hunty, A., Lamberg-Allardt, C., & Ashwell, M. 2010. Measurement of 25-hydroxyvitamin D in the clinical laboratory: current procedures, performance characteristics and limitations. *Steroids*, 75, (7) 477-488

Wallis, D.E., Penckofer, S., & Sizemore, G.W. 2008. The "sunshine deficit" and cardiovascular disease. *Circulation*, 118, (14) 1476-1485

Wang, L., Manson, J.E., Song, Y., & Sesso, H.D. 2010a. Systematic review: Vitamin D and calcium supplementation in prevention of cardiovascular events. *Ann.Intern.Med.*, 152, (5) 315-323

Wang, T.J., Pencina, M.J., Booth, S.L., Jacques, P.F., Ingelsson, E., Lanier, K., Benjamin, E.J., D'Agostino, R.B., Wolf, M., & Vasan, R.S. 2008. Vitamin D Deficiency and Risk of Cardiovascular Disease. *Circulation*, 117, (4) 503-511

Wang, T.J., Zhang, F., Richards, J.B., Kestenbaum, B., van Meurs, J.B., Berry, D., Kiel, D.P., Streeten, E.A., Ohlsson, C., Koller, D.L., Peltonen, L., Cooper, J.D., O'Reilly, P.F., Houston, D.K., Glazer, N.L., Vandenput, L., Peacock, M., Shi, J., Rivadeneira, F., McCarthy, M.I., Anneli, P., de, B., I, Mangino, M., Kato, B., Smyth, D.J., Booth, S.L., Jacques, P.F., Burke, G.L., Goodarzi, M., Cheung, C.L., Wolf, M., Rice, K., Goltzman, D., Hidiroglou, N., Ladouceur, M., Wareham, N.J., Hocking, L.J., Hart, D., Arden, N.K., Cooper, C., Malik, S., Fraser, W.D., Hartikainen, A.L., Zhai, G., Macdonald, H.M., Forouhi, N.G., Loos, R.J., Reid, D.M., Hakim, A., Dennison, E., Liu, Y., Power, C., Stevens, H.E., Jaana, L., Vasan, R.S., Soranzo, N., Bojunga, J., Psaty, B.M., Lorentzon, M., Foroud, T., Harris, T.B., Hofman, A.,

Jansson, J.O., Cauley, J.A., Uitterlinden, A.G., Gibson, Q., Jarvelin, M.R., Karasik, D., Siscovick, D.S., Econs, M.J., Kritchevsky, S.B., Florez, J.C., Todd, J.A., Dupuis, J., Hyppönen, E., & Spector, T.D. 2010b. Common genetic determinants of vitamin D insufficiency: a genome-wide association study. *Lancet*, 376, (9736) 180-188

Wang, W.Y., Barratt, B.J., Clayton, D.G., & Todd, J.A. 2005. Genome-wide association studies: theoretical and practical concerns. *Nat.Rev.Genet.*, 6, (2) 109-118

Wannamethee, S.G., Whincup, P.H., Shaper, A.G., Rumley, A., Lennon, L., & Lowe, G.D. 2009. Circulating inflammatory and hemostatic biomarkers are associated with risk of myocardial infarction and coronary death, but not angina pectoris, in older men. *J.Thromb.Haemost.*, 7, (10) 1605-1611

Ward, M., Berry, D.J., Power, C., & Hyppönen, E. 2011. Working patterns and vitamin D status in mid-life: a cross-sectional study of the 1958 British birth cohort. *Occup.Environ.Med.*, 68, (12) 902-907

Wassernam, R. H. 2005, "Vitamin D and the intestinal absorption of calcuim: a view and overview," *In Vitamin D*, 2nd ed ed. D. Feldman, F. H. Glorieux, & J. W. Pike, eds., Amsterdam: Elsevier Academic Press, pp. 411-428.

Webb, A.R. 2006. Who, what, where and when-influences on cutaneous vitamin D synthesis. *Prog.Biophys.Mol.Biol.*, 92, (1) 17-25

Webb, A.R., DeCosta, B.R., & Holick, M.F. 1989. Sunlight regulates the cutaneous production of vitamin D3 by causing its photodegradation. *J.Clin.Endocrinol.Metab*, 68, (5) 882-887

Webb, A.R., Kift, R., Durkin, M.T., O'Brien, S.J., Vail, A., Berry, J.L., & Rhodes, L.E. 2010. The role of sunlight exposure in determining the vitamin D status of the U.K. white adult population. *Br.J.Dermatol.*, 163, (5) 1050-1055

Weedon, M.N., Lettre, G., Freathy, R.M., Lindgren, C.M., Voight, B.F., Perry, J.R., Elliott, K.S., Hackett, R., Guiducci, C., Shields, B., Zeggini, E., Lango, H., Lyssenko, V., Timpson, N.J., Burtt, N.P., Rayner, N.W., Saxena, R., Ardlie, K., Tobias, J.H., Ness, A.R., Ring, S.M., Palmer, C.N., Morris, A.D., Peltonen, L., Salomaa, V., Davey, S.G., Groop, L.C., Hattersley, A.T., McCarthy, M.I., Hirschhorn, J.N., & Frayling, T.M. 2007. A common variant of HMGA2 is associated with adult and childhood height in the general population. *Nat.Genet.*, 39, (10) 1245-1250

Wellcome Trust Case Control Consortium 2007a. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*, 447, (7145) 661-678

Wellcome Trust Case Control Consortium 2007b. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*, 447, (7145) 661-678

Wellcome Trust Case Control Consortium: Wellcome Trust Case Control Consortium website. Website: http://www.wtccc.org.uk/. Last date accessed: 12 Dec. 2011

Wexler, L., Brundage, B., Crouse, J., Detrano, R., Fuster, V., Maddahi, J., Rumberger, J., Stanford, W., White, R., & Taubert, K. 1996. Coronary artery calcification: pathophysiology, epidemiology, imaging methods, and clinical implications. A statement for health professionals from the American Heart Association. Writing Group. *Circulation*, 94, (5) 1175-1192

Whitlock, G., Lewington, S., Sherliker, P., Clarke, R., Emberson, J., Halsey, J., Qizilbash, N., Collins, R., & Peto, R. 2009. Body-mass index and cause-specific mortality in 900 000 adults: collaborative analyses of 57 prospective studies. *Lancet*, 373, (9669) 1083-1096

Wieberdink, R.G., van Schie, M.C., Koudstaal, P.J., Hofman, A., Witteman, J.C., de Maat, M.P., Leebeek, F.W., & Breteler, M.M. 2010. High von Willebrand factor levels increase the risk of stroke: the Rotterdam study. *Stroke*, 41, (10) 2151-2156

Wilkinson, P., Pattenden, S., Armstrong, B., Fletcher, A., Kovats, R.S., Mangtani, P., & McMichael, A.J. 2004. Vulnerability to winter mortality in elderly people in Britain: population based study. *BMJ*, 329, (7467) 647

Willer, C.J., Li, Y., & Abecasis, G.R. 2010. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics.*, 26, (17) 2190-2191

Wjst, M., Altmuller, J., Braig, C., Bahnweg, M., & Andre, E. 2007. A genome-wide linkage scan for 25-OH-D(3) and 1,25-(OH)2-D3 serum levels in asthma families. *J.Steroid Biochem.Mol.Biol.*, 103, (3-5) 799-802

Wjst, M., Altmuller, J., Faus-Kessler, T., Braig, C., Bahnweg, M., & Andre, E. 2006. Asthma families show transmission disequilibrium of gene variants in the vitamin D metabolism and signalling pathway. *Respir.Res.*, 7, 60

Wolfram Research, I. Mathematica Edition: Version 6.0. 2007. Champaign, Illinois, Wolfram Research, Inc.
Ref Type: Computer Program

Woodhouse, P.R., Khaw, K.T., Plummer, M., Foley, A., & Meade, T.W. 1994. Seasonal variations of plasma fibrinogen and factor VII activity in the elderly: winter infections and death from cardiovascular disease. *Lancet*, 343, (8895) 435-439

Woodward, M., Rumley, A., Tunstall-Pedoe, H., & Lowe, G.D. 2003. Does sticky blood predict a sticky end? Associations of blood viscosity, haematocrit and fibrinogen with mortality in the West of Scotland. *Br.J.Haematol.*, 122, (4) 645-650

Woodward, M., Rumley, A., Welsh, P., MacMahon, S., & Lowe, G. 2007. A comparison of the associations between seven hemostatic or inflammatory variables and coronary heart disease. *J.Thromb.Haemost.*, 5, (9) 1795-1800

Wooldridge, J. M. 2009, "Instrumental Variables Estimation and Two Stage Least Squares," *In Introductory Econometrics: A modern approach*, Fourth ed. South-Western CENGAGE Learning, pp. 506-545.

Wu, C.C., Chang, J.H., Chen, C.C., Su, S.B., Yang, L.K., Ma, W.Y., Zheng, C.M., Diang, L.K., & Lu, K.C. 2011. Calcitriol treatment attenuates inflammation and oxidative stress in hemodialysis patients with secondary hyperparathyroidism. *Tohoku J.Exp.Med.*, 223, (3) 153-159

Wu-Wong, J.R., Nakane, M., & Ma, J. 2006. Effects of vitamin D analogs on the expression of plasminogen activator inhibitor-1 in human vascular cells. *Thromb.Res.*, 118, (6) 709-714

Wu-Wong, J.R., Nakane, M., & Ma, J. 2007a. Vitamin D analogs modulate the expression of plasminogen activator inhibitor-1, thrombospondin-1 and thrombomodulin in human aortic smooth muscle cells. *J.Vasc.Res.*, 44, (1) 11-18

Wu-Wong, J.R., Nakane, M., Ma, J., Ruan, X., & Kroeger, P.E. 2007b. VDR-mediated gene expression patterns in resting human coronary artery smooth muscle cells. *J.Cell Biochem.*, 100, (6) 1395-1405

Xiang, W., Kong, J., Chen, S., Cao, L.P., Qiao, G., Zheng, W., Liu, W., Li, X., Gardner, D.G., & Li, Y.C. 2005. Cardiac hypertrophy in vitamin D receptor knockout mice: role of the systemic and cardiac renin-angiotensin systems. *Am.J.Physiol Endocrinol.Metab*, 288, (1) E125-E132

Yamamoto, N. & Homma, S. 1991. Vitamin D3 binding protein (group-specific component) is a precursor for the macrophage-activating signal factor from lysophosphatidylcholine-treated lymphocytes. *Proc.Natl.Acad.Sci.U.S.A*, 88, (19) 8539-8543

Yanoff, L.B., Parikh, S.J., Spitalnik, A., Denkinger, B., Sebring, N.G., Slaughter, P., McHugh, T., Remaley, A.T., & Yanovski, J.A. 2006. The prevalence of hypovitaminosis D and secondary hyperparathyroidism in obese Black Americans. *Clin.Endocrinol.(Oxf)*, 64, (5) 523-529

Yin, M., Wei, S., & Wei, Q. 2009. Vitamin D Receptor Genetic Polymorphisms and Prostate Cancer Risk: A Meta-analysis of 36 Published Studies. *Int J.Clin.Exp.Med.*, 2, (2) 159-175

Yuan, Y. & Mackinnon, D.P. 2009. Bayesian Mediation Analysis. *Psychological Methods*, 14, (4) 301-322

Zeggini, E. & Ioannidis, J.P. 2009. Meta-analysis in genome-wide association studies. *Pharmacogenomics.*, 10, (2) 191-201

Zella, L.A., Shevde, N.K., Hollis, B.W., Cooke, N.E., & Pike, J.W. 2008. Vitamin D-binding protein influences total circulating levels of 1,25-dihydroxyvitamin D3 but does not directly modulate the bioactive levels of the hormone in vivo. *Endocrinology*, 149, (7) 3656-3667

Zhao, G., Ford, E.S., Li, C., Kris-Etherton, P.M., Etherton, T.D., & Balluz, L.S. 2010. Independent associations of serum concentrations of 25-hydroxyvitamin D and parathyroid hormone with blood pressure among US adults. *J.Hypertens.*, 28, (9) 1821-1828

Zipes, D.P. 1999. Warning: the short days of winter may be hazardous to your health. *Circulation*, 100, (15) 1590-1592

Zittermann, A. & Gummert, J.F. 2010. Sun, vitamin D, and cardiovascular disease. *J.Photochem.Photobiol.B*, 101, (2) 124-129

Zittermann, A., Schleithoff, S.S., & Koerfer, R. 2005. Putting cardiovascular disease and vitamin D insufficiency into perspective. *Br.J.Nutr.*, 94, (4) 483-492

Appendix 1: Publications

	Reprint page	Thesis chapter
Paper I Hyppönen, E., **Berry, D**., Cortina-Borja, M. & Power, C. 2010. 25-Hydroxyvitamin D and pre-clinical alterations in inflammation and hemostatic markers. *PLoS One*, 5, (5) e10801	304	3
Paper II Wang, T.J.*, Zhang, F.*, Richards, J.B.*, Kestenbaum, B.*, van Meurs, J.B.*, **Berry, D**.*, Kiel, D.P., Streeten, E.A., Ohlsson, C., Koller, D.L., Peltonen, L., Cooper, J.D., O'Reilly, P.F., Houston, D.K., Glazer, N.L., Vandenput, L., Peacock, M., Shi, J., Rivadeneira, F., McCarthy, M.I., Anneli, P., de, B., I, Mangino, M., Kato, B., Smyth, D.J., Booth, S.L., Jacques, P.F., Burke, G.L., Goodarzi, M., Cheung, C.L., Wolf, M., Rice, K., Goltzman, D., Hidiroglou, N., Ladouceur, M., Wareham, N.J., Hocking, L.J., Hart, D., Arden, N.K., Cooper, C., Malik, S., Fraser, W.D., Hartikainen, A.L., Zhai, G., Macdonald, H.M., Forouhi, N.G., Loos, R.J., Reid, D.M., Hakim, A., Dennison, E., Liu, Y., Power, C., Stevens, H.E., Jaana, L., Vasan, R.S., Soranzo, N., Bojunga, J., Psaty, B.M., Lorentzon, M., Foroud, T., Harris, T.B., Hofman, A., Jansson, J.O., Cauley, J.A., Uitterlinden, A.G., Gibson, Q., Jarvelin, M.R., Karasik, D., Siscovick, D.S., Econs, M.J., Kritchevsky, S.B., Florez, J.C., Todd, J.A., Dupuis, J.*, Hyppönen, E.*, & Spector, T.D.* 2010. Common genetic determinants of vitamin D insufficiency: a genome-wide association study. *Lancet*, 376, (9736) 180-188 * Authors contributed equally	312	4

Paper	Page	
Paper III **Berry, D** & Hyppönen, E. 2011. Determinants of vitamin D status: focus on genetic variations. *Curr.Opin.Nephrol.Hypertens.*, 20, (4), 331-336	321	5
Paper IV **Berry, D.J***., Vimaleswaran, K.S.*, Whittaker, J.C., Hingorani, A.D., & Hypponen, E. 2012. Evaluation of genetic markers as instruments for mendelian randomization studies on vitamin d. *PLoS.ONE.*, 7, (5) e37465 * Authors contributed equally	327	5

OPEN ACCESS Freely available online PLoS one

25-Hydroxyvitamin D and Pre-Clinical Alterations in Inflammatory and Hemostatic Markers: A Cross Sectional Analysis in the 1958 British Birth Cohort

Elina Hyppönen*, Diane Berry, Mario Cortina-Borja, Chris Power

Medical Research Council Centre for Epidemiology of Child Health and Centre for Paediatric Epidemiology and Biostatistics, University College London Institute of Child Health, London, United Kingdom

Abstract

Background: Vitamin D deficiency has been suggested as a cardiovascular risk factor, but little is known about underlying mechanisms or associations with inflammatory or hemostatic markers. Our aim was to investigate the association between 25-hydroxyvitamin D [25(OH)D, a measure for vitamin D status] concentrations with pre-clinical variations in markers of inflammation and hemostasis.

Methodology/Principal Findings: Serum concentrations of 25(OH)D, C-reactive protein (CRP), fibrinogen, D-dimer, tissue plasminogen activator (tPA) antigen, and von Willebrand factor (vWF) were measured in a large population based study of British whites (aged 45y). Participants for the current investigation were restricted to individuals free of drug treated cardiovascular disease (n = 6538). Adjusted for sex and month, 25(OH)D was inversely associated with all outcomes ($p \leq 0.015$ for all), but associations with CRP, fibrinogen, and vWF were explained by adiposity. Association with tPA persisted after full adjustment (body mass index, waist circumference, physical activity, TV watching, smoking, alcohol consumption, social class, sex, and month), and average concentrations were 18.44% (95% CI 8.13, 28.75) lower for 25(OH)D ≥75 nmol/l compared to <25 nmol/l. D-dimer concentrations were lower for participants with 25(OH)D 50–90nmol/l compared to others (quadratic term $p = 0.01$). We also examined seasonal variation in hemostatic and inflammatory markers, and evaluated 25(OH)D contribution to the observed patterns using mediation models. TPA concentrations varied by season ($p = 0.02$), and much of this pattern was related to fluctuations in 25(OH)D concentrations ($p \leq 0.001$). Some evidence of a seasonal variation was observed also for fibrinogen, D-dimer and vWF ($p < 0.05$ for all), with 25(OH)D mediating some of the pattern for fibrinogen and D-dimer, but not vWF.

Conclusions: Current vitamin D status was associated with tPA concentrations, and to a lesser degree with fibrinogen and D-dimer, suggesting that vitamin D status/intake may be important for maintaining antithrombotic homeostasis.

Citation: Hyppönen E, Berry D, Cortina-Borja M, Power C (2010) 25-Hydroxyvitamin D and Pre-Clinical Alterations in Inflammatory and Hemostatic Markers: A Cross Sectional Analysis in the 1958 British Birth Cohort. PLoS ONE 5(5): e10801. doi:10.1371/journal.pone.0010801

Editor: Pieter H. Reitsma, Leiden University Medical Center, Netherlands

Received February 25, 2010; **Accepted** April 27, 2010; **Published** May 24, 2010

Copyright: © 2010 Hyppönen et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This study was supported by the British Heart Foundation (PG/09/023). Data collection at age 44–46 years and statistical analyses were funded by the UK Medical Research Council (MRC grants G0601653, G0000934), and 25(OH)D assays by the BUPA Foundation. E.H. is a Department of Health (UK) Public Health Career Scientist. This work was undertaken at the Centre for Paediatric Epidemiology and Biostatistics which benefits from funding support from the MRC in its capacity as the MRC Centre of Epidemiology for Child Health. Research at the University College London Institute of Child Health and Great Ormond Street Hospital for Children NHS Trust benefits from R&D funding received from the NHS Executive. The funders had no role in study design, data collection and analyses, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: e.hypponen@ich.ucl.ac.uk

Introduction

Vitamin D deficiency has been suggested to contribute to the high and rising worldwide prevalence of cardiovascular disease (CVD) [1]. Vitamin D is a hormone precursor, which before exerting its metabolic effects undergoes two successive hydroxylations. The first hydroxylation converts vitamin D to 25-hydroxyvitamin D [25(OH)D, which provides an indicator for nutritional vitamin D status] [2] and the second to the main active hormonal form, 1,25-dihydroxyvitamin D [1,25(OH)$_2$D]. Hormonal vitamin D activity is found throughout human circulatory tissue and 1,25(OH)$_2$D production has been demonstrated in endothelial cells of blood vessels [3] . Vitamin D receptors (VDR, mediating the genomic hormonal actions) are expressed in endothelial cells, cardiomyocytes and vascular smooth muscle cells, including those in the coronary arteries [4,5]. VDR knockout mice show signs of enhanced thrombogenicity [6].

The strongest evidence for a relation between vitamin D metabolism and CVD risk has been obtained from clinical studies reporting a marked reduction in mortality following administration of vitamin D analogues to patients with end-stage renal disease [7,8,9]. Evidence for an association between vitamin D status and subsequent risk of CVD was found in recent prospective studies on myocardial infarction [10] and cardiovascular mortality [11] ; both of these studies reported a two-fold increase in the risk of CVD for vitamin D insufficient participants compared to others. Concentration of the active hormone has been related to the degree of arterial calcification in individuals at increased risk of

myocardial infarction [12] and an inverse association has been reported between serum 25(OH)D concentration with carotid artery intima-media thickness [13], myocardial infarction [14], metabolic syndrome [15,16], and CVD [17,18].

To date there is relatively little evidence from population-based studies on the associations of 25(OH)D with indicators of inflammation or hemostasis, and to what extent possible co-variation is affected by adiposity. Obesity is a key determinant for the circulating 25(OH)D concentrations [19] and also an important cardiovascular risk factor. Consequently, it is difficult to separate the effects of vitamin D status and adiposity when evaluating their influences on cardiovascular risk. In order to obtain further evidence for a possible independent contribution of current vitamin D status on pre-clinical alterations in markers of inflammation and hemostasis, we evaluated seasonal patterns in inflammatory and hemostatic markers and the strength of the effect mediation by 25(OH)D. This approach is likely to be informative, as due to the strong influence of sun induced skin synthesis, 25(OH)D concentrations vary greatly by season [19], while little variation would be expected for adiposity. Our aim was to investigate the association between 25(OH)D, adiposity (body mass index, waist circumference) and pre-clinical variations in the available risk markers (namely CRP, fibrinogen, D-dimer, tPA, and von Willebrand factor). In these analyses, we used information from the nationwide 1958 British birth cohort (1958BC) on over 6500 middle aged participants. We hypothesized that if vitamin D intake affects the markers under investigation then further evidence for an association should be obtained through analysing the contribution of 25(OH)D to the seasonal variation in markers of inflammation and hemostasis.

Results

The geometric mean of 25(OH)D concentration was 52.77 nmol/l (95% CI 52.18, 53.36). Table 1 shows the distribution of 25(OH)D concentrations by social and lifestyle characteristics. For both BMI and waist circumference (the available adiposity indicators) the association with 25(OH)D was non-linear (LRT curvature $p \leq 0.0001$ and $p = 0.04$, respectively), with the highest 25-hydoxyvitamin D concentrations observed for individuals with normal weight (Figure 1). There was a steep decline in the average 25(OH)D concentration by increasing adiposity and a smaller reduction for the very lean.

Adjusted for sex and month of measurement only, 25(OH)D was associated with all inflammatory and hemostatic outcomes ($p \leq 0.01$ for CRP, fibrinogen, D-dimer and tPA, $p = 0.015$ for vWF, Figure 2). Associations between 25(OH)D with CRP, fibrinogen, and vWF were strongly attenuated after adjustment for lifestyle and social indicators, and no evidence for an independent inverse association remained after further adjustment for adiposity. For fibrinogen, after full adjustment for adiposity, lifestyle and social indicators, there was a curved association with some suggestion for increased levels at 25(OH)D concentrations of ≥125 nmol/l (LRT quadratic term $p = 0.06$ Figure 2B). 25(OH)D had a curved association with D-dimer, and after full adjustment participants with 25(OH)D between 50–90 nmol/l tended to have lower levels (LRT quadratic term $p = 0.01$ Figure 2C). The association between 25(OH)D and tPA was not strongly affected by adjustment for lifestyle factors, and a significant inverse trend was apparent after further adjustment for adiposity although the effect size was halved (Figure 2D). After full adjustment, participants with 25(OH)D ≥75 nmol/l had on average 18.44% (95% CI 8.13, 28.75) lower tPA concentrations compared to those with <25 nmol/l. There was no evidence for effect modification by obesity on the association between 25(OH)D and the inflammatory or hemostatic outcomes ($p > 0.13$ for all comparisons).

Given the strong influence of season on 25(OH)D concentrations [19], we evaluated seasonal variation in hemostatic and inflammatory markers, and tested mediation effects of 25(OH)D in the observed patterns. Fibrinogen, tPA, D-dimer, and vWF but not CRP had significant seasonal patterns ($p = 0.03$, $p = 0.02$, $p = 0.02$, $p = 0.01$ and $p = 0.8$, respectively, Figure 3). The strongest effect mediation by 25(OH)D was seen in the pattern of tPA ($p < 0.001$), with 25(OH)D contributing to a lesser extent to seasonal variation in D-dimer and fibrinogen. The seasonal variation seen in vWF was not affected by 25(OH)D ($p = 0.99$).

Discussion

We observed a strong cross-sectional association between circulating 25(OH)D and tPA concentrations in participants free of clinical CVD, and a seasonal pattern for tPA that was largely mediated by 25(OH)D in this population. These findings, together with the weaker evidence observed for a relation of 25(OH)D with D-dimer and fibrinogen, suggest a role for current vitamin D status in determining thrombolytic profile before progression to CVD.

A specific methodological challenge for these cross-sectional analyses arose from the strong association of adiposity both with 25(OH)D concentrations and the inflammatory/hemostatic markers under study. In addition to the conventional approach of evaluating the direct association between 25(OH)D and the outcomes adjusting for potential confounders (most importantly, body mass index and waist circumference), we evaluated seasonal variation in the outcomes and the mediating influence of 25(OH)D on the observed patterns. These analyses supported a relation of 25(OH)D with tPA, and interestingly, also to lesser extent with D-dimer and fibrinogen. The seasonal pattern seen in vWF was not affected by 25(OH)D, nor did we observe evidence for a direct cross-sectional association, hence, this confirms the lack of evidence for any association between vitamin D status and circulating vWF concentrations in our study.

Comparison with other studies

Risk of myocardial infarction and other thrombotic complications is typically higher during the winter months than during the summer [20,21], and in line with our study, fibrinogen (but not CRP) has been reported to vary by season [22]. However, there is little information on the direct associations between 25(OH)D and circulating markers of hemostasis in apparently healthy adults. In line with our findings, an earlier smaller study found the association between 25(OH)D and CRP to be explained by adiposity, while an independent relation persisted between 25(OH)D and tPA [23]. The lack of evidence for an association between current vitamin D status and CRP agrees with recent observations in other populations [24], including intervention studies where vitamin D supplementation at varying dosages (700–3332 IU per day) have failed to achieve changes in CRP concentrations [25,26,27]. There is, however, one earlier intervention which showed a significant (22%) fall in CRP following vitamin D supplementation (50,000IU every three months over one year) in vitamin D deficient British Bangladeshi adults [28]. It is possible that associations between 25(OH)D and CRP might not be detected in general population studies if effects are confined to extreme vitamin D deficient groups.

Explanations

Increased concentrations of tPA and D-dimer are thought to serve as markers for aggravated fibrinolytic activity reflecting

Table 1. Distribution of 25-hydroxyvitamin D concentration by background, lifestyle and social characteristics in the 1958 British birth cohort.

	Number	25-hydroxyvitamin D, nmol/l		
	Number (%)	Geometric Mean* (95%CI)	<25 nmol/l %* (n)	>125 nmol/l %* (n)
Sex				
Men	3270 (50.0)	53.6 (52.8, 54.5)	6.2 (203)	1.4 (46)
Women	3268 (50.0)	51.9 (51.1, 52.8)	8.4 (273)	1.4 (45)
p-value		0.003	0.0007	0.9
Body Mass index				
<25	2361 (36.1)	55.1 (54.0, 56.2)	7.5 (178)	2.2 (52)
25–30	2737 (41.9)	54.2 (53.3, 55.1)	5.8 (159)	1.2 (32)
>30	1440 (22.0)	46.8 (45.7, 47.9)	9.7 (139)	0.5 (7)
p-value		≤0.0001	0.04	≤0.0001
Waist circumference**				
Quartile 1	1642 (25.1)	57.4 (56.0, 58.7)	6.8 (111)	2.8 (46)
Quartile 2	1639 (25.1)	55.2 (54.0, 56.4)	6.2 (102)	1.8 (29)
Quartile 3	1627 (24.9)	52.6 (51.4, 53.7)	6.1 (100)	0.7 (11)
Quartile 4	1618 (24.8)	46.6 (45.6, 47.6)	9.9 (160)	0.2 (4)
Unknown	0.2 (12)	44.8 (29.4, 68.3)	25.0 (3)	8.3 (1)
p-value		≤0.0001	0.003	≤0.0001
Vigorous activity				
No	3206 (49.0)	49.6 (48.8, 50.4)	9.4 (300)	1.2 (37)
Yes	3236 (49.5)	56.3 (55.4, 57.2)	5.0 (163)	1.7 (54)
Unknown	96 (1.5)	48.0 (43.2, 53.2)	13.5 (13)	0.0 (0)
p-value		≤0.0001	≤0.0001	0.07
TV watching/use of PC				
<1 hours/day	745 (11.4)	56.3 (54.5, 58.2)	5.8 (43)	2.3 (17)
1–2 hours/day	3455 (52.8)	54.9 (54.1, 55.7)	5.5 (191)	1.6 (54)
≥3 hours/day	2056 (31.4)	48.9 (47.9, 49.9)	9.9 (204)	0.9 (19)
Unknown	282 (4.3)	47.7 (44.9, 50.5)	13.5 (38)	0.4 (1)
p-value		≤0.0001	≤0.0001	0.01
Smoking				
None	3039 (46.5)	54.2 (53.3, 55.1)	6.1 (185)	1.3 (41)
Ex-smoker	1795 (27.5)	54.8 (53.7, 55.9)	5.1 (91)	1.3 (23)
1–19 per day	762 (11.7)	50.3 (48.5, 52.1)	10.4 (79)	1.7 (13)
≥20 per day	720 (11.0)	45.5 (43.8, 47.2)	14.4 (104)	1.9 (14)
Unknown	222 (3.4)	51.8 (48.8, 55.1)	7.7 (17)	0.0 (0)
p-value		≤0.0001	≤0.0001	0.2
Alcohol consumption				
Non-drinker	378 (5.8)	46.1 (44.0, 48.4)	11.4 (43)	0.0 (0)
Light <7 drinks/wk	3155 (48.3)	52.1 (51.3, 52.9)	7.5 (237)	1.0 (31)
Moderate 7–13 drinks/wk	1651 (25.3)	55.6 (54.4, 56.9)	5.6 (92)	1.8 (30)
Heavy 14–21 drinks/wk	746 (11.4)	55.9 (54.1, 57.8)	4.8 (36)	2.5 (19)
Very heavy >21 drinks/wk	590 (9.0)	49.7 (47.7, 51.8)	10.7 (63)	1.7 (10)
Unknown	18 (0.3)	41.5 (31.5, 54.6)	27.8 (5)	5.6 (1)
p-value		0.01	0.9	0.0003
Adult social class (2000)†				
I & II	2675 (40.9)	53.3 (52.4, 54.2)	6.7 (180)	1.4 (38)
III non-manual	1363 (20.9)	52.2 (50.9, 53.5)	7.9 (107)	1.1 (15)
III manual	1220 (18.7)	54.5 (53.1, 56.0)	6.1 (74)	2.0 (25)

Table 1. Cont.

		25-hydroxyvitamin D, nmol/l		
	Number	Geometric Mean*	<25 nmol/l	>125 nmol/l
	(%)	(95%CI)	%* (n)	%* (n)
IV & IV	1013 (15.5)	51.2 (49.8, 52.7)	8.0 (81)	1.0 (10)
Other	267 (4.1)	48.8 (45.9, 51.8)	12.7 (34)	1.1 (3)
p-value		0.007	0.02	0.8

*Values are *n* (%) or geometric mean. *p*-values from test for trend in linear or logistic regression adjusted for season and sex. Unknown values excluded.
**Waist circumference quartiles: for men; 65.4–90.6, 90.7–96.7, 96.8–103.5, 103.6–151.2 cm; for women; 56.2–75.8, 75.9–82.6, 82.7–91.6, 91.7–138.3 cm.
†Classes I&II are managerial/professional, IV/V unskilled manual. "Other" includes cohort members who are institutionalised, retired, unemployed and other unclassifiable.
doi:10.1371/journal.pone.0010801.t001

increased future burden of CVD [29]. Hence, the inverse associations of 25(OH)D with tPA and D-dimer observed in our study support the role of vitamin D metabolism in maintaining antithrombotic homeostasis. The direct influences of hormonal vitamin D axis on hemostasis are not well established, although recent gene-expression studies suggest that vitamin D analogues may suppress thrombogenicity and enhance fibrinolysis thereby reducing intimal plaque formation [30,31]. Vitamin D analogues have been observed to suppress PAI-1 expression in human coronary artery smooth muscle cells [32]. Up regulation of PAI-1 has been associated with increased risk of CVD, and it has been suggested that suppression of PAI-1 expression may contribute to the observations on improved survival among patients with chronic kidney disease who are taking vitamin D analogues [7,32].

Hypovitaminosis D is believed to have wide-ranging influences on vascular physiology, which include both direct (e.g. influences on endothelial cells) and indirect pathways (endocrine, immunomodulatory) [33]. Vitamin D toxicity has been associated with adverse effects on vascular calcification, but available data indicates that calcification is increased also in hypovitaminosis D [34,35]. Associations with increased blood pressure are believed to be mediated through decreased renin production [36] and it has been suggested that (independently of blood pressure), this could also affect vascular stiffness [33]. Vitamin D might also exert anti-proliferative effects on vascular smooth muscle cells, thereby affecting myocardial cell hypertrophy and proliferation.

Discussion regarding optimal status for 25(OH)D concentration is ongoing, and there is some debate about whether a threshold exists [37,38,39,40]. The curved association between 25(OH)D and D-dimer, together with the suggestive elevations in fibrinogen and CRP at the extreme of high concentrations observed in this study, could support a threshold effect with the optimal concentration being between 60 and 120nmol/l. These results are in line with an evaluation using data on multiple health outcomes (including bone mineral density, lower extremity function, risk of falls, fractures and colorectal cancer) [40], which corresponds to an earlier consensus statement on osteoporosis [37] suggesting 75nmol/l as the lower reference value. Earlier studies have provided tentative evidence for a trend towards lower bone mineral density and prolonged sit-to-stand time at a higher extreme of 25(OH)D concentrations (>140 nmol/l) [40].

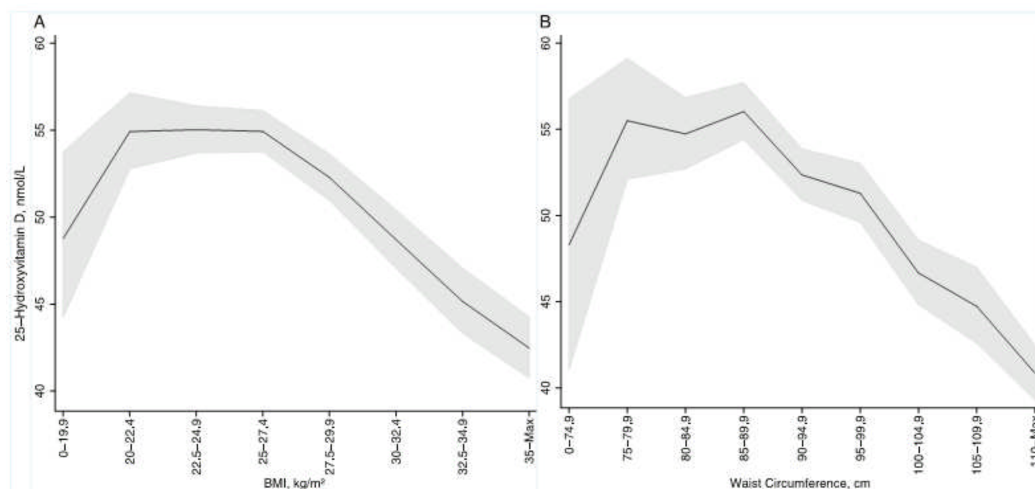


Figure 1. Variation in the average 25(OH)D concentration by body mass index (A) and waist circumference (B). Values are geometric means (95% confidence intervals) standardized by sex.
doi:10.1371/journal.pone.0010801.g001

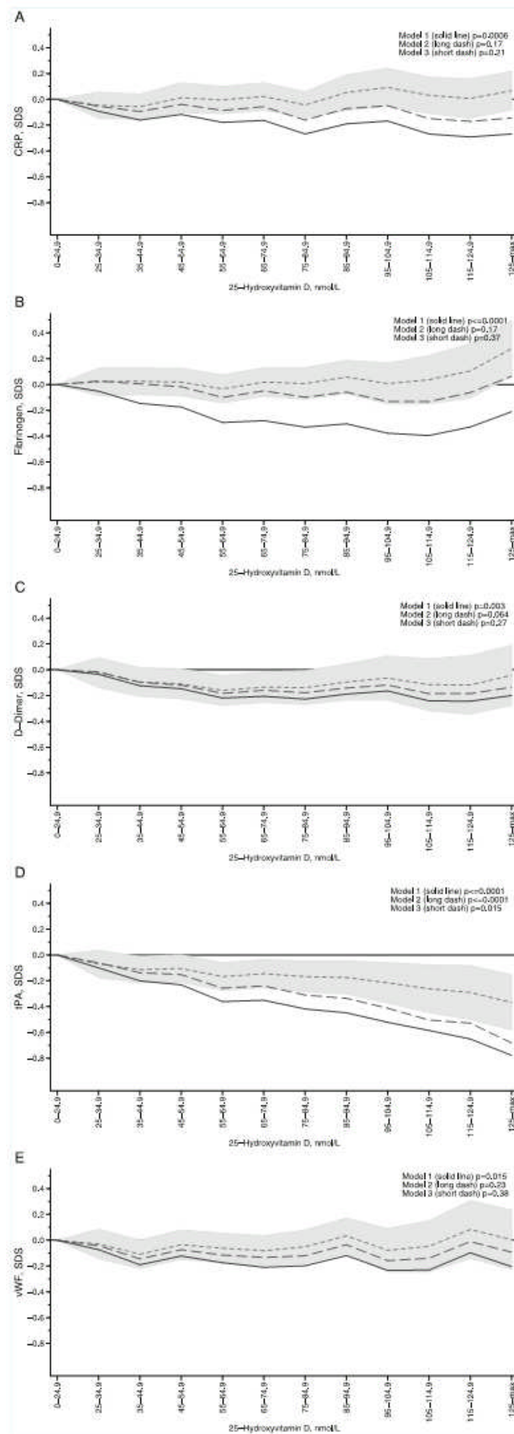


Figure 2. Variation in C-reactive protein (A), fibrinogen (B), D-dimer (C), tissue plasminogen activator (D), and von Willebrand factor (E) by 25(OH)D concentration. Model 1 (solid line): adjusted for month of measurement and sex. Model 2 (dashed, short): adjusted for lifestyle and social indicators (physical activity, time spent watching TV/using PC, smoking, alcohol consumption and birth and adult social class), month of measurement and sex. Model 3 (dashed, long): adjusted for adiposity (BMI and waist circumference), lifestyle/social indicators, month of measurement, and sex. Values are coefficients from linear regression (reference <25nmol/l), 95% confidence intervals presented for Model 3 by the shaded area.
doi:10.1371/journal.pone.0010801.g002

Methodological considerations

The main strength of this study lies in the large sample of participants, which provided adequate power for detailed investigation of the associations between these inter-related health indicators. Moreover, as data collection covered the full seasonal range, we were able to obtain further support for the key findings from the independent evaluation of seasonal patterns in inflammatory and hemostatic markers. Given the exceptional information available from the 1958BC, we were able to adjust for multiple factors in our analyses thereby controlling for confounding introduced by demographic, lifestyle or social variations. Final models evaluating the independent effect of 25(OH)D on inflammatory and hemostatic outcomes were adjusted for quadratic terms in both BMI and waist circumference in order to control for adiposity as fully as possible. The full attenuation of the association between 25(OH)D with CRP and fibrinogen after adjustment for the available indicators suggests that these measures were sufficient for this purpose.

Comparison between the effect of adjustment for 25(OH)D concentrations in the observed seasonal patterns in the inflammatory/hemostatic factors, and the direct associations between 25(OH)D and these outcomes, demonstrates the limitations of cross-sectional analysis of data and the problem of possible over/under adjustment. Given the strong influence of obesity on 25(OH)D concentrations, the latter would be expected to be associated with any factor that is strongly related to obesity (given a tolerable degree of measurement error and sufficient sample size). This argues for the need to adjust for obesity fully to reduce the likelihood of a false positive association due to confounding. However, it could also be argued that adjustment for adiposity may lead to an underestimation of associations between 25(OH)D and inflammatory/hemostatic markers given that adiposity is a key determinant for 25(OH)D [19]. Possible over-adjustment could explain why we observed some evidence for effect mediation by 25(OH)D on our seasonal modeling of fibrinogen, while the inverse relation seen in the unadjusted cross-sectional analyses between these two factors was fully attenuated by the adjustment for indicators of adiposity and lifestyle/social class.

Some further limitations need to be considered in relation to these findings. Given the observational design, we cannot prove causality or fully discount residual confounding by unmeasured variations. Residual confounding may also affect our seasonal mediation models; however, relevant confounders are likely to differ given that potential factors presumed important for the direct associations (such as adiposity) would not necessarily have seasonal patterns. Although 25(OH)D is the best indicator for vitamin D status in humans [2] information on serum parathyroid levels or $1,25(OH)_2D$ is not available and this precludes more detailed investigation of vitamin D metabolism. Furthermore, information was available only for the most commonly used CVD risk markers, while other, perhaps more relevant, indicators were not measured. Finally, although the 1958BC has been reported to

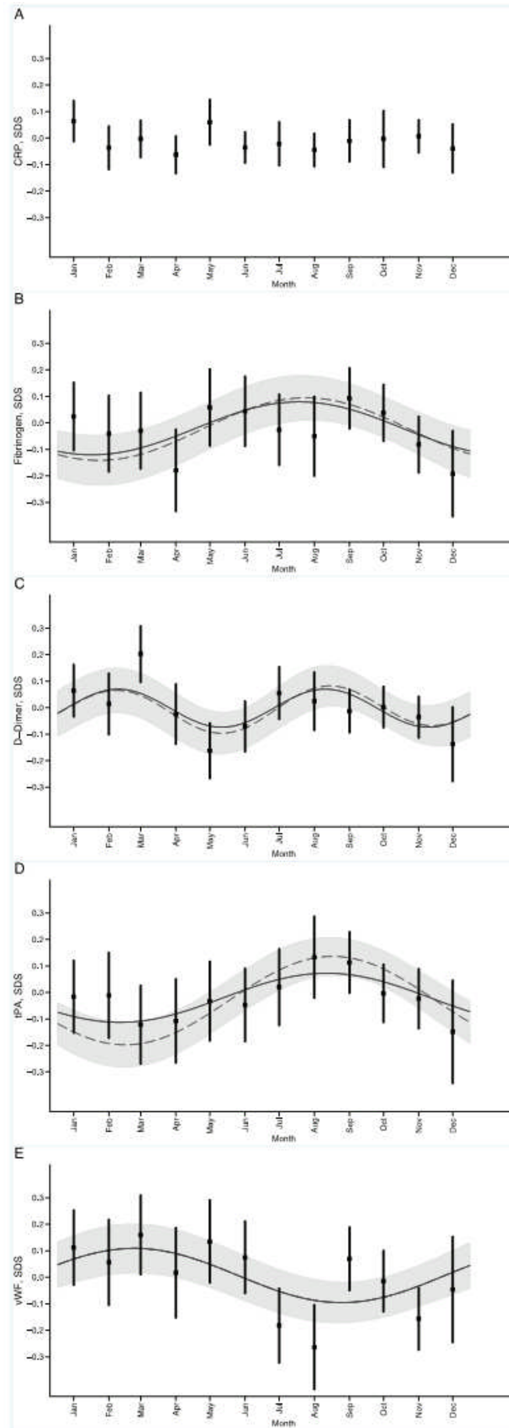


Figure 3. Seasonal variation in C-reactive protein (A), fibrinogen (B), D-dimer (C), tissue plasminogen activator (D), and von Willebrand factor (E). Values are from the partial regression of the harmonic components; Model 1 (solid line) adjusted for respiratory infections, alcohol consumption, PC/TV time, physical activity and social class at birth and adulthood, and Model 2 (dashed line, shown with 95% confidence intervals) in addition to above adjusted for 25-hydroxyvitamin D. Tick marks denote average concentrations (SDS, predicted from random effects models) with 95% confidence intervals shown by error bars. Predicted means for CRP from linear models, no seasonal pattern observed ($p>0.8$). *p-values from the product of coefficient mediation test used to assess the 25(OH)D mediation effect on the seasonal patterns in the outcomes.
doi:10.1371/journal.pone.0010801.g003

remain generally representative of the current UK population, there is an underestimation of some minority groups [41]. These analyses were restricted to individuals of white Caucasian origin, which will reduce population stratification, however, generalization of these findings to other ethnic groups should be done with caution.

