APPLICATIONS OF MENDELIAN RANDOMIZATION TO THE DISCOVERY AND VALIDATION OF BLOOD BIOMARKERS IN CARDIOMETABOLIC DISEASE

APPLICATIONS OF MENDELIAN RANDOMIZATION TO THE DISCOVERY AND VALIDATION OF BLOOD BIOMARKERS IN CARDIOMETABOLIC DISEASE

By PEDRUM MOHAMMADI-SHEMIRANI, B.Sc. (Hons.)

A Thesis Submitted to the School of Graduate Studies in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

McMaster University DOCTOR OF PHILOSOPHY (2022) Hamilton, Ontario, Canada.

(Medical Sciences)

TITLE: Applications of Mendelian randomization to the discovery and validation of blood biomarkers in cardiometabolic disease

AUTHOR: Pedrum Mohammadi-Shemirani, BSc (Western University)

SUPERVISOR: Dr. Guillaume Paré, MD, MSc

NUMBER OF PAGES: xvii, 238

LAY ABSTRACT

Biological markers associated with disease can inform novel therapeutics or diagnostics but distinguishing causation from correlation is challenging. Mendelian randomization – a technique that leverages random inheritance of genetic variation to infer causality – was used to examine the role of biomarkers in cardiometabolic diseases. First, we implicated lipoprotein(a) as a risk factor for atrial fibrillation that acts independent of atherosclerotic cardiovascular disease. Second, we comprehensively characterized the lifelong effects of testosterone on health outcomes in males, where we found evidence of both beneficial and adverse effects on disease. Finally, we discovered trefoil factor 3 as a diagnostic marker for early-stage chronic kidney disease. Altogether, this thesis demonstrated different applications of Mendelian randomization that showcase its utility as a complementary tool to reveal causal biomarkers, and served to identify biomarkers for cardiometabolic diseases that merit further studies to evaluate their potential benefit on patient care.

ABSTRACT

Peripheral blood biomarkers can inform clinical care and drug development. Establishing causality between biomarker and disease is often critical for such applications, but epidemiological studies are limited due to biases from confounding and reverse causation. Mendelian randomization analysis leverages random inheritance of genetic variants at conception to mimic properties of randomized studies and estimate unconfounded effects between biomarker and disease, or vice-versa. This thesis demonstrates the utility of Mendelian randomization as a complementary tool to elucidate observational studies, predict drug safety and repurposing opportunities, and improve diagnostic biomarkers for cardiometabolic diseases. First, we characterized the hypothesized relationship between lipoprotein(a) and atrial fibrillation. We demonstrated both observed and genetically predicted lipoprotein(a) levels were associated with higher risk of atrial fibrillation across multiple independent cohorts. Importantly, risk was partly mediated independent of atherosclerotic cardiovascular disease, a known consequence of elevated lipoprotein(a) and itself a risk factor for atrial fibrillation. Next, we explored the lifelong effects of endogenous testosterone across a comprehensive set of 461 health outcomes in 161,268 males from the UK Biobank cohort. Using Mendelian randomization analysis, we found higher testosterone had beneficial effects on body composition and bone mineral density but adverse effects on prostate cancer, androgenic alopecia, spinal stenosis, and hypertension. Finally, we applied Mendelian randomization with the intention of discovering biomarkers caused by disease, which are expected to represent markers of early disease. As a proof-of-concept, we applied this framework to identify biomarkers

associated with genetic predisposition to kidney function among 238 biomarkers measured in the ORIGIN trial. We discovered reduced kidney function caused increased trefoil factor 3 and showed its addition to models with known risk factors improved discrimination of incident early-stage chronic kidney disease. Taken together, Mendelian randomization identified biomarkers that warrant further study, with promising implications for screening, prevention, and treatment of different cardiometabolic diseases.

ACKNOWLEDGEMENTS

Success is never achieved alone.

I must first thank my advisor – Dr. Guillaume Paré. I am forever grateful for the opportunities that you have provided me as well as the independence to explore and forge my own path. You are a wise mentor and inspirational role model, and I am privileged to be your student. Likewise, I am fortunate to have two phenomenal committee members, Dr. Hertzel Gerstein and Dr. Gregory Steinberg, who have taught me invaluable lessons and provided generous opportunities to broaden my research experience. I would also like to thank the students, staff, and alumni of GMEL. Amanda, Reina, and GMEL-CRLB staff for your tireless work running lab operations. Dr. Marie Pigeyre, Dr. Matt Lanktree, and other co-authors for offering your insights and constructive feedback. And a special thanks to Jenny Sjaarda, Ricky Lali, and Mike Chong who were generous teachers and made me feel welcome when I first joined the lab.

Finally, I wouldn't be here without the support of my family and friends. Both new friendships that were formed during my studies, as well as old friends and family that have supported me through decades of ups and downs in life. Most importantly, I want to dedicate this work to my parents to whom I owe everything. My father – who instilled in me the importance of education and a desire to seek new challenges throughout life. And my mother – who inspires me to work hard and have confidence in myself, and always strive to be a better person each day. Without your influence, I wouldn't be where I am, or the person I am, today.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ACE – angiotensin converting enzyme
ACR – albumin-to-creatinine ratio
AF – atrial fibrillation
ARB – angiotensin receptor blocker
ASCVD – atherosclerotic cardiovascular disease
AUC – area under the curve
BMD – bone mineral density
BMI – body mass index
BPH – benign prostate hyperplasia
CFT – calculated free testosterone
CI – confidence interval
CKD – chronic kidney disease
CKDGen – Chronic Kidney Disease Genetics
CRP – C-reactive protein
CVD – cardiovascular disease
DNA – deoxyribonucleic acid
EA – effect allele
eGFR – estimated glomerular filtration rate
eGFRcrea – estimated glomerular filtration rate based on creatinine
ELISA – enzyme-linked immunosorbent assay
FDR – false discovery rate
FT – free testosterone
GLP-1 – glucagon-like peptide 1
GRS – genetic risk score
GWAS – genome-wide association study
HbA1c – glycated haemoglobin
HDL – high density lipoprotein
HMG-CoA – 3-hydroxy-3-methyl-glutaryl-coenzyme A
HR – hazard ratio
ICD – International Classification of Diseases
INTERVAL – Efficiency and safety of varying the frequency of whole blood donation
IVW – inverse variance weighted
Kb – kilobases
kDa – kilodalton
KORA – cooperative health research in the region of Augsburg F4
LD – linkage disequilibrium
LDL – low density lipoprotein
Lp(a) – lipoprotein(a)
Lp(a)HORIZON – Assessing the Impact of Lipoprotein (a) Lowering With TQJ230 on Major Cardiovascular Events in Patients With CVD
MI – myocardial infarction
MR – Mendelian randomization

MR-PRESSO – Mendelian Randomization Pleiotropy Residual Sum and Outlier
NRI – net reclassification index
NT-proBNP – N-terminal-pro-hormone brain natriuretic peptide
OA – other allele
OR – odds ratio
ORIGIN – outcome reduction with initial glargine intervention
PCSK9 – proprotein convertase subtilin/kexin type 9
pQTL – protein quantitative trait loci
PRS – polygenic risk score
QMDiab – Qatar Metabolomics Study of Diabetes
RAPS – robust adjusted profile score
RCT – randomized controlled trial
SCALLOP – Systematic and Combined Analysis of Olink Proteins
SD – standard deviation
SE – standard error
SGLT-2 – sodium-glucose transport protein 2
SHBG – sex hormone-binding globulin
SNP – single nucleotide polymorphism
SOMAmer – slow off-rate modified aptamer
T2D – type 2 diabetes
TFF3 – trefoil factor 3
TRAVERSE – A Study to Evaluate the Effect of Testosterone Replacement Therapy on the Incidence of Major Adverse Cardiovascular Events and Efficacy Measures in Hypogonadal Men
TT – total testosterone
UACR – urinary albumin-to-creatinine ratio
UKB – UK Biobank
UMOD – uromodulin

DECLARATION OF ACADEMIC ACHIEVEMENT

FORMAT AND ORGANIZATION OF THESIS

This thesis is prepared in the "sandwich" format as outlined in the School of Graduate Studies' Guide for the Preparation of Theses. It includes a general introduction, an overview of hypotheses and objectives, three independent studies prepared in journal article format, and an overall discussion. The candidate is the first author on all manuscripts. At the time of thesis preparation, Chapter 3 was under review in a peer-reviewed journal, and Chapters 4 and 5 were published.

CONTRIBUTION TO PAPERS WITH MULTIPLE AUTHORSHIP

Chapter 3 (Study 1)

Mohammadi-Shemirani P, Chong M, Narula S, Perrot N, Conen D, Roberts JD, Thériault S, Bossé Y, Lanktree MB, Pigeyre M, Paré G. Elevated lipoprotein(a) as a risk factor for atrial fibrillation: an observational and Mendelian randomization study. Journal of the American College of Cardiology. In press. (2022)

Author Contributions:

PM and GP contributed to the conception and design of the study. **PM** and GP contributed to methodological development. **PM** and MC conducted bioinformatic and/or statistical analyses. **PM**, MC, SN, NP, DC, JDR, ST, YB, MBL, MP, and GP contributed to data acquisition, analysis, and/or interpretation of data. GP contributed to funding or data acquisition. GP facilitated project administration and supervision. **PM** was the principal writer of the manuscript. All authors contributed to the drafting and revision of the final article. All authors approved the final submitted version of the manuscript.

Chapter 4 (Study 2)

Mohammadi-Shemirani P, Chong M, Pigeyre M, Morton RW, Gerstein HC, Paré G. Effects of lifelong testosterone exposure on health and disease using Mendelian randomization. eLife. 9:e58914. (2020)

Author Contributions:

PM and GP contributed to the conception and design of the study. **PM** and GP contributed to methodological development. **PM** and MC conducted bioinformatic and/or statistical analyses. **PM**, MC, MP, RWM, HCG, and GP contributed to data acquisition, analysis, and/or interpretation of data. GP contributed to funding or data acquisition. GP facilitated project administration and supervision. **PM** was the principal writer of the manuscript. All authors contributed to the drafting and revision of the final article. All authors approved the final submitted version of the manuscript.