Conclusions

Current vitamin D status was associated with circulating concentrations of tPA and D-dimer, which may suggest a role for vitamin D in maintaining antithrombotic homeostasis. Further studies, including randomised controlled trials, are needed to demonstrate the role of vitamin D metabolism in cardiovascular health, and whether vitamin D supplementation or improved vitamin D status could have beneficial effects.

Methods

Written consent for the use of information in medical studies was obtained from the cohort members. The 45y biomedical survey of the 1958BC was approved by the South-East Multi-Centre Research Ethics Committee (ref: 01/1/44).

Participants in this study are from the 1958 British birth cohort, which included all births in England, Scotland, and Wales during one week in March 1958 ($n=17,416$) [41,42]. Between September 2002 and April 2004 a target population of 11,971 individuals currently living in Britain were contacted aged 44y (31.1%) to 46y (0.4%): 78% ($n=9377$) participated in the biomedical survey and 7591 (80%) also provided blood samples from which 25(OH)D was measured [19,42]. This sample is representative of the surviving cohort; however, as we have reported previously, there is some under-representation of specific minority groups [41]. The 1958BC is largely a white European population (98%); for these analyses 154 individuals of other ethnicity groups were excluded. We further excluded one participant who was pregnant at the time of survey. As the main focus of the analyses was to evaluate the association of 25(OH)D concentrations on pre-clinical alterations in inflammatory and hemostatic markers, we excluded all participants ($n=532$) who used any type of medication used to treat cardiovascular problems (BNF code 2: Cardiovascular systems). We further excluded participants with missing data on BMI or inflammatory/hemostatic markers ($n=366$) leaving 6538 individuals for the main analyses.

Laboratory analyses

Venous blood samples were obtained without prior fasting and posted to collaborating laboratories. Fibrinogen was determined by the Clauss method and CRP assayed by nephelometry (Dade Behring) on citrated plasma samples after one thaw cycle. vWF antigen was measured by Decollates enzyme-linked immunosor-

bent assay (elisa) and tPA antigen by Biopool elisa. 25(OH)D was measured using automated application of the IDS OCTEIA elisa on the Dade-Behring BEP2000 analyser (sensitivity of 5.0 nmol/l, linearity ≤155 nmol/l, intra-assay variation CV 5.3–7.4% and inter-assay variation CV 7.7–8.5%) [41,43]. Heterogeneity of 25(OH)D concentrations measured by different assays is well-known, therefore levels were standardized according to the mean of the Vitamin D External Quality Assessment Scheme (DEQAS) [43].

Demographic, lifestyle and social factors have been described in detail previously [19,42,44,45,46]. In brief, weight, height and waist circumference were measured at 45y. Socio-economic position at birth (1958) and at age 42y was assessed using the Registrar General's occupational classification categorised as I &II (professional/managerial), III non-manual, III manual, IV & V (unskilled manual) [44]. Recreation Metabolic Equivalent of Task (MET) hours per week at 45y was derived from reported frequencies and usual durations for up to 37 activities, and published MET scores. We divided recreational activity into gender-specific quartiles. An additional category was created for implausibly high values (participants with weekly recreation hours of 3 standard deviations (SD) above the gender mean). Participants engaging in vigorous activity were those who recorded an activity with a MET score greater than six. Time spent watching a television or using a computer was reported at age 45y [19]. Information on smoking was based on smoking history recorded at ages 23y, 33y and 42y [46] and alcohol consumption on report from age 45y.

Statistical Analysis

To describe the distribution of 25(OH)D concentration we used dichotomous indicators for levels below 25 nmol/l and above 125 nmol/l and a categorised factor into 25 nmol/l divisions with minimum <25 nmol/l and maximum ≥125 nmol/l tails. The natural logarithmic transformation was used to calculate geometric means to adjust the skewed distribution.

Inflammatory and hemostatic markers were transformed to gender-specific standard deviation scores (SDS) to compare variation across models. The SDS values were used as response variables in linear mixed effects regression models. Initial analyses included validation and graphical examination of data, statistical evaluation of linear and quadratic terms for 25(OH)D, adiposity measures (BMI and waist circumference), and single and joint effects of these measures on the inflammatory/hemostatic outcomes. Continuous measures were used in testing for interactions between the adiposity measures and 25(OH)D on the outcomes. For D-dimer three outlying observations were identified from graphical examination and model diagnostics (leverage and/or influence >2SD), and excluded from further analysis. We fitted linear regression models in three stages, starting with simple associations between 25(OH)D (categorized into 10 nmol/l groups, minimum <25 nmol/l and maximum ≥125 nmol/l tails) and the SDS inflammatory/hemostatic outcomes (model 1), next adjusting in addition for demographic, lifestyle and social factors (model 2), and finally adjusting for adiposity in addition to lifestyle and social factors (model 3). Models (1–3) included covariates gender and month of measurement and models with fibrinogen SDS as the outcome included laboratory assay batch. We also created an additional 25(OH)D category variable of <25, 25–74.9 and ≥75 nmol/l to summarize the effect size and repeated analyses for model 3. Missing information on the lifestyle factors ($n=514$ with missing information on one or more factors) was imputed using the multiple imputation chained equations [47]. The models (1–3) were run on 10 imputed datasets and repeated for the sample restricted to participants with complete data on all confounders. The results were similar with both approaches; hence, results are only presented from the imputed models.

Seasonal variations were modeled using sine and cosine functions [48] with laboratory assay batch as a random effect on the intercept where appropriate (Likelihood Ratio Test (LRT) on between-batch variation $p<0.05$). The model equation was:

$$Y_{ijt} = (\beta_0 + \mu_{0j}) + \sum_{k=1}^{n_c} \sin\left(\frac{2\pi\, t\, p_k}{T}\right) \beta_{1k} + \sum_{k=1}^{n_c} \cos\left(\frac{2\pi\, t\, p_k}{T}\right) \beta_{2k} + \mathbf{X}_i \boldsymbol{\beta}_3 + \varepsilon_{ijt}$$

Where Y is the response variable, i denotes the individual within the assay batch j , T is the time period ($T=365$ days), n_c is the number of seasonal patterns with sequence ($n_c=\{1,2,3\}$) or combination of them, p_k is the period for the season ($p_k=\{1,2,4\}$) where $1 \le t \le T$, $\mathbf{X}_{i3}$ is a matrix of linear predictors for the vector of parameters $\boldsymbol{\beta}_3$ (inclusive of the mediator variable and seasonal confounders). The random effect in the intercept, μ_{0j}, is defined by the laboratory assay batch as the dates of the blood samples were associated with the assay batches and we did not wish to estimate the effect of batch on the response; we assume $\mu_{0j} \sim N\left(0, \sigma_\mu^2\right)$. Finally, the observational error terms ε_{ijt} were assumed to be normally distributed with mean 0 and variance σ_ε^2.

Without the hierarchical structure implied by the random effects terms we may have under-estimated the standard error on the intercept. The n_c seasonal patterns investigated were yearly, half yearly, quarterly, and all their combinations of were included if appropriate. Models were adjusted for the potential seasonal confounders of respiratory infections, alcohol consumption, PC/TV time, physical activity and social class at birth and adulthood. These resulting models were further adjusted for 25(OH)D, its dominant yearly pattern and then re-tested for subsidiary cycles.

To quantify the seasonal effect of 25(OH)D on the outcomes we used the concept of mediation analysis [47,49] where season, as modeled with the sine/cosine transformation, was the independent variable and 25(OH)D acted as its mediator to the outcomes. The product of coefficients test [50] used in mediation analysis was extended to allow for the amplitude (the seasonal variation around the mean) as derived from the sine/cosine transformation. The final seasonal models predict the mean levels of the outcomes from the partial regression coefficients of the seasonal functions. All analyses were carried out using STATA, version 10.0 (StataCorp LP, College Station, TX).

Acknowledgments

We thank Dr Ian Gibb and Steve Turner (Royal Victoria Infirmary, Newcastle-upon-Tyne) for carrying out the 25(OH)D assays; and Professor Gordon Lowe and Dr Ann Rumley (Department of Medicine, Glasgow Royal Infirmary University NHS Trust, Glasgow) for processing measures of fibrinogen, C-reactive protein, tPA, vWF and D-dimer. Data were provided by the Centre for Longitudinal Studies, Institute of Education, University of London (original data producers).

Author Contributions

Conceived and designed the experiments: EH MCB CP. Performed the experiments: DJB. Analyzed the data: DJB. Contributed reagents/materials/analysis tools: EH DJB MCB CP. Wrote the paper: EH DJB. Obtained funding: EH. Made figures: DJB.

References

1. Zittermann A, Schleithoff SS, Koerfer R (2005) Putting cardiovascular disease and vitamin D insufficiency into perspective. Br J Nutr 94[4]: 483–492.
2. Standing Comittee on the Scientific Evaluation of Dietary Reference (1997) Dietary reference intakes:calcium, phosphorus, magnesium, vitamin D and fluoride. National Academy Press.
3. Zehnder D, Bland R, Chana RS, Wheeler DC, Howie AJ, et al. (2002) Synthesis of 1,25-dihydroxyvitamin D(3) by human endothelial cells is regulated by inflammatory cytokines: a novel autocrine determinant of vascular cell adhesion. J Am Soc Nephrol 13[3]: 621–629.
4. Towler DA, Clemens TL (2005) Vitamin D and Cardiovascular Medicine. Feldman D, Pike JW, Glorieux FH, eds. Vitamin D. 2nd[56]. pp 899–910. Stanford, Elsevier Academic Press.
5. Wu-Wong JR, Nakane M, Ma J, Ruan X, Kroeger PE (2007) VDR-mediated gene expression patterns in resting human coronary artery smooth muscle cells. J Cell Biochem 100[6]: 1395–1405.
6. Aihara K, Azuma H, Akaike M, Ikeda Y, Yamashita M, et al. (2004) Disruption of nuclear vitamin D receptor gene causes enhanced thrombogenicity in mice. J Biol Chem 279[34]: 35798–35802.
7. Shoji T, Shinohara K, Kimoto E, Emoto M, Tahara H, et al. (2004) Lower risk for cardiovascular mortality in oral 1alpha-hydroxy vitamin D3 users in a haemodialysis population. Nephrol Dial Transplant 19[1]: 179–184.
8. Teng M, Wolf M, Lowrie E, Ofsthun N, Lazarus JM, et al. (2003) Survival of patients undergoing hemodialysis with paricalcitol or calcitriol therapy. New England Journal of Medicine 349[5]: 446–456.
9. Teng M, Wolf M, Ofsthun MN, Lazarus JM, Hernan MA, et al. (2005) Activated injectable vitamin D and hemodialysis survival: a historical cohort study. J Am Soc Nephrol 16[4]: 1115–1125.
10. Giovannucci E, Liu Y, Hollis BW, Rimm EB (2008) 25-hydroxyvitamin D and risk of myocardial infarction in men: a prospective study. Arch Intern Med 168[11]: 1174–1180.
11. Dobnig H, Pilz S, Scharnagl H, Renner W, Seelhorst U, et al. (2008) Independent association of low serum 25-hydroxyvitamin d and 1,25-dihydroxyvitamin d levels with all-cause and cardiovascular mortality. Arch Intern Med 168[12]: 1340–1349.
12. Watson KE, Abrolat ML, Malone LL, Hoeg JM, Doherty T, et al. (1997) Active serum vitamin D levels are inversely correlated with coronary calcification. Circulation 96[6]: 1755–1760.
13. Targher G, Bertolini L, Padovani R, Zenari L, Scala L, et al. (2006) Serum 25-hydroxyvitamin D3 concentrations and carotid artery intima-media thickness among type 2 diabetic patients. Clin Endocrinol. (Oxf) 65[5]: 593–597.
14. Scragg R, Jackson R, Holdaway IM, Lim T, Beaglehole R (1990) Myocardial infarction is inversely associated with plasma 25-hydroxyvitamin D3 levels: a community-based study. Int J Epidemiol 19[3]: 559–563.
15. Pittas AG, Lau J, Hu FB, Dawson-Hughes B (2007) The role of vitamin D and calcium in type 2 diabetes. A systematic review and meta-analysis. J Clin Endocrinol Metab 92[6]: 2017–2029.
16. Hyppönen E, Boucher BJ, Berry D, Power C (2008) Insulin-like growth factor I, 25-hydroxyvitamin D and metabolic syndrome mid-life: the British 1958 birth cohort study. Diabetes 57[2]: 298–305.
17. Cigolini M, Iagulli MP, Miconi V, Galiotto M, Lombardi S, et al. (2006) Serum 25-hydroxyvitamin D3 concentrations and prevalence of cardiovascular disease among type 2 diabetic patients. Diabetes Care 29[3]: 722–724.
18. Wang TJ, Pencina MJ, Booth SL, Jacques PF, Ingelsson E, et al. (2008) Vitamin D deficiency and risk of cardiovascular disease. Circulation 117[4]: 503–511.
19. Hyppönen E, Power C (2007) Hypovitaminosis D in British adults at age 45y: nationwide cohort study on dietary and lifestyle predictors. American Journal of Clinical Nutrition 85[3]: 860–868.
20. Manfredini R, Manfredini F, Boari B, Bergami E, Mari E, et al. (2009) Seasonal and weekly patterns of hospital admissions for nonfatal and fatal myocardial infarction. Am J Emerg Med 27[9]: 1097–1103.
21. Manfredini R, Boari B, Salmi R, Gallerani M (2004) Seasonal variation of venous thromboembolic disease. Am J Cardiol 94[2]: 276.
22. Kelly GS (2005) Seasonal variations of selected cardiovascular risk factors. Altern Med Rev 10[4]: 307–320.
23. Jorde R, Haug E, Figenschau Y, Hansen JB (2006) Serum Levels of Vitamin D and Haemostatic Factors in Healthy Subjects: The Tromso Study. Acta Haematol 117[2]: 91–97.
24. Shea MK, Booth SL, Massaro JM, Jacques PF, D'Agostino RB, Sr, et al. (2008) Vitamin K and vitamin D status: associations with inflammatory markers in the Framingham Offspring Study. Am J Epidemiol 167[3]: 313–320.
25. Pittas AG, Harris SS, Stark PC, Dawson-Hughes B (2007) The effects of calcium and vitamin D supplementation on blood glucose and markers of inflammation in nondiabetic adults. Diabetes Care 30[4]: 980–986.
26. Schleithoff SS, Zittermann A, Tenderich G, Berthold HK, Stehle P, et al. (2006) Vitamin D supplementation improves cytokine profiles in patients with congestive heart failure: a double-blind, randomized, placebo-controlled trial. American Journal of Clinical Nutrition 83[4]: 754–759.
27. Zittermann A, Frisch S, Berthold HK, Gotting C, Kuhn J, et al. (2009) Vitamin D supplementation enhances the beneficial effects of weight loss on cardiovascular disease risk markers. American Journal of Clinical Nutrition 89[5]: 1321–1327.
28. Timms PM, Mannan N, Hitman GA, Noonan K, Mills PG, et al. (2002) Circulating MMP9, vitamin D and variation in the TIMP-1 response with VDR genotype: mechanisms for inflammatory damage in chronic disorders? QJM 95[12]: 787–796.
29. Pradhan AD, LaCroix AZ, Langer RD, Trevisan M, Lewis CE, et al. (2004) Tissue plasminogen activator antigen and D-dimer as markers for atherothrombotic risk among healthy postmenopausal women. Circulation 110[3]: 292–300.
30. Wu-Wong JR, Nakane M, Ma J, Ruan X, Kroeger PE (2006) Effects of Vitamin D analogs on gene expression profiling in human coronary artery smooth muscle cells. Atherosclerosis 186[1]: 20–28.
31. Koyama T, Shibakura M, Ohsawa M, Kamiyama R, Hirosawa S (1998) Anticoagulant effects of 1alpha, 25-dihydroxyvitamin D3 on human myelogenous leukemia cells and monocytes. Blood 92[1]: 160–167.
32. Wu-Wong JR, Nakane M, Ma J (2006) Effects of vitamin D analogs on the expression of plasminogen activator inhibitor-1 in human vascular cells. Thromb Res 118[6]: 709–714.
33. Towler DA (2007) Calciotropic hormones and arterial physiology: "D"-lightful insights. J Am Soc Nephrol 18[2]: 369–373.
34. Norman PE, Powell JT (2005) Vitamin D, shedding light on the development of disease in peripheral arteries. Arterioscler Thromb Vasc Biol 25[1]: 39–46.
35. Zittermann A, Schleithoff SS, Koerfer R (2007) Vitamin D and vascular calcification. Curr Opin Lipidol 18[1]: 41–46.
36. Li YC, Kong J, Wei M, Chen ZF, Liu SQ, et al. (2002) 1,25-Dihydroxyvitamin D(3) is a negative endocrine regulator of the renin-angiotensin system. J Clin Invest 110[2]: 229–238.
37. Dawson-Hughes B, Heaney RP, Holick MF, Lips P, Meunier PJ, et al. (2005) Estimates of optimal vitamin D status. Osteoporos Int 16[7]: 713–716.
38. Vieth R, El Hajj FG (2005) There is no lower threshold level for parathyroid hormone as 25-hydroxyvitamin D concentrations increase. J Endocrinol Invest 28[2]: 183–186.
39. Heaney RP (2005) Serum 25-hydroxyvitamin D and parathyroid hormone exhibit threshold behavior. J Endocrinol Invest 28[2]: 180–182.
40. Bischoff-Ferrari HA, Giovannucci E, Willett WC, Dietrich T, Dawson-Hughes B (2006) Estimation of optimal serum concentrations of 25-hydroxyvitamin D for multiple health outcomes. American Journal of Clinical Nutrition 84[1]: 18–28.
41. Atherton K, Fuller E, Shepherd P, Strachan D, Power C (2008) Loss and representativeness in a biomedical survey at age 45 years: 1958 British birth cohort. Journal of Epidemiology and Community Health 62[3]: 216–223.
42. Power C, Elliott J (2006) Cohort profile: 1958 British birth cohort (National Child Development Study). International Journal of Epidemiology 35[1]: 34–41.
43. Hyppönen E, Turner S, Cumberland P, Power C, Gibb I (2007) Serum 25-hydroxyvitamin D measurement in a large population survey with statistical harmonization of assay variation to an international standard. Journal of Clinical Endocrinology and Metabolism 92[12]: 4615–4622.
44. Atherton K, Power C (2007) Health inequalities with the National Statistics-Socioeconomic classification: disease risk factors and health in the 1958 British birth cohort. Eur J Public Health 17[5]: 486–491.
45. Parsons TJ, Power C, Manor O (2006) Longitudinal physical activity and diet patterns in the 1958 British Birth Cohort. Med Sci Sports Exerc 38[3]: 547–554.
46. Jefferis BJ, Power C, Graham H, Manor O (2004) Changing social gradients in cigarette smoking and cessation over two decades of adult follow-up in a British birth cohort. J Public Health (Oxf) 26[1]: 13–18.
47. Royston P, Carlin JB, White IR (2009) Multiple imputation of missing values: New features for mim. Stata Journal 9[2]: 252–264.
48. Refinetti R, Cornelissen G, Halberg F (2007) Procedures for numerical analysis of circadian rhythms. Biological Rhythm Research 38[4]: 275–325.
49. Mackinnon DP, Fairchild AJ, Fritz MS (2007) Mediation analysis. Annu Rev Psychol 58: 593–614.
50. Sobel ME (1982) Asymptotic Confidence Intervals for Indirect Effects in Structural Equation Models. Sociological Methodology 13: 290–312. American Sociological Association.

Common genetic determinants of vitamin D insufficiency: a genome-wide association study

Thomas J Wang*, Feng Zhang*, J Brent Richards*, Bryan Kestenbaum*, Joyce B van Meurs*, Diane Berry*, Douglas P Kiel, Elizabeth A Streeten, Claes Ohlsson, Daniel L Koller, Leena Peltonen†, Jason D Cooper, Paul F O'Reilly, Denise K Houston, Nicole L Glazer, Liesbeth Vandenput, Munro Peacock, Julia Shi, Fernando Rivadeneira, Mark I McCarthy, Pouta Anneli, Ian H de Boer, Massimo Mangino, Bernet Kato, Deborah J Smyth, Sarah L Booth, Paul F Jacques, Greg L Burke, Mark Goodarzi, Ching-Lung Cheung, Myles Wolf, Kenneth Rice, David Goltzman, Nick Hidiroglou, Martin Ladouceur, Nicholas J Wareham, Lynne J Hocking, Deborah Hart, Nigel K Arden, Cyrus Cooper, Suneil Malik, William D Fraser, Anna-Liisa Hartikainen, Guangju Zhai, Helen M Macdonald, Nita G Forouhi, Ruth J F Loos, David M Reid, Alan Hakim, Elaine Dennison, Yongmei Liu, Chris Power, Helen E Stevens, Laitinen Jaana, Ramachandran S Vasan, Nicole Soranzo, Jörg Bojunga, Bruce M Psaty, Mattias Lorentzon, Tatiana Foroud, Tamara B Harris, Albert Hofman, John-Olov Jansson, Jane A Cauley, Andre G Uitterlinden, Quince Gibson, Marjo-Riitta Järvelin, David Karasik, David S Siscovick, Michael J Econs, Stephen B Kritchevsky, Jose C Florez, John A Todd*, Josee Dupuis*, Elina Hyppönen*, Timothy D Spector*

Summary

Background Vitamin D is crucial for maintenance of musculoskeletal health, and might also have a role in extraskeletal tissues. Determinants of circulating 25-hydroxyvitamin D concentrations include sun exposure and diet, but high heritability suggests that genetic factors could also play a part. We aimed to identify common genetic variants affecting vitamin D concentrations and risk of insufficiency.

Methods We undertook a genome-wide association study of 25-hydroxyvitamin D concentrations in 33 996 individuals of European descent from 15 cohorts. Five epidemiological cohorts were designated as discovery cohorts (n=16 125), five as in-silico replication cohorts (n=9367), and five as de-novo replication cohorts (n=8504). 25-hydroxyvitamin D concentrations were measured by radioimmunoassay, chemiluminescent assay, ELISA, or mass spectrometry. Vitamin D insufficiency was defined as concentrations lower than 75 nmol/L or 50 nmol/L. We combined results of genome-wide analyses across cohorts using *Z*-score-weighted meta-analysis. Genotype scores were constructed for confirmed variants.

Findings Variants at three loci reached genome-wide significance in discovery cohorts for association with 25-hydroxyvitamin D concentrations, and were confirmed in replication cohorts: 4p12 (overall $p=1 \cdot 9\times10^{-109}$ for rs2282679, in *GC*); 11q12 ($p=2 \cdot 1\times10^{-27}$ for rs12785878, near *DHCR7*); and 11p15 ($p=3 \cdot 3\times10^{-20}$ for rs10741657, near *CYP2R1*). Variants at an additional locus (20q13, *CYP24A1*) were genome-wide significant in the pooled sample ($p=6 \cdot 0\times10^{-10}$ for rs6013897). Participants with a genotype score (combining the three confirmed variants) in the highest quartile were at increased risk of having 25-hydroxyvitamin D concentrations lower than 75 nmol/L (OR 2·47, 95% CI 2·20–2·78, $p=2 \cdot 3\times10^{-48}$) or lower than 50 nmol/L (1·92, 1·70–2·16, $p=1 \cdot 0\times10^{-26}$) compared with those in the lowest quartile.

Interpretation Variants near genes involved in cholesterol synthesis, hydroxylation, and vitamin D transport affect vitamin D status. Genetic variation at these loci identifies individuals who have substantially raised risk of vitamin D insufficiency.

Funding Full funding sources listed at end of paper (see Acknowledgments).

Published Online June 10, 2010 DOI:10.1016/S0140-6736(10)60588-0

See Online/Comment DOI:10.1016/S0140-6736(10)60635-6

*Authors contributed equally

†Prof Peltonen died in March, 2010

Division of Cardiology, Department of Medicine (T J Wang MD), Diabetes Research Center (Diabetes Unit) (J C Florez PhD), and Center for Human Genetic Research (J C Florez), Massachusetts General Hospital, Boston, MA, USA; Hebrew SeniorLife, Institute for Aging Research, Genetic Epidemiology Program (D P Kiel MD, C-L Cheung PhD, D Karasik PhD), Harvard Medical School, Boston, MA, USA (T J Wang, J C Florez); Framingham Heart Study, Framingham, MA, USA (T J Wang, D P Kiel, Prof R S Vasan DM, Prof J Dupuis PhD); Department of Twin Research and Genetic Epidemiology, King's College London, London, UK (F Zhang PhD, M Mangino PhD, B Kato PhD, D Hart PhD, G Zhai PhD, N Soranzo PhD, Prof T D Spector MD); Jewish General Hospital (J Brent Richards MD, M Ladouceur PhD), McGill University Health Centre (Prof D Goltzman MD), Department of Medicine (J B Richards, M Ladouceur, Prof D Goltzman), Department of Human Genetics (J B Richards, M Ladouceur), and Department of Epidemiology and Biostatistics (J B Richards, M Ladouceur), McGill University, Montreal, QC, Canada;

Introduction

Vitamin D insufficiency affects as many as half of otherwise healthy adults in developed countries.[1] The musculoskeletal consequences of inadequate vitamin D concentrations are well established, and include childhood rickets, osteomalacia, and fractures.[2] A growing number of other disorders have also been linked to vitamin D insufficiency, although causal associations have not yet been established in randomised trials. These extraskeletal disorders include type 1 and type 2 diabetes,[2–4] cardiovascular disease,[5,6] increased risk of falls,[7] and cancers of the breast, colon, and prostate.[8–10] Results of a 2007 meta-analysis suggested that vitamin D supplementation substantially reduced mortality.[11]

Personal, social, and cultural factors are important determinants of vitamin D availability via their effects on sun exposure and diet. Sufficient exposure to ultraviolet light or adequate intake from diet or supplements is needed to maintain vitamin D status. Concentrations of the widely accepted biomarker for vitamin D, 25-hydroxyvitamin D, are highest in the summer and lowest in the winter in northern latitudes. However, only about a quarter of the interindividual variability in 25-hydroxyvitamin D concentration is attributable to season of measurement, geographical latitude, or reported vitamin D intake.[12,13] Results of previous twin and family studies suggest that genetic factors contribute substantially to this variability,[13,14] with estimates of

	Chromosome	Position	Nearest gene(s)	MAF	Combined p value for discovery samples (up to n=16 124)	Combined p value for replication samples (up to n=17 744)	Overall p value
rs2282679	4	72827247	*GC*	0·29	$4{\cdot}57\times10^{-63}$	$2{\cdot}88\times10^{-48}$	$1{\cdot}9\times10^{-109}$
rs3755967	4	72828262	*GC*	0·29	$7{\cdot}41\times10^{-53}$	$3{\cdot}00\times10^{-24}$	$2{\cdot}42\times10^{-75}$
rs17467825	4	72824381	*GC*	0·29	$3{\cdot}85\times10^{-52}$	$1{\cdot}61\times10^{-25}$	$6{\cdot}75\times10^{-76}$
rs1155563	4	72862352	*GC*	0·30	$4{\cdot}70\times10^{-55}$	$4{\cdot}26\times10^{-30}$	$2{\cdot}37\times10^{-73}$
rs2298850	4	72833131	*GC*	0·28	$8{\cdot}94\times10^{-49}$	$2{\cdot}12\times10^{-28}$	$2{\cdot}03\times10^{-71}$
rs7041	4	72837198	*GC*	0·44	$3{\cdot}74\times10^{-40}$	$1{\cdot}78\times10^{-19}$	$6{\cdot}31\times10^{-59}$
rs12785878	11	70845097	*DHCR7/NADSYN1*	0·23	$1{\cdot}27\times10^{-10}$	$2{\cdot}39\times10^{-16}$	$2{\cdot}12\times10^{-27}$
rs7944926	11	70843273	*DHCR7/NADSYN1*	0·23	$1{\cdot}56\times10^{-13}$	$7{\cdot}57\times10^{-4}$	$8{\cdot}96\times10^{-16}$
rs12800438	11	70848651	*DHCR7/NADSYN1*	0·23	$5{\cdot}98\times10^{-13}$	$6{\cdot}39\times10^{-4}$	$2{\cdot}54\times10^{-15}$
rs3794060	11	70865327	*DHCR7/NADSYN1*	0·23	$8{\cdot}09\times10^{-13}$	$6{\cdot}44\times10^{-4}$	$3{\cdot}38\times10^{-15}$
rs4945008	11	70898896	*DHCR7/NADSYN1*	0·24	$8{\cdot}98\times10^{-13}$	$6{\cdot}11\times10^{-4}$	$4{\cdot}55\times10^{-15}$
rs4944957	11	70845683	*DHCR7/NADSYN1*	0·23	$1{\cdot}43\times10^{-12}$	$7{\cdot}36\times10^{-4}$	$8{\cdot}70\times10^{-15}$
rs10741657	11	14871454	*CYP2R1*	0·40	$3{\cdot}91\times10^{-8}$	$2{\cdot}09\times10^{-14}$	$3{\cdot}27\times10^{-20}$
rs2060793	11	14871886	*CYP2R1*	0·40	$2{\cdot}69\times10^{-6}$	$2{\cdot}36\times10^{-7}$	$1{\cdot}73\times10^{-11}$
rs1993116	11	14866810	*CYP2R1*	0·40	$2{\cdot}94\times10^{-6}$	$1{\cdot}28\times10^{-6}$	$6{\cdot}25\times10^{-11}$
rs12794714	11	14870151	*CYP2R1*	0·43	$6{\cdot}24\times10^{-5}$	$8{\cdot}71\times10^{-7}$	$1{\cdot}84\times10^{-9}$
rs10500804	11	14866849	*CYP2R1*	0·43	$7{\cdot}43\times10^{-5}$	$1{\cdot}12\times10^{-6}$	$2{\cdot}67\times10^{-9}$
rs7116978	11	14838347	*CYP2R1*	0·36	$1{\cdot}17\times10^{-5}$	$7{\cdot}59\times10^{-5}$	$4{\cdot}99\times10^{-9}$

Results within each locus are ordered by strength of association with 25-hydroxyvitamin D concentration. MAF=minor allele frequency.

Table 1: **Single nucleotide polymorphisms identified in genome-wide association analyses for 25-hydroxyvitamin D concentrations**

Kidney Research Institute, Division of Nephrology, Harborview Medical Center (B Kestenbaum MD, I H de Boer MD), **Department of Medicine** (N L Glazer PhD, K Rice PhD, Prof B M Psaty PhD, Prof D S Siscovick MPH), **Cardiovascular Health Research Unit** (N L Glazer, K Rice, Prof D S Siscovick), **Department of Epidemiology** (Prof B M Psaty, Prof D S Siscovick), **and Department of Health Services** (Prof B M Psaty), **University of Washington, Seattle, WA, USA; Department of Internal Medicine** (J B van Meurs PhD, F Rivadeneira PhD, Q Gibson MBA, Prof A G Uitterlinden PhD), **Department of Epidemiology** (Prof A Hofman MD, Prof A G Uitterlinden), **and Department of Klinical Genetics** (Prof A G Uitterlinden), **Erasmus Medical Center, Rotterdam, Netherlands; Netherlands Genomics Initiative (NGI)-sponsored Netherlands Consortium for Healthy Aging (NCHA), Rotterdam, Netherlands** (J B van Meurs, F Rivadeneira, Prof A Hofman, Prof A G Uitterlinden); **UCL Institute of Child Health, MRC Centre of Epidemiology for Child Health and Centre for Paediatric Epidemiology and Biostatistics, London, UK** (D Berry MSc, Prof C Power PhD, E Hyppönen PhD); **Division of Endocrinology, University of Maryland School of Medicine, Baltimore, MD, USA** (E A Streeten MD, J Shi MSc); **Institute of Medicine, Department of Internal Medicine** (Prof C Ohlsson PhD, L Vandenput PhD, M Lorentzon PhD), **and Institute of Neuroscience and Physiology, Department of Physiology** (Prof J-O Jansson PhD), **Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden; School of Medicine, Indiana University, Indianapolis, IN, USA** (D L Koller PhD, Prof M Peacock DSc(Med), Prof T Foroud PhD, Prof M J Econs MD); **Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, UK** (Prof L Peltonen PhD, N Soranzo); **University of Helsinki and National Institute for Health and Welfare, Partnership for Molecular Medicine,**

heritability as high as 53%. Although several rare mendelian disorders cause functional vitamin D insufficiency, data for the effect of common genetic variation on vitamin D status are scarce. Candidate gene studies have been done to examine the effect of specific vitamin D pathway genes, but these studies have been limited by small sample sizes and the small numbers of variants examined.[13–18]

The SUNLIGHT consortium (Study of Underlying Genetic Determinants of Vitamin D and Highly Related Traits) was formed in 2008. It represents a collaboration of cohorts from the UK, USA, Canada, Netherlands, Sweden, and Finland. We aimed to identify common genetic variants affecting vitamin D concentrations and risk of vitamin D insufficiency.

Methods

Participants

We undertook a large, multicentre, genome-wide association study of 15 cohorts in Europe, Canada, and the USA. The discovery sample consisted of 16 125 individuals of European descent drawn from five epidemiological cohorts: the Framingham Heart Study, TwinsUK, the Rotterdam Study, the 1958 British Birth Cohort (1958BC), and the Amish Family Osteoporosis Study (AFOS). Five additional cohorts (n=9367) with genome-wide association data were used for in-silico replication: the Cardiovascular Health Study, the North Finland Birth Cohort 1966 (NFBC1966), the Indiana cohort, the Health, Aging, and Body Composition study (Health ABC), and the Gothenburg Osteoporosis and Obesity Determinants study (GOOD). We also undertook genotyping of selected variants in 5789 participants from four additional epidemiological cohorts (Canadian Multicentre Osteoporosis Study [CaMos], Chingford, Hertfordshire, and the Aberdeen Prospective Osteoporosis Screening Study [APOSS]), and 2715 additional participants from one of the discovery cohorts (1958BC). Full descriptions of all participating cohorts are shown in the webappendix (pp 1–7). Written informed consent was obtained from all participants in the included cohorts, and the study protocols were reviewed and approved by local institutional review boards.

Procedures

Details of genotyping methods, quality control, and imputation procedures used in all participating cohorts are shown in the webappendix (pp 7–14). 25-hydroxyvitamin D concentrations were measured by radioimmunoassay or chemiluminescent assay (DiaSorin Inc, Stillwater, MN, USA) in the Framingham Heart Study, TwinsUK, Rotterdam Study, Health ABC, AFOS, the GOOD cohort, and CaMoS. Detection limits ranged from 4 nmol/L to 10 nmol/L. In the 1958BC samples, 25-hydroxyvitamin D was measured with automated application of the ImmunoDiagnostic Systems OCTEIA ELISA on a Dade-Behring BEP2000 analyser (sensitivity of 5·0 nmol/L; Marburg, Germany).[19] In the Cardiovascular Health Study, NFBC1966, the Indiana cohort, Chingford, Hertfordshire, and APOSS, total 25-hydroxyvitamin D was measured with high-performance liquid chromatography-tandem mass

www.thelancet.com **Published online June 10, 2010 DOI:10.1016/S0140-6736(10)60588-0**

spectrometry. Serum concentrations of vitamin D binding protein were measured with immunonephelometric assay in the TwinsUK cohort.[20] The detection limit was 50 mg/L.

Statistical analyses

At the threshold $\alpha=5\times10^{-8}$, with a conservative discovery sample size of 14 000, our study had 80% power to detect single nucleotide polymorphisms accounting for 0·28% of the total variance in 25-hydroxyvitamin D concentrations, and 90% power to detect polymorphisms accounting for 0·32% of the total variance.

Genome-wide analyses were done within every cohort. In the Framingham Heart Study, TwinsUK, the Rotterdam Study, 1958BC, AFOS, NFBC1966, the Indiana cohort, Health ABC, and the GOOD study, linear regression models were used to generate cohort-specific residuals of naturally log transformed 25-hydroxyvitamin D concentrations adjusted for age, sex, body-mass index (BMI), and season. Log transformation was used to reduce skewness in the distribution of 25-hydroxyvitamin D. We modelled season using categorical variables for summer (July–September), autumn (October–December), winter (January–March), and spring (April–June). One set of definitions was used for season because most cohorts were at similar latitudes, and all were in the northern hemisphere.

In cohorts that included related individuals (Framingham, TwinsUK, AFOS, Indiana Women), we assessed association between additively coded single nucleotide polymorphism genotypes and standardised 25-hydroxyvitamin D residuals using either linear mixed-effect models or the score test implemented in MERLIN (version 1.1.2).[21] For imputed single nucleotide polymorphisms, expected number of minor alleles (ie, dose) was used in assessments of association between genotype and 25-hydroxyvitamin D residuals. In the Cardiovascular Health Study, analyses were adjusted for age, sex, and study site by inclusion of these factors as covariates in the model. In all samples, the genomic control approach was used to adjust p values for potential effects of mild population stratification and to prevent inflation of type I error occurring from any departure from normality of the trait variable.

A priori, we designated the first five genome-wide association studies, all of which used immunoassays to measure 25-hydroxyvitamin D concentrations, as discovery samples. The remaining five studies, three of which measured 25-hydroxyvitamin D by mass spectrometry and two by immunoassay, were designated as in-silico replication samples. We selected single nucleotide polymorphisms for replication if they had meta-analytic p values for association with 25-hydroxyvitamin D concentrations that were lower than 5×10^{-8} in the discovery samples. Additionally, we considered polymorphisms at or near six prespecified vitamin D pathway candidate genes: vitamin D receptor (*VDR*), 1-α-hydroxylase (*CYP27B1*), 25-hydroxylase (*CYP2R1*), 24-hydroxylase (*CYP24A1*), vitamin D binding protein (*GC*), and 27-hydroxylase and 25-hydroxylase (*CYP27A1*). These polymorphisms were tested in the replication samples if they met a p value threshold of 10^{-3} in the discovery samples. Lastly, we assessed selected polymorphisms for 25-hydroxyvitamin D association in the de-novo replication samples, using the same analytic approach. We then generated combined p values across the 15 studies.[22]

We undertook the meta-analysis using a weighted *Z*-score-based approach, as implemented in the software METAL (version 2009-10-10). In this approach, association p values were converted to signed *Z* statistics, for which the sign showed the direction of effect with respect to a reference allele. All *Z* scores were assigned a weight proportional to the square root of the sample size. Weighted *Z* statistics were summed across studies to obtain a global *Z* score and a corresponding two-sided p value. We regarded p values lower than 5×10^{-8} as genome-wide significant.[23]

We also assessed whether selected genetic variants from the continuous trait analyses were associated with vitamin D insufficiency in the Framingham Heart Study, TwinsUK, CaMoS, and 1958BC. We used two thresholds

	Framingham Heart Study (n=5656)	1958 British Birth Cohort (n=6552)
GC*		
Major homozygotes (nmol/L)	82·6 (0·73)	61·9 (0·34)
Heterozygotes (nmol/L)	74·8 (0·81)	57·0 (0·32)
Minor homozygotes (nmol/L)	64·6 (1·79)	52·8 (0·28)
***DHCR7*†**		
Major homozygotes (nmol/L)	79·7 (0·71)	59·6 (0·32)
Heterozygotes (nmol/L)	76·3 (0·86)	56·3 (0·31)
Minor homozygotes (nmol/L)	71·7 (2·01)	55·7 (0·31)
***CYP2R1*‡**		
Major homozygotes (nmol/L)	75·4 (0·87)	56·8 (0·31)
Heterozygotes (nmol/L)	78·6 (0·76)	60·2 (0·34)
Minor homozygotes (nmol/L)	81·6 (1·26)	61·1 (0·36)
Season		
Winter (nmol/L)	61·6 (1·00)	43·2 (0·26)
Spring/autumn (nmol/L)	77·4 (0·68)	57·1 (0·30)
Summer (nmol/L)	95·8 (1·00)	71·7 (0·31)
Supplementation		
Yes (nmol/L)	83·4 (0·80)	65·9 (0·32)
No (nmol/L)	74·7 (0·69)	56·9 (0·30)

Data are mean (SE). Sample from 1958 British Birth Cohort (1958BC) consists of a combination of the genome-wide association study sample and the de novo genotyping sample (webappendix p 2). *rs2282679 in Framingham cohort, rs4588 in 1958BC (r^2 between single nucleotide polymorphisms >0·99). †rs7944926 in Framingham cohort, rs12785878 in 1958BC (r^2 between polymorphisms >0·99). ‡rs10741657 in Framingham cohort and 1958BC.