Chapter 5 (Study 3)

Mohammadi-Shemirani P, Sjaarda J, Gerstein HC, Treleaven DJ, Walsh M, Mann JF, McQueen MJ, Hess S, Paré G. A Mendelian Randomization-Based Approach to Identify Early and Sensitive Diagnostic Biomarkers of Disease. Clinical Chemistry. 65(3)427-436. (2019).

Author Contributions:

PM and GP contributed to the conception and design of the study. **PM** and GP contributed to methodological development. **PM** and JS conducted bioinformatic and/or statistical analyses. **PM**, JS, HCG, DJT, MW, JFM, MJM, SH, and GP contributed to data acquisition, analysis, and/or interpretation of data. HCG, MJM and GP contributed to funding or data acquisition. GP facilitated project administration and supervision. **PM** was the principal writer of the manuscript. All authors contributed to the drafting and revision of the final article. All authors approved the final submitted version of the manuscript.

CHAPTER 1:

INTRODUCTION

CHAPTER 1: INTRODUCTION

1.1 CARDIOMETABOLIC DISEASE OVERVIEW

Multimorbidity, the co-occurrence of multiple chronic diseases in a single patient, is growing in prevalence and often observed among type 2 diabetes, cardiovascular diseases, chronic kidney disease, and their clinical correlates (Han et al., 2021) (Mechanick et al., 2020) (Busija et al., 2019). Indeed, there is ample evidence supporting a causal relationship between dysglycemia and cardiovascular diseases (Gan et al., 2019) (Ross et al., 2015). Therefore, cardiometabolic disease is a broad categorization that encompasses these interrelated cardiovascular and metabolic diseases and their sequelae.

1.1.1 EPIDEMIOLOGY AND BURDEN OF DISEASE

Ischaemic heart disease, stroke, and type 2 diabetes are among the leading causes of death and disability globally. Cardiovascular diseases were the leading cause of global disease burden with 523 million cases and 18.6 million deaths in 2019 (Roth et al., 2020), while type 2 diabetes was a top 10 cause of global disease burden with 449 million cases and 1.47 million deaths in 2019 (Vos et al., 2020). Importantly, clinical sequelae of uncontrolled diabetes, such as kidney disease, reflect indirect effects of type 2 diabetes but are themselves major contributors to disease burden as well (Lin et al., 2020). In recent years, the mortality burden has shifted from high-income countries to low- and middle-income countries stressing the growing need for effective interventions (Danaei et al., 2014) (Roth et al., 2020)

1.1.2 RISK FACTORS

In the absence of public health measures, the prevalence of cardiometabolic disease and its risk factors are expected to rise as global populations grow older. Risk factors for cardiometabolic diseases, which include hypertension, elevated blood glucose, elevated cholesterol, and elevated body mass index (BMI), are often comorbid (Danaei et al., 2014). Altogether, these four modifiable risk factors account for 63% of deaths due to cardiovascular disease, diabetes, or chronic kidney disease, or 20% of all deaths, globally (Danaei et al., 2014). However, each cardiometabolic disease has its own specific profile of clinical risk factors or at least their relative importance may differ. For instance, LDL cholesterol is a well-characterized risk factor for myocardial infarction and ischemic stroke, but its role in chronic kidney disease is still unclear (Lanktree et al., 2018) (Rasheed et al., 2021) (Su et al., 2016).

Nevertheless, lifestyle and pharmacological interventions targeting these risk factors are expected to lower cardiometabolic risk. Lipid-lowering therapies are used to reduce risk of atherosclerotic cardiovascular diseases, such as myocardial infarction and ischaemic stroke. Statins function by blocking HMG-CoA reductase and are recommended as first-line therapies to lower LDL cholesterol, followed by ezetimibe or PCSK9 inhibitors in patients that show resistance to statins (Grundy et al., 2019). Meanwhile, newer pharmacological agents are highlighting the role of other lipid fractions, such as lipoprotein(a) and triglycerides, in the development of atherosclerotic disease as well (Tsimikas et al., 2020) (Bhatt et al., 2019). In addition to lifestyle interventions focused on weight loss, several glucose-lowering therapies exist, including metformin,

thiazolidinediones, and sulfonylureas, which lower glucose levels through different mechanisms of action. Among pharmacological agents that reduce risk of type 2 diabetes, metformin is a first-line therapy due to its effectiveness and long-term safety profile (American Diabetes Association, 2018). Furthermore, newer glucose-lowering therapies, such as SGLT-2 inhibitors or GLP-1 receptor agonists, are gaining favour for certain patient groups as they demonstrate ancillary benefits on both glycaemic control and cardiorenal outcomes (Palmer et al., 2021). Finally, antihypertensive agents, such as thiazides, ACE inhibitors, ARBs, beta blockers, and calcium channel blockers, act to lower blood pressure and reduce risks of myocardial infarction, stroke, and heart failure (Reboussin et al., 2018). In patients with chronic kidney disease, ACE inhibitors and ARBs reduce the risk of both kidney failure and cardiovascular events, further highlighting the importance of medications for disease with ancillary benefits on its comorbidities (Xie et al., 2016).

1.2 PERIPHERAL BLOOD BIOMARKERS

Although these risk factors are important and readily available in clinical settings, technological advances are enabling researchers to dive deeper into pathophysiology by exploring beyond traditional risk factors and examining individual biological characteristics. A biomarker can be defined as "a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention" (Atkinson et al., 2001) (Califf, 2018). Naturally, such a broad definition can encompass diverse measurements including molecular, histologic, or radiographic data. For the context of this dissertation, biomarker will refer to circulating biomarkers in peripheral blood, unless otherwise specified.

Peripheral blood is a useful source of biomarkers because it is non-invasive, cost-effective, and reflects both systemic and some tissue-specific conditions (Williams et al., 2019) (Võsa et al., 2021) (Ganz et al., 2016). Biomarkers can provide clinical utility by diagnosing or subtyping disease, monitoring disease status or pharmacological response, informing disease prognosis, and predicting risk of disease or pharmacological response (Califf, 2018).

1.2.1 APPLICATIONS OF BIOMARKERS

Diagnostic biomarkers can reflect both acute and chronic physiological states. Cardiac troponin is an acute marker that is used to rule-out myocardial infarction in patients presenting with symptoms, as it is a structural protein that is unique to the heart that leaks into circulation on myocardial injury (Twerenbold et al., 2017). Likewise, C-reactive protein is an acute phase reactant that increases considerably in response to bacterial infections (Jaye & Waites, 1997), though it often lacks specificity as mild elevations may reflect other inflammatory processes (Kaptoge et al., 2010). On the other hand, HbA1c is a type of hemoglobin molecule that is linked to a sugar and reflects long-term blood glucose concentrations over the past 3 months, such that it is used to diagnose and monitor diabetes (American Diabetes Association, 2012). Likewise, creatinine is a ubiquitous circulating protein that gets filtered by the kidneys and elevated serum creatinine can indicate poor kidney health (Levey et al., 2020). Accordingly, glomerular filtration rate can be estimated from serum creatinine, and it is used to diagnose and stage chronic kidney disease (Levey et al., 2009). In order to better reflect complex disease and improve performance, scores are being developed that incorporate multiple biomarkers (Gerstein et al., 2017) (Ganz et

al., 2016) (Hijazi et al., 2016). However, overfitting is an important limitation that researchers must take care to avoid, so independent replications are needed to evaluate generalizability (Ganz et al., 2016).

Two common metrics that quantify accuracy of a diagnostic biomarker are: sensitivity and specificity. Sensitivity, or true positive rate, is the proportion of individuals with disease that are correctly identified by the test. Specificity, or true negative rate, is the proportion of individuals without disease that are correctly identified by the test (Altman & Bland, 1994). The optimal balance of sensitivity and specificity will depend on various factors such as the intended utility of the test. For instance, a biomarker for a rare condition like cancer may elect to have high specificity and few false positives limiting emotional distress (Califf, 2018). On the other hand, a routine screening test that can occur repeatedly may choose to have high sensitivity limiting false negatives followed by a confirmatory test.

Beyond diagnostics, biomarkers can inform various aspects of therapeutic development as well. Lipoprotein(a) was first described in 1963, and it has since been appreciated as a lipoprotein structurally resembling atherogenic LDL cholesterol (Berg, 1963). Epidemiological evidence has confirmed suspicions regarding its role as a bona fide risk factor for atherosclerotic cardiovascular disease (Paré et al., 2019). However, challenges using traditional therapeutic modalities to modulate lipoprotein(a) levels partly hampered its integration into clinical practice (Berglund & Ramakrishnan, 2004). The recent development of ribonucleic acid-based therapeutics led to antisense oligonucleotides against apolipoprotein(a) that reduced lipoprotein(a) levels in a phase 2 trial and it is

currently being investigated in the Lp(a)HORIZON phase 3 trial for prevention of major cardiovascular events (Tsimikas et al., 2020) (ClinicalTrials.gov Identifier: NCT04023552).

It is important to note that the application of biomarkers to drug target identification has an added challenge in that biomarkers should be causally involved in disease. The correlation of any individual protein with disease may be a spurious association rather than an indication of causality (Davey Smith & Hemani, 2014). For instance, low levels of HDL cholesterol are consistently correlated with higher risk of cardiovascular disease in observational studies (The Emerging Risk Factors Collaboration*, 2009), but randomized controlled trials of pharmacologically increasing HDL showed no benefit for stroke, cardiovascular mortality, or all-cause mortality (Riaz et al., 2019). Therefore, it has been suggested that HDL cholesterol is a marker of cardiovascular health rather than a true causal mediator of cardiovascular disease. (Holmes et al., 2015).

1.2.2 LIMITATIONS OF OBSERVATIONAL STUDIES

Drug development is expensive such that each new drug being brought to market costs $1.3 billion on average (Wouters et al., 2020). The expensive cost is partly owing to the high failure rate in drug development, where approximately 90% of preclinical compounds entering phase 1 clinical trials are unsuccessful in reaching market (Smietana et al., 2016). The majority of failed clinical trials cited either lack of efficacy (52%) or poor safety (24%) as the reason for failure (Harrison, 2016). As clinical trials are expensive and time-consuming efforts, preclinical evidence is critical in reducing costs but current data sources that inform preclinical decision-making have inherent limitations. Although *in*

vitro cell lines and *in vivo* animal models are informative, difficulties recapitulating physiology, biological context, and life course of disease can limit external validity in humans (Pound & Ritskes-Hoitinga, 2018). Additionally, observational studies are limited by inherent biases such as reverse causality or confounding that can result in spurious findings (Davies et al., 2018) (Davey Smith & Hemani, 2014).