Table 2: Mean 25-hydroxyvitamin D concentrations by genotype, season, and supplementation status

Institute for Molecular Medicine Finland FIMM, Helsinki, Finland (Prof L Peltonen); **National Institute for Health and Welfare, Helsinki, Finland** (Prof L Peltonen); **JDRF/WT Diabetes and Inflammation Laboratory, University of Cambridge, Cambridge, UK** (J D Cooper PhD, D J Smyth BSc, H E Stevens HNC, Prof J A Todd PhD); **Faculty of Medicine, Department of Epidemiology and Public Health, Imperial College, London, UK** (P F O'Reilly PhD, Prof M-R Järvelin PhD); **Sticht Center on Aging** (D K Houston PhD, Y Liu PhD, Prof S B Kritchevsky PhD) **and Division of Public Health Sciences** (Prof G L Burke MD), **School of Medicine, Wake Forest University, Winston-Salem, NC, USA; Oxford Centre for Diabetes, Endocrinology and Metabolism (OCDEM)** (Prof M I McCarthy FMedSci), **Wellcome Trust Centre for Human Genetics** (Prof M I McCarthy), **NIHR Musculoskeletal Biomedical Research Unit** (Prof N K Arden MD, Prof C Cooper FMedSci), **University of Oxford, Oxford, UK; Oxford NIHR Biomedical Research Centre, Churchill Hospital, Oxford, UK** (Prof M I McCarthy); **National Institute of Health and Welfare, Oulu, Finland** (P Anneli PhD, Prof M-R Järvelin); **Jean Mayer USDA Human Nutrition Research Center on Aging, Tufts University, Boston, MA, USA** (Prof S L Booth PhD, Prof P F Jacques ScD); **Department of Medicine, Cedars-Sinai Medical Center, Los Angeles, CA, USA** (Prof M Goodarzi PhD); **Genome Institute of Singapore, Computational and Mathematical Biology, A*STAR (Agency for Science, Technology and Research), Biopolis, Singapore** (C-L Cheung); **Division of Nephrology and Hypertension, University of Miami Miller School of Medicine, Miami, FL, USA** (M Wolf MD); **Health Canada, Ottawa, ON, Canada** (N Hidiroglou PhD); **MRC Epidemiology Unit, Institute of Metabolic Science, Addenbrooke's Hospital, Cambridge, UK** (Prof N J Wareham PhD, N G Forouhi FFPH, R J F Loos PhD); **Division of Applied Medicine,**

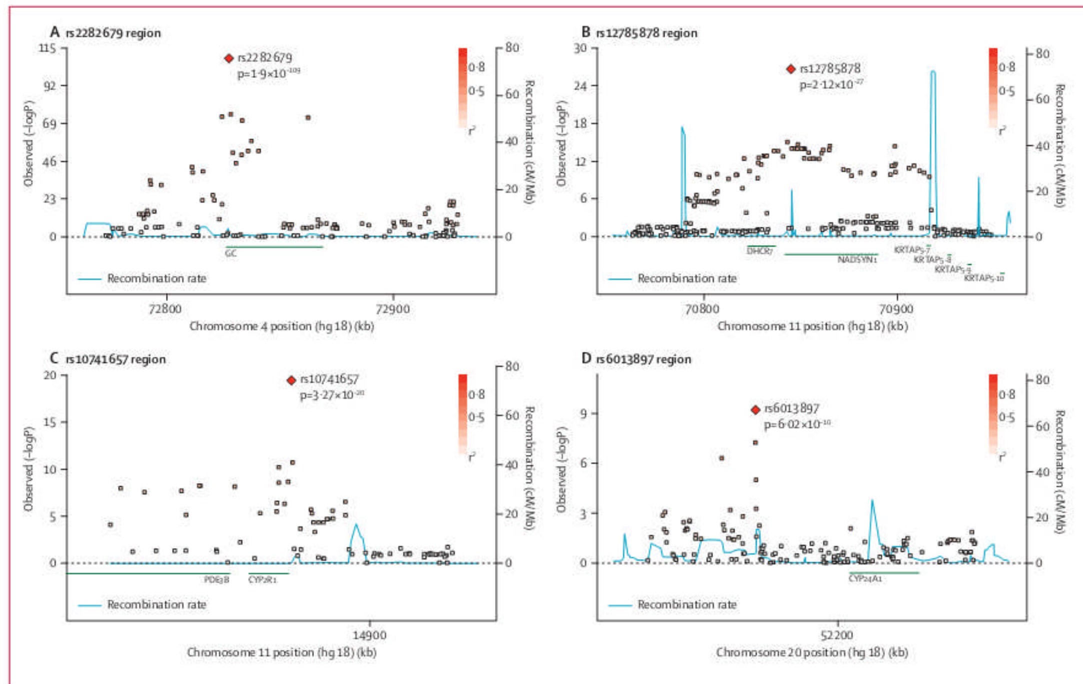


Figure 1: **Regional linkage disequilibrium plots for single nucleotide polymorphisms at *GC* (A), *DHCR7/NADSYN1* (B), *CYP2R1* (C), and *CYP24A1* (D)**

Bone and Musculoskeletal Research Programme, University of Aberdeen, Aberdeen, UK (L J Hocking PhD, H M Macdonald PhD, Prof D M Reid MD); **MRC Epidemiology Resource Centre, University of Southampton, Southampton, UK** (Prof N K Arden, Prof C Cooper, E Dennison PhD); **Office of Biotechnology, Genomics and Population Health, Public Health Agency of Canada, Toronto, ON, Canada** (S Malik PhD); **Unit of Clinical Chemistry, School of Clinical Sciences, University of Liverpool, Liverpool, UK** (Prof W D Fraser MD); **Department of Obstetrics and Gynaecology** (Prof A-L Hartikainen PhD), **Institute of Health Sciences** (L Jaana PhD, Prof M-R Järvelin), **and Biocenter Oulu** (Prof M-R Järvelin), **University of Oulu, Oulu, Finland; Rheumatology Department, Whipps Cross University**

for vitamin D insufficiency: 25-hydroxyvitamin D concentrations lower than 75 nmol/L (30 ng/mL) and lower than 50 nmol/L (20 ng/mL).[1] Covariates were age, sex, season, and BMI. We combined effect estimates from the logistic regression analysis across cohorts by meta-analysis using an inverse-variance weighting approach. We also did analyses using a 25 nmol/L (10 ng/mL) threshold, to examine whether genetic variants were associated with severe vitamin D deficiency.

Additionally, we constructed a genotype score by taking a weighted average of the number of risk alleles for members of a cohort, with weights established using β coefficients from the meta-analysis. Logistic regression was used to calculate the odds of vitamin D insufficiency according to quartile of the genotype score. For this analysis, we combined data from the Framingham Heart Study, TwinsUK, and 1958BC using a multivariate approach, with β coefficients for each quartile of genotype score meta-analysed jointly, as previously described.[24]

Role of the funding source

The sponsors of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Results

Characteristics of the study cohorts are summarised in the webappendix (pp 15–17). Table 1 shows the results of genome-wide association analyses. In analysis of data from the five discovery samples, single nucleotide polymorphisms at three unique loci met the prespecified threshold for genome-wide significance: 4p12, 11q12, and 11p15. The 4p12 polymorphisms were within or near the *GC* gene, and the results included a non-synonymous polymorphism in this gene, rs7041. The 11q12 polymorphisms were near *DHCR7/NADSYN1* (7-dehydrocholesterol reductase/NAD synthetase 1) and the 11p15 polymorphisms near *CYP2R1* (cytochrome P450, subfamily IIR).

Associations at all three loci were confirmed in replication samples. The polymorphism at *GC* with the lowest p value in discovery samples, rs2282679, had a combined p value of $2{\cdot}9\times10^{-48}$ in in-silico replication samples, with a consistent direction of effect. Additional genotyping was not done for this polymorphism. Polymorphism rs10741657 at *CYP2R1* had a p value of

www.thelancet.com **Published online June 10, 2010 DOI:10.1016/S0140-6736(10)60588-0**

$2 \cdot 1 \times 10^{-14}$ in in-silico and de-novo replication samples, also with a consistent direction of effect. At the *DHCR7/NADSYN1* locus, a perfect proxy for rs7944926 (rs12785878, $r^2=1 \cdot 0$) was genotyped in de-novo replication samples, and had a combined replication p value (in-silico and de-novo samples) of $2 \cdot 4 \times 10^{-16}$. Overall p values (discovery and replication samples) for the three confirmed single nucleotide polymorphisms ranged from $3 \cdot 3 \times 10^{-20}$ to $1 \cdot 9 \times 10^{-109}$ (table 1). Figure 1 shows regional plots for the results at each locus. In the discovery cohorts, single nucleotide polymorphisms at the three confirmed loci (*GC*, *DHCR7/NADSYN1*, and *CYP2R1*) accounted for 1–4% of the variation in 25-hydroxyvitamin D concentrations.

We compared mean concentrations of 25-hydroxyvitamin D by genotype category at the three loci in the two largest cohorts (combined n=12 208) with mean concentrations by supplementation status and season (table 2). Differences in mean 25-hydroxyvitamin D concentrations between minor and major homozygotes for the strongest genetic variants were similar to those seen with supplementation in these cohorts, and were nearly as large as differences recorded for a one season change.

In the candidate gene analysis, polymorphism rs6013897 near *CYP24A1* (cytochrome P450, family 24, subfamily A) had a p value of $7 \cdot 2 \times 10^{-4}$ in the discovery cohorts, and was tested for replication. The p value was $8 \cdot 4 \times 10^{-8}$ in the replication cohorts, resulting in an overall p value (discovery and replication) of $6 \cdot 0 \times 10^{-10}$. Figure 1 shows a regional plot for the results at the *CYP24A1* locus. An additional candidate polymorphism, rs2544037 near *VDR*, had a p value of $6 \cdot 2 \times 10^{-4}$ in the discovery cohorts, but was not confirmed in replication samples. No polymorphisms were identified near *CYP27B1* or *CYP27A1* with p values less than 10^{-3} in the discovery cohorts.

We did additional analyses to assess effects of the three confirmed variants on risk of clinical vitamin D insufficiency (25-hydroxyvitamin D concentrations <75 nmol/L or <50 nmol/L). Table 3 shows results for the variants individually and in combination. Participants with a genotype score (combining the three variants) in the top quartile had increased odds of vitamin D insufficiency (figure 2). Genotype score was also associated with risk of severe vitamin D deficiency (25-hydroxyvitamin D concentration <20 nmol/L), with an adjusted odds ratio for participants in the top quartile of 1·43 (95% CI 1·13–1·79; p=0·002).

In view of the strong association of genetic variants at *GC* with 25-hydroxyvitamin D concentrations, we also examined whether these variants were associated with serum concentrations of vitamin D binding protein, which was measured in 1674 individuals in the TwinsUK cohort. The single nucleotide polymorphism rs2282679 was strongly associated with concentrations of vitamin D binding protein ($p=4 \cdot 0 \times 10^{-42}$), with the minor allele related to reduced protein concentrations.

	25-hydroxyvitamin D concentration <75 nmol/L		25-hydroxyvitamin D concentration <50 nmol/L	
	Odds ratio (95% CI)	p value	Odds ratio (95% CI)	p value
Individual variants				
GC (rs2282679)	1·63 (1·53–1·73)	$3 \cdot 5 \times 10^{-50}$	1·49 (1·40–1·59)	$7 \cdot 5 \times 10^{-33}$
DHCR7 (rs7944926)	1·21 (1·14–1·29)	$4 \cdot 1 \times 10^{-10}$	1·21 (1·14–1·29)	$4 \cdot 7 \times 10^{-09}$
CYP2R1 (rs10741657)	1·21 (1·14–1·29)	$9 \cdot 4 \times 10^{-11}$	1·06 (1·00–1·13)	0·06
Genotype score				
Quartile 1	1·0 (Reference)	..	1·0 (Reference)	..
Quartile 2	1·29 (1·15–1·46)	..	1·10 (0·97–1·25)	..
Quartile 3	1·56 (1·39–1·75)	..	1·38 (1·22–1·57)	..
Quartile 4	2·47 (2·20–2·78)*	..	1·92 (1·70–2·16)*	..

For individual variants, odds ratios are per copy of the risk allele. All logistic regressions were adjusted for age, sex, body-mass index, and season. *p values for trends in odds ratios for genotype scores were $2 \cdot 3 \times 10^{-48}$ for 25-hydroxyvitamin D concentrations lower than 75 nmol/L and $1 \cdot 0 \times 10^{-26}$ for lower than 50 nmol/L.

Table 3: **Genetic variants and risk of vitamin D insufficiency**

Discussion

Vitamin D insufficiency has been implicated in many musculoskeletal and extraskeletal diseases,[1,2] which has led to substantial interest in the determinants of vitamin D status. Our findings establish a role for common genetic variants in regulation of circulating 25-hydroxyvitamin D concentrations. The presence of harmful alleles at the three confirmed loci more than doubled the risk of vitamin D insufficiency. These findings improve our understanding of vitamin D homoeostasis and could assist identification of a subgroup of the white population who are at risk of vitamin D insufficiency.

DHCR7/NADSYN1 is a novel locus for association with vitamin D status, but one with compelling biological plausibility. *DHCR7* encodes the enzyme 7-dehydrocholesterol (7-DHC) reductase, which converts 7-DHC to cholesterol, thereby removing the substrate from the synthetic pathway of vitamin D_3, a precursor of 25-hydroxyvitamin D_3. Rare mutations in *DHCR7* lead to Smith-Lemli-Opitz syndrome, which is characterised by reduced activity of 7-DHC reductase, accumulation of 7-DHC, low cholesterol, and many congenital abnormalities.[25] Mutations in *DHCR7* might also confer a competitive advantage to heterozygous carriers, because high concentrations of 7-DHC could provide protection against rickets and osteomalacia from hypovitaminosis D.[26] However, few data exist for vitamin D status in individuals with Smith-Lemli-Opitz syndrome or carriers of mutations.[27] The finding that common variants at *DHCR7* are strongly associated with circulating 25-hydroxyvitamin D concentrations suggests that this enzyme could have a larger role in regulation of vitamin D status than has previously been recognised.

The gene at the second locus, *CYP2R1*, encodes a hepatic microsomal enzyme. CYP2R1 could be the enzyme underlying 25-hydroxylation of vitamin D in the liver, but this suggestion is uncertain because many other

Hospital, London, UK (A Hakim MA); **Finnish Institute of Occupational Health, Oulu, Finland** (L Jaana); **Division of Preventive Medicine, School of Medicine** (Prof R S Vasan), **and Department of Biostatistics, School of Public Health** (Prof J Dupuis), **Boston University, Boston, MA, USA; Klinikum der Johann Wolfgang Goethe University, Frankfurt, Germany** (J Bojunga MD); **National Institutes of Health, National Institute on Aging, Bethesda, MD, USA** (T B Harris MD); **Department of Epidemiology, University of Pittsburgh, Pittsburgh, PA, USA** (J A Cauley PhD); **and Broad Institute, Program in Medical and Population Genetics, Cambridge, MA, USA** (J C Florez)

Correspondence to: Prof Timothy D Spector, Department of Twin Research and Genetic Epidemiology, King's College, London, St Thomas' Hospital Campus, 1st Floor South Wing Block 4, Westminster Bridge Road, London SE1 7EH, UK
tim.spector@kcl.ac.uk

Elina Hyppönen, MRC Centre of Epidemiology for Child Health and Centre for Paediatric Epidemiology and Biostatistics, UCL Institute of Child Health, 30 Guilford Street, London WC1N 1EH
e.hypponen@ich.ucl.ac.uk

Dr Thomas J Wang, Cardiology Division, GRB-800, Massachusetts General Hospital, 55 Fruit Street, Boston, MA 02114, USA
tjwang@partners.org

See Online for webappendix

For **MERLIN** see http://www.sph.umich.edu/csg/abecasis/Merlin/

For **METAL** see www.sph.umich.edu/csg/abecasis/metal/

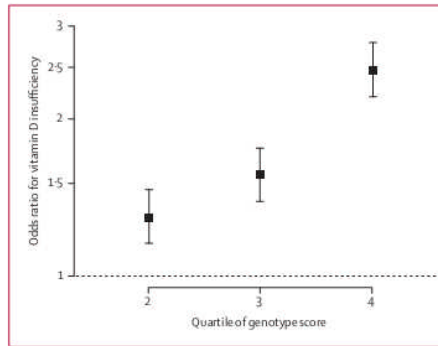


***Figure 2:* Risk of vitamin D insufficiency*, by quartile of genotype score**
*25-hydroxyvitamin D concentration lower than 75 nmol/L. Error bars show 95% CIs.

enzymes with 25-hydroxylase activity in vitro have been described.[28] Previous clinical studies have been limited to a case report of a Nigerian man with a point mutation in *CYP2R1* who had a history of rickets,[28] and a previous candidate gene study in 133 individuals with type 1 diabetes.[18] Because affected individuals with *CYP2R1* polymorphisms have been difficult to identify, redundancy in the enzymes involved in the 25-hydroxylation step has been proposed. Thus, our finding that common variants at the *CYP2R1* locus are associated with circulating 25-hydroxyvitamin D concentrations is the strongest evidence so far that CYP2R1 is the enzyme underlying the crucial first step in vitamin D metabolism.

The third gene, *GC*, encodes vitamin D binding protein, which is a 52–59 kDA protein synthesised in the liver that binds and transports vitamin D and its metabolites (including 25-hydroxyvitamin D and 1,25-dihydroxyvitamin D).[29] A few investigators have reported associations between non-synonymous single nucleotide polymorphisms in this gene[15–17,30,31] and 25-hydroxyvitamin D concentrations. However, their studies were small (≤1500 participants) and results were not replicated. The most widely studied *GC* variants are the non-synonymous polymorphisms rs7041 (Asp→Glu) and rs4588 (Thr→Lys). The previous nomenclature for *GC* haplotypes (GC1S, GC1F, and GC2) was based on specific combinations of alleles at these non-synonymous polymorphisms.[15] Our data strongly confirm the association of rs7041 with circulating 25-hydroxyvitamin D. The other variant, rs4588, is not in the HapMap dataset and is thus not part of our imputed results. However, rs4588 is only 11 bp away from rs7041, and direct genotyping of rs4588 in one of our samples (TwinsUK) confirms that it is in linkage disequilibrium (r^2>0·99) with several associated variants from our genome-wide association study.

We also showed that *GC* variants associated with low 25-hydroxyvitamin D concentrations were strongly related to reduced concentrations of vitamin D binding protein. Whether variation in the amount of circulating binding protein affects metabolism and availability of vitamin D is not well established. Concentrations of the binding protein have been postulated to affect delivery of 25-hydroxyvitamin D and activated vitamin D (1,25-dihydroxyvitamin D) to target organs, as well as clearance of vitamin D metabolites from the circulation.[15,16] Alternatively, changes in quantity or function of the binding protein could be accompanied by changes in the relative proportions of free and bound 25-hydroxyvitamin D, with the free proportion being the potential rate-limiting factor for 1,25-dihydroxyvitamin D production. Further studies are needed to assess the effects of variation in serum concentrations of vitamin D binding protein.

In a screen of candidate gene variants, we noted an additional association at the locus containing *CYP24A1* that was genome-wide significant in pooled analyses of the discovery and replication samples. *CYP24A1* encodes 24-hydroxylase, which initiates degradation of both 25-hydroxyvitamin D and 1,25-dihydroxyvitamin D. In previous candidate gene and linkage studies, investigators have not shown an association of variants at this locus with 25-hydroxyvitamin D concentrations, but these studies were relatively small.[10,32]

A high genotype score for the three variants identified in our genome-wide association study conferred roughly a two-fold increase in risk of vitamin D insufficiency (25-hydroxyvitamin D concentrations <50 nmol/L or <75 nmol/L) compared with a score in the lowest quartile, after we accounted for environmental factors. This result suggests that variation at a few genetic loci could have a clinically important effect on risk of vitamin D insufficiency. High genotype score was associated with a 1·4-fold raised risk of severe vitamin D deficiency (<25 nmol/L). Whether the reduced odds ratio for the 25 nmol/L threshold shows an increased contribution of environmental factors to the most severe forms of vitamin D deficiency is unclear, because severe deficiency was rare in our community-based cohorts.

Whether genetic predisposition modifies response to sun exposure or dietary supplementation warrants further study, especially in view of the large interindividual differences that have been reported in response to treatment with identical doses of vitamin D.[33] Furthermore, these variants might provide useful genetic approaches to investigate the role of vitamin D insufficiency in several chronic diseases with which this disorder has been epidemiologically linked.

The validity of our findings is lent support by the large study sample (more than 30 000 participants combined in discovery and validation samples), consistent results across several standard assays for 25-hydroxyvitamin D, and the strong biological plausibility of genes at the principal loci. Several limitations of the study also deserve mention, however. The study was not designed to identify

www.thelancet.com **Published online June 10, 2010 DOI:10.1016/S0140-6736(10)60588-0**

uncommon or rare variants. Resequencing at selected loci, partly on the basis of our results, could be used to identify uncommon variants with potentially large effects.

We used a multistage design to achieve maximum homogeneity of the assays used in the discovery analyses. We might have identified more genome-wide significant associations had we combined all study cohorts into one stage, but we would not have had a large replication sample. Other factors that might have contributed to reduced statistical power are second-order interactions (eg, with age) and the use of a stringent p-value threshold in the discovery stage.[34] Accordingly, the absence of specific candidate genes, such as those affecting vitamin D action or skin pigmentation, from our most significant results does not exclude an effect of genetic variation at these loci on vitamin D concentrations, but their contributions might be small compared with those of the genes that we identified.

Assays used to measure 25-hydroxyvitamin D concentrations varied between cohorts. To keep potential variability introduced by cohort-specific measurement techniques to a minimum, we standardised 25-hydroxyvitamin D concentrations within cohorts and analysed this variable as a continuous trait. Furthermore, primary results were meta-analysed with a *Z*-score-weighted approach, which is not scale-dependent. Specific information about dietary intake and sunlight exposure was not available from all cohorts. Such factors probably contribute to non-genetic variability in 25-hydroxyvitamin D concentrations, which would reduce the effect noted in our analyses.

The single nucleotide polymorphisms that we have identified might not be causal variants, but rather be in linkage disequilibrium with these variants. We did not examine downstream markers of vitamin D status, because 25-hydroxyvitamin D concentration is regarded as the most reliable indicator of vitamin D status. Other molecules, such as 1,25-dihydroxyvitamin D or parathyroid hormone, have greater intraindividual variability than does 25-hydroxyvitamin D and are affected by several determinants other than vitamin D status. Lastly, we studied only white individuals of European descent. Whether the genetic variants we identified affect vitamin D status in other racial or ethnic groups is unknown and warrants further study.

Contributors

JD, FZ, JBR, BK, JBM, DB, CO, DLK, JDC, PFO, NLG, LV, JS, MM, BK, KR, ML(1), ML(2), LJH, ALH, GZ, RJFL, and TF took part in data analysis. TJW, JBR, BK, JBM, DB, DPK, CO, MRJ, FR, DG, NJW, NKA, CC, ALH, ED, CP, NS, ML(2), TBH, AH, AGU, LP, DK, SBK, JCF, JAT, EH, TDS contributed to study design. TJW, FZ, DB, DPK, EAS, CO, DLK, MRJ, PFO, DKH, LV, MP, MIM, PA, MM, DJS, GLB, DG, NH, NJW, DH, NKA, CC, WDF, GZ, HMM, RJFL, DMR, AH, ED, YL, CP, HES, LJ, NS, JB, ML(2), TBH, JJ, JAC, LP, DSS, MJE, SBK, JAT, EH, TDS, SLH, QG, and SLB contributed to data collection. TJW, FZ, JBR, BK, JBM, DB, DPK, EAS, CO, DLK, MRJ, JDC, PFO, DKH, NLG, LV, MP, FR, MIM, IHB, DJS, SLB, PFJ, GLB, CLC, MW, KR, DG, NH, ML(1), NJW, LJH, NKA, CC, SM, ALH, HMM, RJFL, AH, ED, CP, HES, BMP, ML(2), TBH, AH, JJ, JAC, AGU, LP, DK, MJE, SBK, JCF, JAT, JD, EH, and TDS interpreted results. FZ, JBR, JBM, DB, DPK, EAS, CO, JDC, PFO, DKH, NLG, LV, MP, FR, MIM, PA, IHB, DJS, SLB, PFJ, GLB, MG, CLC, MW, KR, DG, NH, ML(1), NJW, LJH, NKA, CC, SM, WDF, ALH, HMM, RJFL, ED, YL, CP, HES, BMP, ML(2), TF, TBH, AH, JJ, JAC, AGU, DK, DSS, MJE, SBK, JCF, JAT, JD, EH, and TDS read the manuscript critically. The writing group consisted of TJW, FZ, JBR, BK, EAS, DLK, MRJ, MG, JCF, JAT, JD, EH, and TDS. ML(1)=Martin Ladouceur. ML(2)=Mattias Lorentzon.

Conflicts of interest

TJW has served on the scientific advisory board of Diasorin. DKH has received honoraria from Abbott Nutrition. MW has received consultancy fees, honoraria, and speakers' fees from Abbott and Genzyme. DMR has acted as a consultant for Novartis, Roche, Pfizer, Amgen, Shire, Merck, and Servier, has received speakers' fees from Novartis, Roche, and Amgen, and owns stock in GlaxoSmithKline and Astra Zeneca. ML(2) has received lecture fees from Novartis and Sanofi-Aventis. All other authors declare that they have no conflicts of interest.

Acknowledgments

Framingham Heart Study The Framingham Heart Study of the National Heart, Lung and Blood Institute (NHLBI) of the US National Institutes of Health (NIH) and Boston University School of Medicine is supported by the NIH/NHLBI contract N01-HC-25195. The present study received support from the American Heart Association, the US Department of Agriculture, Agricultural Research Service (under Cooperative Agreement No 58-1950-7-707), and the National Institute of Aging (AG14759). DK was supported by a grant from the National Institute of Arthritis, Musculoskeletal, and Skin Diseases and the National Institute on Aging (R01 AR/AG 41398). The analyses reflect intellectual input and resource development from the Framingham Heart Study investigators participating in the SNP Health Association Resource project. This work was partly supported by a contract with Affymetrix Inc for genotyping services (Contract No N02-HL-6-4278). A portion of this research used the Linux Cluster for Genetic Analysis, which is funded by the Robert Dawson Evans Endowment of the Department of Medicine at Boston University School of Medicine and Boston Medical Center. *TwinsUK and Chingford* The study was funded by the Wellcome Trust, Arthritis Research Campaign, European Community's Seventh Framework Programme (FP7/2007-2013)/grant agreement HEALTH-F2-2008-201865-GEFOS and Seventh Framework Programme grant 200800 Treat OA /(FP7/2007-2013), ENGAGE project grant agreement HEALTH-F4-2007-201413, and the FP-5 GenomEUtwin Project (QLG2-CT-2002-01254). The study also receives support from the UK Department of Health via the National Institute for Health Research (NIHR) comprehensive Biomedical Research Centre award to Guy's & St Thomas' NHS Foundation Trust in partnership with King's College London. TDS is an NIHR senior investigator. The project also received support from a Biotechnology and Biological Sciences Research Council project grant. (G20234). The authors acknowledge the funding and support of the National Eye Institute (NEI) via an NIH/Center for Inherited Disease Research (CIDR) genotyping project. We thank the staff from the Genotyping Facilities at the Wellcome Trust Sanger Institute for sample preparation, quality control, and genotyping; Le Centre National de Génotypage, France, for genotyping; Duke University, NC, USA, for genotyping; and the Finnish Institute of Molecular Medicine, Finnish Genome Center, University of Helsinki. Genotyping was also done by CIDR as part of an NEI/NIH project grant. *The Rotterdam Study* This study was funded by the Netherlands Organization of Scientific Research (NWO) Investments (175.010.2005.011, 911-03-012), the Research Institute for Diseases in the Elderly (014-93-015; RIDE2), the Netherlands Genomics Initiative/NWO project 050-060-810, and the European Commission (HEALTH-F2-2008-201865-GEFOS, and HEALTH-F2-2008-00-TREAT-OA). The Rotterdam Study is funded by Erasmus Medical Center and Erasmus University, Rotterdam, Netherlands Organization for the Health Research and Development, the Research Institute for Diseases in the Elderly, the Ministry of Education, Culture and Science, the Ministry for Health, Welfare and Sports, the European Commission (DG XII), and the Municipality of Rotterdam. We thank the staff from the Rotterdam Study, and the participating general practitioners and pharmacists.

For the **Framingham Heart Study** see http://www.framinghamheartstudy.org

For the **Wellcome Trust Case-Control Consortium website** see www.wtccc.org.uk

1958 British Birth Cohort The project was funded by the UK Medical Research Council (MRC) (project grant G0601653), and 25-hydroxyvitamin D assays by the BUPA foundation. EH is funded by a UK Department of Health Public Health Career Scientist Award. Use of DNA from this cohort was funded by MRC grant G0000934 and Wellcome Trust grant 068545/Z/02. This research used resources provided by the Type 1 Diabetes Genetics Consortium, a collaborative clinical study sponsored by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), National Institute of Allergy and Infectious Diseases, National Human Genome Research Institute, National Institute of Child Health and Human Development, and Juvenile Diabetes Research Foundation International (JDRF) and supported by U01 DK062418. This study makes use of data generated by the Wellcome Trust Case-Control Consortium. A full list of investigators who contributed to generation of the data is available from the Wellcome Trust Case-Control Consortium website. Funding for the project was provided by the Wellcome Trust under award 076113. The MRC Centre of Epidemiology for Child Health is funded by the MRC. Great Ormond Street Hospital/University College London, Institute of Child Health receives a proportion of funding from the Department of Health's NIHR ('Biomedical Research Centres' funding).
Health, Aging and Body Composition Study This research was supported by the Intramural Research Program of the NIH National Institute on Aging and National Institute on Aging contracts N01-AG-6-2101, N01-AG-6-2103, and N01-AG-6-2106. Assessment of 25-hydroxyvitamin D concentrations was funded by a National Institute on Aging grant, R01-AG029364. The genome-wide association study was funded by a National Institute on Aging grant, R01-AG032098, and genotyping services were provided by CIDR. CIDR is fully funded through a federal contract from NIH to The Johns Hopkins University (contract number HHSN268200782096C).
The Amish Family Osteoporosis Study The Amish Family Osteoporosis Study was funded by a grant from the National Institute of Arthritis, Musculoskeletal and Skin Diseases (R01 AR46838).
Gothenberg Osteoporosis and Obesity Determinants (GOOD) Study Financial support was received from the Swedish Research Council, the Swedish Foundation for Strategic Research, The ALF/LUA research grant in Gothenburg, the Lundberg Foundation, the Emil and Vera Cornell Foundation, the Torsten and Ragnar Söderberg's Foundation, Petrus and Augusta Hedlunds Foundation, the Västra Götaland Foundation, the Göteborg Medical Society, and the Sahlgrenska Center for Cardiovascular and Metabolic Research (CMR, no A305:188), which is supported by the Swedish Strategic Foundation.
Study of Indiana Women This work was supported by NIH grants P01 AG-18397 and M01 RR-00750. Genotyping services were provided by CIDR. CIDR is fully funded through a federal contract from NIH to The Johns Hopkins University, contract number HHSN268200782096C. This research was supported in part by the Intramural Research Program of the NIH, National Library of Medicine.
North Finland Birth Cohort 1966 The project was funded by the MRC (project grant G0601653). Financial support was received from the Academy of Finland (project grants 104781, 120315, 1114194 and Center of Excellence in Complex Disease Genetics), University Hospital Oulu, Biocenter, University of Oulu, Finland, NHLBI grant 5R01HL087679-02 through the STAMPEED program (1RL1MH083268-01), ENGAGE project and grant agreement HEALTH-F4-2007-201413, MRC (studentship grant G0500539, centre grant G0600705), the Wellcome Trust (project grant GR069224), UK. DNA extractions, sample quality controls, biobank up-keeping, and aliquotting were done at the National Public Health Institute, Biomedicum Helsinki, Finland and supported financially by the Academy of Finland and Biocentrum Helsinki.
Cardiovascular Health Study Research for this cohort that was reported in this article was supported by contract numbers N01-HC-85079 to N01-HC-85086, N01-HC-35129, N01 HC-15103, N01-HC-55222, N01-HC-75150, N01-HC-45133, grant numbers U01 HL080295, R01 HL084443, R01 HL087652, and R01 AG027002 from NHLBI, with additional contribution from the National Institute of Neurological Disorders and Stroke. A full list of principal investigators and institutions is available at the Cardiovascular Health Study website.

For the **Cardiovascular Health Study website** see http://www.chs-nhlbi.org/pi.htm

DNA handling and genotyping was supported in part by National Center for Research Resources grant M01RR00069 to the Cedars-Sinai General Clinical Research Center Genotyping core and NIDDK grant DK063491 to the Southern California Diabetes Endocrinology Research Center.
Hertfordshire This study was funded by the UK MRC and the Arthritis Research Campaign.
Aberdeen Prospective Osteoporosis Screening Study Funding for cohort sample collection and analysis was supported in part by grants from the European Commission (QLRT-2001-02629) and the UK Food Standards Agency. We thank the clinical research staff in the Bone and Musculoskeletal Research Programme for all their hard work, and all the women who participated in the study.
Canadian Multicentre Osteoporosis Study CaMoS is funded by the Canadian Institutes of Health Research (CIHR). JBR is a clinical investigator of the CIHR and an Osteoporosis Canada New Investigator. This work was supported by grants to JBR from the CIHR and Canadian Foundation for Innovation. We acknowledge the support and funding of the Public Health Agency of Canada to assay vitamin D levels, the expert technical support of Kurtis Sarafin and Jenn Kreiger of Health Canada (vitamin D measurements) and the input of Linda S Greene-Finestone and Ross Duncan.
Other The Cambridge Institute for Medical Research is in receipt of a Wellcome Trust Strategic Award (079895). This work was also supported by JDRF, the Wellcome Trust, and the NIHR Cambridge Biomedical Centre.

References

1 Holick MF. Vitamin D deficiency. *N Engl J Med* 2007; **357:** 266–81.
2 Bouillon R, Carmeliet G, Verlinden L, et al. Vitamin D and human health: lessons from vitamin D receptor null mice. *Endocr Rev* 2008; **29:** 726–76.
3 Hyppönen E, Läärä E, Reunanen A, Järvelin MR, Virtanen SM. Intake of vitamin D and risk of type 1 diabetes: a birth-cohort study. *Lancet* 2001; **358:** 1500–03.
4 Zipitis CS, Akobeng AK. Vitamin D supplementation in early childhood and risk of type 1 diabetes: a systematic review and meta-analysis. *Arch Dis Child* 2008; **93:** 512–17.
5 Wang TJ, Pencina MJ, Booth SL, et al. Vitamin D deficiency and risk of cardiovascular disease. *Circulation* 2008; **117:** 503–11.
6 Giovannucci E, Liu Y, Hollis BW, Rimm EB. 25-hydroxyvitamin D and risk of myocardial infarction in men: a prospective study. *Arch Intern Med* 2008; **168:** 1174–80.
7 Bischoff-Ferrari HA, Dawson-Hughes B, Staehelin HB, et al. Fall prevention with supplemental and active forms of vitamin D: a meta-analysis of randomised controlled trials. *BMJ* 2009; **339:** b3692.
8 Martinez ME, Giovannucci EL, Colditz GA, et al. Calcium, vitamin D, and the occurrence of colorectal cancer among women. *J Natl Cancer Inst* 1996; **88:** 1375–82.
9 Garland CF, Comstock GW, Garland FC, Helsing KJ, Shaw EK, Gorham ED. Serum 25-hydroxyvitamin D and colon cancer: eight-year prospective study. *Lancet* 1989; **334:** 1176–78.
10 John EM, Schwartz GG, Dreon DM, Koo J. Vitamin D and breast cancer risk: the NHANES I Epidemiologic follow-up study, 1971–1975 to 1992. National Health and Nutrition Examination Survey. *Cancer Epidemiol Biomarkers Prev* 1999; **8:** 399–406.
11 Autier P, Gandini S. Vitamin D supplementation and total mortality: a meta-analysis of randomized controlled trials. *Arch Intern Med* 2007; **167:** 1730–37.
12 Livshits G, Karasik D, Seibel MJ. Statistical genetic analysis of plasma levels of vitamin D: familial study. *Ann Hum Genet* 1999; **63:** 429–39.
13 Shea MK, Benjamin EJ, Dupuis J, et al. Genetic and non-genetic correlates of vitamins K and D. *Eur J Clin Nutr* 2009; **63:** 458–64.
14 Hunter D, De Lange M, Snieder H, et al. Genetic contribution to bone metabolism, calcium excretion, and vitamin D and parathyroid hormone regulation. *J Bone Miner Res* 2001; **16:** 371–78.
15 Lauridsen AL, Vestergaard P, Hermann AP, et al. Plasma concentrations of 25-hydroxy-vitamin D and 1,25-dihydroxy-vitamin D are related to the phenotype of Gc (vitamin D-binding protein): a cross-sectional study on 595 early postmenopausal women. *Calcif Tissue Int* 2005; **77:** 15–22.

16. Sinotte M, Diorio C, Berube S, Pollak M, Brisson J. Genetic polymorphisms of the vitamin D binding protein and plasma concentrations of 25-hydroxyvitamin D in premenopausal women. *Am J Clin Nutr* 2009; **89:** 634–40.
17. Engelman CD, Fingerlin TE, Langefeld CD, 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:** 3381–88.
18. Ramos-Lopez E, Bruck P, Jansen T, Herwig J, Badenhoop K. CYP2R1 (vitamin D 25-hydroxylase) gene is associated with susceptibility to type 1 diabetes and vitamin D levels in Germans. *Diabetes Metab Res Rev* 2007; **23:** 631–36.
19. Hyppönen E, Turner S, Cumberland P, Power C, Gibb I. Serum 25-hydroxyvitamin D measurement in a large population survey with statistical harmonization of assay variation to an international standard. *J Clin Endocrinol Metab* 2007; **92:** 4615–22.
20. Haughton MA, Mason RS. Immunonephelometric assay of vitamin D-binding protein. *Clin Chem* 1992; **38:** 1796–801.
21. Chen WM, Abecasis GR. Family-based association tests for genomewide association scans. *Am J Hum Genet* 2007; **81:** 913–26.
22. Skol AD, Scott LJ, Abecasis GR, Boehnke M. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nat Genet* 2006; **38:** 209–13.
23. Pe'er I, Yelensky R, Altshuler D, Daly MJ. Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. *Genet Epidemiol* 2008; **32:** 381–85.
24. Becker BJ, Wu M. The synthesis of regression slopes in meta-analysis. *Stat Sci* 2007; **3:** 414.
25. Tint GS, Irons M, Elias ER, et al. Defective cholesterol biosynthesis associated with the Smith-Lemli-Opitz syndrome. *N Engl J Med* 1994; **330:** 107–13.
26. Porter FD. Malformation syndromes due to inborn errors of cholesterol synthesis. *J Clin Invest* 2002; **110:** 715–24.
27. Rossi M, Federico G, Corso G, et al. Vitamin D status in patients affected by Smith-Lemli-Opitz syndrome. *J Inherit Metab Dis* 2005; **28:** 69–80.
28. Cheng JB, Levine MA, Bell NH, Mangelsdorf DJ, Russell DW. Genetic evidence that the human CYP2R1 enzyme is a key vitamin D 25-hydroxylase. *Proc Natl Acad Sci USA* 2004; **101:** 7711–15.
29. Speeckaert M, Huang G, Delanghe JR, Taes YE. Biological and clinical aspects of the vitamin D binding protein (Gc-globulin) and its polymorphism. *Clin Chim Acta* 2006; **372:** 33–42.
30. Ahn J, Albanes D, Berndt SI, et al. Vitamin D-related genes, serum vitamin D concentrations and prostate cancer risk. *Carcinogenesis* 2009; **30:** 769–76.
31. Janssens W, Bouillon R, Claes B, et al. Vitamin D deficiency is highly prevalent in COPD and correlates with variants in the vitamin D binding gene. *Thorax* 2010; **65:** 215–20.
32. Wjst M, Altmuller J, Braig C, Bahnweg M, Andre E. A genome-wide linkage scan for 25-OH-D(3) and 1,25-(OH)2-D3 serum levels in asthma families. *J Steroid Biochem Mol Biol* 2007; **103:** 799–802.
33. Heaney RP. Vitamin D and calcium interactions: functional outcomes. *Am J Clin Nutr* 2008; **88:** 541S–44S.
34. Lasky-Su J, Lyon HN, Emilsson V, et al. On the replication of genetic associations: timing can be everything! *Am J Hum Genet* 2008; **82:** 849–58.

Determinants of vitamin D status: focus on genetic variations

Diane Berry and Elina Hyppönen

Centre for Paediatric Epidemiology and Biostatistics and Medical Research Council Centre for the Epidemiology of Child Health, UCL Institute of Child Health, London, UK

Correspondence to Dr Elina Hyppönen, Centre for Paediatric Epidemiology and Biostatistics, Institute of Child Health, 30 Guilford Street, London WC1N 1EH, UK
Tel: +44 20 7905 2622; fax: +44 20 7905 2793; e-mail: e.hypponen@ich.ucl.ac.uk

Current Opinion in Nephrology and Hypertension 2011, 20:331–336

Purpose of review

The role of vitamin D beyond its importance for bone health is under much debate. In this article, we review recent evidence for genetic influences on 25-hydroxyvitamin D [25(OH)D] and discuss the uses of this information and its importance for public health.

Recent findings

Findings from large-scale genome-wide association meta-analyses on 25(OH)D confirmed the associations for loci nearby genes encoding vitamin D binding protein (GC, group component), 7-dehydrochlesterol reductase (*DHCR7*), 25-hydroxylase (*CYP2R1*) and 24-hydroxylase (*CYP24A1*), all influencing key sites for vitamin D metabolism. Findings from candidate gene studies have been inconsistent, with some implicating an association with 25(OH)D for loci near the gene encoding the hormonal vitamin D activation enzyme (*CYP27B1*).

Summary

The amount of variation in 25(OH)D explained by genetic determinants is small compared with environmental exposures. Information on genetic variants affecting 25(OH)D can be used as tools for Mendelian randomization analyses on vitamin D, and they provide some potential for the use as drug targets.

Keywords

genes, 25-hdyroxyvitamin D, metabolism, vitamin D

Curr Opin Nephrol Hypertens 20:331–336
© 2011 Wolters Kluwer Health | Lippincott Williams & Wilkins
1062-4821

Introduction

Much of the nutritional vitamin D status [measured by 25-hydroxyvitamin D, 25(OH)D] is determined by season or by other largely modifiable determinants which affect either the synthesis in the skin or dietary intake. However, research on the genetic influences on vitamin D status has also been active during the past couple of years. Heritability of 25(OH)D concentrations is believed to be considerable; however, there is wide variability in estimated rates (range from 29 to 80%) [1–3]. The first large genome-wide association studies (GWASs) on 25(OH)D were published in 2010, which together with an increasing number of candidate gene studies have provided important insights into the influence of the common genetic variations on vitamin D status. In this article, we review recent evidence for genetic influences on 25(OH)D and discuss the uses of this information and its importance for public health.

Vitamin D intake and metabolism

Vitamin D is mostly obtained through skin synthesis initiated by ultraviolet B (UVB) radiation exposure [4], and typically only smaller amounts are obtained through diet [5]. Due to the strong role of sunlight-induced skin synthesis, 25(OH)D demonstrates strong seasonal patterns, with the highest concentrations seen late summer and the lowest in late winter/early spring [6]. 25(OH)D concentrations are affected by latitude, time of day and climate [7]. There are ethnic differences in vitamin D status [8] and the more efficient vitamin D synthesis in individuals with lighter skin pigmentation is believed to be an adaptation resulting from migration to more northern and less sunny climates during evolution [9]. Oily fish is the most abundant natural dietary source, with eggs and some types of mushrooms containing smaller amounts of vitamin D [5]. In some countries, food (most commonly milk, breakfast cereals or margarine) is fortified with vitamin D, which can increase the importance of dietary intake. There are two major forms of vitamin D, vitamin D_2 (ergocalciferol, the plant form) and vitamin D_3 (cholecalciferol, synthesized in skin or obtained from animal-based sources in the diet). The bioavailability of D_3 is in some studies suggested to be better than that of D_2 [10,11]; however, the metabolism of these two forms is considered to be similar. Often a distinction between these two forms is not made in related literature, and also here for simplicity we will largely refer to them jointly as vitamin D.