Advances in high-throughput DNA sequencing and genotyping have revolutionized genetic epidemiology and the study of genetic determinants of human disease. Drug targets with genetic support for an indication are two-fold more likely to be approved suggesting human genetics is informative for target prioritization (King et al., 2019). Indeed, there are well-known examples of monogenic diseases, such as gain-of-function mutations in *PCSK9* causing familial hypercholesterolemia, that have informed the development of approved therapies (Abifadel et al., 2003). Polygenic diseases can inform drug discovery as well, and genome-wide association studies are instrumental in studying these complex traits by agnostically surveying the human genome for disease-associated genetic variants. The exponential growth in genome-wide association studies has been paralleled by a similar increase in the discovery of novel trait-associated genetic loci (Welter et al., 2014). In the coming years, large-scale national biobanks studies with genetic data linked to electronic healthcare records will result in unprecedented opportunities to discover novel genetic variants associated with many complex diseases and traits (Zhou et al., 2021).

1.3 MENDELIAN RANDOMIZATION

Mendelian randomization is a technique in genetic epidemiology that seeks to estimate the unbiased causal effect of an exposure on an outcome, such as a biomarker and

risk of cardiometabolic disease (Davies et al., 2018). The underlying principle is akin to a randomized controlled trial – the gold-standard for causal inference in health sciences – where randomization of an intervention balances known and unknown confounders across the two arms of the trial (Gerstein et al., 2019). Likewise, genetic alleles have the unique property of being randomly assorted during meiosis in accordance with Mendel's laws of segregation and independent assortment, and genetic variants are fixed at conception providing a safeguard against reverse causation (Davey Smith & Hemani, 2014). Therefore, a genetic variant associated with a biomarker can be thought of as a proxy for that biomarker that is independent of other confounding traits akin to a natural randomized trial for that biomarker (Figure 1.1) (Gill et al., 2021). In fact, Mendelian randomization has replicated numerous findings from randomized controlled trials, such as demonstrating causal roles for LDL cholesterol and dysglycemia on cardiovascular disease risk (Holmes et al., 2015) (Ross et al., 2015).

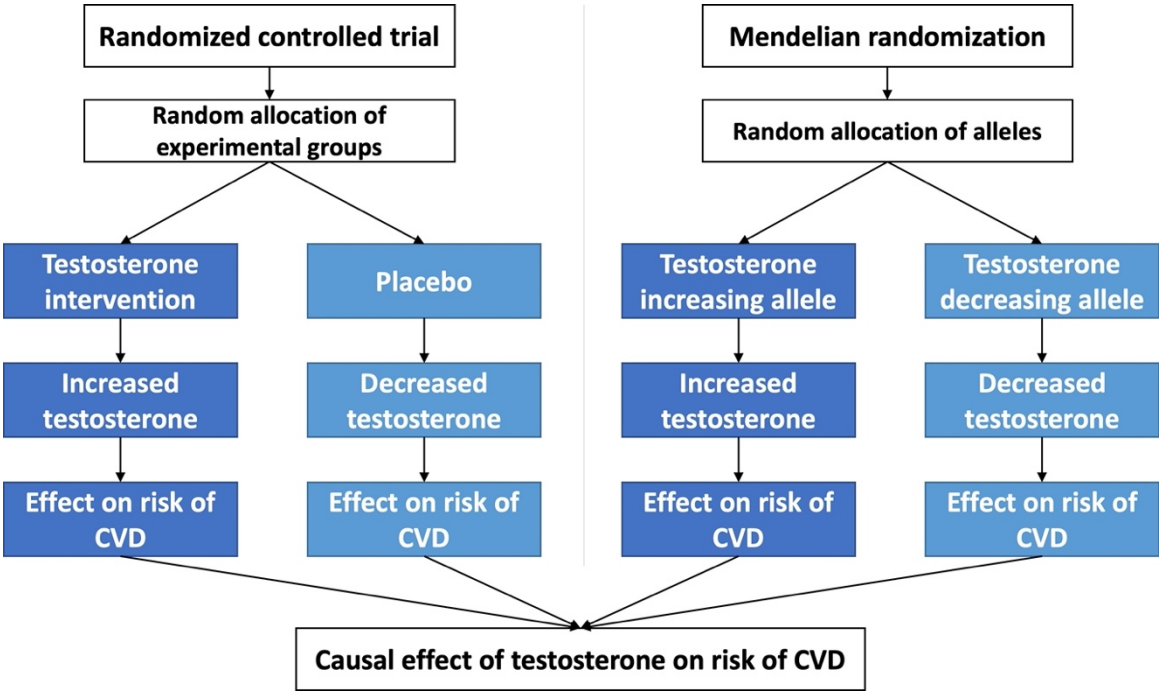


Figure 1.1. Comparison of randomized controlled trial and Mendelian randomization study designs demonstrating the common foundation behind interpretation of a causal effect of testosterone on cardiovascular disease. Due to the random inheritance of genetic alleles, a genetic variant associated with altered testosterone expression or quality is akin to a "natural randomized trial" for testosterone modulation. By the same reasoning, if Mendelian randomization finds genetic variants affecting testosterone are associated with a difference in cardiovascular risk, it provides evidence that testosterone causally affects cardiovascular disease.

1.3.1 LIMITATIONS OF MENDELIAN RANDOMIZATION

In Mendelian randomization, genetic variants are treated as instrumental variables, which are broadly defined as any variables associated with an exposure, unrelated to confounders, and affect the outcome only through the exposure. Therefore, there are

important limitations of Mendelian randomization that must be considered when interpreting findings (Figure 1.2) (Davey Smith & Hemani, 2014). First, genetic variants included in Mendelian randomization analysis must be robustly associated with the exposure. Weak instrument bias can occur if genetic variants do not explain enough variation in the exposure. In a one-sample design, where all genetic effect estimates are derived from the same study participants, results will be biased towards the confounded estimate of the exposure-outcome relationship. In a two-sample design, where genetic estimates of the exposure and outcome are derived from separate samples of participants, results will be biased towards the null (Davies et al., 2018). Even in the absence of weak instrument bias, sufficient statistical power is an important consideration and it is partly a function of the variance explained by the genetic variant. As a result, methods have been developed that combine multiple genetic variants to increase statistical power (Burgess et al., 2016). Second, there must be no confounding of the genetic variant with the outcome. Spurious genetic associations can be confounded by differences in ancestry, a major source of genetic diversity, rather than true genetic effects, so statistical models in genetics adjust for principal components as a covariate. Third, genetic variants must affect the outcome exclusively through the exposure. Horizontal pleiotropy violates this assumption as it describes a state where a genetic variant has an effect on an outcome through a pathway independent of the exposure of interest. However, it is important to note that vertical pleiotropy, where a genetic variant exhibits pleiotropic effects on intermediate exposures that mediate the primary exposure-outcome relationship, does not violate any assumptions because the genetic variant still affects the outcome exclusively through the primary

exposure. The second and third assumptions are difficult to rule-out conclusively, as there can exist unmeasured confounders, but there are some strategies that manage these violations. When investigating protein biomarkers, the likelihood of horizontal pleiotropy can be reduced by choosing genetic variants near the gene encoding for the protein (Gill et al., 2021). Moreover, there is a growing body of statistical literature regarding Mendelian randomization methods that are robust to horizontal pleiotropy and violations of these assumptions (Burgess et al., 2019). Given each method is robust to certain scenarios (e.g., majority of genetic variants must be valid, plurality of genetic variants must be valid, etc.), it is recommended that researchers perform multiple approaches and interpret results holistically. Finally, certain limitations of Mendelian randomization affect its generalizability when predicting efficacy and safety of pharmacological interventions, including: inability to assess off-target effects of drugs, reflecting lifelong changes in endogenous biomarker levels, and assumption of linearity in the exposure-outcome relationship though some methods can allow for non-linearity (Staley & Burgess, 2017).

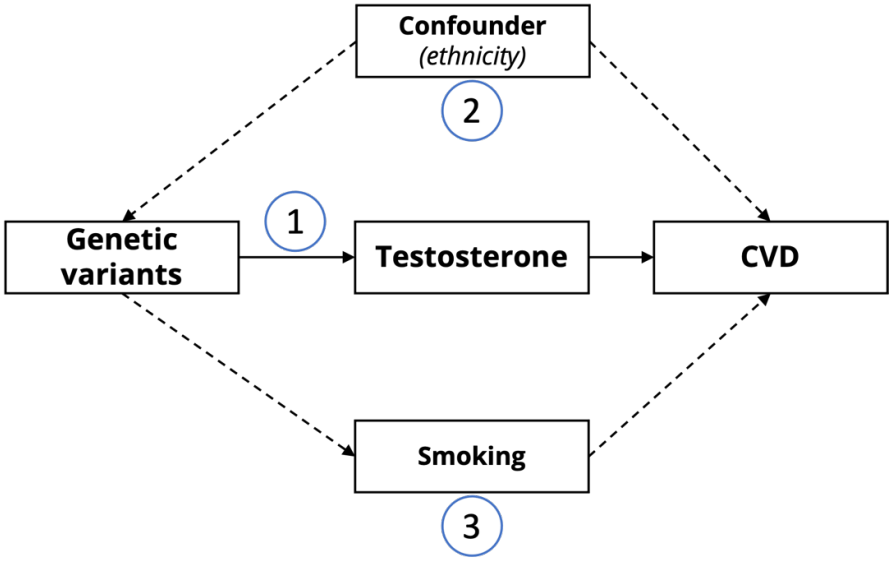


Figure 1.2. Major assumptions of Mendelian randomization analysis. 1) Relevance assumption: genetic variants must be associated with the exposure or biomarker (e.g., testosterone). 2) Independence assumption: there must be no confounders between genetic variant and outcome. 3) Exclusion restriction assumption: genetic variants must affect outcome only through exposure or biomarker (i.e., no horizontal pleiotropy).

1.3.2 GENOME-WIDE ASSOCIATION STUDIES OF BIOMARKERS

In Mendelian randomization analysis, there is evidence of a causal relationship if the effects of genetic variants associated with the exposure are proportional to effects on the outcome. Therefore, Mendelian randomization analysis investigating the causal relationship between a biomarker and outcome requires estimates of the effect of genetic variants on both the biomarker and the outcome.

Genome-wide association studies have been traditionally focused on diseases and complex traits. Early genome-wide association studies of biomarkers were focused on individual biomarkers with clinical relevance, like fibrinogen or uric acid, but they

represented a miniscule fraction of the 20,000 to 30,000 proteins comprising the human proteome (Dehghan et al., 2008) (Dehghan et al., 2009) (Williams et al., 2019). Recent technological advances in quantitative proteomics have enabled high-throughput, multiplexed, and cost-effective quantification for thousands of proteins simultaneously (Correa Rojo et al., 2021). When applied to large-scale studies or clinical trials with genetic data, these advancements have enabled genome-wide association studies of an exploratory nature that are designed to discover novel genetic determinants for circulating protein levels, or protein quantitative trait loci (pQTL). For instance, an early study in this domain measured 284 serum biomarkers relevant to cardiovascular disease in 5,078 ORIGIN trial participants using a multiplex ELISA array, the Human Discovery Multi-Analyte Profile 250+ panel on the LUMINEX 100/200 platforms from Myriad RBM Incorporated (Sjaarda et al., 2018) (Gerstein et al., 2015).

The SOMAScan assay developed by SOMALogic Incorporated is a popular technology for multiplex protein quantification that uses a DNA-based Slow Off-Rate Modified Aptamer (SOMAmer) to bind to proteins and can be quantified using DNA microarrays (Gold et al., 2010). To improve specificity and sensitivity, their modified oligonucleotides undergo an iterative process of enrichment to select for aptamers with highest affinity and slowest dissociation from a target protein. In 2017, this technology was used to measure 1,124 proteins in 1,338 individuals from the KORA and QMDiab studies (Suhre et al., 2017). Then, a larger study in 3,301 individuals from the INTERVAL study used an updated SOMAScan assay that interrogated 3,662 plasma proteins (Sun et al., 2018). Most recently, deCODE genetics performed a monumental study by testing 4,907

plasma proteins in 35,559 Icelanders identifying 18,084 pQTL (Ferkingstad et al., 2021). The current SOMAScan Discovery assay measures 7,000 proteins.

Olink Proteomics invented another popular antibody-based technology with high specificity called Proximity Extension Assay that features two matched antibodies conjugated to complementary single-stranded DNA. Only when both matched antibodies bind to a target protein, the single-stranded DNA can hybridize providing a unique signature for that protein that can be quantified by quantitative PCR (Assarsson et al., 2014). The SCALLOP consortium first published a genome-wide association study of 90 cardiovascular-related plasma proteins from the Olink cardiovascular I panel in 30,931 individuals (Folkersen et al., 2020), while a more recent preprint broadened the scope to 184 proteins plasma proteins from the Olink cardiovascular II and III panels in up to 26,494 individuals (Macdonald-Dunlop et al., 2021). The current Olink Explore assay measures up to 3,072 proteins across 8 panels designed for specific therapeutic areas.

As with genome-wide association studies of complex disease, broader adoption of proteomic technologies will foster the international collaborations and consortia necessary to facilitate future large-scale population studies. Indeed, biobank studies have already begun incorporating proteomic measurements in their studies (Sudlow et al., 2015). However, differences between quantitative proteomic platforms may pose a challenge for meta-analysis of genome-wide association studies of protein concentration. Although nearly 65% of protein quantitative trait loci replicate across Olink and SOMALogic, there are some protein characteristics that affect cross-platform correlation, such as protein-altering variants, glycosylation, and presence of transmembrane domains among others

(Pietzner et al., 2021) (Ferkingstad et al., 2021). Cross-platform differences may have implications for downstream analyses such as Mendelian randomization and highlight the importance of identifying shared signals in future efforts.

1.4 REFERENCES

Abifadel, M., Varret, M., Rabès, J.-P., Allard, D., Ouguerram, K., Devillers, M., Cruaud, C., Benjannet, S., Wickham, L., Erlich, D., Derré, A., Villéger, L., Farnier, M., Beucler, I., Bruckert, E., Chambaz, J., Chanu, B., Lecerf, J.-M., Luc, G., … Boileau, C. (2003). Mutations in PCSK9 cause autosomal dominant hypercholesterolemia. *Nature Genetics*, *34*(2), 154–156. https://doi.org/10.1038/ng1161

Altman, D. G., & Bland, J. M. (1994). Diagnostic tests 1: sensitivity and specificity. *BMJ*, *308*(6943), 1552–1552. https://doi.org/10.1136/bmj.308.6943.1552

American Diabetes Association. (2012). Diagnosis and Classification of Diabetes Mellitus. *Diabetes Care*, *35*(Supplement_1), S64–S71. https://doi.org/10.2337/dc12-s064

American Diabetes Association. (2018). 5. Prevention or Delay of Type 2 Diabetes: Standards of Medical Care in Diabetes—2018. *Diabetes Care*, *41*(Supplement_1), S51–S54. https://doi.org/10.2337/dc18-S005

Assarsson, E., Lundberg, M., Holmquist, G., Björkesten, J., Bucht Thorsen, S., Ekman, D., Eriksson, A., Rennel Dickens, E., Ohlsson, S., Edfeldt, G., Andersson, A.-C., Lindstedt, P., Stenvang, J., Gullberg, M., & Fredriksson, S. (2014). Homogenous 96-Plex PEA Immunoassay Exhibiting High Sensitivity, Specificity, and Excellent Scalability. *PLoS ONE*, *9*(4), e95192. https://doi.org/10.1371/journal.pone.0095192

Atkinson, A. J., Colburn, W. A., DeGruttola, V. G., DeMets, D. L., Downing, G. J., Hoth, D. F., Oates, J. A., Peck, C. C., Schooley, R. T., Spilker, B. A., Woodcock, J., & Zeger, S. L. (2001). Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework. *Clinical Pharmacology & Therapeutics*, *69*(3), 89–95. https://doi.org/10.1067/mcp.2001.113989

Berg, K. (1963). A NEW SERUM TYPE SYSTEM IN MAN--THE LP SYSTEM. *Acta Pathologica et Microbiologica Scandinavica*, *59*(3), 369–382. https://doi.org/10.1111/j.1699-0463.1963.tb01808.x

Berglund, L., & Ramakrishnan, R. (2004). Lipoprotein(a): an elusive cardiovascular risk factor. *Arteriosclerosis, Thrombosis, and Vascular Biology*, *24*(12), 2219–2226. https://doi.org/10.1161/01.ATV.0000144010.55563.63

Bhatt, D. L., Steg, P. G., Miller, M., Brinton, E. A., Jacobson, T. A., Ketchum, S. B., Doyle, R. T., Juliano, R. A., Jiao, L., Granowitz, C., Tardif, J.-C., & Ballantyne, C. M. (2019). Cardiovascular Risk Reduction with Icosapent Ethyl for Hypertriglyceridemia. *New England Journal of Medicine*, *380*(1), 11–22. https://doi.org/10.1056/nejmoa1812792

Burgess, S., Davey Smith, G., Davies, N. M., Dudbridge, F., Gill, D., Glymour, M. M., Hartwig, F. P., Holmes, M. V., Minelli, C., Relton, C. L., & Theodoratou, E. (2019). Guidelines for performing Mendelian randomization investigations. *Wellcome Open Research*, *4*, 186. https://doi.org/10.12688/wellcomeopenres.15555.1

Burgess, S., Dudbridge, F., & Thompson, S. G. (2016). Combining information on multiple instrumental variables in Mendelian randomization: Comparison of allele score and summarized data methods. *Statistics in Medicine*, *35*(11), 1880–1906. https://doi.org/10.1002/sim.6835

Busija, L., Lim, K., Szoeke, C., Sanders, K. M., & McCabe, M. P. (2019). Do replicable profiles of multimorbidity exist? Systematic review and synthesis. *European Journal of Epidemiology*, *34*(11), 1025–1053. https://doi.org/10.1007/s10654-019-00568-5

Califf, R. M. (2018). Biomarker definitions and their applications. *Experimental Biology and Medicine*, *243*(3), 213–221. https://doi.org/10.1177/1535370217750088

Correa Rojo, A., Heylen, D., Aerts, J., Thas, O., Hooyberghs, J., Ertaylan, G., & Valkenborg, D. (2021). Towards Building a Quantitative Proteomics Toolbox in Precision Medicine: A Mini-Review. *Frontiers in Physiology*, *12*(August). https://doi.org/10.3389/fphys.2021.723510

Danaei, G., Lu, Y., Singh, G. M., Carnahan, E., Stevens, G. A., Cowan, M. J., Farzadfar, F., Lin, J. K., Finucane, M. M., Rao, M., Khang, Y. H., Riley, L. M., Arian, D. M., Lim, S. S., Ezzati, M., Aamodt, G., Abdeen, Z., Abdella, N. A., Rahim, H. F. A., … Zhou, M. (2014). Cardiovascular disease, chronic kidney disease, and diabetes mortality burden of cardiometabolic risk factors from 1980 to 2010: a comparative risk assessment. *The Lancet Diabetes & Endocrinology*, *2*(8), 634–647. https://doi.org/10.1016/S2213-8587(14)70102-0

Davey Smith, G., & Hemani, G. (2014). Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Human Molecular Genetics*, *23*(R1), R89–R98. https://doi.org/10.1093/hmg/ddu328

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

Dehghan, A., Köttgen, A., Yang, Q., Hwang, S. J., Kao, W. L., Rivadeneira, F., Boerwinkle, E., Levy, D., Hofman, A., Astor, B. C., Benjamin, E. J., van Duijn, C. M., Witteman, J. C., Coresh, J., & Fox, C. S. (2008). Association of three genetic loci with uric acid concentration and risk of gout: a genome-wide association study. *The Lancet*, *372*(9654), 1953–1961. https://doi.org/10.1016/S0140-6736(08)61343-4

Dehghan, A., Yang, Q., Peters, A., Basu, S., Bis, J. C., Rudnicka, A. R., Kavousi, M., Chen, M. H., Baumert, J., Lowe, G. D. O., McKnight, B., Tang, W., De Maat, M., Larson, M. G., Eyhermendy, S., McArdle, W. L., Lumley, T., Pankow, J. S., Hofman, A., … Folsom, A. R. (2009). Association of novel genetic loci with circulating fibrinogen levels a genome-wide association study in 6 population-based cohorts. *Circulation: Cardiovascular Genetics*, *2*(2), 125–133. https://doi.org/10.1161/CIRCGENETICS.108.825224

Ferkingstad, E., Sulem, P., Atlason, B. A., Sveinbjornsson, G., Magnusson, M. I., Styrmisdottir, E. L., Gunnarsdottir, K., Helgason, A., Oddsson, A., Halldorsson, B.