UVB radiation from the sun induces vitamin D synthesis by converting 7-dehydrocholesterol (7-DHC, present in the skin) to vitamin D_3 (Fig. 1) [12]. Most of the vitamin

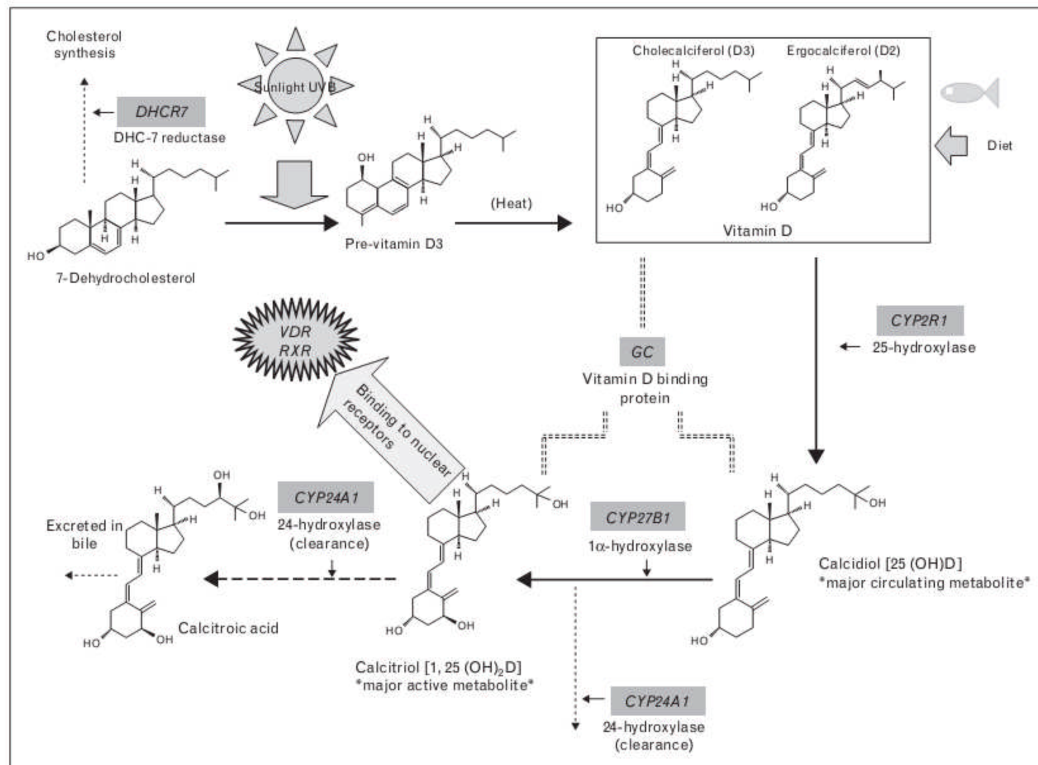
Copyright © Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.

D obtained from skin synthesis and through diet is transported to the liver for hydroxylation to 25(OH)D, but the 25-hydroxylation also occurs in some extrahepatic tissues [13]. Hydroxylation of vitamin D to 25(OH)D in the liver has little regulation [12] and circulating concentrations of 25(OH)D increase in proportion with vitamin D intake. A further hydroxylation of 25(OH)D in the kidneys (or in extrarenal tissues) leads to the active hormonal form of 1,25-dihydroxyvitamin D [1,25$(OH)_2$D, calcitriol] [12]. In the circulation, most vitamin D metabolites [including vitamin D, 25(OH)D and 1,25$(OH)_2$D] are transported bound to the vitamin D binding protein (VDBP, or GC-group component) [14,15]. Circulating concentrations of 1,25$(OH)_2$D are tightly regulated and the half-life of 1,25$(OH)_2$D is only 10–20 h [16]. In contrast, the half-life of 25(OH)D is about 15 days [16], and hence it is the preferred metabolite for measuring nutritional vitamin D status [12]. Excess concentrations of both 25(OH)D and 1,25$(OH)_2$D are cleared in a process starting with 24-hydroxylation, which leads to inactive metabolites and eventually extraction from the system [13]. The genomic actions of 1,25$(OH)_2$D are mediated through binding to nuclear vitamin D receptors (VDRs). The VDR forms

Key points

- Loci near genes influencing cholesterol synthesis, hydroxylation and vitamin D transport have been shown to influence 25(OH)D concentrations (vitamin D status) in genome-wide association studies.
- The known genetic loci account for a small amount of variation in 25(OH)D concentrations and exposure to sunlight together with other lifestyle factors remain as the key determinants of vitamin D status.
- Information on genetic variants affecting 25(OH)D will be invaluable as tools for Mendelian randomization analyses on vitamin D, and they provide some potential for use as drug targets.

Figure 1 Vitamin D metabolism and closely related genetic variations



CYP24A1, 24-hydroxylase; *CYP27B1*, vitamin D activation enzyme; *CYP2R1*, 25-hydroxylase; *DHCR7*, 7-dehydrocholesterol reductase; GC, group component; RXR, retinoid-X receptor; UVB, ultraviolet B; VDR, vitamin D receptors.

opyright © Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.

a heterodimer structure with the retinoid-X receptor (RXR) and binds to vitamin D response elements in the regulatory element region of the vitamin D target gene [17]. Over 2700 VDR-binding sites exist in the genome, which respond to the active 1,25$(OH)_2$D hormone, providing support for wide-ranging influences for hormonal vitamin D in human metabolism [18••].

Genetic influences on 25-hydroxyvitamin D

Recent GWASs on 25(OH)D have been published as the meta-analysis from consortia of several large studies, with key discoveries published in two independent meta-analyses, both on participants from European ancestry [19••,20••]. The first consisted of nine cohorts mainly from the USA, with 4501 participants in the discovery phase and an additional 2221 in the replication analysis [20••]. The larger SUNLIGHT (Study of Underlying Genetic Determinants of Vitamin D and Highly Related Traits) meta-analysis included in aggregate 14 different population cohorts, with 16 125 participants in the discovery and 17 871 in the replication stage [19••]. Both meta-analyses identified three single nucleotide polymorphisms (SNPs) [7-dehydrocholesterol reductase (*DHCR7*), 25-hydroxylase (*CYP2R1*) and GC], which were all in or near genes encoding well established steps in vitamin D metabolism or associated with its transport (Fig. 1). The SUNLIGHT meta-analysis also confirmed a variant affecting clearance of 25(OH)D (*CYP24A1*). In addition to the two large meta-analyses, a third smaller GWAS on a Hispanic population was published in 2010. However, this study was based on a very small discovery sample of 229 and found no associations at the GWAS significance level [21]. Alongside GWASs there has been an expansion in the number of candidate gene analyses on 25(OH)D, and all but *DHCR7* had been picked up as relevant for vitamin D in these type of studies.

DHCR7 expresses a reductase catalysing the conversion of 7-DHC to cholesterol (Fig. 1). Effectively, *DHCR7* removes the precholesterol from the vitamin D pathway reducing the substrate available for 25(OH)D. Mutations in *DHCR7* are known to lead to a rare Smith–Lemli–Opitz syndrome, in which impaired activity of the gene leads to accumulation of 7-DHC and a deficiency of cholesterol along with congenital abnormalities and intellectual disabilities [22]. Limited information exists on individuals with the syndrome and whether their vitamin D status is affected [23]. However, in earlier animal studies, the use of *DHCR7* inhibitors has led to increased 7-DHC and 25(OH)D concentrations [24,25].

CYP2R1 was one of the top hits in both GWA meta-analyses [19••,20••] and there was prior evidence for an association with 25(OH)D concentrations also from a candidate gene study [26]. *CYP2R1* is a member of the cytochrome P450 superfamily of enzymes, which are the catalysts of many reactions in the synthesis of cholesterol, lipids and steroids [27•]. GWAS findings confirmed the role of *CYP2R1* as the enzyme primarily responsible for the hydroxylation of vitamin D to 25(OH)D. Four cytochrome P450 forms had been considered candidates for the expression of 25-hydroxylase in the liver [27•,28]. However, out of these four, *CYP2R1* has the highest affinity and specificity to vitamin D [27•]. Furthermore, a genetic mutation in the *CYP2R1* is known to lead to vitamin D deficiency [29].

Loci from the gene region GC (group-specific component or Gc globulin), which encodes the VDBP, have been consistently associated with 25(OH)D concentrations in candidate gene studies [26,30•,31–34] and were the 'top hits' in the two genome-wide studies [19••,20••]. Most of the circulating 25(OH)D is bound by the protein (83–85%; high affinity), less by albumin (12–15%; low affinity) and 0.04% of 25(OH)D is free [35]. In an in-vitro study, genetic variations in GC were associated with 25(OH)D bioavailability to target cells notably monocytes, and the cells' subsequent response [36••]. No individuals have been found to be deplete in VDBP, but GC null type animals are viable [37,38]. Compared with the wild type, the GC null animals are more likely to develop secondary hyperparathyroidism and hypocalcaemia when subjected to the vitamin D-deficient diet, but they are also likely to be more resistant to hypercalcaemia and vitamin D toxicity [37]. Protein bound vitamin D metabolites have a longer half-life in circulation, as they are less susceptible to hydroxylation and degradation [14]. This suggests that an important function for VDBP is to stabilize and maintain circulating concentrations of 25(OH)D and other vitamin D metabolites [37]. There are three common isoforms of the VDBP, namely *GC1F*, *GC1S* and *GC2*. These protein isoforms can be based on the combination of alleles from SNPs rs4588 and rs7041 [14,32]. The common variants differ by amino acid substitutions and/or by the degree of glycosylation [39]. In populations with European ancestry, either for SNPs rs4588 (or rs2282679, a close proxy with $r^2 > 0.99$) and/or rs7041, the allele with a lower frequency is consistently associated with lower 25(OH)D concentrations [30•,32–34]. Concentrations of the circulating VDBP also tend to be lower with carriers of *GC2* compared to those with *GC1* [40]. There is some evidence for ethnic and geographical differences in the GC variants, suggesting that *GC1F* type is the more common isoform in people with dark skin compared with those with pale skin, whereas *GC1S* and *GC2* are more common in those with pale skin compared with dark [41]. This can be confirmed by the genetic information available for the Hapmap populations [42]. This suggests that compared to *GC2*, VDBP in people with *GC1* may have a higher affinity to vitamin D metabolites [14,36••], and, potentially, that there may be

Copyright © Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibitec

variations in the bioavailability of 25(OH)D between the different ethnicities.

A locus near the *CYP24A1* achieved the genome-wide significance threshold ($P < 5 \times 10^{-8}$) with 25(OH)D concentrations after combining the discovery and replication studies in the SUNLIGHT meta-analysis [19••]. The 24-hydroxylation is crucial in the regulation of the concentrations of the active 1,25(OH)$_2$D hormone in renal and in extrarenal sites. It is also known to degrade 25(OH)D into an inactive compound that can be excreted from the system [43]. The expression of *CYP24A1* has been detected at some level in most of the active vitamin D target cells and tissues [27•]. *CYP24A1* is currently considered as a drug target and inhibitors for the enzyme are under development [27•]. In the future, these might be used with individuals who have chronic kidney disease and who are susceptible to vitamin D deficiency [27•].

Other vitamin D-related genetic variations have been considered in candidate gene studies. A large population-based cohort with 6288 participants [44] has reported an association between the gene coding 1α-hydroxylase (*CYP27B1*) and 25(OH)D, and this finding has been supported by two other smaller studies on gestational diabetes [45] and multiple sclerosis [46]. It is biologically plausible for *CYP27B1* to influence 25(OH)D concentrations as it converts 25(OH)D to its active hormone, 1,25(OH)$_2$D. However, given the tight regulation of circulating 1,25(OH)$_2$D concentrations and 1α-hydroxylation, it is no surprise that *CYP27B1* did not reach genome-wide significance in the SUNLIGHT analysis [19••] or that the association with 25(OH)D in candidate gene studies has not been consistent [26,30•]. Rare mutations in *CYP27B1* prevent the renal hydroxylation of 25(OH)D and cause a deficiency in 1,25(OH)$_2$D, a disorder known as vitamin D-dependent rickets type 1 [47].

The VDR, mediating the hormonal actions of 1,25(OH)$_2$D, is the classical gene for vitamin D metabolism, and until recently most vitamin D relevant genetic studies were focused on VDR. There has been great interest in assessing VDR disease associations, and, among others, the recent meta-analyses suggest links between selected VDR variants and the risk of fractures [48], rheumatoid arthritis [49], tuberculosis [50] and the risk of various types of cancers [51,52,53•,54–56]. Given the distance of VDR from 25(OH)D in the metabolic pathway, strong associations between these two would not be expected. Typically, SNPs in the VDR have not been associated with 25(OH)D concentrations [26,57–59], although some evidence has been found for Fok-1 polymorphisms in studies on multiple sclerosis [46,60]. Additionally, there has been no association found for the heterodimer partner of VDR, RXR, with 25(OH)D [59].

Discussion

Recent GWASs have provided an important validation for what was already largely known about the key influences on vitamin D metabolism, confirming *CYP2R1* as the primary 25-hydroxylase and highlighting the role of *DHCR7*, the locus previously least well appreciated. Importantly for public health, these studies have shown that despite variation in 25(OH)D by common genetic loci, exposure to sunlight and other modifiable lifestyle-related factors remain as the key determinants of vitamin D status.

As demonstrated in this review, genetic influences on vitamin D status are likely to be complex. Studies on the strongest determinant, GC, suggest that genetic variations may lead to differences in the bioavailability of 25(OH)D and that these differences may be associated with ethnicity. Furthermore, GWASs and candidate gene studies alike have so far only investigated direct SNP effects on 25(OH)D, whereas it is possible that genetic factors might affect the response to variations in vitamin D intake or synthesis. These epistatic effects could explain part of ‘the missing heritability’ for 25(OH)D [61•]; that is the fact that the key common loci identified in the GWAS still only explain about 1–4% of the variation in 25(OH)D [19••], whereas heritability estimates from twin studies are markedly higher [1–3].

Despite accumulating evidence for an association between vitamin D deficiency and various diseases [5,62•], the recent Institute of Medicine (IOM) report on ‘dietary reference intakes for calcium and vitamin D’ concluded that evidence about these potential health benefits of vitamin D was from studies that could not be considered reliable [63••]. We believe that the recent genetic discoveries can help to address the apparent need for further evidence on the causal association between vitamin D and various disease end points. In so-called Mendelian randomization analyses, genetic variants that mimic the effects of an environmental exposure (and thereby provide a proxy estimate for lifelong differences in the exposure) are used to assess the causal relationship between the exposure and the health outcome of interest [64•]. In the context of vitamin D, variations especially in *DHCR1* and *CYP2R1* are likely to be helpful as they affect the first stage of vitamin D metabolism [prior to the hydroxylation of 25(OH)D], and we can quantify the effect that these genetic variations have on 25(OH)D and use this information to predict their associations with the outcome. If these genetic variants (*DHCR1*/*CYP2R1*) are associated with the health outcome of interest [in proportion to their estimated effect on 25(OH)D], this would then provide strong evidence for a casual relationship between vitamin D and the outcome [65,66]. There are various challenges with this approach, including the

opyright © Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.

need of very large sample sizes [67]; however, the various large global collaborative efforts are likely to prove very helpful in enabling this type of study in the future.

In conclusion, despite undisputable genetic influences on vitamin D status, modifiable environmental determinants have a key role in defeating vitamin D deficiency. The role of genetic factors is only partially understood, and more work is needed to establish gene–environment interactions and the potential for ethnic variations in the response to and requirement for vitamin D. Through the use as drug targets and as instruments in Mendelian randomization analysis, recent discoveries about vitamin D-associated genes are likely to offer insights into both causality and treatment of various diseases.

Acknowledgements

The review was funded by the UK Medical Research Council (grant G0601653). E.H. holds a Department of Health (UK) Public Health Career Scientist Award, with further support from the British Heart Foundation, Biotechnology and Biological Sciences Research Council (UK), and the Academy of Finland. This work was undertaken at the Centre for Paediatric Epidemiology and Biostatistics, which benefits from funding support from the MRC in its capacity as the MRC Centre of Epidemiology for Child Health. Research at the University College London Institute of Child Health and Great Ormond Street Hospital for Children National Health Service (NHS) Trust benefits from R&D funding received from the NHS Executive.

There are no conflicts of interest.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

Additional references related to this topic can also be found in the Current World Literature section in this issue (p. 433).

1 Shea MK, Benjamin EJ, Dupuis J, *et al.* Genetic and nongenetic correlates of vitamins K and D. Eur J Clin Nutr 2009; 63:458–464.

2 Hunter D, De Lange M, Snieder H, *et al.* Genetic contribution to bone metabolism, calcium excretion, and vitamin D and parathyroid hormone regulation. J Bone Miner Res 2001; 16:371–378.

3 Wjst M, Altmuller J, Braig C, *et al.* A genome-wide linkage scan for 25-OH-D(3) and 1,25-(OH)2-D3 serum levels in asthma families. J Steroid Biochem Mol Biol 2007; 103:799–802.

4 Webb AR. Who, what, where and when-influences on cutaneous vitamin D synthesis. Prog Biophys Mol Biol 2006; 92:17–25.

5 Holick MF. Vitamin D deficiency. N Engl J Med 2007; 357:266–281.

6 Hypponen E, Power C. Hypovitaminosis D in British adults at age 45 y: nationwide cohort study of dietary and lifestyle predictors. Am J Clin Nutr 2007; 85:860–868.

7 Holick MF, Chen TC, Lu Z, *et al.* Vitamin D and skin physiology: a D-lightful story. J Bone Miner Res 2007; 22 (Suppl 2):V28–V33.

8 Looker AC, Dawson-Hughes B, Calvo MS, *et al.* Serum 25-hydroxyvitamin D status of adolescents and adults in two seasonal subpopulations from NHANES III. Bone 2002; 30:771–777.

9 Jablonski NG, Chaplin G. Colloquium paper: human skin pigmentation as an adaptation to UV radiation. Proc Natl Acad Sci U S A 2010; 107 (Suppl 2):8962–8968.

10 Trang HM, Cole DE, Rubin LA, *et al.* Evidence that vitamin D3 increases serum 25-hydroxyvitamin D more efficiently than does vitamin D2. Am J Clin Nutr 1998; 68:854–858.

11 Armas LA, Hollis BW, Heaney RP. Vitamin D2 is much less effective than vitamin D3 in humans. J Clin Endocrinol Metab 2004; 89:5387–5391.

12 Horst R, Reinhardt T, Satyanarayana Reddy G. Vitamin D metabolism. In: Feldman D, Glorieux FH, Pike JW, editors. Vitamin D. 2nd ed. Amsterdam: Elsevier; 2005. pp. 15–36.

13 DeLuca HF. Evolution of our understanding of vitamin D. Nutr Rev 2008; 66:S73–S87.

14 Speeckaert M, Huang G, Delanghe JR, *et al.* Biological and clinical aspects of the vitamin D binding protein (Gc-globulin) and its polymorphism. Clin Chim Acta 2006; 372:33–42.

15 Daiger SP, Schanfield MS, Cavalli-Sforza LL. Group-specific component (Gc) proteins bind vitamin D and 25-hydroxyvitamin D. Proc Natl Acad Sci U S A 1975; 72:2076–2080.

16 Jones G. Pharmacokinetics of vitamin D toxicity. Am J Clin Nutr 2008; 88:582S–586S.

17 Pike JW, Shevde NK. The vitamin D receptor. In: Feldman D, Glorieux FH, Pike JW, editors. Vitamin D. 2nd ed. Amsterdam: Elsevier; 2005. pp. 167–192.

18 Ramagopalan SV, Heger A, Berlanga AJ, *et al.* A ChIP-seq defined genome-
•• wide map of vitamin D receptor binding: associations with disease and evolution. Genome Res 2010; 20:1352–1360.
This article evaluated the degree of VDR binding throughout the human genome, with evidence for enrichment in genomic regions near autoimmune and cancer-associated genes.

19 Wang TJ, Zhang F, Richards JB, *et al.* Common genetic determinants of
•• vitamin D insufficiency: a genome-wide association study. Lancet 2010; 376:180–188.
The largest GWAS to date with some 34 000 participants of European ancestry, identifying four common SNPs in the vitamin D pathway with strong evidence for association with 25(OH)D and the risk of vitamin D insufficiency.

20 Ahn J, Yu K, Stolzenberg-Solomon R, *et al.* Genome-wide association study of
•• circulating vitamin D levels. Hum Mol Genet 2010; 19:2739–2745.
A GWAS on 25(OH)D identifying three SNPs in the vitamin D pathway, including the DHCR7, which was previously considered relatively unimportant.

21 Engelman CD, Meyers KJ, Ziegler JT, *et al.* Genome-wide association study of vitamin D concentrations in Hispanic Americans: the IRAS family study. J Steroid Biochem Mol Biol 2010; 122:186–192.

22 Tint GS, Irons M, Elias ER, *et al.* Defective cholesterol biosynthesis associated with the Smith-Lemli-Opitz syndrome. N Engl J Med 1994; 330:107–113.

23 Rossi M, Federico G, Corso G, *et al.* Vitamin D status in patients affected by Smith-Lemli-Opitz syndrome. J Inherit Metab Dis 2005; 28:69–80.

24 Bonjour JP, Trechsel U, Granzer E, *et al.* The increase in skin 7-dehydrocholesterol induced by an hypocholesterolemic agent is associated with elevated 25-hydroxyvitamin D3 plasma level. Pflugers Arch 1987; 410:165–168.

25 Morris JG. Ineffective vitamin D synthesis in cats is reversed by an inhibitor of 7-dehydrocholestrol-delta7-reductase. J Nutr 1999; 129:903–908.

26 Bu FX, Armas L, Lappe J, *et al.* Comprehensive association analysis of nine candidate genes with serum 25-hydroxy vitamin D levels among healthy Caucasian subjects. Hum Genet 2010; 128:549–556.

27 Schuster I. Cytochromes P450 are essential players in the vitamin D signaling
• system. Biochim Biophys Acta 2010; 1814:186–199.
This is a good review on cytochrome enzymes that are believed to contribute to the vitamin D metabolism pathway.

28 Gascon-Barre M. The vitamin D 25-hydroxylase. In: Feldman D, Glorieux FH, Pike JW, editors. Vitamin D. 2nd ed. Amsterdam: Elsevier Academic Press; 2005. pp. 47–68.

29 Cheng JB, Levine MA, Bell NH, *et al.* Genetic evidence that the human CYP2R1 enzyme is a key vitamin D 25-hydroxylase. Proc Natl Acad Sci U S A 2004; 101:7711–7715.

30 Ahn J, Albanes D, Berndt SI, *et al.* Vitamin D-related genes, serum vitamin D
• concentrations, and prostate cancer risk. Carcinogenesis 2009; 30:769–776.
A comprehensive candidate gene study, evaluating the association between 48 SNPs in the vitamin D pathway and 25(OH)D concentrations, it confirmed the strong association between GC and 25(OH)D.

31 Fang Y, van Meurs JB, Arp P, *et al.* Vitamin D binding protein genotype and osteoporosis. Calcif Tissue Int 2009; 85:85–93.

32 Fu L, Yun F, Oczak M, *et al.* Common genetic variants of the vitamin D binding protein (DBP) predict differences in response of serum 25-hydroxyvitamin D [25(OH)D] to vitamin D supplementation. Clin Biochem 2009; 42:1174–1177.

33 Janssens W, Bouillon R, Claes B, *et al.* Vitamin D deficiency is highly prevalent in COPD and correlates with variants in the vitamin D-binding gene. Thorax 2010; 65:215–220.

34 Sinotte M, Diorio C, Berube S, *et al.* Genetic polymorphisms of the vitamin D binding protein and plasma concentrations of 25-hydroxyvitamin D in premenopausal women. Am J Clin Nutr 2009; 89:634–640.

Copyright © Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.

35 Bikle DD, Gee E, Halloran B, *et al.* Assessment of the free fraction of 25-hydroxyvitamin D in serum and its regulation by albumin and the vitamin D-binding protein. J Clin Endocrinol Metab 1986; 63:954–959.

36 Chun RF, Lauridsen AL, Suon L, *et al.* Vitamin D-binding protein directs
•• monocyte responses to 25-hydroxy- and 1,25-dihydroxyvitamin D. J Clin Endocrinol Metab 2010; 95:3368–3376.
An interesting in-vitro study, suggesting differences in bioavailability of 25(OH)D and 1,25$(OH)_2$D by VDBP polymorphisms, and an influence on vitamin D-dependent antimicrobial responses.

37 Safadi FF, Thornton P, Magiera H, *et al.* Osteopathy and resistance to vitamin D toxicity in mice null for vitamin D binding protein. J Clin Invest 1999; 103:239–251.

38 Zella LA, Shevde NK, Hollis BW, *et al.* Vitamin D-binding protein influences total circulating levels of 1,25-dihydroxyvitamin D3 but does not directly modulate the bioactive levels of the hormone in vivo. Endocrinology 2008; 149:3656–3667.

39 Braun A, Bichlmaier R, Cleve H. Molecular analysis of the gene for the human vitamin-D-binding protein (group-specific component): allelic differences of the common genetic GC types. Hum Genet 1992; 89:401–406.

40 Lauridsen AL, Vestergaard P, Nexo E. Mean serum concentration of vitamin D-binding protein (Gc globulin) is related to the Gc phenotype in women. Clin Chem 2001; 47:753–756.

41 Kamboh MI, Ferrell RE. Ethnic variation in vitamin D-binding protein (GC): a review of isoelectric focusing studies in human populations. Hum Genet 1986; 72:281–293.

42 The International HapMap Project. Nature 2003; 426:789–796.

43 Omdahl JL, May B. The 25-hydroxyvitamin D 24-hydroxylase. In: Feldman D, Glorieux FH, Pike JW, editors. Vitamin D. 2nd ed. Amsterdam: Elsevier Academic Press; 2005. pp. 85–104.

44 Hypponen E, Berry DJ, Wjst M, *et al.* Serum 25-hydroxyvitamin D and IgE: a significant but nonlinear relationship. Allergy 2009; 64:613–620.

45 Ramos-Lopez E, Kahles H, Weber S, *et al.* Gestational diabetes mellitus and vitamin D deficiency: genetic contribution of CYP27B1 and CYP2R1 polymorphisms. Diabetes Obes Metab 2008; 10:683–685.

46 Orton SM, Morris AP, Herrera BM, *et al.* Evidence for genetic regulation of vitamin D status in twins with multiple sclerosis. Am J Clin Nutr 2008; 88:441–447.

47 Glorieux FH, St Arnaud R. Vitamin D pseudodeficiency. In: Feldman D, Glorieux FH, Pike JW, editors. Vitamin D. 2nd ed. Amsterdam: Elsevier Academic Press; 2005. pp. 1197–1198.

48 Ji GR, Yao M, Sun CY, *et al.* BsmI, TaqI, ApaI and FokI polymorphisms in the vitamin D receptor (VDR) gene and risk of fracture in Caucasians: a meta-analysis. Bone 2010; 47:681–686.

49 Lee YH, Bae SC, Choi SJ, *et al.* Associations between vitamin D receptor polymorphisms and susceptibility to rheumatoid arthritis and systemic lupus erythematosus: a meta-analysis. Mol Biol Rep 2010 [Epub ahead of print].

50 Gao L, Tao Y, Zhang L, *et al.* Vitamin D receptor genetic polymorphisms and tuberculosis: updated systematic review and meta-analysis. Int J Tuberc Lung Dis 2010; 14:15–23.

51 Chen L, Davey Smith G, Evans DM, *et al.* Genetic variants in the vitamin D receptor are associated with advanced prostate cancer at diagnosis: findings from the prostate testing for cancer and treatment study and a systematic review. Cancer Epidemiol Biomarkers Prev 2009; 18:2874–2881.

52 Yin M, Wei S, Wei Q. Vitamin D receptor genetic polymorphisms and prostate cancer risk: a meta-analysis of 36 published studies. Int J Clin Exp Med 2009; 2:159–175.

53 Randerson-Moor JA, Taylor JC, Elliott F, *et al.* Vitamin D receptor gene
• polymorphisms, serum 25-hydroxyvitamin D levels, and melanoma: UK case-control comparisons and a meta-analysis of published VDR data. Eur J Cancer 2009; 45:3271–3281.
A meta-analysis for the association between VDR polymorphisms with a type of cancer that is of great interest. This demonstrates the usefulness of genetic variation in overcoming environmental confounders (in this context sunlight exposure) when evaluating associations between vitamin D and disease risk.

54 Tang C, Chen N, Wu M, *et al.* Fok1 polymorphism of vitamin D receptor gene contributes to breast cancer susceptibility: a meta-analysis. Breast Cancer Res Treat 2009; 117:391–399.

55 Raimondi S, Johansson H, Maisonneuve P, *et al.* Review and meta-analysis on vitamin D receptor polymorphisms and cancer risk. Carcinogenesis 2009; 30:1170–1180.

56 Kostner K, Denzer N, Muller CS, *et al.* The relevance of vitamin D receptor (VDR) gene polymorphisms for cancer: a review of the literature. Anticancer Res 2009; 29:3511–3536.

57 Arslan AA, Clendenen TV, Koenig KL, *et al.* Circulating vitamin D and risk of epithelial ovarian cancer. J Oncol 2009; 2009:672492.

58 Balsa JA, Iglesias B, Peromingo R, *et al.* Vitamin D receptor polymorphisms in secondary hyperparathyroidism after scopinaro's biliopancreatic diversion. Obes Surg 2010; 20:1415–1421.

59 Hibler EA, Jurutka PW, Egan JB, *et al.* Association between polymorphic variation in VDR and RXRA and circulating levels of vitamin D metabolites. J Steroid Biochem Mol Biol 2010; 121:438–441.

60 Smolders J, Damoiseaux J, Menheere P, *et al.* Fok-I vitamin D receptor gene polymorphism (rs10735810) and vitamin D metabolism in multiple sclerosis. J Neuroimmunol 2009; 207:117–121.

61 Manolio TA, Collins FS, Cox NJ, *et al.* Finding the missing heritability of
• complex diseases. Nature 2009; 461:747–753.
The potential future directions of genetic research beyond GWASs.

62 Norman AW, Bouillon R. Vitamin D nutritional policy needs a vision for the
• future. Exp Biol Med (Maywood) 2010; 235:1034–1045.
This is a nice overview of the role of vitamin D in health.

63 Institute of Medicine. Dietary reference intakes for calcium and vitamin D. In:
•• Ross AC, Taylor CL, Yaktine AL, Del Valle HB, editors. Washington, DC: The National Academies Press; 2011.
The authoritative IOM report on reference intakes, confirming the importance of vitamin D for bone health, but calling for stronger evidence to establish the link with other health outcomes.

64 Lawlor DA, Harbord RM, Sterne JA, *et al.* Mendelian randomization: using
• genes as instruments for making causal inferences in epidemiology. Stat Med 2008; 27:1133–1163.
This is an excellent outline for the concept of Mendelian randomization with an example application.

65 Bochud M, Rousson V. Usefulness of Mendelian randomization in observational epidemiology. Int J Environ Res Public Health 2010; 7:711–728.

66 Sheehan NA, Meng S, Didelez V. Mendelian randomisation: a tool for assessing causality in observational epidemiology. Methods Mol Biol 2011; 713:153–166.

67 Palmer TM, Lawlor DA, Harbord RM, *et al.* Using multiple genetic variants as instrumental variables for modifiable risk factors. Stat Methods Med Res 2011 [Epub ahead of print].

opyright © Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.

OPEN ACCESS Freely available online

PLoS one

Evaluation of Genetic Markers as Instruments for Mendelian Randomization Studies on Vitamin D

Diane J. Berry[1¶], Karani S. Vimaleswaran[1¶], John C. Whittaker[2], Aroon D. Hingorani[3], Elina Hyppönen[1*]

1 Centre for Paediatric Epidemiology and Biostatistics and MRC Centre for the Epidemiology of Child Health, UCL Institute of Child Health, London, United Kingdom, **2** Department of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London, United Kingdom, **3** Genetic Epidemiology Group, Department of Epidemiology and Public Health and the Centre for Clinical Pharmacology, University College London, London, United Kingdom

Abstract

Background: Mendelian randomization (MR) studies use genetic variants mimicking the influence of a modifiable exposure to assess and quantify a causal association with an outcome, with an aim to avoid problems with confounding and reverse causality affecting other types of observational studies.

Aim: We evaluated genetic markers that index differences in 25-hydroxyvitamin D (25(OH)D) as instruments for MR studies on vitamin D.

Methods and Findings: We used data from up-to 6,877 participants in the 1958 British birth cohort with information on genetic markers and 25(OH)D. As potential instruments, we selected 20 single nucleotide polymorphisms (SNP) which are located in the vitamin D metabolism pathway or affect skin pigmentation/tanning, including 4 SNPs from genome-wide association (GWA) meta-analyses on 25(OH)D. We analyzed SNP associations with 25(OH)D and evaluated the use of allele scores dividing genes to those affecting 25(OH)D synthesis (*DHCR7*, *CYP2R1*) and metabolism (*GC*, *CYP24A1*, *CYP27B1*). In addition to the GWA SNPs, only two SNPs (*CYP27B1*, *OCA2*) showed evidence for association with 25(OH)D, with the *OCA2* association abolished after lifestyle adjustment. Per allele differences varied between −0.02 and −0.08 nmol/L ($P \leq 0.02$ for all), with a 6.1 nmol/L and a 10.2 nmol/L difference in 25(OH)D between individuals with highest compared lowest number of risk alleles in synthesis and metabolism allele scores, respectively. Individual SNPs but not allele scores showed associations with lifestyle factors. An exception was geographical region which was associated with synthesis score. Illustrative power calculations (80% power, 5% alpha) suggest that approximately 80,000 participants are required to establish a causal effect of vitamin D on blood pressure using the synthesis allele score.

Conclusions: Combining SNPs into allele scores provides a more powerful instrument for MR analysis than a single SNP in isolation. Population stratification and the potential for pleiotropic effects need to be considered in MR studies on vitamin D.

Citation: Berry DJ, Vimaleswaran KS, Whittaker JC, Hingorani AD, Hyppönen E (2012) Evaluation of Genetic Markers as Instruments for Mendelian Randomization Studies on Vitamin D. PLoS ONE 7(5): e37465. doi:10.1371/journal.pone.0037465

Editor: Ana Paula Arez, Instituto de Higiene e Medicina Tropical, Portugal

Received January 3, 2012; **Accepted** April 20, 2012; **Published** May 21, 2012

Copyright: © 2012 Berry et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This study was supported by grants from the British Heart Foundation and the UK Medical Research Council [project grant G0601653, PrevMetSyn/SALVE with Academy of Finland]. EH is a Department of Health (UK) Public Health Career Scientist. The 25-hydroxyvitamin D assay in the 1958 British birth cohort was funded by the BUPA foundation. Genotype data from the 1958 British birth cohort DNA collection was funded by MRC grant G0000934 and Wellcome Trust grant 068545/Z/02. This research used resources provided by the Type 1 Diabetes Genetics Consortium, a collaborative clinical study sponsored by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), National Institute of Allergy and Infectious Diseases, National Human Genome Research Institute, National Institute of Child Health and Human Development, and Juvenile Diabetes Research Foundation International (JDRF) and supported by U01 DK062418. This study also makes use of data generated by the Wellcome Trust Case-Control Consortium. A full list of the investigators who contributed to the generation of the data is available from www.wtccc.org.uk. Funding for the project was provided by the Wellcome Trust under award 076113. The work was undertaken at the Centre for Paediatric Epidemiology and Biostatistics which benefits from funding support from the MRC in its capacity as the MRC Centre of Epidemiology for Child Health. Research at the University College London Institute of Child Health and Great Ormond Street Hospital for Children NHS Trust benefits from R&D funding received from the NHS Executive. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: DJB, KSV, ADH, EH have declared that no competing interests exist. JW is the Director of Statistical Genetics at GlaxoSmithKline. The company had no role in the production of the paper, development or interpretation of the study questions or in the decision to submit for publication. This does not alter the authors' adherence to all the PLoS ONE policies on sharing data and materials.

* E-mail: e.hypponen@ucl.ac.uk

¶ These authors contributed equally to this work.

Introduction

There has been much interest in the potential effects of vitamin D on a wide range of health outcomes, and vitamin D deficiency has been suggested to predispose to common chronic diseases such as diabetes, cancer and cardiovascular disease [1–6]. Effects on bone health are undisputed and severe vitamin D deficiency leads to rickets in children and osteomalacia in adults [7]. However, the recent report by the Institute of Medicine on vitamin D, concluded that "with the exception of measures related to bone health, the potential indicators examined are currently not supported by evidence that could be judged either convincing or adequate in

terms of cause and effect, or informative regarding dose–response relationships for determining nutrient requirements." [8]. This reflects the fact that much of the evidence has been obtained from observational studies, with only a handful of randomized controlled trials evaluating the effects of vitamin D supplementation.

Observational associations of 25-hydroxyvitamin D concentration (25(OH)D, a marker for nutritional vitamin D status) and adverse health outcomes may provide evidence of a causal link, but could also arise from limitations with this type of study. For example, studies on vitamin D may be prone to confounding, as status is associated with risk factors for chronic diseases such as obesity [9]. Reverse causality may be a problem as given the strong role of sunlight induced skin synthesis in vitamin D production, lower concentrations of 25(OH)D could be a consequence of less time spent outdoors and hence, caused by an underlying disease rather than being the cause of it. Mendelian randomization (MR) analysis refers to the use of genetic variants that index the exposure of interest (in this case vitamin D intake/status) to gain insight on whether the relationship between an exposure and disease is causal [10,11]. If lower vitamin D status is causally related to an adverse outcome (e.g. cardiovascular disease), a genetic variant associated with lower 25(OH)D concentration should be associated with a higher risk of cardiovascular disease (in relation to its effect on 25(OH)D). The genetic association, unlike the directly observed association of vitamin D intake/status itself, will be less prone to confounding (as the genotype is assigned at random) and free from reverse causation since the genotype is not modifiable by disease [12].

Nevertheless, the MR approach has some limitations [12–14]. An assumption of MR analysis is that the effect of a genetic variant on an outcome functions only via the intermediate exposure, such as a lifestyle factor or biomarker. However, a genetic variant may result in multiple biological alterations (pleiotropy). Hence, if these alterations also independently affect relevant outcomes not via of the intermediate phenotype, this may lead to associations through a mechanism that does not involve the exposure of interest [10]. Genetic confounding may also result in violation of MR assumptions if closely located genetic variants are inherited together (i.e. in the presence of linkage disequilibrium) and affect the outcome outside of the mechanism of the exposure, or if there is evidence for population stratification [10].

In this study, we used information from the large nationwide survey of 1958 British birth cohort to evaluate genetic markers for the use as instruments in MR studies on vitamin D.

Methods

Ethics statement

Ethical approval for the biomedical survey was given by the South-East Multi-Centre Research Ethics Committee. Written informed consent for the use of information in medical studies was obtained from the participants.

Participants

Detailed description of the study has been published previously [15]. In brief, participants are from the 1958 British birth cohort (1958BC), initially including all births in England, Scotland or Wales during one week in March 1958 (n = 17,638) [15]. Between September 2002 and June 2004, 11,971 participants still residing in Britain were invited to participate in a biomedical survey. At the time of data collection participants were aged 44–46 years old, and 9,377 (78%) completed at least one questionnaire and 7,591 (81% of the respondents) had valid 25(OH)D measures. The 1958BC is almost entirely a white European population (98%) [16], and since we utilize genetic data for these analyses, 158 individuals of other ethnic groups and one pregnant participant were excluded. Data was further restricted by availability of genetic data with 4,572–6,877 individuals included in the single nucleotide polymorphism (SNP) analyses. Analyses using the vitamin D allele score were done in participants with full data on 25(OH)D synthesis (n = 5,623) and/or metabolism (n = 5,856) markers (for full description, please see below).

Measurement of biochemical and clinical parameters

Serum 25(OH)D concentrations were measured using an automated IDS OCTEIA enzyme-linked immunosorbent assay (ELISA) (Dade-Behring BEP2000 analyzer), standardized according to the mean from Vitamin D External Quality Assessment Scheme (DEQAS) [17]. Serum IGF-1 concentration was measured using the Nichols Advantage IGF-1 chemiluminescence immunoassay (referenced against World Health Organization 1st International Reference Reagent 1988; IGF-1 87/518) and glycosylated hemoglobin (HbA1C) with high-performance liquid chromatography (as certified by the US National Glycohemoglobin Standardization Program (NGSP) [18]). Triglycerides and total and high-density lipoprotein (HDL) cholesterol were measured by standard autoanalyzer methodology and low-density lipoprotein (LDL) cholesterol was calculated using the Friedewald formula. Fibrinogen was determined by the Clauss method, D-dimer by ELISA assay and C-reactive protein (CRP) was assayed by nephelometry (Dade Behring). Von Willebrand factor (vWF) antigen was measured by Decollates ELISA and tissue plasminogen activator (tPA) antigen by Biopool elisa. Total IgE was assayed using the HYTEC automated enzyme immunoassay [19].

Weight, height and waist circumference were measured at 45 years of age. Blood pressure was determined as an average of three repeated measures (Omron 705CP automated sphygmomanometer). For forced expiratory volume 1 (FEV_1) and forced vital capacity (FVC), the highest technically satisfactory values (three repeated measures) were used [20].

Social, dietary and lifestyle factors

Socioeconomic position was assessed using the Registrar General's occupational classification categorized as I and II (managerial and professional), III (non-manual), III (manual), and IV and V (manual unskilled) [21]. Individuals who were institutionalized, retired, or long-term unemployed were classified separately. Physical activity was determined as recreation Metabolic Equivalent of Task (MET) hours, derived from reported frequencies and usual durations for up to 37 activities [22]. Smoking was recorded as never/ex-smoker vs current smoker based on smoking history recorded at ages 23, 33, and 42 years. Frequency and amounts of alcohol consumption were reported at 45 years. Information on current geographical region of residence was based on Government Office Regions, and categorized as South (South East, South West, and Greater London), Middle (East Anglia, Midlands, and Wales), North (North, North West, and Yorkshire and the Humber), and Scotland. Geographical region was dichotomized when used as an outcome (South/Middle vs North/Scotland).

The following factors (measured at 45 years) were also considered and dichotomized: time spent watching a television/using a PC (coded as <1 h vs ≥ 1 h), time spent outside (coded as <1 h vs ≥ 1 h), protecting skin in the sun (often/sometimes vs rarely/never), oily fish consumption (≥ 3 days/week vs. 3 days/week), use of vitamin D supplement, and season of blood drawn (winter/spring vs. summer/autumn).

Selection of candidate genes and SNPs

We selected 20 SNPs from 12 genes, which are involved in the vitamin D pathway or affect skin pigmentation or the ability to tan [23–26]. Four of the SNPs were identified as hits in the recently published genome-wide association (GWA) meta-analysis for vitamin D insufficiency (n~34,000, including the 1958 British birth cohort) [23] (**Figure 1**). The GWA *GC* SNP rs2282679 was in high linkage disequilibrium with the candidate *GC* SNP rs4588 ($r^2 = 0.98$) and hence due to higher numbers with data available (n = 6,551 vs. 5,224), we have chosen rs4588 as a proxy for the SNP rs2282679 in the present study. Two of the SNPs were identified as candidates based on the available evidence from the literature pertaining to their potential roles in vitamin D metabolic pathway (**Figure 1**). Fourteen SNPs were chosen based on the GWA and candidate studies for skin colour/tanning. A complete list of the selected genes/SNPs is shown in the **Table S1**.