V., Jensson, B. O., Zink, F., Halldorsson, G. H., Masson, G., Arnadottir, G. A., Katrinardottir, H., Juliusson, K., Magnusson, M. K., Magnusson, O. T., … Stefansson, K. (2021). Large-scale integration of the plasma proteome with genetics and disease. *Nature Genetics*, *53*(12), 1712–1721. https://doi.org/10.1038/s41588-021-00978-w

Folkersen, L., Gustafsson, S., Wang, Q., Hansen, D. H., Hedman, Å. K., Schork, A., Page, K., Zhernakova, D. V., Wu, Y., Peters, J., Eriksson, N., Bergen, S. E., Boutin, T. S., Bretherick, A. D., Enroth, S., Kalnapenkis, A., Gådin, J. R., Suur, B. E., Chen, Y., … Mälarstig, A. (2020). Genomic and drug target evaluation of 90 cardiovascular proteins in 30,931 individuals. *Nature Metabolism*, *2*(10), 1135–1148. https://doi.org/10.1038/s42255-020-00287-2

Gan, W., Bragg, F., Walters, R. G., Millwood, I. Y., Lin, K., Chen, Y., Guo, Y., Vaucher, J., Bian, Z., Bennett, D., Lv, J., Yu, C., Mahajan, A., Clarke, R. J., Li, L., Holmes, M. V., McCarthy, M. I., & Chen, Z. (2019). Genetic predisposition to type 2 diabetes and risk of subclinical atherosclerosis and cardiovascular diseases among 160,000 Chinese adults. *Diabetes*, *68*(11), 2155–2164. https://doi.org/10.2337/db19-0224

Ganz, P., Heidecker, B., Hveem, K., Jonasson, C., Kato, S., Segal, M. R., Sterling, D. G., & Williams, S. A. (2016). Development and Validation of a Protein-Based Risk Score for Cardiovascular Outcomes Among Patients With Stable Coronary Heart Disease. *JAMA*, *315*(23), 2532. https://doi.org/10.1001/jama.2016.5951

Gerstein, H. C., McMurray, J., & Holman, R. R. (2019). Real-world studies no substitute for RCTs in establishing efficacy. *The Lancet*, *393*(10168), 210–211. https://doi.org/10.1016/S0140-6736(18)32840-X

Gerstein, H. C., Paré, G., McQueen, M. J., Haenel, H., Lee, S. F., Pogue, J., Maggioni, A. P., Yusuf, S., & Hess, S. (2015). Identifying novel biomarkers for cardiovascular events or death in people with dysglycemia. *Circulation*, *132*(24), 2297–2304. https://doi.org/10.1161/CIRCULATIONAHA.115.015744

Gerstein, H. C., Paré, G., McQueen, M. J., Lee, S. F., & Hess, S. (2017). Validation of the origin cardiovascular biomarker panel and the value of adding troponin i in dysglycemic people. *Journal of Clinical Endocrinology and Metabolism*, *102*(7), 2251–2257. https://doi.org/10.1210/jc.2017-00273

Gill, D., Georgakis, M. K., Walker, V. M., Schmidt, A. F., Gkatzionis, A., Freitag, D. F., Finan, C., Hingorani, A. D., Howson, J. M. M., Burgess, S., Swerdlow, D. I., Davey Smith, G., Holmes, M. V., Dichgans, M., Scott, R. A., Zheng, J., Psaty, B. M., & Davies, N. M. (2021). Mendelian randomization for studying the effects of perturbing drug targets. *Wellcome Open Research*, *6*, 16. https://doi.org/10.12688/wellcomeopenres.16544.2

Gold, L., Ayers, D., Bertino, J., Bock, C., Bock, A., Brody, E. N., Carter, J., Dalby, A. B., Eaton, B. E., Fitzwater, T., Flather, D., Forbes, A., Foreman, T., Fowler, C., Gawande, B., Goss, M., Gunn, M., Gupta, S., Halladay, D., … Zichi, D. (2010). Aptamer-Based Multiplexed Proteomic Technology for Biomarker Discovery. *PLoS ONE*, *5*(12), e15004. https://doi.org/10.1371/journal.pone.0015004

Grundy, S. M., Stone, N. J., Bailey, A. L., Beam, C., Birtcher, K. K., Blumenthal, R. S., Braun, L. T., de Ferranti, S., Faiella-Tommasino, J., Forman, D. E., Goldberg, R.,

Heidenreich, P. A., Hlatky, M. A., Jones, D. W., Lloyd-Jones, D., Lopez-Pajares, N., Ndumele, C. E., Orringer, C. E., Peralta, C. A., … Yeboah, J. (2019). 2018 AHA/ACC/AACVPR/AAPA/ABC/ACPM/ADA/AGS/APhA/ASPC/NLA/PCNA Guideline on the Management of Blood Cholesterol: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. *Journal of the American College of Cardiology*, *73*(24), e285–e350. https://doi.org/10.1016/j.jacc.2018.11.003

Han, Y., Hu, Y., Yu, C., Guo, Y., Pei, P., Yang, L., Chen, Y., Du, H., Sun, Di., Pang, Y., Chen, N., Clarke, R., Chen, J., Chen, Z., Li, L., & Lv, J. (2021). Lifestyle, cardiometabolic disease, and multimorbidity in a prospective Chinese study. *European Heart Journal*, *42*(34), 3374–3384. https://doi.org/10.1093/eurheartj/ehab413

Harrison, R. K. (2016). Phase II and phase III failures: 2013-2015. *Nature Reviews Drug Discovery*, *15*(12), 817–818. https://doi.org/10.1038/nrd.2016.184

Hijazi, Z., Lindbäck, J., Alexander, J. H., Hanna, M., Held, C., Hylek, E. M., Lopes, R. D., Oldgren, J., Siegbahn, A., Stewart, R. A. H., White, H. D., Granger, C. B., & Wallentin, L. (2016). The ABC (age, biomarkers, clinical history) stroke risk score: A biomarker-based risk score for predicting stroke in atrial fibrillation. *European Heart Journal*, *37*(20), 1582–1590. https://doi.org/10.1093/eurheartj/ehw054

Holmes, M. V., Asselbergs, F. W., Palmer, T. M., Drenos, F., Lanktree, M. B., Nelson, C. P., Dale, C. E., Padmanabhan, S., Finan, C., Swerdlow, D. I., Tragante, V., van Iperen, E. P. A., Sivapalaratnam, S., Shah, S., Elbers, C. C., Shah, T., Engmann, J., Giambartolomei, C., White, J., … Casas, J. P. (2015). Mendelian randomization of blood lipids for coronary heart disease. *European Heart Journal*, *36*(9), 539–550. https://doi.org/10.1093/eurheartj/eht571

Jaye, D. L., & Waites, K. B. (1997). Clinical applications of C-reactive protein in pediatrics. *The Pediatric Infectious Disease Journal*, *16*(8), 735–747. https://doi.org/10.1097/00006454-199708000-00003

Kaptoge, S., Di Angelantonio, E., Lowe, G., Pepys, M. B., Thompson, S. G., Collins, R., Danesh, J., Tipping, R. W., Ford, C. E., Pressel, S. L., Walldius, G., Jungner, I., Folsom, A. R., Chambless, L., Ballantyne, C. M., Panagiotakos, D., Pitsavos, C., Chrysohoou, C., Stefanadis, C., … Wood, A. M. (2010). C-reactive protein concentration and risk of coronary heart disease, stroke, and mortality: An individual participant meta-analysis. *The Lancet*, *375*(9709), 132–140. https://doi.org/10.1016/S0140-6736(09)61717-7

King, E. A., Wade Davis, J., & Degner, J. F. (2019). Are drug targets with genetic support twice as likely to be approved? Revised estimates of the impact of genetic support for drug mechanisms on the probability of drug approval. *PLoS Genetics*, *15*(12), 1–20. https://doi.org/10.1371/journal.pgen.1008489

Lanktree, M. B., Thériault, S., Walsh, M., & Paré, G. (2018). HDL Cholesterol, LDL Cholesterol, and Triglycerides as Risk Factors for CKD: A Mendelian Randomization Study. *American Journal of Kidney Diseases*, *71*(2), 166–172. https://doi.org/10.1053/j.ajkd.2017.06.011

Levey, A. S., Coresh, J., Tighiouart, H., Greene, T., & Inker, L. A. (2020). Measured and

estimated glomerular filtration rate: current status and future directions. *Nature Reviews Nephrology*, *16*(1), 51–64. https://doi.org/10.1038/s41581-019-0191-y

Levey, A. S., Stevens, L. A., Schmid, C. H., Zhang, Y. L., Castro, A. F., Feldman, H. I., Kusek, J. W., Eggers, P., Van Lente, F., Greene, T., & Coresh, J. (2009). A New Equation to Estimate Glomerular Filtration Rate. *Annals of Internal Medicine*, *150*(9), 604. https://doi.org/10.7326/0003-4819-150-9-200905050-00006

Lin, X., Xu, Y., Pan, X., Xu, J., Ding, Y., Sun, X., Song, X., Ren, Y., & Shan, P.-F. (2020). Global, regional, and national burden and trend of diabetes in 195 countries and territories: an analysis from 1990 to 2025. *Scientific Reports*, *10*(1), 14790. https://doi.org/10.1038/s41598-020-71908-9