Genotyping

The SNPs (rs4588, rs12785878, rs10741657, rs6013897, rs10877012, rs17470271, rs7495174, rs4778241, rs4778138, rs13289, rs1805005, rs2228479) were genotyped using the Taqman platform (Applied Biosystems, Warrington, UK) [27]. The custom genotype SNPs had passed the inclusion criteria (Hardy-Weinberg Equilibrium (HWE) *P* value >0.01 [28], MAF>0.01 and call rate >80%) (**Table S1**). The remaining SNPs were genotyped on the platforms of Affymetrix 6.0 and Illumina 550 K Infinium by the two sub-studies of WTCCC2 and TIDGC and QC procedures were applied accordingly [29,30]. Imputation of SNPs genotyped by the WTCCC2 and TIDGC was done using the software Impute [31] (as outlined for the 1958BC in the GWAS on 25(OH)D [23]). In the TIDGC sub-study, two SNPs (rs16891982, rs1805007) were imputed, and in the WTCCC2, three SNPs (rs16891982, rs11648785, rs464349) were imputed (a call rate threshold of 0.9 was used for the imputed SNPs).

Statistical methods

Natural log transformation was used for 25(OH)D and the biomarkers (except for lung function) to improve the approximation of the normal distribution. Distributions were assessed before and after transformation using quantile-normal plots. Variation in continuous outcomes was evaluated by linear regression and, in dichotomous outcomes, by logistic regression with the p-values from Wald tests.

The four SNPs in genes with confirmed associations with 25(OH)D (*GC*; rs4588, *CYP2R1*; rs10741657, *DHCR7*; rs12785878, *CYP24A1*; rs6013897) [23] were considered as possible instruments by default, and taken forward to subsequent analyses. For the other genes, we tested their associations with 25(OH)D using linear regression, taking forward all SNPs with a p-value below the Bonferroni corrected threshold of <0.007 (≤0.05/number of candidate genes) to control for multiple testing. The number of candidate genes (rather than number of SNPs) was used in Bonferroni correction to account for moderate linkage disequilibrium between SNPs. To indicate the strength of the SNP as instrument we included the F-statistic from a simple linear regression model with 25(OH)D. The F-statistic in a simple linear regression model is derived from the proportion of the variation explained by the genetic variant in the phenotype given the sample size [32]. As a rule of thumb an F-statistic less than 10 is taken to indicate a weak instrument [33]. Formal MR analysis often uses instrumental variable (IV) regression by two-stage least squares estimator, however this may introduce a bias if sample size is small and there is too much variability in the estimated association between the SNP and intermediate phenotype [34]. The relative bias of IV analysis compared to ordinary least squares (linear regression) can be approximated as the inverse of the strength of the instrument (1/F-statistic) [34]. Interactions between the SNPs were tested by including the interaction terms in the linear regression model on 25(OH)D adjusting for sex with p-values corrected by the number of vitamin D SNPs to account for multiple testing (0.05/5, *P*<0.01).

Relatively small proportion of participants had missing data on confounders/covariates (15% with at least one missing value) after restriction by the availability of genetic information. All main analyses requiring confounder adjustments were run with complete information. To investigate whether the results were sensitive to missing information in covariates, multivariate imputation by chained equations was used to impute missing values [35] and the main analyses were re-run. The results were identical whether based on models run on complete data or on data obtained by multiple imputation. To assess whether the SNP associations with

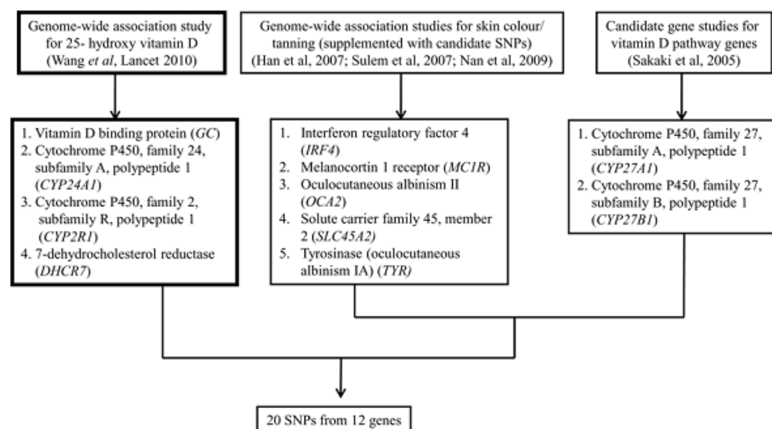


Figure 1. Strategy for SNP selection using genome-wide association and candidate gene studies.
doi:10.1371/journal.pone.0037465.g001

25(OH)D were confounded, we adjusted for lifestyle and social factors in the linear regression model examining the associations of the SNPs with 25(OH)D (namely time spent outside, sun cover, oily fish consumption, vitamin D supplements, season, smoking, alcohol consumption, PC/TV time, recreational metabolic equivalent task (MET) hours, social class, body mass index (BMI), abdominal obesity, geographical region, and sex). We also analysed the direct associations between the SNPs and social, dietary and lifestyle factors using logistic regression. Interactions between the SNPs and lifestyle/social factors were evaluated by including appropriate interaction terms in the model. In these analyses a Bonferroni correction was used to correct for the number of factors tested, with the p-value threshold determined as 0.05/14 ($P<0.004$).

To investigate whether the relationship between the SNPs and 25(OH)D was due to other health factors and differing genetic pathways ("genetic confounding"), we adjusted for the available biomarkers (von Willebrand factor, tPA, D-dimer, fibrinogen, CRP, IgE, triglycerides, low density lipoproteins, high density lipoproteins and total cholesterol, FEV, diastolic and systolic blood pressures, IGF1 and HbA1c) in the linear regression model examining the genetic marker association with 25(OH)D. We also investigated the direct SNP- biomarker associations adjusted for 25(OH)D, with the assumption that if pleiotropy did exist these associations should appear fairly strong, and not be affected by 25(OH)D adjustment. Interactions between 25(OH)D and SNPs with biomarkers were also tested. Bonferroni corrected *P*-value for these analyses was <0.003 (0.05/15, where the denominator is the number of biomarker tests for each SNP).

"Synthesis score" was created using the two SNPs in genes encoding proteins involved in 25(OH)D synthesis (*DHCR7* and *CYP2R1*), both of which had been identified through the GWA meta-analyses on 25(OH)D [23]. Three SNPs in genes encoding proteins involved in 25(OH)D metabolism (*GC*, *CYP24A1* and *CYP27B1*) were included in the "metabolism score", with analyses repeated only including the GWA confirmed SNPs ("metabolism^{GWA} score", *GC* and *CYP24A1*) (**Figure 2**). The allele score was created by counting the number of vitamin D lowering alleles. For alleles scores based on metabolizing genes (*GC*, *CYP24A1* and *CYP27B1*), there was only one subject with six lowering alleles, and 36 subjects with five lowering alleles, these groups were combined with those who had four lowering alleles. Likewise for alleles in genes involved in 25(OH)D synthesis (*CYP2R1* and *DHCR7*) score, only 97 subjects had four lowering alleles, so this group was combined with those who had three lowering alleles. As described above, we examined the associations of allelic score indicators with potential confounders and disease-relevant biomarkers.

We used simulation to carry out illustrative power calculations for the association between vitamin D and systolic blood pressure using single SNPs, separate allele scores and allele scores together as instruments (**Appendix S1**). We assumed effect sizes as observed in the 1958BC, notably a 5% reduction in blood pressure by each 10 nmol/l increase in 25(OH)D, and the observed SNPs/scores effects on 25(OH)D. Data was simulated 1,000 times for a given sample size and two-staged least squares regression was run. The parameter of interest was tested and power was estimated from the proportion of times the test *p*-value was less than the significant level $\alpha = 0.05$. Analyses were carried out using STATA, version 11 [36].

Results

The 4 SNPs that had been chosen on the basis of the 25(OH)D GWAS (in *GC*, *CYP2R1*, *DHCR7*, *CYP24A1*) [23] were associated with 25(OH)D in the 1958BC ($P \leq 0.016$). Of remaining 16 SNPS, one SNP in *OCA2* (rs7495174) had a significant association with 25(OH)D after Bonferroni correction ($P=0.002$) (**Table 1**). These five SNPs were taken forward for further evaluation together with one SNP (*CYP27B1*, rs10877012, $P=0.008$) (**Figure 3**) that fell slightly below the significance threshold, but had previous evidence for replication [37]. There was no evidence for SNP-SNP interactions between any of the six SNPs (*P* for all comparisons ≥0.08, data not presented).

We next examined whether the associations with 25(OH)D observed for these six SNPs, were sensitive to adjustment for geographical region, dietary and lifestyle factors or available biomarkers (**Figure 4**). The missing data for the lifestyle and dietary factors, and biomarkers ranged from 0.2% (BMI) to 7.7% (time spent outside). Most associations between the SNPs and 25(OH)D concentrations were not affected by these adjustments, however, an exception was *OCA2*, as its association was no longer present after adjustment for sex, season, geographical region, time spent outside, sun cover, oily fish consumption, vitamin D supplements, smoking, alcohol consumption, PC/TV time, recreational MET hours, social class at 42 yrs, and BMI (beta −0.06, $P=0.007$ for unadjusted vs. −0.02, $P=0.27$ adjusted, **Figure 4**). Due to the lack of independent association with 25(OH)D, *OCA2* was considered as unsuitable for the use as a

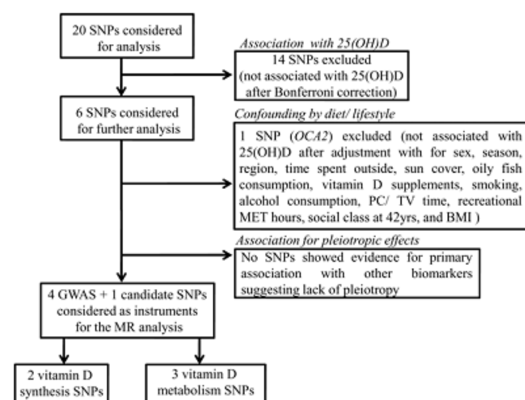


Figure 2. Genetic variation in the vitamin D synthesis and metabolic pathway. Skin exposure to ultraviolet B (UVB) radiation initiates the conversion of 7-dehydrocholesterol to previtamin D3. 7-dehydrocholesterol reductase (***DHCR7***) encodes the enzyme 7-dehydrocholesterol reductase, which converts 7-dehydrocholesterol to cholesterol, thereby removing the substrate from the synthetic pathway of vitamin D3. The previtamin D3 in turn gets converted to vitamin D3 in a heat-dependent process. Vitamin D (represents D2 or D3) is transported to the liver, where it is converted by vitamin D-25-hydroxylase (***CYP2R1***) to 25-hydroxyvitamin D [25(OH)D]. This is the major circulating form of vitamin D that is used by clinicians to determine vitamin D status. This form of vitamin D is biologically inactive; it is bound to the vitamin D-binding protein (***GC***), transported to the kidneys and converted by 25-hydroxyvitamin D-1α- hydroxylase (1-OHase) (***CYP27B1***) to the biologically active form 1,25-dihydroxyvitamin D3 (Calcitriol). Calcitriol increases the expression of 25-hydroxyvitamin D-24- hydroxylase (24-OHase) (***CYP24A1***) to catabolise 25(OH)D to the water-soluble, biologically inactive calcitroic acid, which is excreted in the bile. *DHCR7* and *CYP2R1* function upstream of the production of 25(OH)D and hence, termed as 25(OH)D synthesis indicators, while *GC*, *CYP27B1* and *CYP24A1* function downstream of the 25(OH)D production and hence, termed as 25(OH)D metabolism indicators.
doi:10.1371/journal.pone.0037465.g002

Table 1. Association of SNP with ln 25-hydroxyvitamin D adjusted for sex.

Gene	SNP	*n*	MAF	Beta for minor allele	SE	*P* value	% of variance explained	F-statistic	Relative bias†, %
GWA on 25(OH)D									
GC	rs4588	6027	0.30	−0.08	0.009	**1.48×10^{-17}**	1.18	73.4	1.4
DHCR7/NADSYN1	rs12785878	6504	0.22	−0.05	0.009	**1.2×10^{-6}**	0.35	22.4	4.4
CYP2R1	rs10741657	5968	0.40	0.03	0.009	**0.0003**	0.21	13.3	8.1
CYP24A1	rs6013897	6534	0.20	−0.03	0.010	0.016	0.07	5.4	21.1
Vitamin D pathway genes									
CYP27B1	rs10877012	6877	0.33	−0.02	0.008	**0.008**	0.09	7.2	16.8
CYP27A1	rs17470271	5831	0.44	0.005	0.009	0.55	-	0.4	-
GWA on skin colour/tanning									
OCA2	rs7495174	5013	0.06	−0.06	0.02	**0.002**	0.16	8.5	12.2
OCA2	rs4778241	4961	0.17	−0.01	0.01	0.39	-	0.6	-
OCA2	rs4778138	5036	0.12	−0.02	0.01	0.13	0.03	1.9	75.8
OCA2	rs12913832	4989	0.22	0.003	0.01	0.76	-	0.1	-
SLC45A2	rs13289	5039	0.38	−0.02	0.009	0.013	0.10	6.5	19.3
SLC45A2	rs16891982	4843	0.02	−0.03	0.04	0.48	-	0.5	-
MC1R	rs11648785	4816	0.30	−0.02	0.01	0.037	0.07	4.0	30.0
MC1R	rs1805005	5091	0.12	0.02	0.01	0.14	0.02	2.0	81.6
MC1R	rs464349	4966	0.46	−0.01	0.01	0.25	0.01	1.5	-
MC1R	rs2228479	5084	0.10	0.007	0.02	0.65	-	0.1	-
MC1R	rs1805007	4572	0.10	−0.003	0.02	0.87	-	0.0	-
IRF4	rs12203592	5184	0.22	−0.01	0.01	0.20	0.01	1.6	-
IRF4	rs12210050	4988	0.20	−0.01	0.01	0.58	-	0.3	-
TYR	rs1393350	4992	0.29	0.01	0.01	0.48	-	0.5	-

MAF, minor allele frequency.
†Relative bias has not been estimated where the SNP has an F-statistic less than 1.9.
doi:10.1371/journal.pone.0037465.t001

proxy of 25(OH)D concentrations and removed from subsequent analyses. For the remaining five SNPs, we found no significant interactions by dietary and lifestyle factors on their influence on 25(OH)D ($P>0.08$ for all comparisons, data not presented). We also observed no evidence for pleiotropic effects for these SNPs, as no strong associations were observed with any of the available biomarkers with adjustment for 25(OH)D concentrations (**Table S2**). Furthermore, 25(OH)D did not modify the associations observed between the SNPs and the biomarkers (*P* interaction for all comparisons ≥ 0.07).

The combined 25(OH)D synthesis score (including SNPs in *DHCR7* and *CYP2R1*), explained 0.56% of the variation of 25(OH)D concentrations and there was a 6.1 nmol/l difference in 25(OH)D between top and bottom groups in the allelic score (**Table 2**). The allele score based on 25(OH)D metabolism GWAS SNPs (*GC* and *CYP24A1*) explained 1.04% of the variation of 25(OH)D concentrations and there was a 7.9 nmol/l difference in 25(OH)D between top and bottom groups in the allelic score, while the metabolism score also including *CYP27B1* explained 1.12% of the variation of 25(OH)D concentrations, with a 10.2 nmol/l difference between top and bottom quartile categories. As seen in **Figure 4**, associations between the allele scores and 25(OH)D were unaffected by adjustment for lifestyle and dietary factors or biomarkers. There were some associations between individual SNPs and dietary and lifestyle factors, notably rs10741657 from *CYP2R1* which was associated with social class ($P=0.003$), and *GC* SNP rs4588 which was associated with oily fish consumption ($P=0.002$), while allele scores were not associated with any of the lifestyle factors (**Figure 5**). However, there was an association between 25(OH)D synthesis allele score and geographical region even after applying Bonferroni correction (for the number of dietary and lifestyle factors tested), which reflected the strong association between *DHCR7* SNP rs12785878 with geographical region ($P=3.0\times10^{-5}$). As with the single variants, we observed no evidence for pleiotropic effects as the allele scores were not associated with the biomarkers after adjustment for 25(OH)D (**Table S2**).

Illustrative power calculations for a MR study on blood pressure using single SNPs, allele scores (synthesis, metabolism, metabolism^{GWA}), and allele scores combined as instruments for vitamin D status are presented in **Figure 6A** & **6B**. Allele scores had greater power to detect an effect than the individual SNPs used in the scores. For example, to achieve 80% power using the synthesis allele score as instrument would require a sample size of 76,000 individuals, whilst using *CYP2R1* alone would require 155,000 individuals.

Discussion

Insufficient intake of vitamin D has been proposed to affect up to 50% of the UK population [38,39]. This is a potentially a very important public health issue because vitamin D (intake and/or deficiency) has been linked to several common diseases including cancer, diabetes, and cardiovascular diseases [40]. However, there

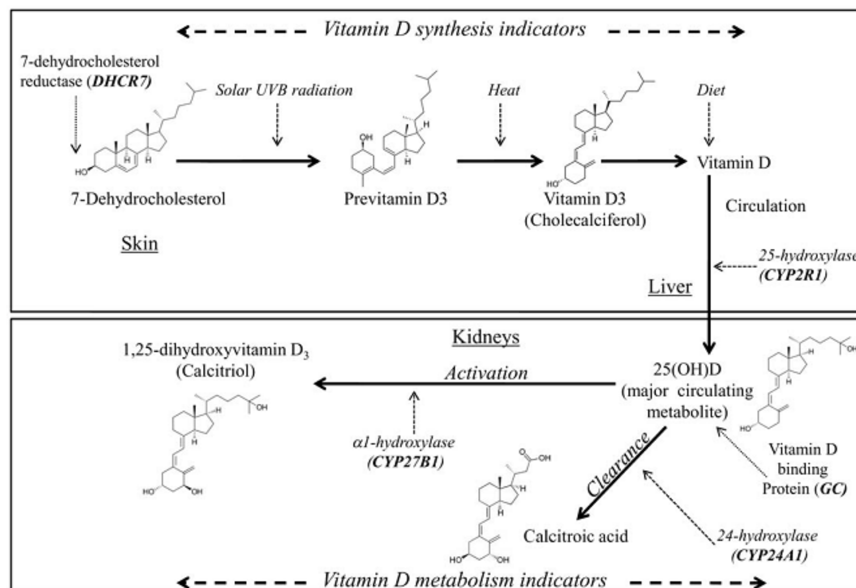


Figure 3. The selection of vitamin D SNPs for the use as instruments in Mendelian Randomization (MR) analysis.
doi:10.1371/journal.pone.0037465.g003

is an urgent need to improve the evidence base for causal relations of vitamin D, as much of the evidence still relies on observational studies where it is difficult to disentangle causation from association. Randomized controlled trials are clearly the gold standard for demonstrating causality, however, they are expensive and time consuming. In this paper, we describe methodological work related to a complementary approach for causal analysis; notably, the rationale for identification and process to evaluate 25(OH)D associated SNPs as tools for future MR analysis of vitamin D.

We identified five SNPs either affecting vitamin D synthesis or metabolism as plausible instruments for MR analyses. We observed some residual co-variation by lifestyle/dietary factors in relation to selected instruments, for example the *GC* SNP was associated with oily fish consumption. The *GC* SNP was also marginally stronger in terms of F-statistic and variation of 25(OH)D concentrations explained than the combined unweighted allelic scores. Our analyses suggested that use of combined allelic scores reduced confounding since the scores were not associated with lifestyle/dietary factors. The associations between selected SNPs and geographical region, and to some extent with lifestyle/dietary factors, may also indicate a potential issue with regional variation/ancestry. Hence, it is important to consider population stratification in the context of MR studies on vitamin D, and to correct for geographical region/population stratification in related analyses. As genome-wide data becomes more readily available, one approach might be to correct for population stratification using principal components [41]. However, further work is required to demonstrate the extent to which genes expressed in the vitamin D pathway are under genetic selection related to geographical region.

The five genes used in the allelic scores have an important role in the vitamin D metabolic pathway. *DHCR7* and *CYP2R1* function upstream of the 25(OH)D production (synthesis), while *GC*, *CYP24A1* and *CYP27B1* function downstream of the 25(OH)D production (metabolism) (**Figure 2**). *CYP2R1* encodes the enzyme that catalyzes the 25-hydroxylation step in the liver leading to the synthesis of 25(OH)D [42], and as such, it is conceptually the best instrument for MR studies on vitamin D. The SNP included in these analyses is not functional, but it was chosen as it showed the strongest association with 25(OH)D in the published GWAs meta-analyses [23]. In the same study, *DHCR7* was identified as a novel locus for association with vitamin D status [23]. *DHCR7* encodes the enzyme 7-dehydrocholesterol reductase, which converts 7-dehydrocholesterol (7-DHC) to cholesterol, removing the compound from the pathway of vitamin D and onto becoming 25(OH)D. In our study, *DHCR7* was not associated with cholesterol (or other lipid makers) and there was no evidence for effect modification by it on the association of 25(OH)D with cholesterol biomarkers. Also the large GWA meta-analyses on lipid traits failed to identify *DHCR7* as a genetic influence on cholesterol [43], suggesting it has a primary role for vitamin D rather than cholesterol metabolism. Nevertheless, given the possibility for pleiotropic associations, MR studies using *DHCR7* to index 25(OH)D should be interpreted with caution if the suggested association is not also seen for *CYP2R1*.

The strongest of the metabolism markers was vitamin D binding protein (*DBP*), also known as group specific component (Gc), which is involved in the transport of vitamin D and its metabolites [44]. The enzyme encoded by the *CYP24A1* gene plays a crucial role in calcium homeostasis and the vitamin D endocrine system, acting at the first stage of 25(OH)D and 1,25(OH)2D catabolism [45]. *CYP27B1* gene is a well-known candidate for vitamin D pathway [45], as it encodes 1α-hydroxylase, the enzyme that converts 25(OH)D into 1α-25(OH)$_2$D (the active hormonal form). Although our findings showed an association of the *CYP27B1* SNP (rs10877012) with 25(OH)D levels just outside the Bonferroni corrected $P<0.008$, we also evaluated rs10877012 as a component

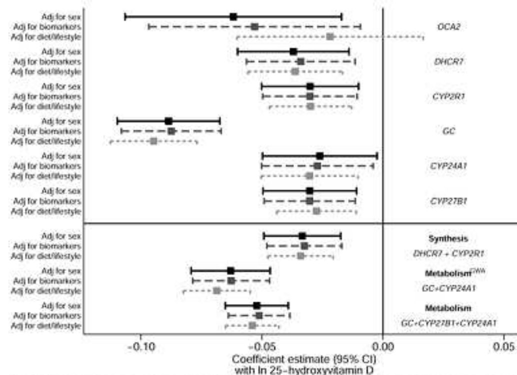


Figure 4. Association between the SNPs, synthesis, metabolism and metabolism^{GWA} allele scores and ln 25(OH)D with and without adjustment for biomarkers, dietary and lifestyle indicators. The bars are the 95% CI. *Biomarkers*: coagulation markers- von Willebrand factor, tPA and D-dimer; Inflammatory markers- fibrinogen and CRP; Lipid marker- Triglycerides, low density lipoproteins, high density lipoproteins and total cholesterol; Lung function marker- FEV; Cardiovascular disease related factors- diastolic and systolic blood pressures, IgE, IGF1 and HbA1c). *Dietary and lifestyle markers*: time spent outside, sun cover, oily fish consumption, vitamin D supplements, season, smoking, alcohol consumption, PC/TV time, recreational MET hours, social class, body mass index, abdominal obesity and geographical region.
doi:10.1371/journal.pone.0037465.g004

of the allele score analysis based on the previous evidence for replication [37]. Conversely, a candidate gene (*OCA2*) based on the association with skin coloration was included in our initial investigations. However, associations between *OCA2* and 25(OH)D have not consistently been observed [27]. In our study evidence for an association between the *OCA2* genotype and 25(OH)D concentrations was abolished by adjustment for lifestyle and social indicators, suggesting that this marker (despite biological plausibility) is not suitable for the use as an instrument in MR studies on vitamin D.

The position of the target gene in the metabolic vitamin D pathway affects the quantification of the expected direction of the SNP-25(OH)D association, which led us to divide the SNPs into those affecting synthesis and metabolism when creating combined allele score indicators. The ability to estimate the magnitude of a possible causal effect for an environmental exposure that can be improved is a key strength with the MR approach [11]. Conceptually, the quantification of the 25(OH)D association for the synthesis markers appears quite straightforward as *DHCR7* and *CYP2R1* contribute to the production of 25(OH)D. Associations are likely to be more complex for the metabolism markers which are involved in the clearance or transport of 25(OH)D (and other vitamin D metabolites). The magnitude of the association between metabolism SNPs and 25(OH)D may depend upon current vitamin D status and requirement, and is likely to be under the influence of (unmeasured, potentially unknown) metabolic feed-back loops. For example, there is evidence to indicate that *GC*, the key 25(OH)D carrier protein, is an important determinant of the bioavailability of vitamin D metabolites to key target cells such as

Table 2. Association of Allele Scores with ln 25-hydroxyvitamin D concentrations adjusted for sex.

	N	Geometric Mean (95% CI)	Beta	SE	*P* value	% of variance explained	F-statistic	Relative Bias, %
*Synthesis score**								
0	587	56.2 (54.0, 58.5)	Reference	Reference	-	-	-	
1	2025	55.1 (54.0, 56.2)	−0.02	0.02	-	-	-	
2	2276	53.0 (52.1, 54.0)	−0.06	0.02	-	-	-	
3,4	968	50.2 (48.8, 51.6)	−0.11	0.02	-	-	-	
Synthesis score, per allele	5856	-	−0.04	0.007	6.1×10^{-9}	0.56	33.4	3.1
Metabolism^{GWA} score†								
0	1902	56.7 (55.5, 57.9)	Reference	Reference	-	-		
1	2489	53.2 (52.3, 54.2)	−0.06	0.01	-	-		
2	1305	50.3 (49.1, 51.5)	−0.12	0.02	-	-		
3, 4	240	48.8 (46.0, 51.7)	−0.15	0.03	-	-		
Metabolism^{GWA} score, per allele	5936	-	−0.06	0.007	1.8×10^{-15}	1.04	63.1	1.6
Metabolism score‡								
0	818	57.8 (56.0, 59.6)	Reference	Reference	-	-		
1	1836	55.6 (54.4, 56.8)	−0.04	0.02	-	-		
2	1767	52.6 (51.5, 53.7)	−0.09	0.02	-	-		
3	925	51.0 (49.5, 52.5)	−0.13	0.02	-	-		
4, 5, 6	277	47.6 (45.2, 50.2)	−0.19	0.03	-	-		
Metabolism score, per allele	5623	-	−0.05	0.006	8.8×10^{-16}	1.12	64.6	1.6

*Synthesis SNPs include *DHCR7* and *CYP2R1*.
†Metabolism^{GWA} SNPs include *GC* and *CYP24A1*.
‡Metabolism SNPs include *GC*, *CYP24A1* and *CYP27B1*.
doi:10.1371/journal.pone.0037465.t002

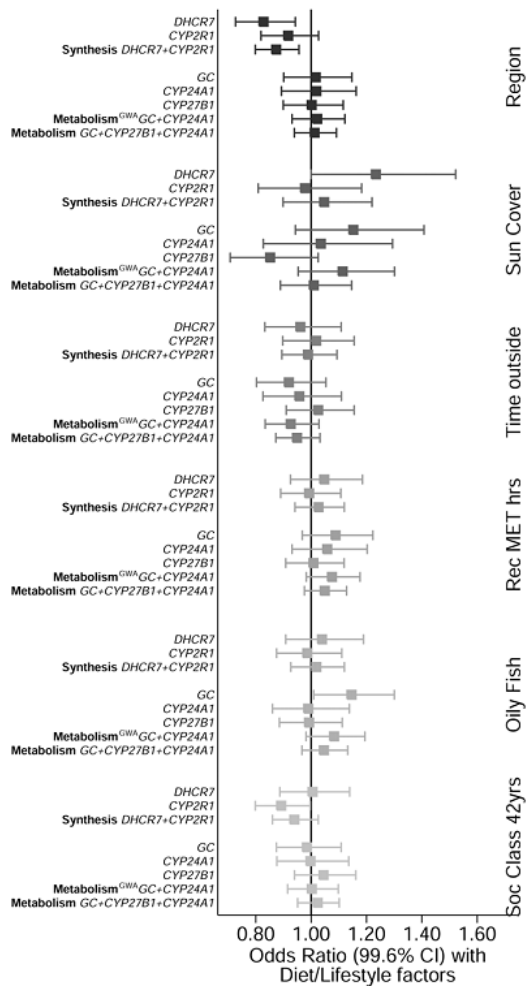


Figure 5. Associations of the five SNPs and allele scores with geographical region, social, dietary and lifestyle factors. The bars are the 99.6% CI. The effects of the allele scores and the individual SNPs for each lifestyle factor can be identified based on the intensity of the coloured boxes.

doi:10.1371/journal.pone.0037465.g005

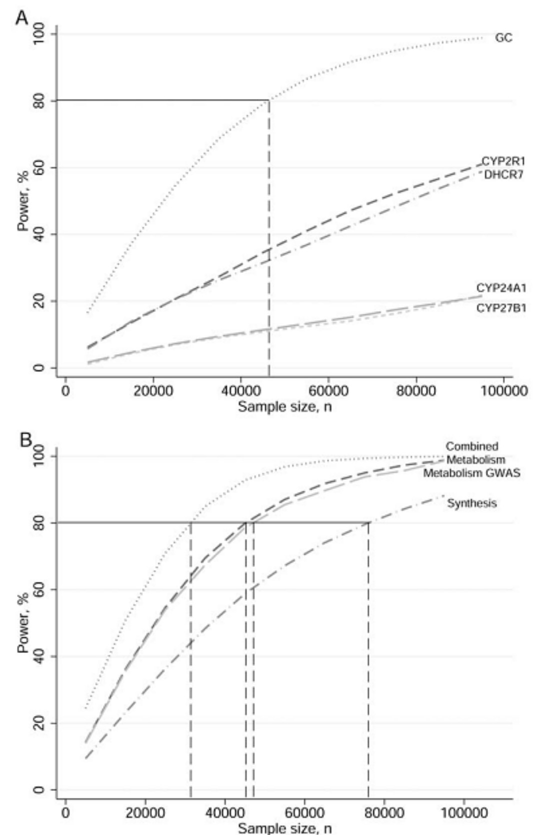


Figure 6. Power and sample size to detect the 5% decrease in blood pressure by 10 nmol/l increase in 25(OH)D observed in the 1958 British birth cohort using genetic proxy indicators (significance level $\alpha = 0.05$). The curves in (**A**) from the bottom to the top of the graph are in the order of min effect size with *CYP27B1* (short dash), *CYP24A1* (long dash), *DHCR7* (dash dot), *CYP2R1* (dash), *GC* (dot). The curves in (**B**) from the bottom to the top of the graph are in the order of min effect size with Synthesis score (dash dot), Metabolism^{GWA} score (long dash) Metabolism score (dash), both scores (dot). The horizontal black line and attached vertical dashed lines indicate the sample size required for a study with 80% power using the genetic proxy.

doi:10.1371/journal.pone.0037465.g006

monocytes [46], which will influence the use of related genetic variants as instruments in MR analyses. Separating the SNPs by function into the two allele scores gives flexibility to the subsequent MR models, providing the ability to use them singularly as one instrument or together as two instruments, whilst still accounting for the complex associations with 25(OH)D.

An important strength with MR approach is that the limitations for this method are by and large independent of those typical for other types of observational studies. However, as we have shown in this study, there remains a possibility of residual confounding when using single SNPs as genetic proxy markers. The single most important potential confounder was geographical region, which also had a borderline association with the allele score based on 25(OH)D synthesis SNPs. These analyses highlight the importance of considering population structure/regional variations in MR studies for vitamin D, which given the strong influence of sunlight induced synthesis on serum concentrations, may be particularly vulnerable to this source of confounding.

MR relies on the assumption that genetic variants used as instruments are uncorrelated with other variants that are associated with the outcome outside of the exposure pathway [12]. This assumption may be violated due to the presence of linkage disequilibrium, where variants which are located close to each other are inherited together. It is also possible that the synthesis or metabolism SNPs could have led to biological adaptations during the development (i.e. canalisation) [14]. It is

also possible that SNPs used as instruments could have pleiotropic effects where they influence other metabolic pathways independently of the influence on 25(OH)D concentrations. In this study, we found no evidence for strong associations between the SNPs of interest with a range of biomarkers, suggesting specificity for their association with 25(OH)D. An important methodological limitation for MR analyses, including those done in the context of vitamin D, relates to the requirement of very large sample sizes. In the illustrative power calculations included in this paper, we showed how even by a combination of two SNPs into an allele score we were able to half the sample size required. In the MR analysis as modelled by IV regression using a two-stage-least-squares estimator, the association can be biased when both the variance explained by the instruments (in the intermediate phenotype) and the sample size are small [33,34]. As indicated by the large F-statistics for all allele scores, none of the composite instruments were deemed to be "weak" in our study [33]. However, given the small amount of variation in 25(OH)D explained by these genetic variants, it is clear that the application of MR analyses in the context of vitamin D is resource intensive and as shown here, ~80,000 participants would be required to detect expected influences on blood pressure using the synthesis SNPs. If both allele scores are combined, the sample size requirement is reduced to ~40,000 individuals. However, as noted above, quantification of the association for the metabolism SNPs is difficult, and related power calculations will not be correct if there are variations in 25(OH)D bioavailability by GC genotypes as has been suggested [46].

There is great promise in the use of genetic variants as instruments for modifiable exposures, given their ability to avoid some of the limitations of observational epidemiology in making causal inferences. At a public health level the benefits with the approach are evident, as MR studies can be used to imply reductions in disease risk that can be achieved by improving vitamin D status (which in turn, can be done for example by the use of vitamin D supplementation). However, in this paper we demonstrate the complexities of using MR in the context of vitamin D research, most notably the requirement of very large samples, possibility for pleiotropic effects, and the potential of confounding by population stratification. Informative MR studies on vitamin D are likely to be feasible in the context of large international consortia, with the issues on population structure duly considered at the analysis stage. However, within that type of context, MR is highly likely to serve as a useful first-stage approach to testing causality between vitamin D and various health outcomes.

Supporting Information

Table S1 Call rates and *P* values for Tests of Hardy-Weinberg Equilibrium for Vitamin D Polymorphisms Identified from Candidate Gene and Genome-wide Association Studies.
(DOC)

Table S2 The SNP association with biomarkers adjusted for 25(OH)D, sex and region.
(DOC)

Appendix S1 Sample size calculations.
(DOC)

Acknowledgments

We are deeply grateful to the participants of the 1958 British birth cohort. We thank Professor Chris Power (UCL Institute of Child Health) for support in vitamin D studies; Professor Ian Gibb, Dr Steve Turner, and Marie-Claude Fawcett (Royal Victoria Infirmary, Newcastle-upon-Tyne, U.K.) for carrying out the laboratory assays; and the Centre for Longitudinal studies, Institute of Education (original data producers) for providing the data. Professor John Todd, Dr Jason Cooper and Debbie Smyth are acknowledged for genotyping (Juvenile Diabetes Research Foundation/Wellcome Trust Diabetes and Inflammation Laboratory, Cambridge Institute for Medical Research, University of Cambridge, UK).

Author Contributions

Conceived and designed the experiments: EH. Performed the experiments: DJB KSV EH. Analyzed the data: DJB. Contributed reagents/materials/analysis tools: DJB KSV JCW ADH EH. Wrote the paper: KSV EH DJB ADH. Supervision of the statistical analysis: EH JCW ADH.

References

1. Wortsman J, Matsuoka LY, Chen TC, Lu Z, Holick MF (2000) Decreased bioavailability of vitamin D in obesity. Am J Clin Nutr 72: 690–693.
2. Hyppönen E, Laara E, Reunanen A, Jarvelin MR, Virtanen SM (2001) Intake of vitamin D and risk of type 1 diabetes: a birth-cohort study. Lancet 358: 1500–1503.
3. Lappe JM, Travers-Gustafson D, Davies KM, Recker RR, Heaney RP (2007) Vitamin D and calcium supplementation reduces cancer risk: results of a randomized trial. Am J Clin Nutr 85: 1586–1591.
4. Lee W, Kang PM (2010) Vitamin D deficiency and cardiovascular disease: Is there a role for vitamin D therapy in heart failure? Curr Opin Investig Drugs 11: 309–314.
5. Pittas AG, Chung M, Trikalinos T, Mitri J, Brendel M, et al. (2010) Systematic review: Vitamin D and cardiometabolic outcomes. Ann Intern Med 152: 307–314.
6. Liu E, Meigs JB, Pittas AG, Economos CD, McKeown NM, et al. (2010) Predicted 25-hydroxyvitamin D score and incident type 2 diabetes in the Framingham Offspring Study. Am J Clin Nutr 91: 1627–1633.
7. Stern PH (2005) Bone. In: Feldman D, Glorieux FH, Pike JW, eds. Vitamin D. Amsterdam: Elsevier Academic Press. pp 565–573.
8. Institute of Medicine (2011) Dietary Reference Intakes for Calcium and Vitamin D.
9. Hyppönen E, Power C (2006) Vitamin D status and glucose homeostasis in the 1958 British birth cohort: the role of obesity. Diabetes Care 29: 2244–2246.
10. Sheehan NA, Didelez V, Burton PR, Tobin MD (2008) Mendelian randomisation and causal inference in observational epidemiology. PLoS Med 5: e177.
11. Davey Smith G, Ebrahim S (2005) What can mendelian randomisation tell us about modifiable behavioural and environmental exposures? BMJ 330: 1076–1079.
12. Lawlor DA, Harbord RM, Sterne JA, Timpson N, Davey SG (2008) Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. Stat Med 27: 1133–1163.
13. Nitsch D, Molokhia M, Smeeth L, DeStavola BL, Whittaker JC, et al. (2006) Limits to causal inference based on Mendelian randomization: a comparison with randomized controlled trials. Am J Epidemiol 163: 397–403.
14. Davey Smith G, Ebrahim S (2003) Mendelian randomization: can genetic epidemiology contribute to understanding environmental determinants of disease? Int J Epidemiol 32: 1–22.
15. Power C, Elliott J (2006) Cohort profile: 1958 British birth cohort (National Child Development Study). Int J Epidemiol 35: 34–41.
16. Atherton K, Fuller E, Shepherd P, Strachan DP, Power C (2008) Loss and representativeness in a biomedical survey at age 45 years: 1958 British birth cohort. J Epidemiol Community Health 62: 216–223.
17. Hyppönen E, Turner S, Cumberland P, Power C, Gibb I (2007) Serum 25-hydroxyvitamin D measurement in a large population survey with statistical harmonization of assay variation to an international standard. J Clin Endocrinol Metab 92: 4615–4622.
18. Gibb I, Parnham A, Fonfrede M, Lecock F (1999) Multicenter evaluation of Tosoh glycohemoglobin analyzer. Clin Chem 45: 1833–1841.
19. Nolte H, DuBuske LM (1997) Performance characteristics of a new automated enzyme immunoassay for the measurement of allergen-specific IgE. Summary of the probability outcomes comparing results of allergen skin testing to results obtained with the HYTEC system and CAP system. Ann Allergy Asthma Immunol 79: 27–34.
20. American Thoracic Society (1987) Standardization of spirometry–1987 update. Am Rev Respir Dis 136: 1285–1298.

21. Power C, Atherton K, Strachan DP, Shepherd P, Fuller E, et al. (2007) Life-course influences on health in British adults: effects of socio-economic position in childhood and adulthood. Int J Epidemiol 36: 532–539.
22. Parsons TJ, Thomas C, Power C (2009) Estimated activity patterns in British 45 year olds: cross-sectional findings from the 1958 British birth cohort. Eur J Clin Nutr 63: 978–985.
23. Wang TJ, Zhang F, Richards JB, Kestenbaum B, van Meurs JB, et al. (2010) Common genetic determinants of vitamin D insufficiency: a genome-wide association study. Lancet 376: 180–188.
24. Han J, Kraft P, Nan H, Guo Q, Chen C, et al. (2008) A genome-wide association study identifies novel alleles associated with hair color and skin pigmentation. PLoS Genet 4: e1000074.
25. Sulem P, Gudbjartsson DF, Stacey SN, Helgason A, Rafnar T, et al. (2007) Genetic determinants of hair, eye and skin pigmentation in Europeans. Nat Genet 39: 1443–1452.
26. Nan H, Kraft P, Qureshi AA, Guo Q, Chen C, et al. (2009) Genome-wide association study of tanning phenotype in a population of European ancestry. J Invest Dermatol 129: 2250–2257.
27. Cooper JD, Smyth DJ, Walker NM, Stevens H, Burren OS, et al. (2011) Inherited variation in vitamin D genes is associated with predisposition to autoimmune disease type 1 diabetes. Diabetes 60: 1624–1631.
28. Cleeves M (1999) Hardy-Weinberg equilibrium test and allele frequency estimation. Stata Technical Bulletin 48: 34–38.
29. Sawcer S, Hellenthal G, Pirinen M, Spencer CC, Patsopoulos NA, et al. (2011) Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. Nature 476: 214–219.
30. Barrett JC, Clayton DG, Concannon P, Akolkar B, Cooper JD, et al. (2009) Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. Nat Genet.
31. Marchini J, Howie B, Myers S, McVean G, Donnelly P (2007) A new multipoint method for genome-wide association studies by imputation of genotypes. Nat Genet 39: 906–913.
32. Rice AJ (1995) Mathematical Statistics and Data Analysis. International Thomson Publishing.
33. Staiger D, Stock JH (1997) Instrumental Variables Regression with Weak Instruments. Econometrica 65: 557.
34. Palmer TM, Lawlor DA, Harbord RM, Sheehan NA, Tobias JH, et al. (2011) Using multiple genetic variants as instrumental variables for modifiable risk factors. Stat Methods Med Res.
35. Royston P, Carlin JB, White IR (2009) Multiple imputation of missing values: New features for mim. Stata Journal 9: 252–264.
36. StataCorp (2009) Stata Statistical Software: Release 11., version StataCorp LP.
37. McGrath JJ, Saha S, Burne TH, Eyles DW (2010) A systematic review of the association between common single nucleotide polymorphisms and 25-hydroxyvitamin D concentrations. J Steroid Biochem Mol Biol 121: 471–477.
38. Hyppönen E, Power C (2007) Hypovitaminosis D in British adults at age 45 y: nationwide cohort study of dietary and lifestyle predictors. Am J Clin Nutr 85: 860–868.
39. Lanham-New SA, Buttriss JL, Miles LM, Ashwell M, Berry JL, et al. (2011) Proceedings of the Rank Forum on Vitamin D. Br J Nutr 105: 144–156.
40. Holick MF (2007) Vitamin D deficiency. N Engl J Med 357: 266–281.
41. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, et al. (2006) Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 38: 904–909.
42. Cheng JB, Levine MA, Bell NH, Mangelsdorf DJ, Russell DW (2004) Genetic evidence that the human CYP2R1 enzyme is a key vitamin D 25-hydroxylase. Proc Natl Acad Sci U S A 101: 7711–7715.
43. Teslovich TM, Musunuru K, Smith AV, Edmondson AC, Stylianou IM, et al. (2010) Biological, clinical and population relevance of 95 loci for blood lipids. Nature 466: 707–713.
44. Speeckaert M, Huang G, Delanghe JR, Taes YE (2006) Biological and clinical aspects of the vitamin D binding protein (Gc-globulin) and its polymorphism. Clin Chim Acta 372: 33–42.
45. Sakaki T, Kagawa N, Yamamoto K, Inouye K (2005) Metabolism of vitamin D3 by cytochromes P450. Front Biosci 10: 119–134.
46. Chun RF, Lauridsen AL, Suon L, Zella LA, Pike JW, et al. (2010) Vitamin D-binding protein directs monocyte responses to 25-hydroxy- and 1,25-dihydroxyvitamin D. J Clin Endocrinol Metab 95: 3368–3376.

Table S1. Call rates and *P* values for Tests of Hardy-Weinberg Equilibrium for Vitamin D Polymorphisms Identified from Candidate Gene and Genome-wide Association Studies

Gene symbol	Chromosome position:	Gene names	SNP	Call rate (%)	HWE *P* value
GWA on 25(OH)D					
GC	4q12-q13	Vitamin D binding protein	rs4588[*]	83.4	0.24
DHCR7/ NADSYN1	11q13.4	7-dehydrocholesterol reductase/ NAD synthetase 1	rs12785878[†]	100	0.80
CYP2R1	11p15.2	Cytochrome P450, family 2, subfamily R, polypeptide 1	rs10741657[†]	82.5	0.39
CYP24A1	20q13	Cytochrome P450, family 24, subfamily A, polypeptide 1	rs6013897[†]	98.9	0.64
Vitamin D pathway genes					
CYP27B1	12q13.1-q13.3	Cytochrome P450, family 27, subfamily B, polypeptide 1	rs10877012	95.1	0.17
CYP27A1	2q33	Cytochrome P450, family 27, subfamily A, polypeptide 1	rs17470271	80.6	0.98

GWA on skin colour/ tanning

OCA2	15q	Oculocutaneous albinism II	rs7495174	96.9	0.42
OCA2	15q	Oculocutaneous albinism II	rs4778241	95.9	0.48
OCA2	15q	Oculocutaneous albinism II	rs4778138	97.4	0.09
OCA2	15q	Oculocutaneous albinism II	rs12913832‡	99.9	0.72
SLC45A2	5p13.2	Solute carrier family 45, member 2	rs13289	97.4	0.90
SLC45A2	5q13.2	Solute carrier family 45, member 2	rs16891982‡	97.1	0.14
MC1R	16q24.3	Melanocortin 1 receptor	rs11648785‡	96.5	0.18
MC1R	16q24.3	Melanocortin 1 receptor	rs1805005	98.4	0.27
MC1R	16q24.3	Melanocortin 1 receptor	rs464349‡	99.5	0.05
MC1R	16q24.3	Melanocortin 1 receptor	rs2228479	98.3	0.26
MC1R	16q24.3	Melanocortin 1 receptor	rs1805007 ‡	91.6	0.17
IRF4	6p23-p25	Interferon regulatory factor 4	rs12203592‡	99.8	0.01

IRF4	6p23-p25	Interferon regulatory factor 4	rs12210050‡	99.9	0.73
TYR	11q14-q21	Tyrosinase	rs1393350‡	100	0.50

HWE, Hardy-Weinberg equilibrium *P* value; MAF, minor allele frequency

*SNP rs4588 used as a proxy for the GWAS SNP rs2282679 (r^2= 0.98)

†SNPs identified from genome-wide association study (GWAS)

‡SNPs are from gene-chips and QC procedures are outlined under Genotyping in the Methods section

Table S2. The SNP association with biomarkers adjusted for 25(OH)D, sex and region

Biomarker*	Gene/Count	SNP	Coefficient (95% CI)	*P*-value†	Interaction *P*-value†
Coagulation markers					
vWF	*DHCR7*	rs12785878	-0.008 (-0.022, 0.007)	0.30	0.88
	CYP2R1	rs10741657	0.011 (-0.002, 0.023)	0.10	0.16
	Synthesis count		0.005 (-0.005, 0.015)	0.32	0.72
	CYP27B1	rs10877012	0.005 (-0.007, 0.017)	0.44	0.89
	CYP24A1	rs6013897	0.000 (-0.015, 0.014)	0.97	0.88
	GC	rs4588	-0.005 (-0.019, 0.008)	0.43	0.33
	Metabolism count		0.001 (-0.008, 0.009)	0.89	0.37
	Metabolism^{GWA}		-0.004 (-0.014, 0.007)	0.50	0.80
tPA	*DHCR7*	rs12785878	-0.016 (-0.038, 0.007)	0.17	0.009(0.14)
	CYP2R1	rs10741657	-0.009 (-0.029, 0.010)	0.34	0.32
	Synthesis count		-0.012 (-0.028, 0.003)	0.13	0.13
	CYP27B1	rs10877012	-0.009 (-0.028, 0.01)	0.38	0.46
	CYP24A1	rs6013897	-0.013 (-0.036, 0.01)	0.27	0.075
	GC	rs4588	0.007 (-0.014, 0.028)	0.51	0.34
	Metabolism count		-0.001 (-0.014, 0.012)	0.88	0.04(0.51)
	Metabolism^{GWA}		0.004 (-0.020, 0.012)	0.60	0.55
D-dimer	*DHCR7*	rs12785878	0.005 (-0.018, 0.028)	0.66	0.40
	CYP2R1	rs10741657	0.005 (-0.016, 0.025)	0.65	0.005(0.08)
	Synthesis count		0.006 (-0.01, 0.022)	0.48	0.25
	CYP27B1	rs10877012	0.007	0.48	0.91

			(-0.012, 0.026)		
	CYP24A1	rs6013897	-0.006 (-0.03, 0.018)	0.63	0.64
	GC	rs4588	0.003 (-0.019, 0.025)	0.80	0.79
	Metabolism count		0.003 (-0.010, 0.017)	0.65	0.37
	Metabolism^{GWA}		-0.001 (-0.017, 0.016)	0.95	0.80
Fibrinogen	*DHCR7*	rs12785878	0.001 (-0.008, 0.009)	0.90	0.42
	CYP2R1	rs10741657	0.003 (-0.004, 0.011)	0.40	0.28
	Synthesis count		0.001 (-0.005, 0.007)	0.83	0.81
	CYP27B1	rs10877012	0.002 (-0.005, 0.009)	0.57	0.43
	CYP24A1	rs6013897	-0.002 (-0.011, 0.007)	0.61	0.72
	GC	rs4588	-0.002 (-0.010, 0.006)	0.66	0.60
	Metabolism count		-0.001 (-0.006, 0.004)	0.74	0.71
	Metabolism^{GWA}		-0.003 (-0.009, 0.003)	0.37	0.65
Inflammatory marker					
CRP	*DHCR7*	rs12785878	0.020 (-0.031, 0.071)	0.44	0.40
	CYP2R1	rs10741657	0.000 (-0.045, 0.045)	0.99	0.07
	Synthesis count		0.004 (-0.032, 0.039)	0.83	0.26
	CYP27B1	rs10877012	0.017 (-0.026, 0.06)	0.44	0.87
	CYP24A1	rs6013897	-0.046 (-0.098, 0.006)	0.08	0.78
	GC	rs4588	-0.051 (-0.099, -0.003)	0.04	0.96
	Metabolism count		-0.025 (-0.055, 0.005)	0.10	0.34
	Metabolism^{GWA}		-0.051 (-0.088, -0.015)	0.006 (0.09)	0.99
Lipid markers					
Triglycerides	*DHCR7*	rs12785878	-0.010 (-0.033, 0.013)	0.38	0.94
	CYP2R1	rs10741657	0.008	0.47	0.14

			(-0.013, 0.028)		
	Synthesis count		0.000 (-0.016, 0.017)	0.97	0.43
	CYP27B1	rs10877012	0.007 (-0.013, 0.027)	0.48	0.18
	CYP24A1	rs6013897	0.007 (-0.017, 0.031)	0.54	0.22
	GC	rs4588	-0.003 (-0.025, 0.019)	0.81	0.92
	Metabolism count		0.006 (-0.008, 0.019)	0.42	0.22
	Metabolism^{GWA}		0.000 (-0.017, 0.017)	0.96	0.49
LDL	*DHCR7*	rs12785878	-0.011 (-0.022, 0.001)	0.07	0.61
	CYP2R1	rs10741657	-0.006 (-0.016, 0.005)	0.28	0.80
	Synthesis count		-0.007 (-0.015, 0.001)	0.08	0.76
	CYP27B1	rs10877012	0.004 (-0.006, 0.013)	0.47	0.41
	CYP24A1	rs6013897	-0.003 (-0.015, 0.009)	0.62	0.07
	GC	rs4588	-0.004 (-0.015, 0.007)	0.50	0.89
	Metabolism count		0.001 (-0.006, 0.008)	0.82	0.48
	Metabolism^{GWA}		-0.003 (-0.011, 0.005)	0.48	0.56
HDL	*DHCR7*	rs12785878	-0.002 (-0.012, 0.007)	0.62	0.13
	CYP2R1	rs10741657	0.000 (-0.008, 0.009)	0.95	0.23
	Synthesis count		-0.002 (-0.008, 0.005)	0.63	0.39
	CYP27B1	rs10877012	0.001 (-0.007, 0.009)	0.77	0.15
	CYP24A1	rs6013897	-0.003 (-0.013, 0.007)	0.52	0.27
	GC	rs4588	0.000 (-0.010, 0.009)	0.94	0.33
	Metabolism count		-0.002 (-0.007, 0.004)	0.55	0.31
	Metabolism^{GWA}		-0.001 (-0.008, 0.007)	0.88	0.17
Cholesterol	*DHCR7*	rs12785878	-0.006	0.10	0.26

			(-0.014, 0.001)		
	CYP2R1	rs10741657	-0.003 (-0.009, 0.004)	0.39	0.90
	Synthesis count		-0.005 (-0.01, 0.001)	0.082	0.64
	CYP27B1	rs10877012	0.003 (-0.003, 0.009)	0.37	0.53
	CYP24A1	rs6013897	-0.002 (-0.009, 0.006)	0.70	0.15
	GC	rs4588	-0.001 (-0.008, 0.006)	0.73	0.70
	Metabolism count		0.001 (-0.004, 0.005)	0.71	0.54
	Metabolism^{GWA}		0.001 (-0.007, 0.004)	0.65	0.56
Lung function marker					
FEV	*DHCR7*	rs12785878	0.001 (-0.023, 0.025)	0.95	0.04 (0.54)
	CYP2R1	rs10741657	0.014 (-0.007, 0.035)	0.20	0.44
	Synthesis count		0.010 (-0.007, 0.027)	0.23	0.73
	CYP27B1	rs10877012	0.001 (-0.02, 0.021)	0.95	0.70
	CYP24A1	rs6013897	0.004 (-0.021, 0.029)	0.76	0.97
	GC	rs4588	0.007 (-0.016, 0.03)	0.56	0.51
	Metabolism count		0.006 (-0.008, 0.02)	0.43	0.77
	Metabolism^{GWA}		-0.006 (-0.011, 0.024)	0.47	0.03 (0.48)
Cardiovascular disease related markers					
Diastolic BP	*DHCR7*	rs12785878	0.003 (-0.003, 0.008)	0.34	0.68
	CYP2R1	rs10741657	-0.004 (-0.008, 0.001)	0.12	0.04 (0.62)
	Synthesis count		-0.002 (-0.006, 0.002)	0.29	0.46
	CYP27B1	rs10877012	-0.002 (-0.006, 0.003)	0.50	0.11
	CYP24A1	rs6013897	-0.005 (-0.011, 0.000)	0.065	0.35
	GC	rs4588	-0.004 (-0.009, 0.001)	0.17	0.004 (0.07)
	Metabolism		-0.003	0.09	0.69

	count		(-0.006, 0)		
	Metabolism[GWA]		-0.004 (-0.008, 0.000)	0.03 (0.50)	0.33
Systolic BP	*DHCR7*	rs12785878	0.003 (-0.002, 0.007)	0.30	0.57
	CYP2R1	rs10741657	0.000 (-0.004, 0.005)	0.89	0.072
	Synthesis count		0.001 (-0.003, 0.004)	0.67	0.52
	CYP27B1	rs10877012	-0.001 (-0.005, 0.003)	0.67	0.49
	CYP24A1	rs6013897	-0.005 (-0.01, 0)	0.03	0.38
	GC	rs4588	-0.003 (-0.007, 0.002)	0.26	0.058
	Metabolism count		-0.002 (-0.005, 0.001)	0.15	0.77
	Metabolism[GWA]		-0.004 (-0.008, 0.000)	0.03 (0.39)	0.16
IgE	*DHCR7*	rs12785878	-0.010 (-0.069, 0.05)	0.75	0.47
	CYP2R1	rs10741657	0.058 (0.005, 0.111)	0.03	0.39
	Synthesis count		0.037 (-0.005, 0.079)	0.08	0.79
	CYP27B1	rs10877012	0.047 (-0.003, 0.098)	0.07	0.72
	CYP24A1	rs6013897	0.016 (-0.045, 0.078)	0.60	0.88
	GC	rs4588	-0.016 (-0.072, 0.041)	0.59	0.03 (0.44)
	Metabolism count		0.019 (-0.016, 0.054)	0.29	0.59
	Metabolism[GWA]		0.005 (-0.039, 0.048)	0.84	0.58
IGF-1	*DHCR7*	rs12785878	0.001 (-0.012, 0.013)	0.90	0.091
	CYP2R1	rs10741657	-0.002 (-0.013, 0.009)	0.70	0.21
	Synthesis count		-0.001 (-0.010, 0.008)	0.77	0.99
	CYP27B1	rs10877012	-0.008 (-0.019, 0.002)	0.12	0.31
	CYP24A1	rs6013897	-0.001 (-0.014, 0.012)	0.84	0.15
	GC	rs4588	0.012	0.06	0.61

			(-0.001, 0.024)		
	Metabolism count		-0.002 (-0.01, 0.005)	0.52	0.70
	Metabolism^{GWA}		0.004 (-0.005, 0.013)	0.39	0.62
HbA1c	*DHCR7*	rs12785878	0.000 (-0.004, 0.005)	0.94	0.13
	CYP2R1	rs10741657	0.000 (-0.004, 0.004)	0.90	0.63
	Synthesis count		0.000 (-0.003, 0.003)	0.82	0.60
	CYP27B1	rs10877012	0.000 (-0.003, 0.004)	0.88	0.39
	CYP24A1	rs6013897	0.001 (-0.004, 0.005)	0.72	0.69
	GC	rs4588	-0.001 (-0.005, 0.003)	0.56	0.46
	Metabolism count		0.000 (-0.002, 0.003)	0.84	0.17
	Metabolism^{GWA}		0.000 (-0.003, 0.003)	0.90	0.51

[*]Where required the biomarker has been natural log transformed to achieve normal distribution

[†]In brackets, the *p*-value adjusted for multiple testing.

Appendix S1 – Sample size calculations

Mendelian randomisation (MR) studies can be modelled using two-stage least squares (2SLS) regression for instrumental variable (IV) analysis. In this setting the endogenous variable is environmental exposure (in our example 25(OH)D concentrations), exogenous /instrumental variable is the genetic marker (in our example SNPs or allele count from multiple SNPs), and outcome is a quantitative bio-marker (in our example the standardized systolic BP). For 2SLS regression there is no standard approach for estimating power and sample size, however one way is through repetitive simulation of a dataset. This was done by simulating a dataset of a fixed sample size (n) by imposing constraints on the known relationships, running the regression model and then testing the parameter of interest at a fixed significant level α. The dataset was re-simulated a 1000 times for the same sample size (n) and the proportion of times the parameter of interest was less than the significant level α was used to indicate the power of the study with the sample size in question [1].

To create a dataset for the simulation, the models are defined as:

$$\begin{aligned} x_i &= \alpha_0 + \alpha_1 g_i + \varepsilon_{xi} \\ y_i &= \beta_0 + \beta_1 x_i + \varepsilon_{yi} \end{aligned},$$

where i is the number of individuals, x_i, g_i, y_i are the exposure, genetic variant, and outcome for each individual, respectively, with $\varepsilon_{xi} \sim N\left(0, \sigma_1^2\right)$ and $\varepsilon_{yi} \sim N\left(0, \sigma_2^2\right)$. The genetic variant g_i takes the values for the single marker $\{0,1,2\}$ and joint allele count $\{0,1,2,3\}$. We assumed an effect of a 5% reduction in a standardized blood pressure outcome by 10nmol/l increase in 25(OH)D based on effect estimates

from the 1958BC. The MAF of the SNPs were also based on the proportions from the 1958BC.

Reference List

1. Feiveson AH (2002) Power by simulation. Stata Journal 2: 107-124

Appendix 2: Additional tables for Chapter 3

Table 1 Results from fitting the mediation models using a Frequentist approach in the 1958BC (aged 45 yrs); iterative inclusion of harmonic terms with 25(OH)D (mediator) and individually with the inflammatory/haemostatic markers (outcomes) . Greyed boxes highlight models with the significant harmonic terms *p*-values and lowest Bayesian Information Criterion.

Likelihood ratio test (LRT) *p*-value						
Harmonic term(s) in the model	**Mediator**[*]	**Outcome**[†]				
	25(OH)D	**CRP**	**D-Dimer**	**Fibrinogen**	**tPA**	**vWF**
Yearly *p*-value	<0.001	0.77	0.76	0.021	0.010	0.012
Semestral *p*-value	<0.001	0.82	0.016	0.77	0.227	0.23
Quarterly *p*-value	<0.001	0.91	0.19	0.89	0.732	0.42
Semestral *p*-value, and adjusted for yearly	<0.001	0.85	0.015	0.71	0.14	0.10
Quarterly *p*-value, and adjusted for yearly	<0.001	0.89	0.14	0.97	0.84	0.27
Quarterly *p*-value, and adjusted for semestral	<0.001	0.91	0.40	0.92	0.82	0.54
Quarterly *p*-value, and adjusted for yearly and semestral	<0.001	0.88	0.33	0.94	0.90	0.42

Bayesian Information Criterion (BIC)						
Harmonic term(s) in the model	Mediator[1]	Outcome[2]				
	25(OH)D	CRP	D-Dimer	Fibrinogen	tPA	vWF
Yearly	56421	17479	17589	17210	17429	17522
Semestral	57616	17479	17582	17217	17435	17527
Quarterly	57697	17480	17586	17217	17437	17528
Yearly and semestral	56353	17496	17599	17227	17443	17534
Yearly and quarterly	56400	17497	17603	17228	17447	17536
Semestral and quarterly	57588	17497	17598	17234	17452	17543
Yearly, semestral and quarterly	56344	17514	17614	17244	17460	17550

[*] Linear regression model adjusted for sex

[†] Mixed effects model with sex-specified standard deviation scores where variability among laboratory batches was modelled as a random effect on the intercept adjusted for seasonal confounders (chest infection, alcohol consumption, PC/TC leisure time, recreation MET hours, social class at adulthood), except for C-reactive protein that was model using linear regression.

Table 2 Results from fitting the mediation models using a Bayesian approach in the 1958BC (aged 45 yrs); iterative inclusion of harmonic terms with 25(OH)D (mediator) and individually with the inflammatory/haemostatic markers (outcomes) . Greyed boxes highlight models with the harmonic terms parameters credible intervals that do not overlap zero and lowest Deviance Information Criterion (DIC).

	Parameter estimates with 95% credible interval					
	Mediator*	**Outcome†**				
Harmonic term(s) in the model	**25(OH)D, nmol/L**	**CRP, SDS**	**D-Dimer, SDS**	**Fibrinogen, SDS**	**tPA, SDS**	**vWF, SDS**
Yearly terms						
Cosine	-7.02 (-7.86, -6.18)	-0.007 (-0.044, 0.028)	0.019 (-0.039, 0.081)	-0.089 (-0.164, -0.011)	-0.052 (-0.117, 0.015)	0.039 (-0.032, 0.111)
Sine	-14.7 (-15.5, -13.9)	-0.009 (-0.043, 0.025)	0.012 (-0.043, 0.068)	-0.051 (-0.12, 0.022)	-0.076 (-0.137, -0.015)	0.097 (0.029, 0.167)
Semestral terms						
Cosine	-5.15 (-6.02, -4.23)	-0.007 (-0.042, 0.028)	-0.021 (-0.073, 0.03)	-0.013 (-0.089, 0.063)	0.002 (-0.062, 0.064)	-0.06 (-0.133, 0.007)
Sine	2.43 (1.51, 3.34)	-0.004 (-0.038, 0.031)	0.069 (0.018, 0.118)	0.02 (-0.055, 0.096)	0.058 (-0.007, 0.121)	0.003 (-0.068, 0.074)
Quarterly terms						
Cosine	2.84 (1.97, 3.72)	0.001 (-0.033, 0.034)	0.046 (-0.005, 0.099)	0.011 (-0.055, 0.079)	0.022 (-0.045, 0.084)	0.025 (-0.04, 0.095)
Sine	-2.70 (-3.61, -1.78)	0.008 (-0.027, 0.042)	0.003 (-0.048, 0.055)	-0.008 (-0.075, 0.057)	0.013 (-0.051, 0.077)	-0.036 (-0.106, 0.032)
Semestral terms, adjusted for yearly						
Cosine	-2.87	-0.007	-0.021	-0.011	0.018	-0.073

Harmonic term(s) in the model	Mediator* 25(OH)D, nmol/L	Outcome† CRP, SDS	D-Dimer, SDS	Fibrinogen, SDS	tPA, SDS	vWF, SDS
	Parameter estimates with 95% credible interval					
	(-3.72, -2.07)	(-0.043, 0.029)	(-0.073, 0.032)	(-0.083, 0.064)	(-0.043, 0.079)	(-0.143, -0.005)
Sine	2.58 (1.76, 3.40)	-0.004 (-0.04, 0.03)	0.069 (0.017, 0.123)	0.025 (-0.046, 0.096)	0.06 (-0.001, 0.121)	-0.006 (-0.078, 0.063)
Quarterly terms, adjusted for yearly						
Cosine	1.31 (0.508, 2.12)	-0.001 (-0.036, 0.033)	0.052 (0, 0.108)	-0.001 (-0.066, 0.065)	0.015 (-0.045, 0.075)	0.039 (-0.029, 0.105)
Sine	-2.29 (-3.12, -1.47)	0.009 (-0.026, 0.044)	-0.001 (-0.053, 0.052)	-0.002 (-0.067, 0.065)	0.014 (-0.048, 0.076)	-0.039 (-0.11, 0.025)
Quarterly terms, adjusted for semestral						
Cosine	1.74 (0.863, 2.65)	0.000 (-0.036, 0.034)	0.034 (-0.016, 0.088)	0.008 (-0.055, 0.076)	0.017 (-0.045, 0.082)	0.019 (-0.052, 0.088)
Sine	-2.56 (-3.45, -1.67)	0.008 (-0.027, 0.044)	0.004 (-0.046, 0.054)	-0.008 (-0.074, 0.06)	0.014 (-0.047, 0.081)	-0.03 (-0.102, 0.039)
Quarterly terms, adjusted for yearly and semestral						
Cosine	0.479 (-0.329, 1.31)	-0.002 (-0.037, 0.033)	0.04 (-0.012, 0.094)	-0.006 (-0.07, 0.062)	0.008 (-0.053, 0.07)	0.032 (-0.034, 0.099)

Parameter estimates with 95% credible interval						
	Mediator*	Outcome†				
Harmonic term(s) in the model	25(OH)D, nmol/L	CRP, SDS	D-Dimer, SDS	Fibrinogen, SDS	tPA, SDS	vWF, SDS
Sine	-2.15 (-2.96, -1.33)	0.009 (-0.027, 0.044)	0.001 (-0.052, 0.051)	-0.002 (-0.064, 0.063)	0.014 (-0.047, 0.079)	-0.033 (-0.102, 0.034)

Deviance Information Criterion (DIC)						
Harmonic term(s) in the model	Mediator*	Outcome†				
	25(OH)D	CRP	D-Dimer	Fibrinogen	tPA	vWF
Yearly	56396	17362	17424	17042	17273	17321
Semestral	57591	17362	17420	17044	17273	17320
Quarterly	57672	17362	17422	17045	17274	17322
Yearly and semestral	56315	17365	17421	17043	17272	17318
Yearly and quarterly	56361	17365	17422	17044	17274	17320
Semestral and quarterly	57550	17366	17420	17046	17274	17321
Yearly, semestral and quarterly	56292	17369	17421	17045	17273	17318

* Linear model adjusted for sex with non-informative priors for the parameters and model error.

† Mixed effects model with sex-specified standard deviation scores where variability among laboratory batches was modelled as a random effect on the intercept adjusted for seasonal confounders (chest infection, alcohol consumption, PC/TC leisure time, recreation MET hours, social class

at adulthood), except for C-reactive protein that was model using linear regression. All models had non-informative priors for the parameters and model error terms.

Table 3 Results from using the Frequentist approach to fit models adjusted for 25-hydroxyvitamin D (mediator) in the 1958BC (aged 45 yrs); iterative inclusion of harmonic terms individually with the haemostatic markers (outcomes). Greyed boxes highlight models with the significant harmonic terms *p*-values and lowest BIC

Likelihood ratio test (LRT) *p*-value				
Harmonic term(s) in model	**Outcome***			
	D-Dimer	**Fibrinogen**	**tPA**	**vWF**
Yearly *p*-value	0.81	0.004	<0.001	0.019
Semestral *p*-value	0.007	0.71	0.15	0.20
Quarterly *p*-value	0.13	0.86	0.70	0.40
Semestral *p*-value, adjusted for yearly	0.006	0.60	0.059	0.10
Quarterly *p*-value, adjusted for yearly	0.12	0.95	0.84	0.27
Quarterly *p*-value, adjusted for Semestral	0.33	0.90	0.85	0.51
Quarterly *p*-value, adjusted for yearly and semestral	0.31	0.92	0.95	0.41

Bayesian Information Criterion (BIC)				
Harmonic term(s) in model	**Outcome***			
	D-Dimer	**Fibrinogen**	**tPA**	**vWF**
Yearly	17588	17211	17361	17530
Semestral	17579	17220	17381	17535
Quarterly	17584	17220	17384	17536
Yearly and semestral	17595	17227	17373	17543
Yearly and quarterly	17601	17228	17378	17545
Semestral and quarterly	17594	17237	17398	17551
Yearly, semestral and quarterly	17611	17244	17390	17559

* Mixed effects model with sex-specified standard deviation scores where variability among laboratory batches was modelled as a random effect on the intercept adjusted for seasonal confounders (chest infection, alcohol consumption, PC/TC leisure time, recreation MET hours, social class at adulthood) and 25(OH)D.

Table 4 Results from using the Bayesian approach to fit models adjusted for 25-hydroxyvitamin D (mediator) in the 1958BC (aged 45 yrs); iterative inclusion of harmonic terms individually with the haemostatic markers (outcomes) . Greyed boxes highlight models whose harmonic terms parameters' 95% credible intervals do not include zero and had lowest DIC.

Parameter estimates with 95% credible interval				
Harmonic term(s) in the model	**Outcome***			
	D-Dimer	**Fibrinogen**	**tPA**	**vWF**
Yearly terms				
Cosine	0.008 (-0.051, 0.067)	-0.100 (-0.178, -0.022)	-0.087 (-0.155, -0.018)	0.040 (-0.032, 0.113)
Sine	-0.016 (-0.071, 0.040)	-0.076 (-0.145, -0.004)	-0.149 (-0.212, -0.085)	0.092 (0.023, 0.159)
Semestral terms				
Cosine	-0.029 (-0.080, 0.020)	-0.017 (-0.094, 0.07)	-0.022 (-0.093, 0.046)	-0.062 (-0.129, 0.007)
Sine	0.073 (0.023, 0.125)	0.025 (-0.052, 0.099)	0.066 (-0.004, 0.137)	0.003 (-0.065, 0.076)
Quarterly terms				
Cosine	0.051 (0.003, 0.103)	0.014 (-0.051, 0.078)	0.030 (-0.042, 0.096)	0.027 (-0.040, 0.095)
Sine	-0.002 (-0.056, 0.052)	-0.010 (-0.076, 0.062)	0.007 (-0.058, 0.076)	-0.036 (-0.103, 0.029)
Semestral terms, adjusted for yearly				
Cosine	-0.025 (-0.078, 0.027)	-0.017 (-0.089, 0.056)	0.001 (-0.061, 0.063)	-0.073 (-0.140, -0.007)
Sine	0.074 (0.024, 0.127)	0.030 (-0.039, 0.104)	0.074 (-0.012, 0.136)	-0.006 (-0.074, 0.063)
Quarterly terms, adjusted for yearly				
Cosine	0.054 (0.002, 0.106)	0.000 (-0.061, 0.062)	0.020 (-0.043, 0.080)	0.038 (-0.026, 0.103)

Parameter estimates with 95% credible interval				
Harmonic term(s) in the model	Outcome*			
	D-Dimer	Fibrinogen	tPA	vWF
Sine	-0.004 (-0.057, 0.05)	-0.004 (-0.069, 0.062)	0.002 (-0.060, 0.066)	-0.040 (-0.105, 0.029)
Quarterly terms, adjusted for semestral				
Cosine	0.037 (-0.011, 0.087)	0.010 (-0.056, 0.076)	0.017 (-0.052, 0.085)	0.022 (-0.043, 0.091)
Sine	0.002 (-0.048, 0.052)	-0.009 (-0.076, 0.059)	0.012 (-0.056, 0.084)	-0.033 (-0.101, 0.034)
Quarterly terms, adjusted for yearly and semestral				
Cosine	0.039 (-0.010, 0.091)	-0.004 (-0.070, 0.063)	0.010 (-0.052, 0.068)	0.031 (-0.037, 0.097)
Sine	-0.001 (-0.053, 0.050)	-0.005 (-0.070, 0.063)	0.004 (-0.057, 0.070)	-0.033 (-0.100, 0.033)

Deviance Information Criterion (DIC)				
Harmonic term(s) in the model	Outcome*			
	D-Dimer	Fibrinogen	tPA	vWF
Yearly	17414	17034	17190	17323
Semestral	17410	17037	17195	17322
Quarterly	17412	17038	17197	17324
Yearly and semestral	17412	17035	17190	17321
Yearly and quarterly,	17413	17037	17192	17322
Semestral and quarterly	17410	17039	17196	17322

Yearly, semestral and quarterly	17412	17038	17191	17320

[*] Mixed effects model with sex-specified standard deviation scores where variability among laboratory batches was modelled as a random effect on the intercept adjusted for seasonal confounders (chest infection, alcohol consumption, PC/TC leisure time, recreation MET hours, social class at adulthood) and 25(OH)D. All models had non-informative priors for the parameters and model error terms.

Table 5 The parametric bootstrap mediation product of coefficients test for the product of the best fit harmonic function with 25(OH)D and the 25(OH)D associations with the haemostatic markers in the 1958BC (aged 45 years). Greyed boxes highlight models whose harmonic terms' 95% confidence intervals do not include zero.

	Product of the three frequency harmonic function parameters and 25(OH)D parameter for the haemostatic markers (95% CI)			
	D-Dimer, SDS	**Fibrinogen, SDS**	**tPA, SDS**	**vWF, SDS**
Mediation test product of harmonic terms and 25(OH)D parameter				
Yearly cosine	0.012 (0.006, 0.019)	0.009 (0.003, 0.015)	0.032 (0.025, 0.039)	-0.000 (-0.006, 0.006)
Yearly sine	0.026 (0.013, 0.039)	0.019 (0.006, 0.032)	0.067 (0.054, 0.081)	-0.000 (-0.014, 0.013)
Semestral cosine	0.005 (0.002, 0.008)	0.004 (0.001, 0.006)	0.013 (0.009, 0.017)	-0.000 (-0.003, 0.003)
Semestral sine	-0.004 (-0.007, -0.002)	-0.003 (-0.006, -0.001)	-0.012 (-0.016, -0.008)	0.000 (-0.002, 0.002)
Quarterly cosine	-0.001 (-0.002, 0.001)	-0.001 (-0.002, 0.001)	-0.002 (-0.006, 0.002)	0.000 (-0.001, 0.001)
Quarterly sine	0.004 (0.002, 0.006)	0.003 (0.001, 0.005)	0.010 (0.006, 0.014)	-0.000 (-0.021, 0.002)

Table 6 Associations between 25-hydroxyvitamin D and the haemostatic markers adjusted for the best fit (Bayesian) harmonic terms and seasonal confounders in a mixed effect model in the 1958BC (aged 45 years).

	Outcome adjusted for best fit harmonic terms, 25(OH)D and seasonal confounders (95% credible intervals)*			
	D-Dimer, SDS adjusted for semestral pattern	**Fibrinogen, SDS adjusted for yearly pattern**	**tPA, SDS adjusted for yearly pattern**	**vWF, SDS adjusted for yearly pattern**
25(OH)D per 10 nmol/L parameter	-0.019 (-0.029, -0.008)	-0.017 (-0.028, -0.006)	-0.051 (-0.062, -0.040)	-0.002 (-0.013, 0.009)

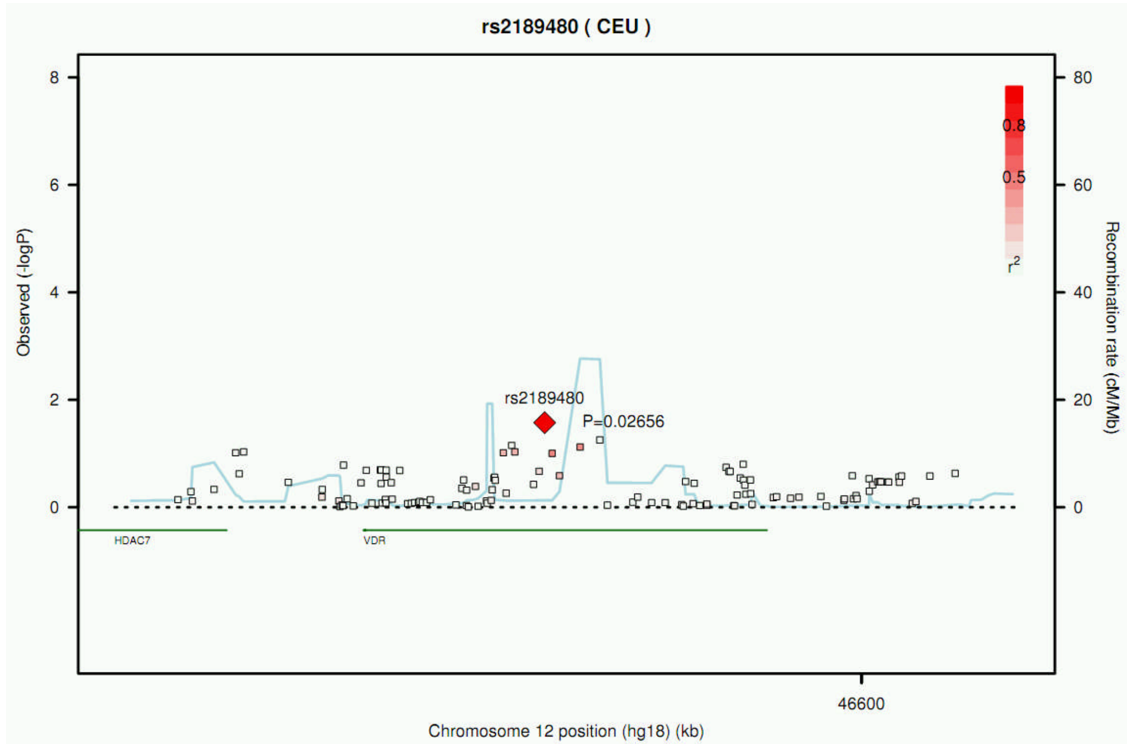
* Seasonal confounders adjusted for as fixed effect chest infection, alcohol consumption, PC/TC leisure time, recreation MET hours, social class at adulthood and random effect was laboratory batch. All models had non-informative priors for the parameters and model error terms.

Table 7 The Bayesian mediation product of coefficients test for the product of the best fit harmonic function parameters with 25(OH)D and 25(OH)D parameters with the haemostatic markers in the 1958BC (aged 45 yrs). Greyed boxes highlight models whose harmonic terms' parameters 95% credible intervals do not include zero.

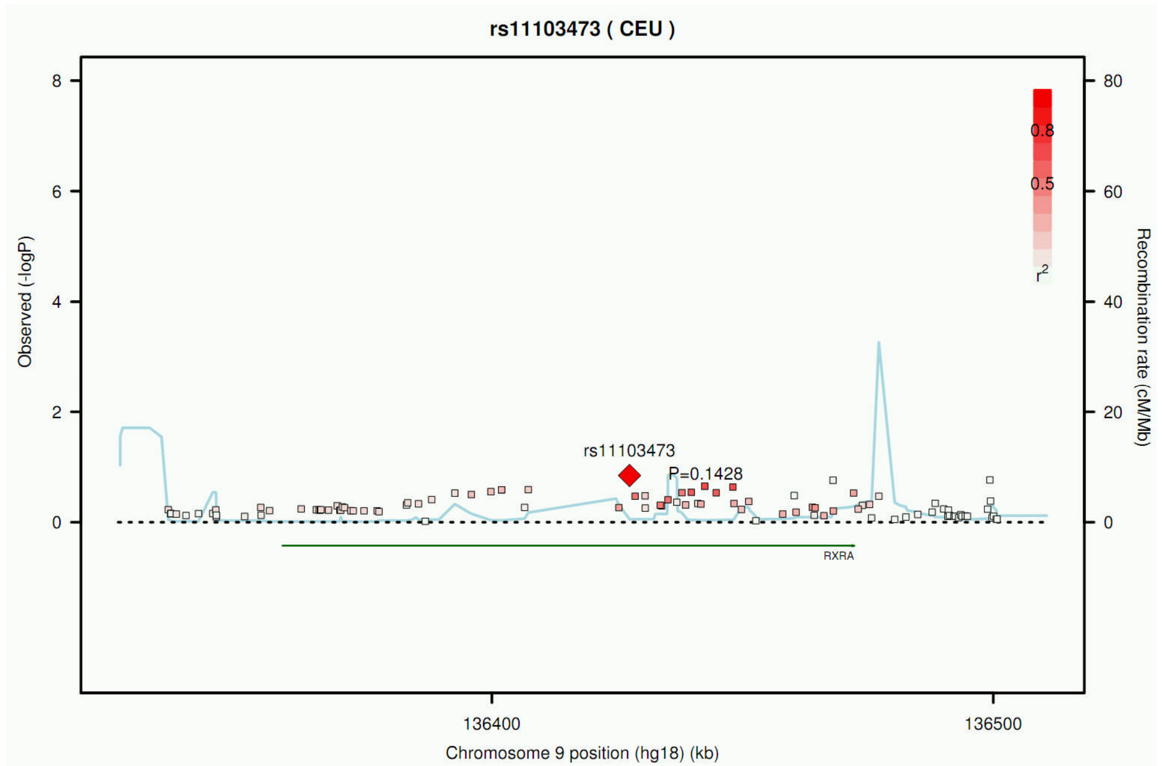
	Product of the three frequency harmonic function parameters and 25(OH)D parameter for the haemostatic markers (95% credible intervals)			
	D-Dimer, SDS	**Fibrinogen, SDS**	**tPA, SDS**	**vWF, SDS**
Mediation test product of harmonic terms and 25(OH)D parameter				
Yearly cosine	0.013 (0.006, 0.02)	0.011 (0.004, 0.019)	0.034 (0.026, 0.043)	0.001 (-0.006, 0.009)
Yearly sine	0.027 (0.012, 0.042)	0.024 (0.009, 0.04)	0.073 (0.056, 0.089)	0.003 (-0.013, 0.019)
Semestral cosine	0.005 (0.002, 0.009)	0.004 (0.002, 0.008)	0.014 (0.009, 0.019)	0.001 (-0.002, 0.006)
Semestral sine	-0.005 (-0.008, -0.002)	-0.004 (-0.008, -0.001)	-0.013 (-0.018, -0.008)	0.001 (-0.003, 0.002)
Quarterly cosine	-0.001 (-0.003, 0.001)	-0.001 (-0.003, 0.001)	-0.002 (-0.007, 0.002)	-0.001 (0.001, 0.001)
Quarterly sine	0.004 (0.002, 0.007)	0.004 (0.001, 0.007)	0.011 (0.006, 0.016)	0.003 (-0.013, 0.019)