Macdonald-Dunlop, E., Klarić, L., Folkersen, L., Timmers, P. R. H. J., Gustafsson, S., Zhao, J. H., Eriksson, N., Richmond, A., Enroth, S., Mattsson-Carlgren, N., Johansson, Å., Hayward, C., Wallentin, L., Siegbahn, A., Lind, L., Butterworth, A. S., Michaëlsson, K., Peters, J. E., Mälarstig, A., … Wilson, J. F. (2021). Mapping genetic determinants of 184 circulating proteins in 26,494 individuals to connect proteins and diseases. *MedRxiv*, *10*, 39. https://doi.org/10.1101/2021.08.03.21261494

Mechanick, J. I., Farkouh, M. E., Newman, J. D., & Garvey, W. T. (2020). Cardiometabolic-Based Chronic Disease, Adiposity and Dysglycemia Drivers: JACC State-of-the-Art Review. *Journal of the American College of Cardiology*, *75*(5), 525–538. https://doi.org/10.1016/j.jacc.2019.11.044

Palmer, S. C., Tendal, B., Mustafa, R. A., Vandvik, P. O., Li, S., Hao, Q., Tunnicliffe, D., Ruospo, M., Natale, P., Saglimbene, V., Nicolucci, A., Johnson, D. W., Tonelli, M., Rossi, M. C., Badve, S. V., Cho, Y., Nadeau-Fredette, A.-C., Burke, M., Faruque, L. I., … Strippoli, G. F. M. (2021). Sodium-glucose cotransporter protein-2 (SGLT-2) inhibitors and glucagon-like peptide-1 (GLP-1) receptor agonists for type 2 diabetes: systematic review and network meta-analysis of randomised controlled trials. *BMJ*, *372*, m4573. https://doi.org/10.1136/bmj.m4573

Paré, G., Çaku, A., McQueen, M., Anand, S. S., Enas, E., Clarke, R., Boffa, M. B., Koschinsky, M., Wang, X., & Yusuf, S. (2019). Lipoprotein(a) Levels and the Risk of Myocardial Infarction Among 7 Ethnic Groups. *Circulation*, *139*(12), 1472–1482. https://doi.org/10.1161/CIRCULATIONAHA.118.034311

Pietzner, M., Wheeler, E., Carrasco-Zanini, J., Kerrison, N. D., Oerton, E., Koprulu, M., Luan, J., Hingorani, A. D., Williams, S. A., Wareham, N. J., & Langenberg, C. (2021). Synergistic insights into human health from aptamer- and antibody-based proteomic profiling. *Nature Communications*, *12*(1). https://doi.org/10.1038/s41467-021-27164-0

Pound, P., & Ritskes-Hoitinga, M. (2018). Is it possible to overcome issues of external validity in preclinical animal research? Why most animal models are bound to fail. *Journal of Translational Medicine*, *16*(1), 1–8. https://doi.org/10.1186/s12967-018-1678-1

Rasheed, H., Zheng, J., Rees, J., Sanderson, E., Thomas, L., Richardson, T. G., Fang, S., Bekkevold, O.-J., Stovner, E. B., Gabrielsen, M. E., Skogholt, A. H., Romundstad, S., Brumpton, B., Hallan, S., Willer, C., Burgess, S., Hveem, K., Davey Smith, G.,

Gaunt, T. R., & Åsvold, B. O. (2021). The causal effects of serum lipids and apolipoproteins on kidney function: multivariable and bidirectional Mendelian-randomization analyses. *International Journal of Epidemiology*, *50*(5), 1569–1579. https://doi.org/10.1093/ije/dyab014

Reboussin, D. M., Allen, N. B., Griswold, M. E., Guallar, E., Hong, Y., Lackland, D. T., Miller, E. (Pete) R., Polonsky, T., Thompson-Paul, A. M., & Vupputuri, S. (2018). Systematic Review for the 2017 ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ASPC/NMA/PCNA Guideline for the Prevention, Detection, Evaluation, and Management of High Blood Pressure in Adults. *Journal of the American College of Cardiology*, *71*(19), 2176–2198. https://doi.org/10.1016/j.jacc.2017.11.004

Riaz, H., Khan, S. U., Rahman, H., Shah, N. P., Kaluski, E., Lincoff, A. M., & Nissen, S. E. (2019). Effects of high-density lipoprotein targeting treatments on cardiovascular outcomes: A systematic review and meta-analysis. *European Journal of Preventive Cardiology*, *26*(5), 533–543. https://doi.org/10.1177/2047487318816495

Ross, S., Gerstein, H. C., Eikelboom, J., Anand, S. S., Yusuf, S., & Paré, G. (2015). Mendelian randomization analysis supports the causal role of dysglycaemia and diabetes in the risk of coronary artery disease. *European Heart Journal*, *36*(23), 1454–1462. https://doi.org/10.1093/eurheartj/ehv083

Roth, G. A., Mensah, G. A., Johnson, C. O., Addolorato, G., Ammirati, E., Baddour, L. M., Barengo, N. C., Beaton, A. Z., Benjamin, E. J., Benziger, C. P., Bonny, A., Brauer, M., Brodmann, M., Cahill, T. J., Carapetis, J., Catapano, A. L., Chugh, S. S., Cooper, L. T., Coresh, J., … Fuster, V. (2020). Global Burden of Cardiovascular Diseases and Risk Factors, 1990–2019. *Journal of the American College of Cardiology*, *76*(25), 2982–3021. https://doi.org/10.1016/j.jacc.2020.11.010

Sjaarda, J., Gerstein, H. C., Yusuf, S., Treleaven, D., Walsh, M., Mann, J. F. E., Hess, S., & Paré, G. (2018). Blood HER2 and Uromodulin as Causal Mediators of CKD. *Journal of the American Society of Nephrology*, *29*(4), 1326–1335. https://doi.org/10.1681/ASN.2017070812

Smietana, K., Siatkowski, M., & Møller, M. (2016). Trends in clinical success rates. *Nature Reviews Drug Discovery*, *15*(6), 379–380. https://doi.org/10.1038/nrd.2016.85

Staley, J. R., & Burgess, S. (2017). Semiparametric methods for estimation of a nonlinear exposure-outcome relationship using instrumental variables with application to Mendelian randomization. *Genetic Epidemiology*, *41*(4), 341–352. https://doi.org/10.1002/gepi.22041

Su, X., Zhang, L., Lv, J., Wang, J., Hou, W., Xie, X., & Zhang, H. (2016). Effect of statins on kidney disease outcomes: A systematic review and meta-analysis. *American Journal of Kidney Diseases*, *67*(6), 881–892. https://doi.org/10.1053/j.ajkd.2016.01.016

Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., Downey, P., Elliott, P., Green, J., Landray, M., Liu, B., Matthews, P., Ong, G., Pell, J., Silman, A., Young, A., Sprosen, T., Peakman, T., & Collins, R. (2015). UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases

of Middle and Old Age. *PLOS Medicine*, *12*(3), e1001779. https://doi.org/10.1371/journal.pmed.1001779

Suhre, K., Arnold, M., Mukund Bhagwat, A., Cotton, R. J., Engelke, R., Raffler, J., Sarwath, H., Thareja, G., Wahl, A., Kirk DeLisle, R., Gold, L., Pezer, M., Lauc, G., El-Din Selim, M. A., Mook-Kanamori, D. O., Al-Dous, E. K., Kastenmüller, G., Gieger, C., & Graumann, J. (2017). Connecting genetic risk to disease end points through the human blood plasma proteome. *Nature Communications*, *8*(14357), 1–13. https://doi.org/10.1038/ncomms14357

Sun, B. B., Maranville, J. C., Peters, J. E., Stacey, D., Staley, J. R., Blackshaw, J., Burgess, S., Jiang, T., Paige, E., Surendran, P., Oliver-Williams, C., Kamat, M. A., Prins, B. P., Wilcox, S. K., Zimmerman, E. S., Chi, A., Bansal, N., Spain, S. L., Wood, A. M., … Butterworth, A. S. (2018). Genomic atlas of the human plasma proteome. *Nature*, *558*(7708), 73–79. https://doi.org/10.1038/s41586-018-0175-2

The Emerging Risk Factors Collaboration*. (2009). Major Lipids, Apolipoproteins, and Risk of Vascular Disease. *JAMA*, *302*(18), 1993. https://doi.org/10.1001/jama.2009.1619

Tsimikas, S., Karwatowska-Prokopczuk, E., Gouni-Berthold, I., Tardif, J.-C., Baum, S. J., Steinhagen-Thiessen, E., Shapiro, M. D., Stroes, E. S., Moriarty, P. M., Nordestgaard, B. G., Xia, S., Guerriero, J., Viney, N. J., O'Dea, L., & Witztum, J. L. (2020). Lipoprotein(a) Reduction in Persons with Cardiovascular Disease. *New England Journal of Medicine*, *382*(3), 244–255. https://doi.org/10.1056/nejmoa1905239

Twerenbold, R., Boeddinghaus, J., Nestelberger, T., Wildi, K., Rubini Gimenez, M., Badertscher, P., & Mueller, C. (2017). Clinical Use of High-Sensitivity Cardiac Troponin in Patients With Suspected Myocardial Infarction. *Journal of the American College of Cardiology*, *70*(8), 996–1012. https://doi.org/10.1016/j.jacc.2017.07.718

Vos, T., Lim, S. S., Abbafati, C., Abbas, K. M., Abbasi, M., Abbasifard, M., Abbasi-Kangevari, M., Abbastabar, H., Abd-Allah, F., Abdelalim, A., Abdollahi, M., Abdollahpour, I., Abolhassani, H., Aboyans, V., Abrams, E. M., Abreu, L. G., Abrigo, M. R. M., Abu-Raddad, L. J., Abushouk, A. I., … Murray, C. J. L. (2020). Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *The Lancet*, *396*(10258), 1204–1222. https://doi.org/10.1016/S0140-6736(20)30925-9