Appendix 3: Additional figures for Chapter 4

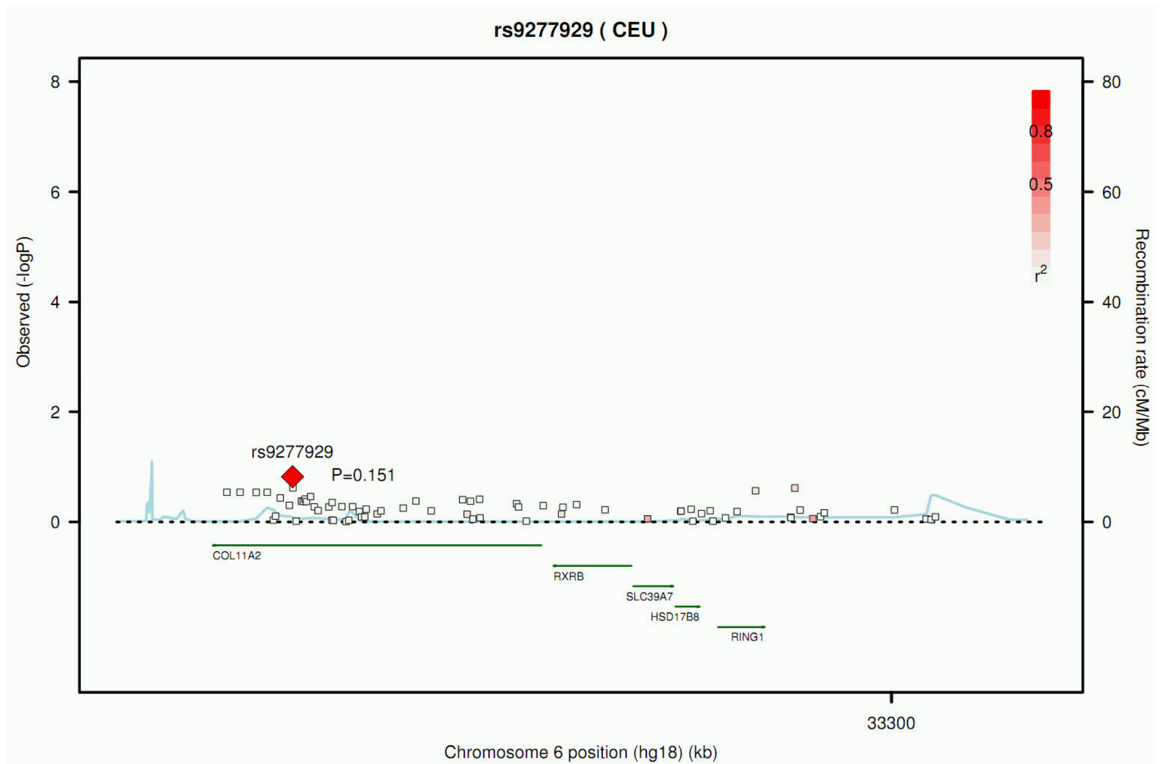
A



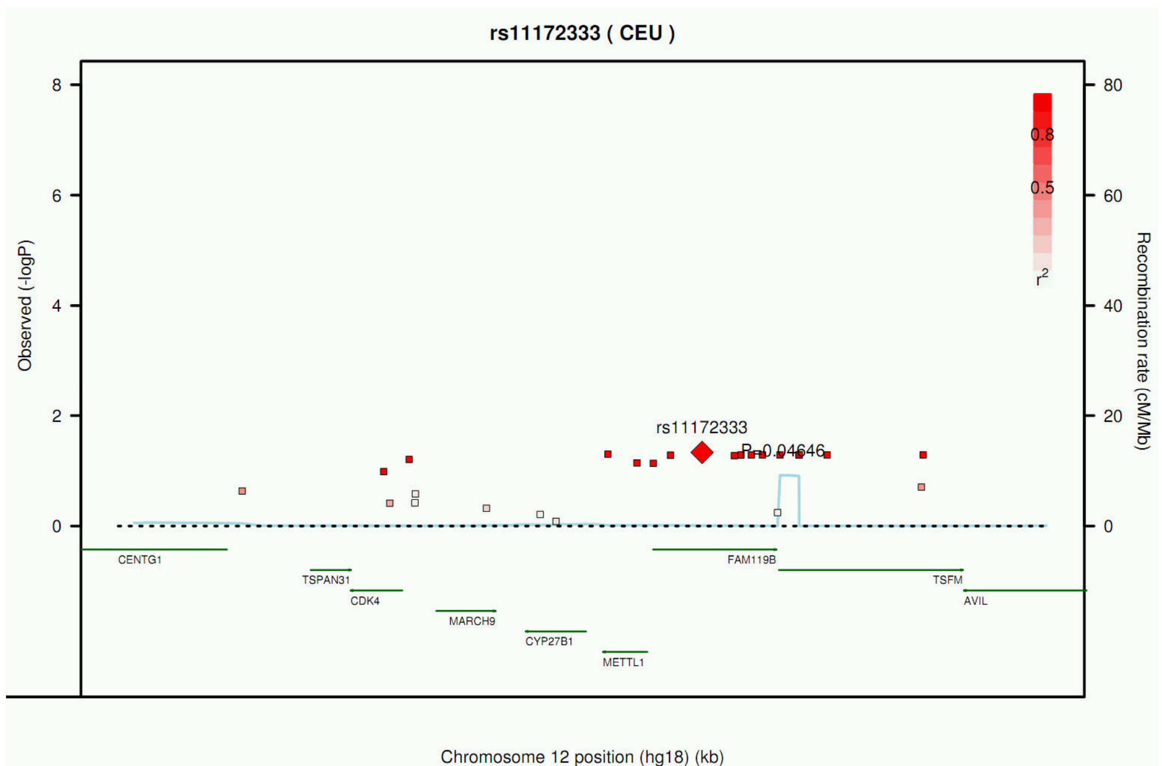
B



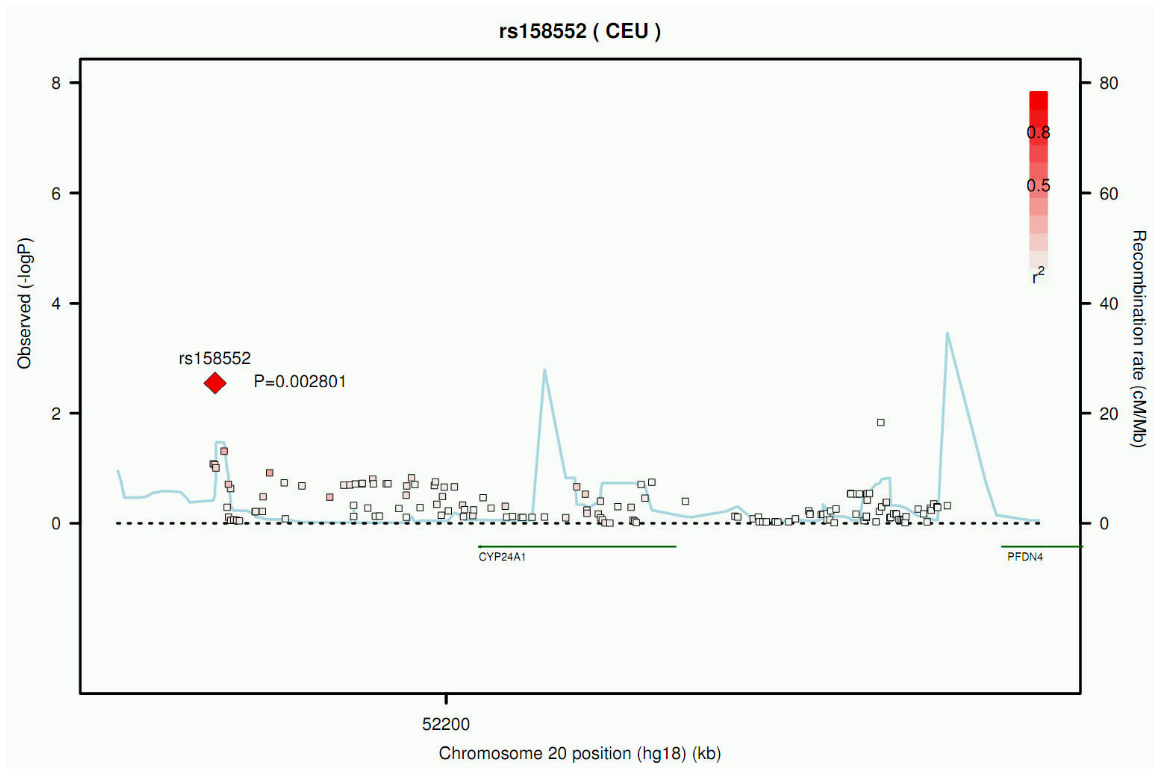
C



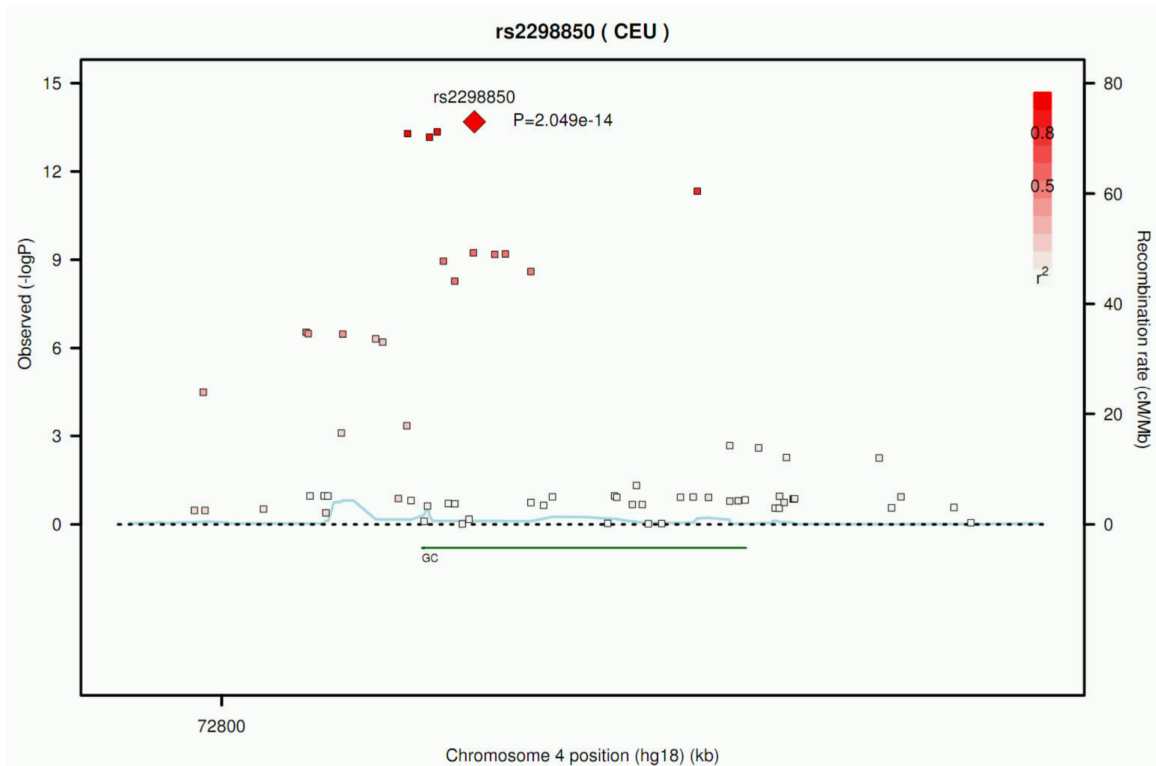
D



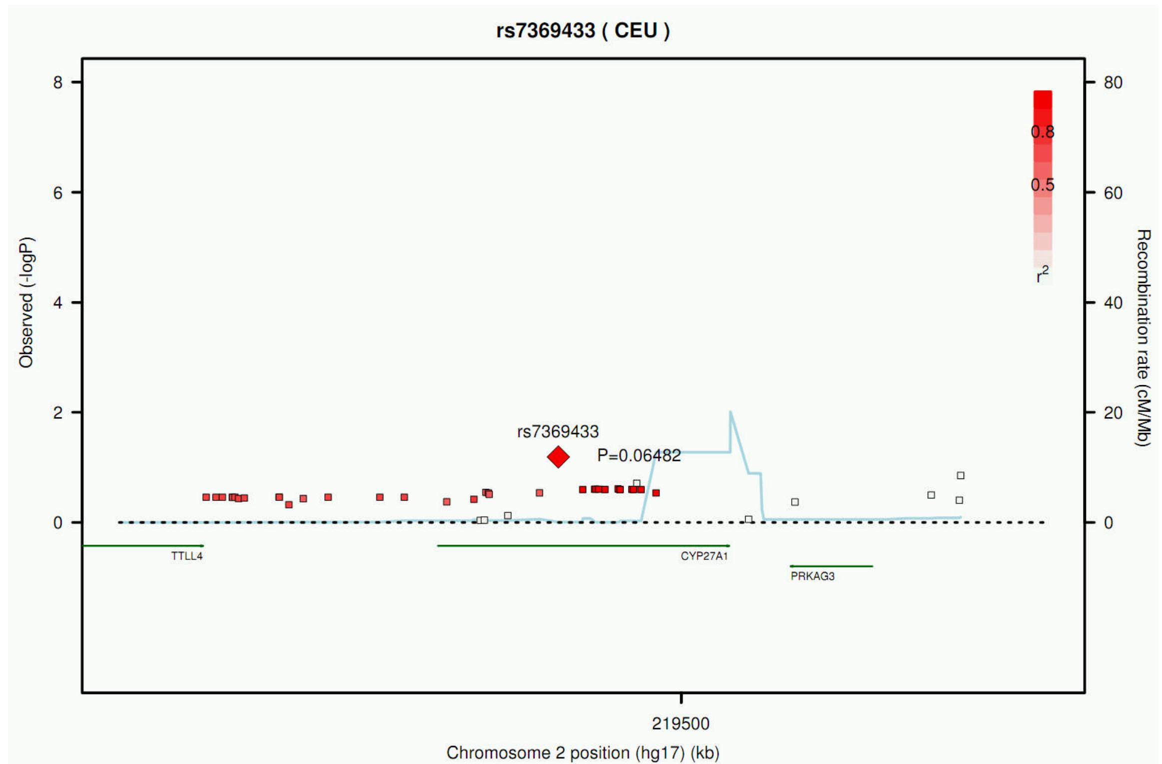
E



F



G



Regional linkage disequilibrium plots for A) *VDR*, B) *RXRA*, C) *RXRB*, D) *CYP27B1*, E) *CYP24A1*, F) *GC* and H) *CYP27A1*. The light blue line on the graphs is the recombination rate for the region. The diamonds indicate the SNPs *p*-value for its association with the 25(OH)D standardised residuals in the 1958BC discovery samples. The colour of the diamond indicates the LD between SNPs with the gradient of the LD on the top right hand side.

Appendix 4: Additional methods for Chapter 5

Estimation of the Gc haplotypes

The two non-synonymous SNPs rs4588 and rs7041 located in exon 11 of gene *GC* were used to estimate the Gc haplotypes (Figure 1). The haplotypes were inferred by SimHap (Carter *et al.* 2008), which uses expectation maximisation methods to estimate the haplotypes from SNP data despite the unknown phase of the alleles. In the instances in which individuals had haplotypes with a posterior probability < 0.95, these haplotypes were excluded from further analysis.

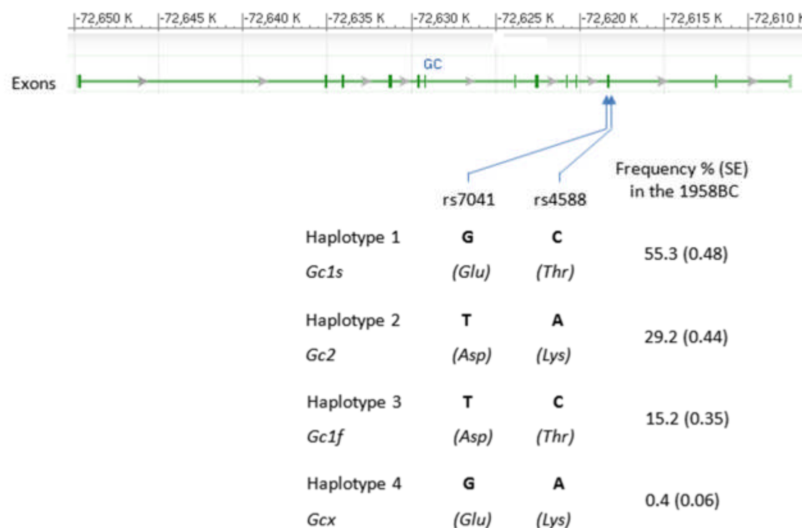


Figure 1 Exon/functional regions (dark green) of *GC* and haplotypes estimated from SNPs rs4588 and rs7041 in the 1958BC (adapted from (Fang *et al.* 2009)).

The haplotypes were categorised in slightly different ways to assess which categorisation gave the best fit with ln 25(OH)D. The categorisations were: *Gc* haplotype variables were created to represent the the number of copies they carried of the *Gc1s*, *Gc1f* and *Gc2* haplotypes (Table 1), by the genotypes of the haplotypes,

and similarly by the haplotype pairs of first haplotype vs second haplotype (Table 2). Each haplotype variable was assessed on balance of the adjusted R^2, F-statistic and the Bayesian Information Criterion (BIC). The R^2 adjusted for the number of parameters was considered, as is the proportion of variation explained in 25(OH)D, calculated from the difference in adjusted R^2 from a linear regression model with ln 25(OH)D including and excluding the *Gc* variable (or additive trend of the variable). The value of F should be greater than 10 to considered as an instrument for an exposure, on the basis of probability of the instrument's association with the exposure and sample size (Staiger & Stock 1997). The BIC was also considered when choosing the haplotype variable(s) as it penalises a model's likelihood by its number of estimated parameters (or degrees of freedom) proportionally to the logarithm of the number of independent observations. This results in selecting models more conservative against over-parametrisation, tha other likelihood-based criteria, which assign less severe penalties to the model's goodness-of-fit (Kuha 2004).

Table 1 Copies of the *Gc* haplotypes with mean 25(OH)D concentrations in the 1958BC

Haplotype copies	*n*	25(OH)D, nmol/l geometric mean (95% CI)
Gc1s		
no copies	1029	50.6 (49.2, 52.0)
1 copy	2701	53.4 (52.5, 54.4)
2 copies	1600	57.1 (55.9, 58.4)
Gc1f		
no copies	3852	53.7 (52.9, 54.4)
1 copy	1346	54.8 (53.5, 56.2)
2 copies	132	52.7 (48.4, 57.5)
Gc2		
no copies	2652	56.6 (55.6, 57.6)
1 copy	2240	51.9 (50.9, 52.9)
2 copies	438	49.2 (47.2, 51.3)

Table 2 Copies of the *Gc* genotypes (haplotype pairs) with mean 25(OH)D concentrations in the 1958BC

Gc* Genotype** ***n **25(OH)D, nmol geometric mean (95% CI)** — First haplotype	***Gc1s***	***Gc1f***	***Gc2***	***Gcx***
Second haplotype				
Gc1s	1600 57.1 (55.9, 58.4)	0 -	0 -	0 -
Gc1f	902 56.4 (54.7, 58.1)	132 52.7 (48.4, 57.5)	0 -	0 -
Gc2	1785 52.0 (50.9, 53.1)	444 51.8 (49.7, 54.1)	438 49.2 (47.2, 51.3)	11 35.9 (25.9, 49.8)
Gcx	14 51.8 (40.9, 65.7)	0 -	0 -	4 41.2 (16.5, 102.7)

The *Gc* haplotype pairs explained the most variation in 25(OH)D concentrations at 1.26%, then the *Gc* genotypes at 1.25% compared to the other *Gc* variables (Table 3). However, since the haplotype pairs and genotypes were fitted as category variables in the models, the BIC statistic was larger than many of the other model fits. Copies of the *Gc2* variable fitted as a trend had the smallest BIC and largest value of *F* compared to all other coding with only a modest decrease in the variation explained of 25(OH)D. On balance of the three statistics used, *Gc2* fitted as trend had the best fit for the purposes of an instrument for 25(OH)D in instrumental variable analysis.

Table 3 Variation explained in 25(OH)D concentrations, *F*-statistic and BIC by combinations of Gc genotypes/haplotypes in 1958BC

Gc* genotype/haplotype variable(s)	**Degrees of freedom for variable(s)**	**R^2, %**	***F*-statistic**	**BIC**
Gc genotypes	8	1.25	56.3	6773.7
Gc genotypes as trend	1	1.01	9.7	6733.4
Gc haplotype pairs mutually adjusted	6	1.26	12.6	6758.1
Gc haplotype pairs as trend mutually adjusted	2	1.20	34.0	6730.9
Copies of *Gc1s*	2	0.82	23.6	6751.3
Copies of *Gc1s* as trend	1	0.84	46.9	6742.8
Copies of *Gc1f*	2	0.01	1.3	6794.9
Copies of Gc1f as trend	1	0.00	0.9	6787.7
Copies of *Gc2*	2	1.14	31.8	6733.9
Copies of *Gc2* as trend	1	1.14	62.6	6726.4

Regression diagnostics for transformed variables

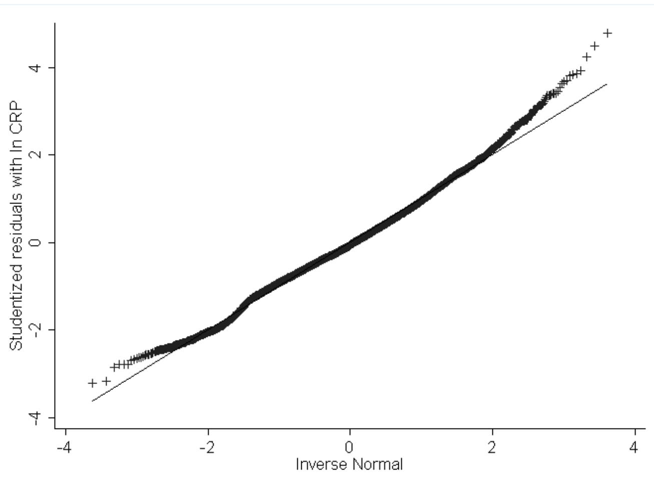
Model diagnostics were used to assess whether it was appropriate to transform the inflammatory/hemostatic outcomes and 25-hydroxyvitamin D with the natural logarithm in the subsequent instrumental variable analysis.

Diagnostics for natural log transformed inflammatory/haemostatic outcomes

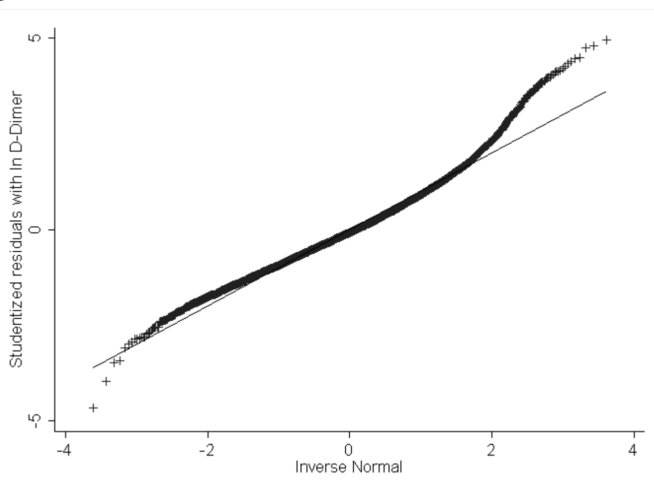
To assess the adequacy of model fit with the natural log transformed outcomes of C-reactive protein (CRP), D-Dimer, fibrinogen, tissue plasminogen activator (tPA) and von Willebrand factor (vWF), the studentised residuals were predicted from linear regression models for each of the outcomes adjusted for sex, month of measure, body mass index and geographical region. As the studentized residuals should be normally distributed with mean 0 and standard deviation 1 (Armitage & Berry 1994), these were plotted in a normal probability plot and inspected for strong deviations from normality.

The normality probability plots for each of the outcomes (Figure 2) showed some departures from normality at the tails; however for all of the outcomes the plots did appear to be satisfactory.

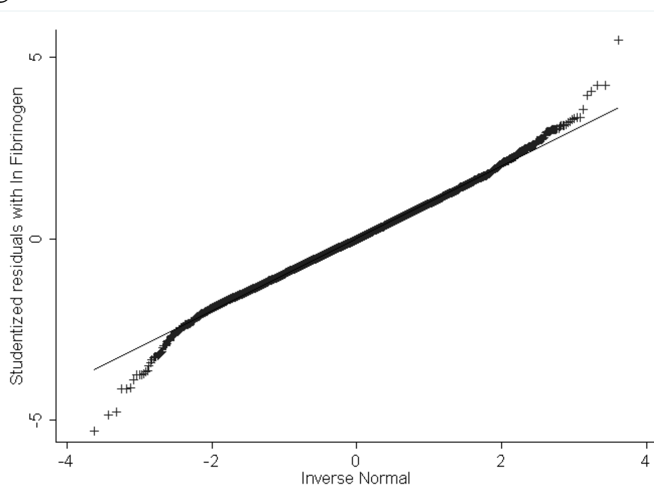
A



B



C



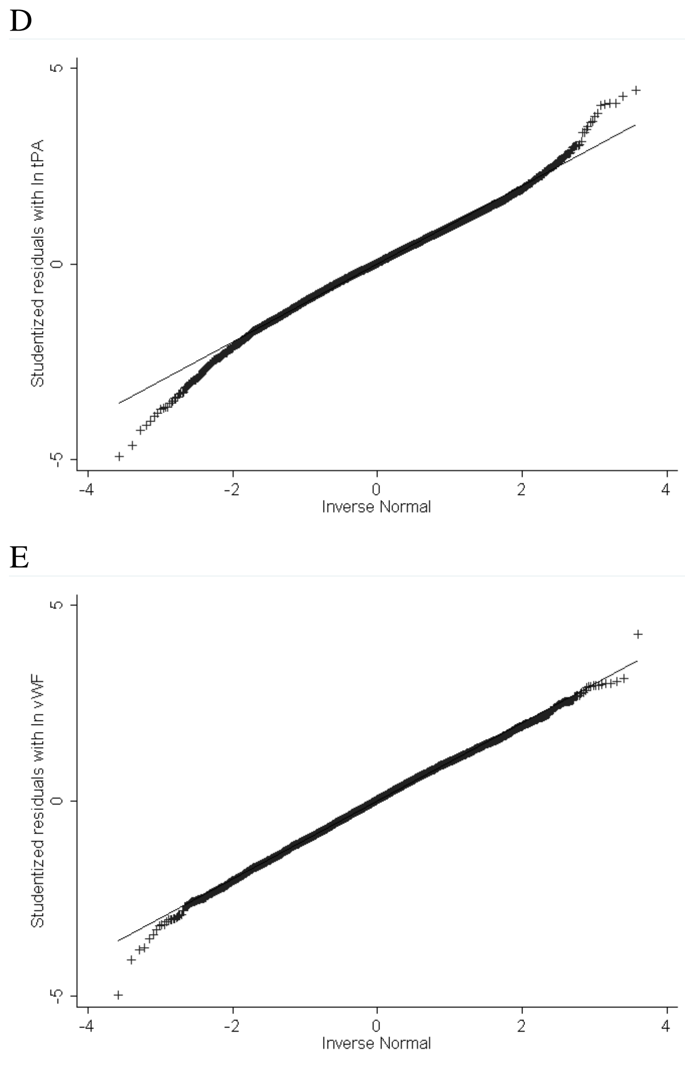


Figure 2 Normal probability plots of the studentised residuals from linear regression models (adjusted for sex, BMI, month and region) with the outcomes A) CRP, B) D-dimer, C) fibrinogen, D) tPA, E) vWF. The black line is the cumulative probability of a standard normal random variable.

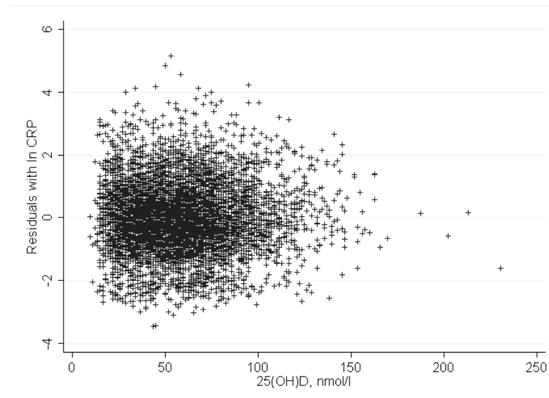
Diagnostics for natural log transformed 25(OH)D vs 25(OH)D covariate

After establishing the satisfactory model fit with the natural log outcomes, linear regression models for each of the outcomes were run twice with the raw 25(OH)D concentrations and natural logarithm transformed 25(OH)D. In both instances, the linear regression models were adjusted for sex, month of measure, body mass index

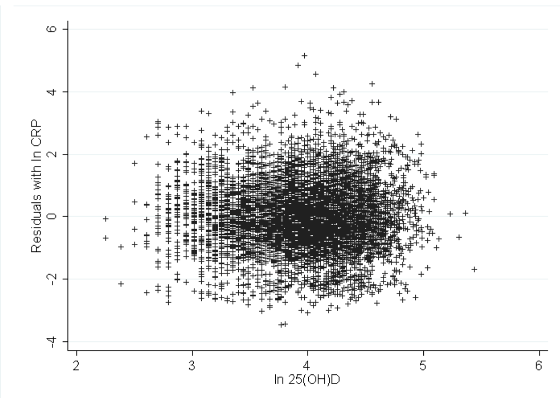
and geographical region. The residuals were predicted from the models and plotted against 25(OH)D covariate used in the model, and inspected for a non-linear trend. Furthermore, a quadratic term of 25(OH)D was created and used as additional covariate in the linear regression models for each of the outcomes in the above model, adjusted for the non-transformed 25(OH)D concentrations. The likelihood ratio test was used to assess whether there was evidence of curvature of the 25(OH)D concentrations as modelled by its quadratic term.

In the residuals plots vs the 25(OH)D covariate (untransformed) for each of the outcomes, there was indication of clustering on the left-hand side of the plots (Figure 3-1). Furthermore, for each of the outcomes there was evidence of a curved 25(OH)D association ($p \leq 0.033$ for all outcomes, adjusted for sex, month of measure, BMI and region). In the residuals plots vs ln 25(OH)D covariate, the clustering had improved for each of the outcomes and the plots appear visually random (Figure 3-2).

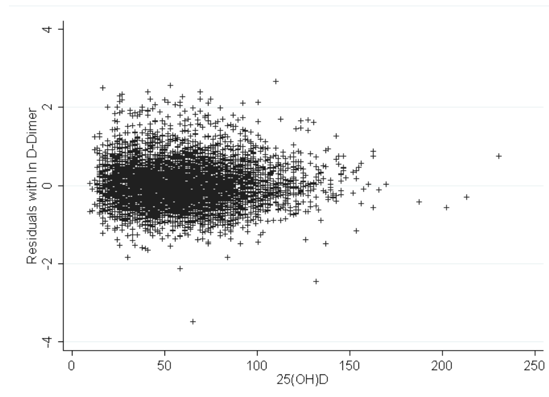
A-1



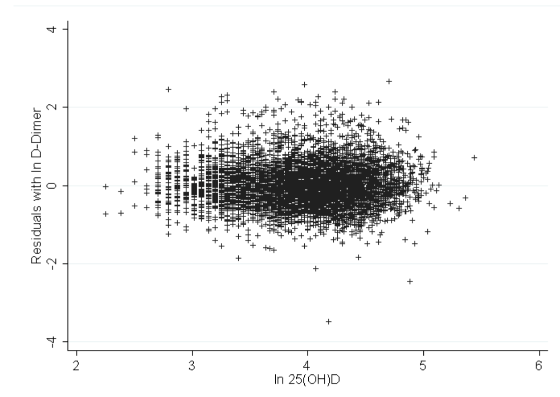
A-2



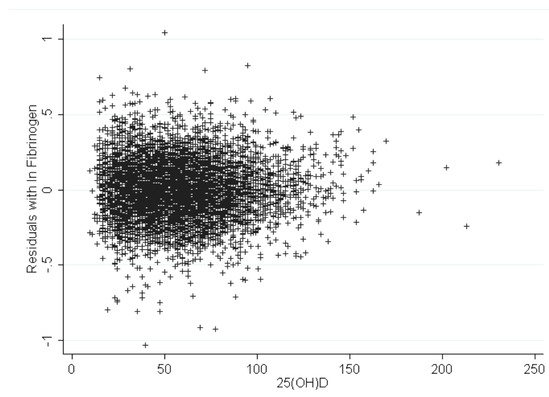
B-1



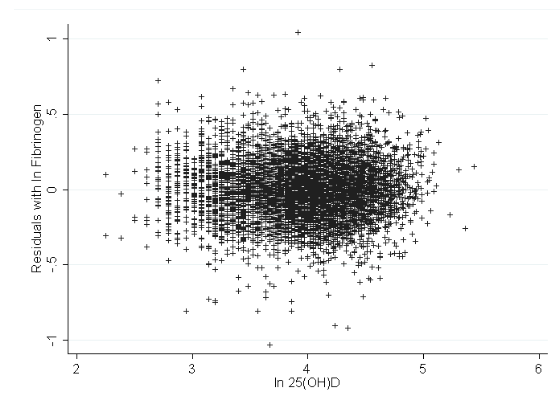
B-2



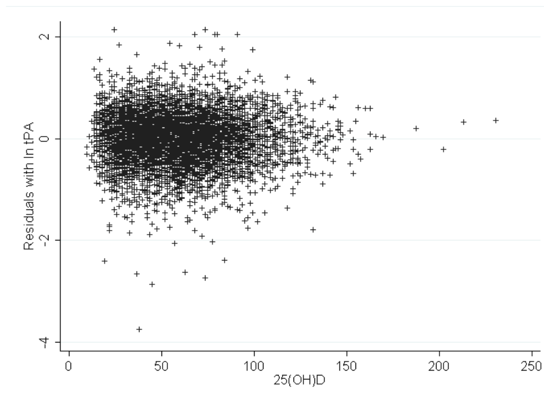
C-1



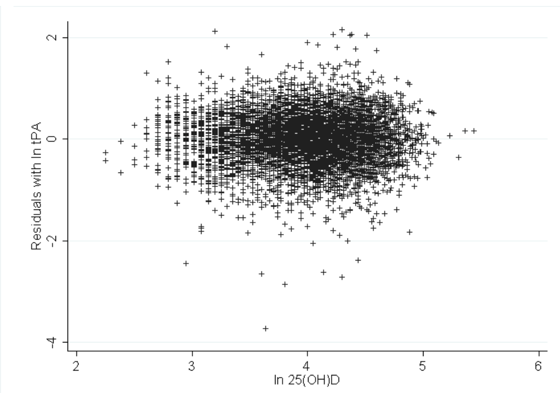
C-2



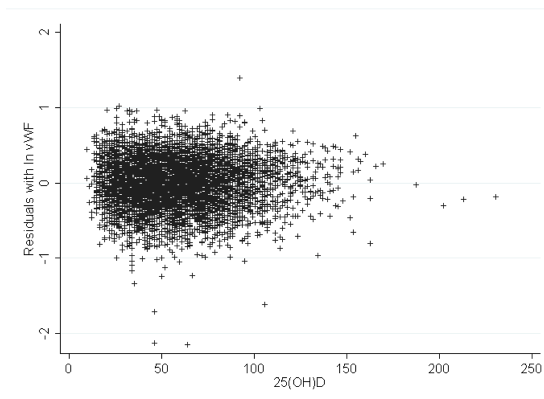
D-1



D-2



E-1



E-2

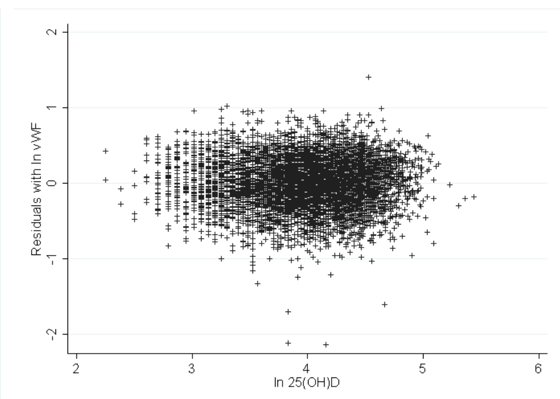


Figure 3 The model residuals of linear regression models (adjusted for sex, BMI, month and region) with the outcomes A) CRP, B) D-dimer, C) fibrinogen, D) tPA, E) vWF vs the 1) 25(OH)D covariate non-transformed 2) and natural logarithm transformed.

Appendix 5: Additional figures for Chapter 5

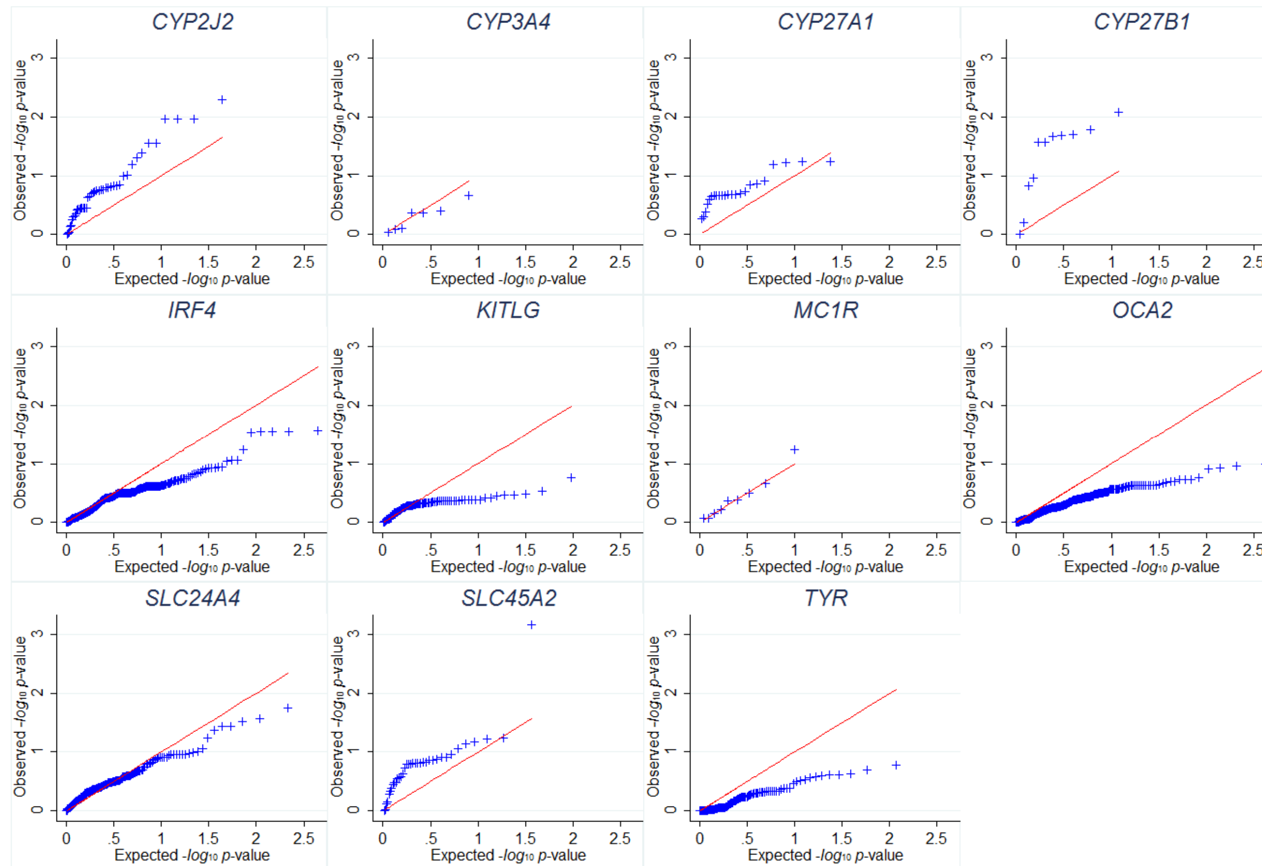
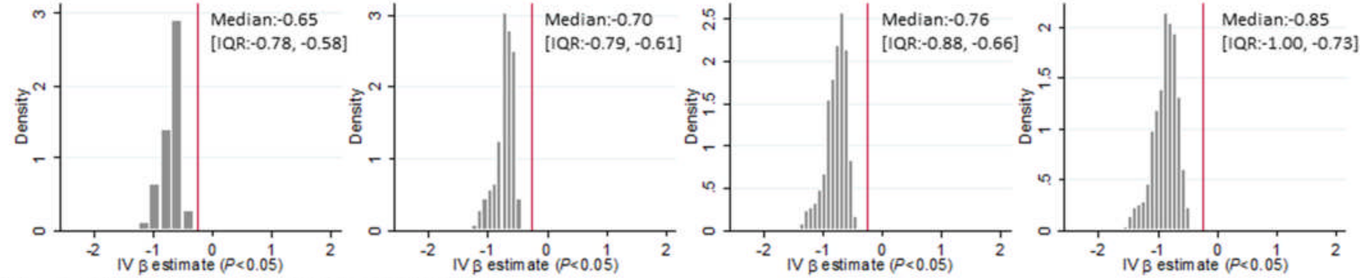


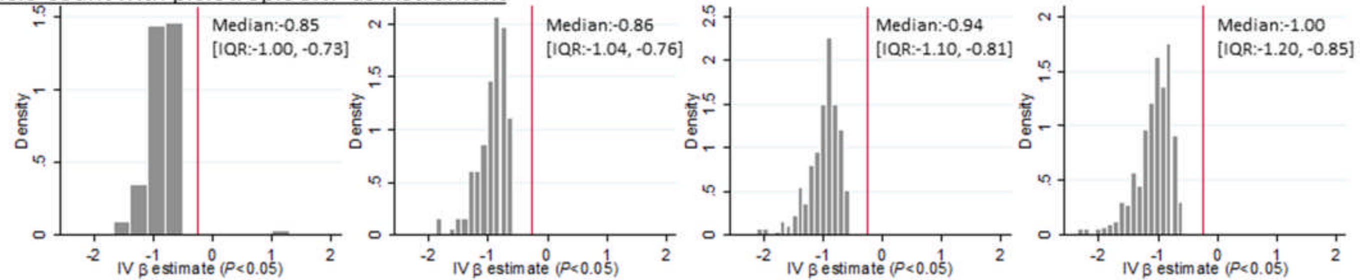
Figure 1 QQ Plots by genes of the SNPs associations with 25(OH)D, where the blue dots are the observed log *p*-values and red line is their expected log values

True β=-0.25 and sample size n=5000

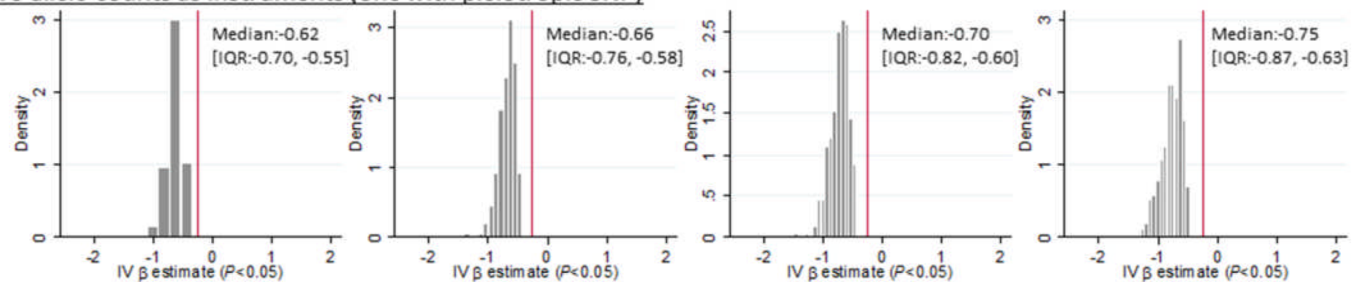
Individual SNP as instrument



Allele count with pleiotropic SNP as instrument



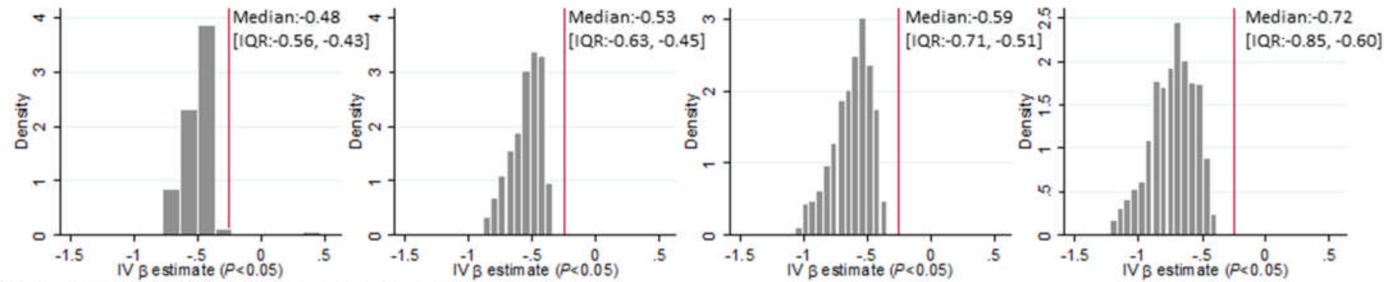
Two allele counts as instruments (one with pleiotropic SNP)