Võsa, U., Claringbould, A., Westra, H.-J., Bonder, M. J., Deelen, P., Zeng, B., Kirsten, H., Saha, A., Kreuzhuber, R., Yazar, S., Brugge, H., Oelen, R., de Vries, D. H., van der Wijst, M. G. P., Kasela, S., Pervjakova, N., Alves, I., Favé, M.-J., Agbessi, M., … Franke, L. (2021). Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. *Nature Genetics*, *53*(9), 1300–1310. https://doi.org/10.1038/s41588-021-00913-z

Welter, D., MacArthur, J., Morales, J., Burdett, T., Hall, P., Junkins, H., Klemm, A., Flicek, P., Manolio, T., Hindorff, L., & Parkinson, H. (2014). The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Research*, *42*(D1), D1001–D1006. https://doi.org/10.1093/nar/gkt1229

Williams, S. A., Kivimaki, M., Langenberg, C., Hingorani, A. D., Casas, J. P., Bouchard,

C., Jonasson, C., Sarzynski, M. A., Shipley, M. J., Alexander, L., Ash, J., Bauer, T., Chadwick, J., Datta, G., DeLisle, R. K., Hagar, Y., Hinterberg, M., Ostroff, R., Weiss, S., … Wareham, N. J. (2019). Plasma protein patterns as comprehensive indicators of health. *Nature Medicine*, *25*(12), 1851–1857. https://doi.org/10.1038/s41591-019-0665-2

Wouters, O. J., McKee, M., & Luyten, J. (2020). Estimated Research and Development Investment Needed to Bring a New Medicine to Market, 2009-2018. *JAMA - Journal of the American Medical Association*, *323*(9), 844–853. https://doi.org/10.1001/jama.2020.1166

Xie, X., Liu, Y., Perkovic, V., Li, X., Ninomiya, T., Hou, W., Zhao, N., Liu, L., Lv, J., Zhang, H., & Wang, H. (2016). Renin-Angiotensin System Inhibitors and Kidney and Cardiovascular Outcomes in Patients With CKD: A Bayesian Network Meta-analysis of Randomized Clinical Trials. *American Journal of Kidney Diseases*, *67*(5), 728–741. https://doi.org/10.1053/j.ajkd.2015.10.011

Zhou, W., Kanai, M., Wu, K. H., Humaira, R., Tsuo, K., & Hirbo, J. B. (2021). Global Biobank Meta-analysis Initiative: powering genetic discovery across human diseases. *MedRxiv*. https://doi.org/10.1101/2021.11.19.21266436

CHAPTER 2:

GENERAL HYPOTHESIS, OBJECTIVE, RATIONALE, AND APPROACH

CHAPTER 2: HYPOTHESIS, OBJECTIVE, RATIONALE, & APPROACH

2.1 GENERAL HYPOTHESIS

We hypothesized that causal blood biomarkers for cardiometabolic disease could be identified that have applications to disease prevention, treatment, and diagnosis.

2.2 GENERAL OBJECTIVE

We sought to use Mendelian randomization to discover blood biomarkers that are causal for, or a consequence of, cardiometabolic disease to inform improved therapies or diagnostic biomarkers of disease.

2.3 RATIONALE AND APPROACH

Blood biomarkers can be helpful in informing novel therapeutic targets or diagnostic markers for cardiometabolic disease (Califf, 2018). Observational studies suffer from inherent biases such as reverse causation and confounding that limit their ability to discern causal biomarkers related to disease. Large-scale randomized controlled trials are the gold-standard for causal inference in health sciences research but they are time-consuming, resource-intensive, or, in some cases, impossible from an ethical or logistical standpoint (Gerstein et al., 2019). Mendelian randomization is a technique that leverages random inheritance of alleles at conception to mimic properties of randomized studies and estimate the causal effect of a biomarker and disease (Davies et al., 2018) (Gill et al., 2021). Therefore, this thesis sought to demonstrate the utility of Mendelian randomization as a complementary tool to explain findings from observational studies, predict drug safety and repurposing opportunities, inform clinical trial decision making, and improve diagnostic biomarkers.

In the context of this theme, we pursued three research projects that both make novel contributions to the scientific literature and provide clinical insights to improve treatment and diagnosis of cardiometabolic diseases.

In chapter 3, "*Elevated lipoprotein(a) and risk of atrial fibrillation: an observational and Mendelian randomization study*", we applied Mendelian randomization analysis to a pre-specified hypothesis. Given ongoing trials of lipoprotein(a) inhibitors for prevention of cardiovascular disease and lack of preventative therapies for atrial fibrillation, we intended to elucidate whether lipoprotein(a) causes atrial fibrillation and whether any effect is mediated independent of atherosclerotic cardiovascular disease. To do so, we tested the association of observed and genetically predicted lipoprotein(a) levels on risk of atrial fibrillation in the UK Biobank (Sudlow et al., 2015) followed by a formal two-sample Mendelian randomization analysis using two independent genome-wide association studies of atrial fibrillation (Nielsen et al., 2018) (Borodulin et al., 2018). The plausibility of an effect of lipoprotein(a) on atrial fibrillation independent of atherosclerotic disease was tested through mediation analysis conditioning on ischemic heart disease and aortic valve stenosis, and comparison between the effect of lipoprotein(a) to other atherogenic lipid fractions.

In chapter 4, "*Effects of lifelong elevated testosterone on health and disease*", we applied Mendelian randomization analysis to broadly survey the causal effects of testosterone – a biomarker of topical interest to the medical community due to its controversial role in cardiometabolic disease (Münzer et al., 2001) (Snyder et al., 2018) (Bhasin et al., 2018). In addition to exploring suspected effects, the exploratory nature of

this study afforded the ability to agnostically identify novel effects that could inform drug repurposing opportunities and adverse effects. We identified genetic determinants of calculated free testosterone in males from the UK Biobank, which were used to test the association of genetically predicted free testosterone with hundreds of diseases and traits in a one-sample Mendelian randomization analysis design (Mohammadi-Shemirani et al., 2020).

In chapter 5, "*A Mendelian randomization-based approach to identify early and sensitive diagnostic biomarkers of disease*", we reimagined traditional Mendelian randomization analysis to discover biomarkers that are caused by disease rather than act as causal mediators themselves (Mohammadi-Shemirani et al., 2019). Such biomarkers that are consequences of disease are expected to serve as sensitive and early markers for diagnosis. We explored this notion through a proof-of-concept study in the ORIGIN trial (Sjaarda et al., 2018), where we identified serum biomarkers that were associated with genetically predicted levels of $eGFR_{crea}$ using genetic variants from the CKDGen consortium (Pattaro et al., 2016). To validate their ability to discriminate early stages of disease, we tested whether significant biomarkers were associated with incident CKD in participants without traditional risk factors of CKD, and whether inclusion of biomarkers improved discriminative ability of stage 3 CKD in models with CKD risk factors.

2.4 REFERENCES

Bhasin, S., Brito, J. P., Cunningham, G. R., Hayes, F. J., Hodis, H. N., Matsumoto, A. M., Snyder, P. J., Swerdloff, R. S., Wu, F. C., & Yialamas, M. A. (2018). Testosterone Therapy in Men With Hypogonadism: An Endocrine Society* Clinical Practice

Guideline. *The Journal of Clinical Endocrinology & Metabolism*, *103*(5), 1715–1744. https://doi.org/10.1210/jc.2018-00229

Borodulin, K., Tolonen, H., Jousilahti, P., Jula, A., Juolevi, A., Koskinen, S., Kuulasmaa, K., Laatikainen, T., Männistö, S., Peltonen, M., Perola, M., Puska, P., Salomaa, V., Sundvall, J., Virtanen, S. M., & Vartiainen, E. (2018). Cohort Profile: The National FINRISK Study. *International Journal of Epidemiology*, *47*(3), 696-696i. https://doi.org/10.1093/ije/dyx239

Califf, R. M. (2018). Biomarker definitions and their applications. *Experimental Biology and Medicine*, *243*(3), 213–221. https://doi.org/10.1177/1535370217750088

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

Gerstein, H. C., McMurray, J., & Holman, R. R. (2019). Real-world studies no substitute for RCTs in establishing efficacy. *The Lancet*, *393*(10168), 210–211. https://doi.org/10.1016/S0140-6736(18)32840-X

Gill, D., Georgakis, M. K., Walker, V. M., Schmidt, A. F., Gkatzionis, A., Freitag, D. F., Finan, C., Hingorani, A. D., Howson, J. M. M., Burgess, S., Swerdlow, D. I., Davey Smith, G., Holmes, M. V., Dichgans, M., Scott, R. A., Zheng, J., Psaty, B. M., & Davies, N. M. (2021). Mendelian randomization for studying the effects of perturbing drug targets. *Wellcome Open Research*, *6*, 16. https://doi.org/10.12688/wellcomeopenres.16544.2

Mohammadi-Shemirani, P., Chong, M., Pigeyre, M., Morton, R. W., Gerstein, H. C., &

Paré, G. (2020). Effects of lifelong testosterone exposure on health and disease using mendelian randomization. *ELife*, *9*, 1–17. https://doi.org/10.7554/eLife.58914

Mohammadi-Shemirani, P., Sjaarda, J., Gerstein, H. C., Treleaven, D. J., Walsh, M., Mann, J. F., McQueen, M. J., Hess, S., & Paré, G. (2019). A Mendelian randomization-based approach to identify early and sensitive diagnostic biomarkers of disease. *Clinical Chemistry*, *65*(3), 427–436. https://doi.org/10.1373/clinchem.2018.291104

Münzer, T., Harman, S. M., Hees, P., Shapiro, E., Christmas, C., Bellantoni, M. F., Stevens, T. E., O'Connor, K. G., Pabst, K. M., St. Clair, C., Sorkin, J. D., & Blackman, M. R. (2001). Effects of GH and/or Sex Steroid Administration on Abdominal Subcutaneous and Visceral Fat in Healthy Aged Women and Men. *The Journal of Clinical Endocrinology & Metabolism*, *86*(8), 3604–3610. https://doi.org/10.1210/jcem.86.8.7773

Nielsen, J. B., Thorolfsdottir, R. B., Fritsche, L. G., Zhou, W., Skov, M. W., Graham, S. E., Herron, T. J., McCarthy, S., Schmidt, E. M., Sveinbjornsson, G., Surakka, I., Mathis, M. R., Yamazaki, M., Crawford, R. D., Gabrielsen, M. E., Skogholt, A. H., Holmen, O. L., Lin, M., Wolford, B. N., … Willer, C. J. (2018). Biobank-driven genomic discovery yields new insight into atrial fibrillation biology. *Nature Genetics*, *50*(9), 1234–1239. https://doi.org/10.1038/s41588-018-0171-3

Pattaro, C., Teumer, A., Gorski, M., Chu, A. Y., Li, M., Mijatovic, V., Garnaas, M., Tin, A., Sorice, R., Li, Y., Taliun, D., Olden, M., Foster, M., Yang, Q., Chen, M. H., Pers, T. H., Johnson, A. D., Ko, Y. A., Fuchsberger, C., … Fox, C. S. (2016).