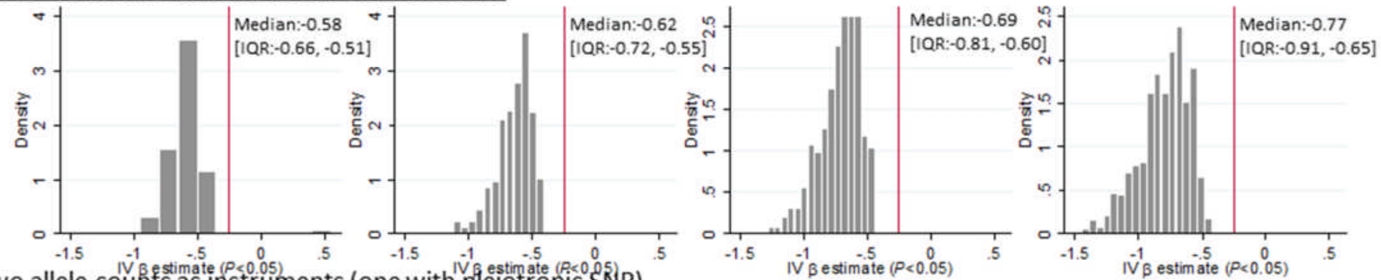
No pleiotropy | Small pleiotropic effect | Moderate pleiotropic effect | Large pleiotropic effect

True β=-0.25 and sample size n=10000

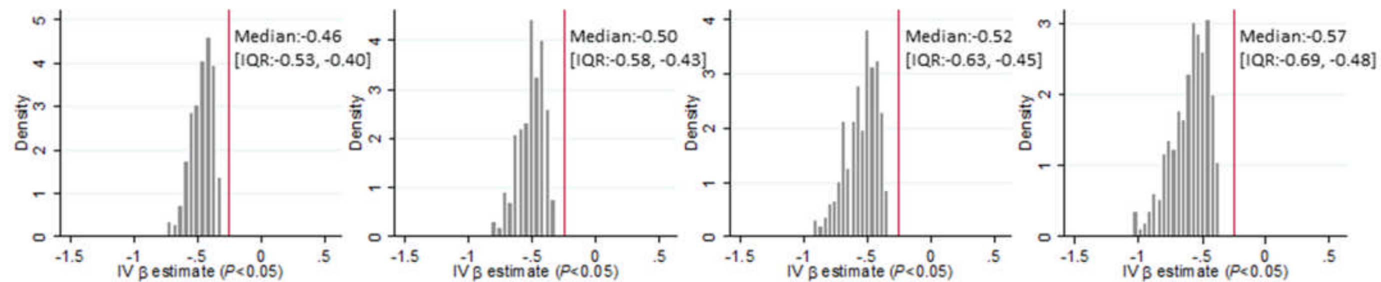
Individual SNP as instrument



Allele count with pleiotropic SNP as instrument



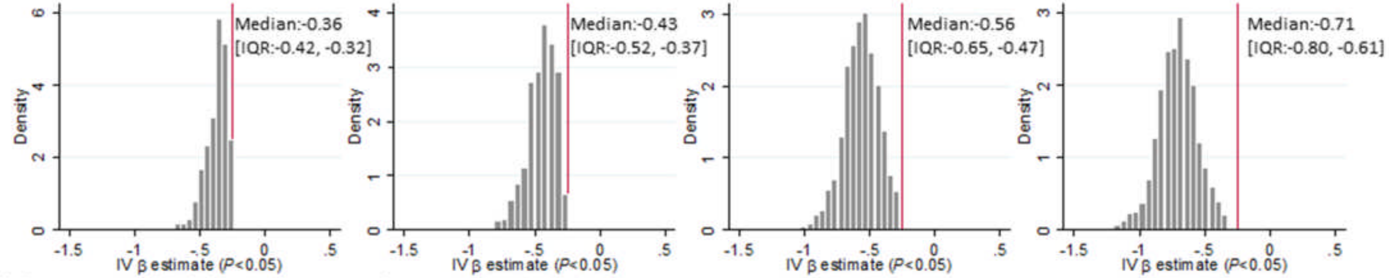
Two allele counts as instruments (one with pleiotropic SNP)



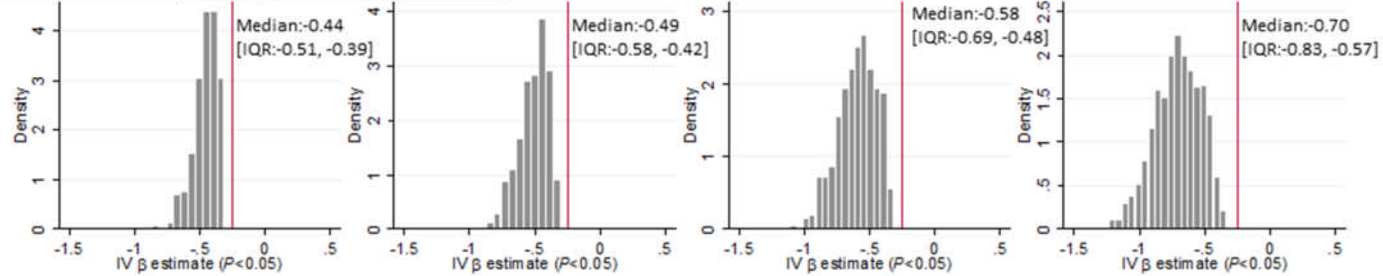
No pleiotropy | Small pleiotropic effect | Moderate pleiotropic effect | Large pleiotropic effect

True β=-0.25 and sample size n=20000

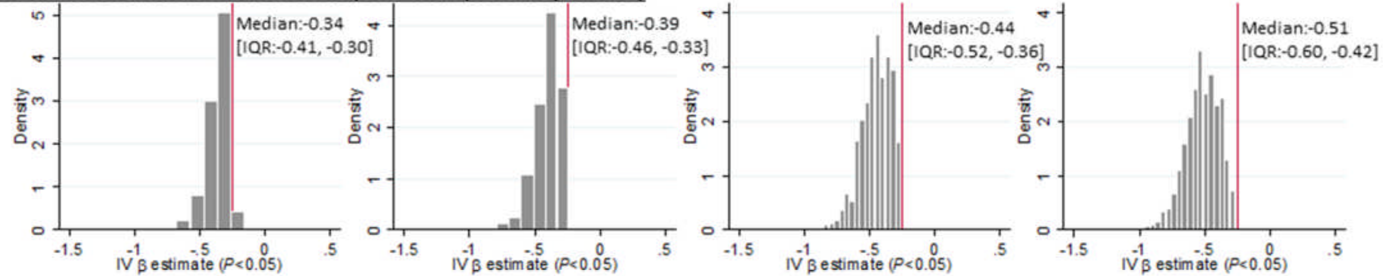
Individual SNP as instrument



Allele count with pleiotropic SNP as instrument



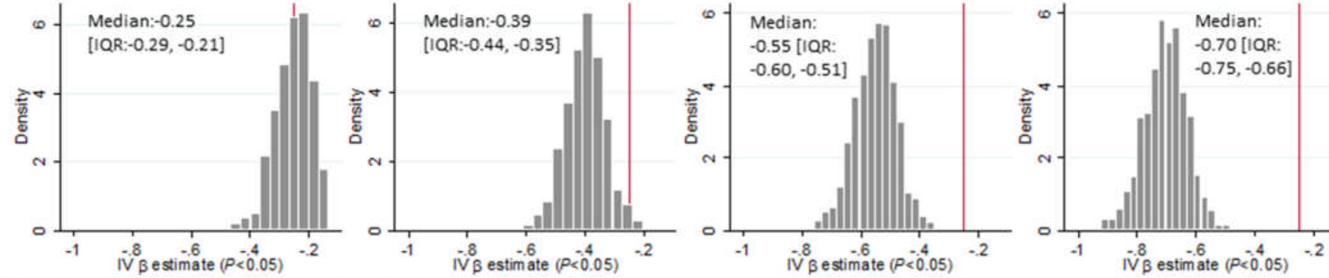
Two allele counts as instruments (one with pleiotropic SNP)



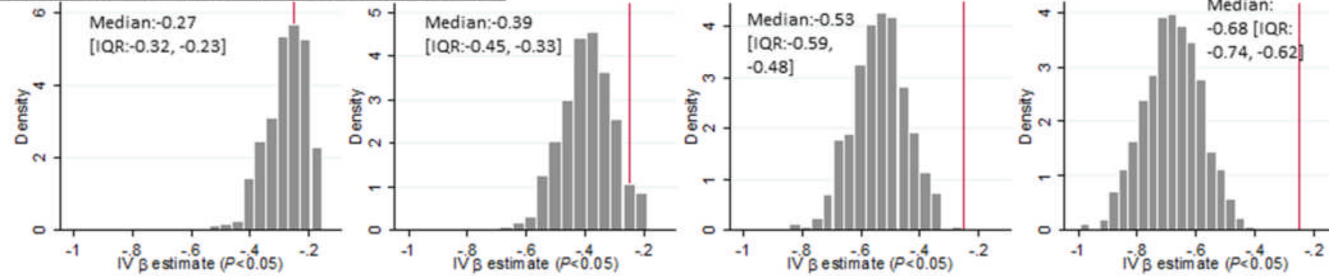
No pleiotropy | Small pleiotropic effect | Moderate pleiotropic effect | Large pleiotropic effect

True β=-0.25 and sample size n=80000

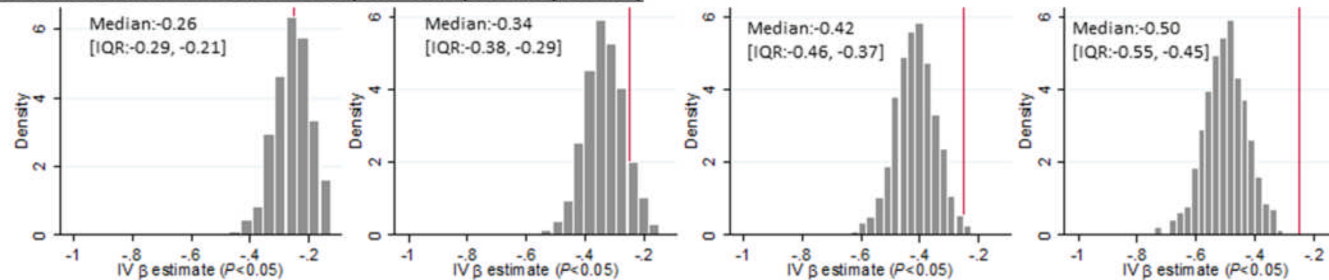
Individual SNP as instrument



Allele count with pleiotropic SNP as instrument



Two allele counts as instruments (one with pleiotropic SNP)



No pleiotropy | Small pleiotropic effect | Moderate pleiotropic effect | Large pleiotropic effect

Figure 2 The effect of a pleiotropic SNP on the instrumental variable (IV) β estimate when used as a a) single instrument, b) in a allele count as in instrument, c) in a allele count as one of two instruments in two-stage least squares (2SLS) regression.

Appendix 6: Additional tables for Chapter 5

Table 1 SNP association with biomarkers adjusted for sex, geographical location and 25(OH)D in the 1958BC

Biomarker	Gene	SNP	SNP Coefficient	Standard Error	Unadjusted *p*-value (adjusted)
Haemostatic markers					
D-Dimer	*SLC45A2*	rs16892096	0.061	0.046	0.18
	DHCR7	rs12785878	0.005	0.012	0.66
	CYP2R1	rs10741657	0.005	0.010	0.65
	GC	rs7041	0.008	0.011	0.44
	GC	rs4588	0.003	0.011	0.80
	CYP27B1	rs10877012	0.007	0.010	0.48
	CYP24A1	rs6013897	-0.006	0.012	0.63
Fibrinogen	*SLC45A2*	rs16892096	0.016	0.017	0.37
	DHCR7	rs12785878	0.001	0.004	0.90
	CYP2R1	rs10741657	0.003	0.004	0.40
	GC	rs7041	-0.001	0.004	0.86
	GC	rs4588	-0.002	0.004	0.66
	CYP27B1	rs10877012	0.002	0.004	0.57
	CYP24A1	rs6013897	-0.002	0.005	0.61
Tissue plasminogen activator	*SLC45A2*	rs16892096	0.000	0.045	0.99
	DHCR7	rs12785878	-0.016	0.011	0.17
	CYP2R1	rs10741657	-0.009	0.010	0.34
	GC	rs7041	-0.003	0.010	0.78
	GC	rs4588	0.007	0.011	0.51
	CYP27B1	rs10877012	-0.009	0.010	0.38
	CYP24A1	rs6013897	-0.013	0.012	0.27
von	*SLC45A2*	rs16892096	0.004	0.028	0.89

Biomarker	Gene	SNP	SNP Coefficient	Standard Error	Unadjusted *p*-value (adjusted)
Willebrand factor					
	DHCR7	rs12785878	-0.008	0.007	0.30
	CYP2R1	rs10741657	0.011	0.006	0.096
	GC	rs7041	-0.006	0.007	0.32
	GC	rs4588	-0.005	0.007	0.43
	CYP27B1	rs10877012	0.005	0.006	0.44
	CYP24A1	rs6013897	0.000	0.007	0.97
Inflammation marker					
C-reactive protein	*SLC45A2*	rs16892096	-0.007	0.101	0.95
	DHCR7	rs12785878	0.020	0.026	0.44
	CYP2R1	rs10741657	0.000	0.023	1.00
	GC	rs7041	-0.038	0.024	0.11
	GC	rs4588	-0.051	0.024	0.0361 (0.61)
	CYP27B1	rs10877012	0.017	0.022	0.44
	CYP24A1	rs6013897	-0.046	0.027	0.084 (0.99)
Lipid markers					
Triglycerides	*SLC45A2*	rs16892096	0.068	0.047	0.15
	DHCR7	rs12785878	-0.010	0.012	0.38
	CYP2R1	rs10741657	0.008	0.010	0.47
	GC	rs7041	-0.008	0.011	0.47
	GC	rs4588	-0.003	0.011	0.81
	CYP27B1	rs10877012	0.007	0.010	0.48
	CYP24A1	rs6013897	0.007	0.012	0.54
Low density lipoproteins	*SLC45A2*	rs16892096	-0.077	0.023	0.001 (0.019)
	DHCR7	rs12785878	-0.011	0.006	0.072 (0.99)
	CYP2R1	rs10741657	-0.006	0.005	0.28

Biomarker	Gene	SNP	SNP Coefficient	Standard Error	Unadjusted *p*-value (adjusted)
	GC	rs7041	0.006	0.005	0.28
	GC	rs4588	-0.004	0.006	0.50
	CYP27B1	rs10877012	0.004	0.005	0.47
	CYP24A1	rs6013897	-0.003	0.006	0.62
High density lipoproteins	*SLC45A2*	rs16892096	-0.014	0.019	0.48
	DHCR7	rs12785878	-0.002	0.005	0.62
	CYP2R1	rs10741657	0.000	0.004	0.95
	GC	rs7041	0.000	0.005	0.99
	GC	rs4588	0.000	0.005	0.94
	CYP27B1	rs10877012	0.001	0.004	0.77
	CYP24A1	rs6013897	-0.003	0.005	0.52
Total cholesterol	*SLC45A2*	rs16892096	-0.030	0.015	0.042 (0.72)
	DHCR7	rs12785878	-0.006	0.004	0.097 (0.99)
	CYP2R1	rs10741657	-0.003	0.003	0.39
	GC	rs7041	0.001	0.003	0.75
	GC	rs4588	-0.001	0.004	0.73
	CYP27B1	rs10877012	0.003	0.003	0.37
	CYP24A1	rs6013897	-0.002	0.004	0.70
Hypertension markers					
Diastolic	*SLC45A2*	rs16892096	-0.004	0.011	0.73
	DHCR7	rs12785878	0.003	0.003	0.34
	CYP2R1	rs10741657	-0.004	0.002	0.12
	GC	rs7041	-0.002	0.003	0.54
	GC	rs4588	-0.004	0.003	0.17
	CYP27B1	rs10877012	-0.002	0.002	0.50
	CYP24A1	rs6013897	-0.005	0.003	0.069 (0.99)
Systolic	*SLC45A2*	rs16892096	-0.004	0.010	0.65
	DHCR7	rs12785878	0.003	0.002	0.30

Biomarker	**Gene**	**SNP**	**SNP Coefficient**	**Standard Error**	**Unadjusted *p*-value (adjusted)**
	CYP2R1	rs10741657	0.000	0.002	0.89
	GC	rs7041	-0.002	0.002	0.46
	GC	rs4588	-0.003	0.002	0.26
	CYP27B1	rs10877012	-0.001	0.002	0.67
	CYP24A1	rs6013897	-0.005	0.003	0.034 (0.58)
Glucose related marker					
HbA1c	*SLC45A2*	rs16892096	-0.002	0.048	0.97
	DHCR7	rs12785878	0.001	0.012	0.95
	CYP2R1	rs10741657	0.014	0.011	0.20
	GC	rs7041	0.006	0.011	0.59
	GC	rs4588	0.007	0.012	0.56
	CYP27B1	rs10877012	0.001	0.010	0.95
	CYP24A1	rs6013897	0.004	0.013	0.76
Lung function marker					
FEV_1	*SLC45A2*	rs16892096	0.032	0.009	0.0002 (0.005)
	DHCR7	rs12785878	0.000	0.002	0.94
	CYP2R1	rs10741657	0.000	0.002	0.90
	GC	rs7041	-0.002	0.002	0.27
	GC	rs4588	-0.001	0.002	0.56
	CYP27B1	rs10877012	0.000	0.002	0.88
	CYP24A1	rs6013897	0.001	0.002	0.72
Allergy marker					
IgE	*SLC45A2*	rs16892096	0.136	0.119	0.25
	DHCR7	rs12785878	-0.010	0.030	0.75
	CYP2R1	rs10741657	0.058	0.027	0.032 (0.45)
	GC	rs7041	0.027	0.028	0.34
	GC	rs4588	-0.016	0.029	0.59

Biomarker	Gene	SNP	SNP Coefficient	Standard Error	Unadjusted *p*-value (adjusted)
	CYP27B1	rs10877012	0.047	0.026	0.065 (0.99)
	CYP24A1	rs6013897	0.016	0.031	0.60
Growth related marker					
IGF-1	*SLC45A2*	rs16892096	-0.016	0.025	0.53
	DHCR7	rs12785878	0.001	0.006	0.90
	CYP2R1	rs10741657	-0.002	0.006	0.70
	GC	rs7041	0.002	0.006	0.69
	GC	rs4588	0.012	0.006	0.061 (0.99)
	CYP27B1	rs10877012	-0.008	0.005	0.12
	CYP24A1	rs6013897	-0.001	0.007	0.84
Adiposity markers					
BMI	*SLC45A2*	rs16892096	-0.001	0.014	0.95
	DHCR7	rs12785878	-0.002	0.004	0.50
	CYP2R1	rs10741657	-0.006	0.003	0.069 (0.99)
	GC	rs7041	-0.001	0.003	0.69
	GC	rs4588	0.000	0.003	0.98
	CYP27B1	rs10877012	0.001	0.003	0.74
	CYP24A1	rs6013897	0.000	0.004	0.99
Waist circumference	*SLC45A2*	rs16892096	0.001	0.010	0.94
	DHCR7	rs12785878	-0.003	0.003	0.25
	CYP2R1	rs10741657	-0.005	0.002	0.047 (0.79)
	GC	rs7041	-0.003	0.002	0.19
	GC	rs4588	-0.003	0.002	0.21
	CYP27B1	rs10877012	0.000	0.002	0.88
	CYP24A1	rs6013897	0.000	0.003	0.89

Table 2 Synthesis count, metabolism scores and *Gc2* haplotype associations with biomarkers adjusted for sex, geographical location and 25(OH)D in the 1958BC

Biomarker	Combined genetic marker	SNP of genes in marker	Combined marker coefficient	Standard Error	Unadjusted *p*-value (adjusted)
Hemostasis markers					
D-Dimer	Synthesis count	*DHCR7, CYP2R1*	0.006	0.008	0.48
	Metabolism score GWAS	*GC, CYP24A1*	0.007	0.027	0.80
	Gc2 haplotype	*GC*	0.006	0.012	0.64
	Metabolism score non-GWAS	*GC, CYP27B1, CYP24A1*	0.017	0.032	0.59
Fibrinogen	Synthesis count	*DHCR7, CYP2R1*	0.001	0.003	0.83
	Metabolism score GWAS	*GC, CYP24A1*	-0.007	0.010	0.48
	Gc2 haplotype	*GC*	-0.001	0.004	0.84
	Metabolism score non-GWAS	*GC, CYP27B1, CYP24A1*	-0.006	0.012	0.60
Tissue plasminogen activator	Synthesis count	*DHCR7, CYP2R1*	-0.012	0.008	0.13
	Metabolism score GWAS	*GC, CYP24A1*	0.007	0.026	0.78
	Gc2 haplotype	*GC*	0.014	0.011	0.21
	Metabolism score non-GWAS	*GC, CYP27B1, CYP24A1*	0.015	0.031	0.63
von Willebrand factor	Synthesis count	*DHCR7, CYP2R1*	0.005	0.005	0.32
	Metabolism score GWAS	*GC, CYP24A1*	-0.012	0.017	0.49
	Gc2 haplotype	*GC*	-0.004	0.007	0.63

Biomarker	Combined genetic marker	SNP of genes in marker	Combined marker coefficient	Standard Error	Unadjusted *p*-value (adjusted)
	Metabolism score non-GWAS	*GC, CYP27B1, CYP24A1*	-0.003	0.020	0.88
Inflammation marker					
C-reactive protein	Synthesis count	*DHCR7, CYP2R1*	0.004	0.018	0.83
	Metabolism score GWAS	*GC, CYP24A1*	-0.134	0.060	0.025 (0.42)
	Gc2 haplotype	*GC*	-0.046	0.026	0.083
	Metabolism score non-GWAS	*GC, CYP27B1, CYP24A1*	-0.133	0.070	0.058 (0.99)
Lipid markers					
Triglycerides	Synthesis count	*DHCR7, CYP2R1*	0.000	0.008	0.97
	Metabolism score GWAS	*GC, CYP24A1*	-0.004	0.028	0.90
	Gc2 haplotype	*GC*	0.001	0.012	0.95
	Metabolism score non-GWAS	*GC, CYP27B1, CYP24A1*	0.008	0.032	0.80
Low density lipoproteins	Synthesis count	*DHCR7, CYP2R1*	-0.007	0.004	0.075 (0.99)
	Metabolism score GWAS	*GC, CYP24A1*	-0.009	0.014	0.50
	Gc2 haplotype	*GC*	0.000	0.006	0.98
	Metabolism score non-GWAS	*GC, CYP27B1, CYP24A1*	-0.005	0.016	0.74
High density lipoproteins	Synthesis count	*DHCR7, CYP2R1*	-0.002	0.003	0.63
	Metabolism score GWAS	*GC, CYP24A1*	0.002	0.011	0.85
	Gc2	*GC*	-0.002	0.005	0.73

Biomarker	Combined genetic marker	SNP of genes in marker	Combined marker coefficient	Standard Error	Unadjusted *p*-value (adjusted)
	haplotype				
	Metabolism score non-GWAS	*GC, CYP27B1, CYP24A1*	-0.001	0.013	0.95
Total cholesterol	Synthesis count	*DHCR7, CYP2R1*	-0.005	0.003	0.082 (0.99)
	Metabolism score GWAS	*GC, CYP24A1*	-0.003	0.009	0.77
	Gc2 haplotype	*GC*	0.000	0.004	0.95
	Metabolism score non-GWAS	*GC, CYP27B1, CYP24A1*	0.001	0.010	0.90
Hypertension markers					
Diastolic	Synthesis count	*DHCR7, CYP2R1*	-0.002	0.002	0.29
	Metabolism score GWAS	*GC, CYP24A1*	-0.011	0.006	0.089 (0.99)
	Gc2 haplotype	*GC*	-0.003	0.003	0.33
	Metabolism score non-GWAS	*GC, CYP27B1, CYP24A1*	-0.010	0.008	0.19
Systolic	Synthesis count	*DHCR7, CYP2R1*	0.001	0.002	0.67
	Metabolism score GWAS	*GC, CYP24A1*	-0.009	0.006	0.12
	Gc2 haplotype	*GC*	-0.003	0.003	0.27
	Metabolism score non-GWAS	*GC, CYP27B1, CYP24A1*	-0.007	0.007	0.28
Glucose related marker					
HbA1c	Synthesis count	*DHCR7, CYP2R1*	0.000	0.002	0.82
	Metabolism score GWAS	*GC, CYP24A1*	-0.002	0.005	0.68
	Gc2 haplotype	*GC*	0.000	0.002	0.88

Biomarker	Combined genetic marker	SNP of genes in marker	Combined marker coefficient	Standard Error	Unadjusted p-value (adjusted)
	Metabolism score non-GWAS	*GC, CYP27B1, CYP24A1*	-0.002	0.006	0.70
Lung function markers					
FEV_1	Synthesis count	*DHCR7, CYP2R1*	0.010	0.009	0.23
	Metabolism score GWAS	*GC, CYP24A1*	0.019	0.029	0.50
	Gc2 haplotype	*GC*	0.010	0.012	0.41
	Metabolism score non-GWAS	*GC, CYP27B1, CYP24A1*	0.028	0.034	0.41
Allergy marker					
IgE	Synthesis count	*DHCR7, CYP2R1*	0.037	0.021	0.082 (0.99)
	Metabolism score GWAS	*GC, CYP24A1*	-0.020	0.071	0.78
	Gc2 haplotype	*GC*	-0.021	0.031	0.50
	Metabolism score non-GWAS	*GC, CYP27B1, CYP24A1*	0.011	0.083	0.90
Growth related marker					
IGF-1	Synthesis count	*DHCR7, CYP2R1*	-0.001	0.005	0.77
	Metabolism score GWAS	*GC, CYP24A1*	0.023	0.015	0.13
	Gc2 haplotype	*GC*	0.008	0.007	0.21
	Metabolism score non-GWAS	*GC, CYP27B1, CYP24A1*	0.015	0.018	0.40
Adiposity markers					
BMI	Synthesis count	*DHCR7, CYP2R1*	-0.005	0.003	0.032 (0.55)
	Metabolism score GWAS	*GC, CYP24A1*	0.000	0.008	0.98
	Gc2	*GC*	0.001	0.004	0.70

Biomarker	Combined genetic marker	SNP of genes in marker	Combined marker coefficient	Standard Error	Unadjusted *p*-value (adjusted)
	haplotype				
	Metabolism score non-GWAS	*GC, CYP27B1, CYP24A1*	0.001	0.010	0.91
Waist circumference	Synthesis count	*DHCR7, CYP2R1*	-0.005	0.002	0.010 (0.16)
	Metabolism score GWAS	*GC, CYP24A1*	-0.008	0.006	0.21
	Gc2 haplotype	*GC*	-0.001	0.003	0.58
	Metabolism score non-GWAS	*GC, CYP27B1, CYP24A1*	-0.009	0.007	0.20

Table 3 Allele score, count and haplotype association with inflammation/hemostatic outcomes adjusted for sex, month of measure, BMI and geographical region in the 1958BC.

	ln CRP		**ln D-dimer**		**ln Fibrinogen**		**ln TPA**		**ln vWF**	
	n	β (95% CI)	*n*	β (95% CI)	*n*	β (95% CI)	*n*	β (95% CI)	*n*	β (95% CI)
Synthesis allele count *(DCHR7, CYP2R1)*	5383	0.016 (-0.017,0.049)	5369	0.007 (-0.009, 0.024)	5379	0.003 (-0.003, 0.009)	5383	-0.002 (-0.017, 0.012)	5384	0.005 (-0.005, 0.015)
p-value		0.34		0.38		0.39		0.74		0.29
Metabolism allele count GWAS *(GC, CYP24A1)*	5457	-0.104 (-0.213, 0.005)	5441	0.021 (-0.033, 0.075)	5454	-0.005 (-0.025, 0.014)	5457	0.028 (-0.021, 0.076)	5458	-0.006 (-0.039, 0.027)
p-value		0.061		0.45		0.61		0.26		0.72
Gc2 Haplotype *(GC)*	4930	-0.037 (-0.085, 0.011)	4916	0.008 (-0.015, 0.031)	4931	-0.001 (-0.010, 0.008)	4930	0.021 (-0.001, 0.042)	4931	-0.002 (-0.016, 0.012)
p-value		0.13		0.51		0.82		0.057		0.79
Metabolism allele count non-GWAS *(GC, CYP27B1, CYP24A1)*	5175	-0.108 (-0.237, 0.020)	5159	0.024 (-0.039, 0.088)	5172	-0.007 (-0.030, 0.016)	5176	0.036 (-0.021, 0.093)	5176	0.003 (-0.035, 0.042)
p-value		0.098		0.45		0.55		0.22		0.87

Table 4 IV analysis using 2SLS regression for the 25(OH)D association using the genetic proxies as instruments for inflammation/hemostatic biomarkers adjusted for BMI, geographical region, month of measurement and sex in the 1958BC

n	**Instrument for ln 25(OH)D**	**ln 25(OH)D IV Beta (95% CI)**	***IV p*-value**	**Homoscedasticity Pagan Hall *p*-value**	**Over-identified Sargan *p*-value**	**OLS vs IV Durbin Wu Hausman *p*-value**
Outcome: ln CRP						
5383	Synthesis SNPs (*DHCR7*, *CYP2R1*)	-0.416 (-1.267,0.434)	0.34	<0.001	0.91	0.46
5383	Synthesis allele count (*DHCR7*, *CYP2R1*)	-0.406 (-1.242,0.430)	0.34	<0.001	-	0.46
5457	Metabolism GWAS SNPs *(GC, CYP24A1)*	0.529 (0.000,1.059)	0.050	<0.001	0.15	0.015
5457	Metabolism allele score GWAS *(GC, CYP24A1)*	0.495 (-0.035,1.025)	0.067	<0.001	-	0.022
4930	Gc2 haplotype	0.434 (-0.135,1.002)	0.14	<0.001	-	0.055
5175	Metabolism non-GWAS SNPs *(GC, CYP27B1, CYP24A1)*	0.451 (-0.059,0.961)	0.083	<0.001	0.33	0.031
5175	Metabolism allele score non-GWAS *(GC, CYP27B1, CYP24A1)*	0.427 (-0.086,0.941)	0.10	<0.001	-	0.041
4989	Synthesis and metabolism non-GWAS SNPs (*DHCR7*,	0.210 (-0.238,0.659)	0.36	<0.001	0.26	0.19

n	Instrument for ln 25(OH)D	ln 25(OH)D IV Beta (95% CI)	*IV* *p*-value	Homoscedasticity Pagan Hall *p*-value	Over-identified Sargan *p*-value	OLS vs IV Durbin Wu Hausman *p*-value
	CYP2R1, GC, CYP27B1, CYP24A1)					
4989	Synthesis allele count and metabolism allele score non-GWAS (*DHCR7*, *CYP2R1, GC, CYP27B1, CYP24A1*)	0.193 (-0.256,0.642)	0.40	<0.001	0.090	0.22
Outcome: ln D-Dimer						
5369	Synthesis SNPs (*DHCR7*, *CYP2R1*)	-0.151 (-0.575,0.273)	0.49	0.043	0.65	0.66
5369	Synthesis allele count (*DHCR7*, *CYP2R1*)	-0.187 (-0.605,0.231)	0.38	0.052	-	0.53
5441	Metabolism GWAS SNPs *(GC, CYP24A1)*	-0.095 (-0.354,0.165)	0.47	0.034	0.63	0.77
5441	Metabolism allele score GWAS *(GC, CYP24A1)*	-0.101 (-0.361,0.160)	0.45	0.025	-	0.74
4916	Gc2 haplotype	-0.094 (-0.370,0.183)	0.51	0.033	-	0.75
5159	Metabolism non-GWAS SNPs *(GC, CYP27B1, CYP24A1)*	-0.090 (-0.340,0.160)	0.48	0.068	0.92	0.78
5159	Metabolism allele score non-GWAS *(GC, CYP27B1,*	-0.097 (-0.350,0.155)	0.45	0.051	-	0.74

n	Instrument for ln 25(OH)D	ln 25(OH)D IV Beta (95% CI)	*IV* *p*-value	Homoscedasticity Pagan Hall *p*-value	Over-identified Sargan *p*-value	OLS vs IV Durbin Wu Hausman *p*-value
	CYP24A1)					
4975	Synthesis and metabolism non-GWAS SNPs (*DHCR7*, *CYP2R1, GC, CYP27B1, CYP24A1*)	-0.099 (-0.322,0.124)	0.39	0.11	0.97	0.68
4975	Synthesis allele count and metabolism allele score non-GWAS (*DHCR7*, *CYP2R1, GC, CYP27B1, CYP24A1*)	-0.112 (-0.336,0.112)	0.34	0.10	0.62	0.60
Outcome: ln fibrinogen						
5379	Synthesis SNPs (*DHCR7*, *CYP2R1*)	-0.060 (-0.212,0.092)	0.44	0.063	0.35	0.68
5379	Synthesis allele count (*DHCR7*, *CYP2R1*)	-0.065 (-0.214,0.084)	0.39	0.049	-	0.62
5454	Metabolism GWAS SNPs *(GC, CYP24A1)*	0.025 (-0.068,0.118)	0.60	0.022	0.94	0.25
5454	Metabolism allele score GWAS *(GC, CYP24A1)*	0.025 (-0.069,0.118)	0.61	0.016	-	0.25
4931	Gc2 haplotype	0.012 (-0.088,0.111)	0.82	0.11	-	0.42
5172	Metabolism non-GWAS SNPs *(GC, CYP27B1,*	0.027 (-0.063,0.118)	0.56	0.031	0.99	0.23

n	Instrument for ln 25(OH)D	ln 25(OH)D IV Beta (95% CI)	*IV* *p*-value	Homoscedasticity Pagan Hall *p*-value	Over-identified Sargan *p*-value	OLS vs IV Durbin Wu Hausman *p*-value
	CYP24A1)					
5172	Metabolism allele score non-GWAS *(GC, CYP27B1, CYP24A1)*	0.028 (-0.064,0.119	0.55	0.016	-	0.24
4986	Synthesis and metabolism non-GWAS SNPs (*DHCR7*, *CYP2R1, GC, CYP27B1, CYP24A1*)	0.015 (-0.065,0.095)	0.72	0.076	0.66	0.33
4986	Synthesis allele count and metabolism allele score non-GWAS (*DHCR7*, *CYP2R1, GC, CYP27B1, CYP24A1*)	0.013 (-0.067,0.094)	0.74	0.035	0.21	0.35
Outcome: ln tPA						
5383	Synthesis SNPs (*DHCR7*, *CYP2R1*)	0.100 (-0.280,0.480)	0.61	<0.001	0.55	0.21
5383	Synthesis allele count (*DHCR7*, *CYP2R1*)	0.062 (-0.310,0.433)	0.75	<0.001	-	0.29
5457	Metabolism GWAS SNPs *(GC, CYP24A1)*	-0.114 (-0.342,0.114)	0.33	<0.001	0.096	0.82
5457	Metabolism allele score GWAS	-0.132 (-0.361,0.098)	0.260	<0.001	-	0.94

n	Instrument for ln 25(OH)D	ln 25(OH)D IV Beta (95% CI)	*IV* *p*-value	Homoscedasticity Pagan Hall *p*-value	Over-identified Sargan *p*-value	OLS vs IV Durbin Wu Hausman *p*-value
	(GC, CYP24A1)					
4930	Gc2 haplotype	-0.240 (-0.487,0.007)	0.056	<0.001	-	0.40
5176	Metabolism non-GWAS SNPs *(GC, CYP27B1, CYP24A1)*	-0.123 (-0.344,0.099)	0.28	<0.001	0.30	0.91
5176	Metabolism allele score non-GWAS *(GC, CYP27B1, CYP24A1)*	-0.140 (-0.363,0.083)	0.22	<0.001	-	0.97
4990	Synthesis and metabolism non-GWAS SNPs (*DHCR7*, *CYP2R1*, *GC*, *CYP27B1*, *CYP24A1*)	-0.059 (-0.256,0.138)	0.56	<0.001	0.67	0.46
4990	Synthesis allele count and metabolism allele score non-GWAS (*DHCR7*, *CYP2R1*, *GC*, *CYP27B1*, *CYP24A1*)	-0.074 (-0.271,0.123)	0.46	<0.001	0.51	0.56
Outcome: ln vWF						
5384	Synthesis SNPs (*DHCR7*, *CYP2R1*)	-0.081 (-0.336,0.175)	0.54	0.011	0.12	0.57
5384	Synthesis allele count (*DHCR7*, *CYP2R1*)	-0.136 (-0.390,0.117)	0.29	0.010	-	0.31

n	Instrument for ln 25(OH)D	ln 25(OH)D IV Beta (95% CI)	*IV* *p*-value	Homoscedasticity Pagan Hall *p*-value	Over-identified Sargan *p*-value	OLS vs IV Durbin Wu Hausman *p*-value
5458	Metabolism GWAS SNPs *(GC, CYP24A1)*	0.031 (-0.125,0.186)	0.70	0.002	0.74	0.60
5458	Metabolism allele score GWAS *(GC, CYP24A1)*	0.029 (-0.128,0.185)	0.72	0.001	-	0.62
4931	Gc2 haplotype	0.023 (-0.143,0.189)	0.79	0.070	-	0.69
5176	Metabolism non-GWAS SNPs *(GC, CYP27B1, CYP24A1)*	-0.013 (-0.163,0.137)	0.87	0.002	0.69	0.99
5176	Metabolism allele score non-GWAS *(GC, CYP27B1, CYP24A1)*	-0.013 (-0.164,0.139)	0.87	0.002	-	0.98
4990	Synthesis and metabolism non-GWAS SNPs (*DHCR7, CYP2R1, GC, CYP27B1, CYP24A1*)	-0.001 (-0.135,0.133)	0.99	0.019	0.32	0.91
4990	Synthesis allele count and metabolism allele score non-GWAS (*DHCR7, CYP2R1, GC, CYP27B1, CYP24A1*)	-0.016 (-0.150,0.118)	0.81	0.010	0.21	0.91

Table 5 IV analysis using a robust generalised methods of moments (GMM) estimator for the 25(OH)D association using allele counts/scores SNPs as instruments for inflammation/hemostatic biomarkers adjusted for BMI, geographical region, month of measurement and sex in the 1958BC

n	Instrument for ln 25(OH)D	ln 25(OH)D IV Beta (95% CI)	*IV p*-value	Over-identified Hansen *J p*-value
Outcome: ln CRP				
5383	Synthesis SNPs (*DHCR7*, *CYP2R1*)	-0.416 (-1.272,0.440)	0.34	0.91
5457	Metabolism GWAS SNPs *(GC, CYP24A1)*	0.526 (0.001,1.052)	0.049	0.14
5175	Metabolism non-GWAS SNPs *(GC, CYP27B1, CYP24A1)*	0.444 (-0.063,0.951)	0.086	0.31
4989	Synthesis and metabolism non-GWAS SNPs (*DHCR7*, *CYP2R1, GC, CYP27B1, CYP24A1*)	0.214 (-0.236,0.665)	0.35	0.26
4989	Synthesis allele count and metabolism allele score non-GWAS (*DHCR7*, *CYP2R1, GC, CYP27B1, CYP24A1*)	0.201 (-0.249,0.651)	0.38	0.090
Outcome: ln D-Dimer				
5369	Synthesis SNPs (*DHCR7*, *CYP2R1*)	-0.152 (-0.577,0.272)	0.48	0.65
5441	Metabolism GWAS SNPs *(GC, CYP24A1)*	-0.093 (-0.351,0.164)	0.48	0.63
5159	Metabolism non-GWAS SNPs *(GC, CYP27B1, CYP24A1)*	-0.090 (-0.343,0.162)	0.48	0.92
4975	Synthesis and metabolism non-GWAS SNPs (*DHCR7*, *CYP2R1, GC, CYP27B1, CYP24A1*)	-0.101 (-0.330,0.129)	0.39	0.98
4975	Synthesis allele count	-0.112	0.34	0.60

n	Instrument for ln 25(OH)D	ln 25(OH)D IV Beta (95% CI)	*IV* *p*-value	Over-identified Hansen *J p*-value
	and metabolism allele score non-GWAS (*DHCR7*, *CYP2R1, GC, CYP27B1, CYP24A1*)	(-0.342,0.119)		
Outcome: ln fibrinogen				
5379	Synthesis SNPs (*DHCR7*, *CYP2R1*)	-0.059 (-0.210,0.093)	0.45	0.36
5454	Metabolism GWAS SNPs *(GC, CYP24A1)*	0.025 (-0.067,0.117)	0.60	0.94
5172	Metabolism non-GWAS SNPs *(GC, CYP27B1, CYP24A1)*	0.027 (-0.063,0.117)	0.56	0.99
4986	Synthesis and metabolism non-GWAS SNPs (*DHCR7*, *CYP2R1, GC, CYP27B1, CYP24A1*)	0.015 (-0.064,0.094)	0.72	0.66
4986	Synthesis allele count and metabolism allele score non-GWAS (*DHCR7*, *CYP2R1, GC, CYP27B1, CYP24A1*)	0.014 (-0.066,0.093)	0.74	0.21
Outcome: ln tPA				
5383	Synthesis SNPs (*DHCR7*, *CYP2R1*)	0.101 (-0.286,0.489)	0.61	0.56
5457	Metabolism GWAS SNPs *(GC, CYP24A1)*	-0.108 (-0.335,0.118)	0.35	0.094
5176	Metabolism non-GWAS SNPs *(GC, CYP27B1, CYP24A1)*	-0.119 (-0.337,0.099)	0.29	0.29
4990	Synthesis and metabolism non-GWAS SNPs (*DHCR7*, *CYP2R1, GC, CYP27B1, CYP24A1*)	-0.056 (-0.248,0.137)	0.57	0.68
4990	Synthesis allele count	-0.074	0.45	0.53

n	Instrument for ln 25(OH)D	ln 25(OH)D IV Beta (95% CI)	*IV* *p*-value	Over-identified Hansen *J p*-value
	and metabolism allele score non-GWAS (*DHCR7*, *CYP2R1, GC, CYP27B1, CYP24A1*)	(-0.267,0.119)		
Outcome: ln vWF				
5384	Synthesis SNPs (*DHCR7*, *CYP2R1*)	-0.084 (-0.343,0.176)	0.53	0.12
5458	Metabolism GWAS SNPs *(GC, CYP24A1)*	0.031 (-0.126,0.189)	0.70	0.74
5176	Metabolism non-GWAS SNPs *(GC, CYP27B1, CYP24A1)*	-0.012 (-0.164,0.139)	0.87	0.68
4990	Synthesis and metabolism non-GWAS SNPs (*DHCR7*, *CYP2R1, GC, CYP27B1, CYP24A1*)	0.005 (-0.127,0.137)	0.94	0.31
4990	Synthesis allele count and metabolism allele score non-GWAS (*DHCR7*, *CYP2R1, GC, CYP27B1, CYP24A1*)	-0.017 (-0.151,0.116)	0.80	0.22