Genetic associations at 53 loci highlight cell types and biological pathways relevant for kidney function. *Nature Communications*, *7*, 1–19. https://doi.org/10.1038/ncomms10023

Sjaarda, J., Gerstein, H. C., Yusuf, S., Treleaven, D., Walsh, M., Mann, J. F. E., Hess, S., & Paré, G. (2018). Blood HER2 and Uromodulin as Causal Mediators of CKD. *Journal of the American Society of Nephrology*, *29*(4), 1326–1335. https://doi.org/10.1681/ASN.2017070812

Snyder, P. J., Bhasin, S., Cunningham, G. R., Matsumoto, A. M., Stephens-Shields, A. J., Cauley, J. A., Gill, T. M., Barrett-Connor, E., Swerdloff, R. S., Wang, C., Ensrud, K. E., Lewis, C. E., Farrar, J. T., Cella, D., Rosen, R. C., Pahor, M., Crandall, J. P., Molitch, M. E., Resnick, S. M., … Ellenberg, S. S. (2018). Lessons from the Testosterone Trials. *Endocrine Reviews*, *39*(3), 369–386. https://doi.org/10.1210/er.2017-00234

Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., Downey, P., Elliott, P., Green, J., Landray, M., Liu, B., Matthews, P., Ong, G., Pell, J., Silman, A., Young, A., Sprosen, T., Peakman, T., & Collins, R. (2015). UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. *PLOS Medicine*, *12*(3), e1001779. https://doi.org/10.1371/journal.pmed.1001779

CHAPTER 3:

Elevated lipoprotein(a) and risk of atrial fibrillation: an observational and Mendelian randomization study

Submitted to *Journal of the American College of Cardiology*. (June 22, 2021)

Elevated lipoprotein(a) and risk of atrial fibrillation: an observational and Mendelian randomization study

Authors: Pedrum Mohammadi-Shemirani BSc[a,b,c], Michael Chong MSc[a,b,d], Sukrit Narula BA[a,b,e], Nicolas Perrot PhD[a,f,g], David Conen MD MPH[a,e,h], Jason D. Roberts MD MAS[a], Sébastien Thériault MD MSc[f,i], Yohan Bossé PhD[f,j], Matthew B. Lanktree MD PhD[a,k], Marie Pigeyre MD PhD[a,h], Guillaume Paré MD MSc[a,b,e,l]

a. Population Health Research Institute, David Braley Cardiac, Vascular and Stroke Research Institute, Hamilton, Ontario, Canada

b. Thrombosis and Atherosclerosis Research Institute, David Braley Cardiac, Vascular and Stroke Research Institute, Hamilton, Ontario, Canada

c. Department of Medical Sciences, McMaster University, Hamilton, Ontario, Canada

d. Department of Biochemistry and Biomedical Sciences, McMaster University, Hamilton, Ontario, Canada

e. Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, Ontario, Canada

f. Centre de Recherche de l'Institut Universitaire de Cardiologie et de Pneumologie de Québec, Québec, Canada

g. Department of Medicine, Faculty of Medicine, Université Laval, Québec, Québec, Canada

h. Department of Medicine, McMaster University, Hamilton Health Sciences, Hamilton, Ontario, Canada

i. Department of Molecular Biology, Medical Biochemistry and Pathology, Faculty of Medicine, Université Laval, Québec, Québec, Canada

j. Department of Molecular Medicine, Faculty of Medicine, Université Laval, Québec, Québec, Canada

k. Division of Nephrology, Department of Medicine, McMaster University, Hamilton, Ontario, Canada

l. Department of Pathology and Molecular Medicine, McMaster University, Michael G. DeGroote School of Medicine, Hamilton, Ontario, Canada

Abstract

Background: Atrial fibrillation (AF) is a cardiac arrythmia associated with elevated risk of stroke, heart failure, and mortality. However, preventative therapies are needed with ancillary benefits on its cardiovascular comorbidities. Lipoprotein(a) (Lp(a)) is a recognized risk factor for atherosclerotic cardiovascular disease (ASCVD), which itself increases AF risk, but it remains unknown whether Lp(a) is a causal mediator of AF independent of ASCVD.

Objectives: This study investigated the role of Lp(a) in AF and whether it is independent of ASCVD.

Methods: Measured and genetically predicted Lp(a) levels were tested for association with 20,432 cases of incident AF in the UK Biobank (UKB) (N=435,579). Mendelian randomization (MR) analyses were performed using summary-level data for AF from Nielsen *et al*. and FinnGen (N=1,145,375).

Results: In the UKB, each 50 nmol/L (23 mg/dL) increase in Lp(a) was associated with increased risk of incident AF using measured (HR=1.03; 95%CI=1.02 to 1.04; $p=1.65x10^{-8}$) and genetically predicted Lp(a) (OR=1.03; 95%CI=1.02 to 1.05; $p=1.33x10^{-5}$). MR analyses using independent data replicated the effect (OR=1.04 per 50 nmol/L Lp(a) increase; 95%CI=1.03 to 1.05; $p=9.23x10^{-10}$). There was no evidence of risk-conferring effect from LDL cholesterol or triglycerides, and only 39% (95%CI=27% to 73%) of Lp(a) risk was mediated through ASCVD, suggesting Lp(a) partly influences AF independent of its known effects on ASCVD.

Conclusions: Our findings implicate Lp(a) as a potential causal mediator in development of AF that demonstrate Lp(a)'s effects extend across myocardial tissues. Ongoing clinical trials for Lp(a)-lowering therapies should evaluate effects on AF prevention.

Funding: P.M. is supported by the Frederick Banting and Charles Best Canada Graduate Scholarship from the Canadian Institute of Health Research. M.C. is supported by the Frederick Banting and Charles Best Canada Graduate Scholarship the Canadian Institute of Health Research. D.C. holds a McMaster University Department of Medicine Mid-Career Research Award. S.T. holds a junior scholar award from the FRQS (Fonds de recherche du Québec - Santé). Y.B. holds a Canada Research Chair in Genomics of Heart and Lung Diseases. M.B.L. is a new investigator in the KRESCENT (Kidney Research Scientist Core Education and National Training) program and is supported by the Canadian Institutes of Health Research, Kidney Foundation of Canada, and the Canadian Society of Nephrology. M.P. is supported by the E.J. Moran Campbell Internal Career Research Award from McMaster University. G.P. is holds the Canada Research Chair in Genetic and Molecular Epidemiology and Cisco Systems Professorship in Integrated Health Biosystems.

Introduction

Atrial fibrillation (AF) is the most common cardiac arrythmia in adults with a prevalence between 0.5-2% (1, 2). Longer lifespan and higher prevalence of AF risk factors are projected to further increase global prevalence of AF (3). Patients with AF have greater risks of heart failure, ischemic stroke, dementia and death, and are more likely to be hospitalized resulting in higher healthcare costs (4–7). Early intervention and prevention are important for management of any disease, but it is critically important for AF as structural and electrical remodelling in the atria begets further remodelling that exacerbates the condition (8, 9). However, preventative therapies repurposed from other cardiovascular diseases (CVD) have shown inconsistent results in AF. There is some evidence of beneficial effect for renin-angiotensin-aldosterone system inhibitors in primary prevention, but inconclusive results for statins or even potentially harmful effects for polyunsaturated fatty acids (10–14). Therefore, there is a lack of preventative therapies for AF with ancillary benefits on cardiovascular comorbidities.

Lipoprotein(a) [Lp(a)] is a particle consisting of an LDL-like core attached to an apolipoprotein(a) chain. Apolipoprotein(a) size is inversely related to circulating Lp(a) concentration, and Lp(a) levels are up to 90% heritable with most of the variance explained by the *LPA* locus alone (15, 16). Lp(a) plays a role in atherosclerotic cardiovascular disease and is a well-established risk factor for coronary artery disease, ischemic stroke, and aortic valve stenosis (15, 17–19). Higher risk of aortic valve stenosis suggests a pathological mechanism that extends beyond the arteries to other types of cardiac tissue and may include the atria (20). Although coronary artery disease is itself a risk factor for AF, Lp(a) particles

have additional thrombogenic and inflammatory properties that could provide other mechanisms, independent of atherosclerotic cardiovascular disease, by which an effect on AF could be mediated (21, 22). As a result, Lp(a) could represent a target that simultaneously prevents or treats both cardiovascular comorbidities of AF and AF itself, where the latter could be mediated as a downstream effect of its beneficial effect on atherosclerotic cardiovascular disease or through independent mechanisms unique to Lp(a) particles.

Unlike coronary artery disease, ischemic stroke, and aortic valve stenosis, an effect of Lp(a) on AF has been suggested but never effectively evaluated. Observational studies have identified Lp(a) as a risk factor for atrial thrombi in patients with AF but no association between Lp(a) and incident AF events (23). However, these studies were limited by the number of AF cases and underpowered to detect smaller effects of Lp(a) on AF (24, 25). A recent Mendelian randomization study explored multiple lipid traits and identified an effect of Lp(a) on AF, but it was only nominally significant without adjustment for multiple hypothesis testing (26). By integrating observational and genetic epidemiology evidence across multiple independent cohorts representing a 20-fold increase in AF cases over previous observational studies, we sought to conclusively answer whether Lp(a) has a causal role in AF development, and to demonstrate for the first time whether this role is independent of its effects in atherosclerotic cardiovascular disease (Figure 1).

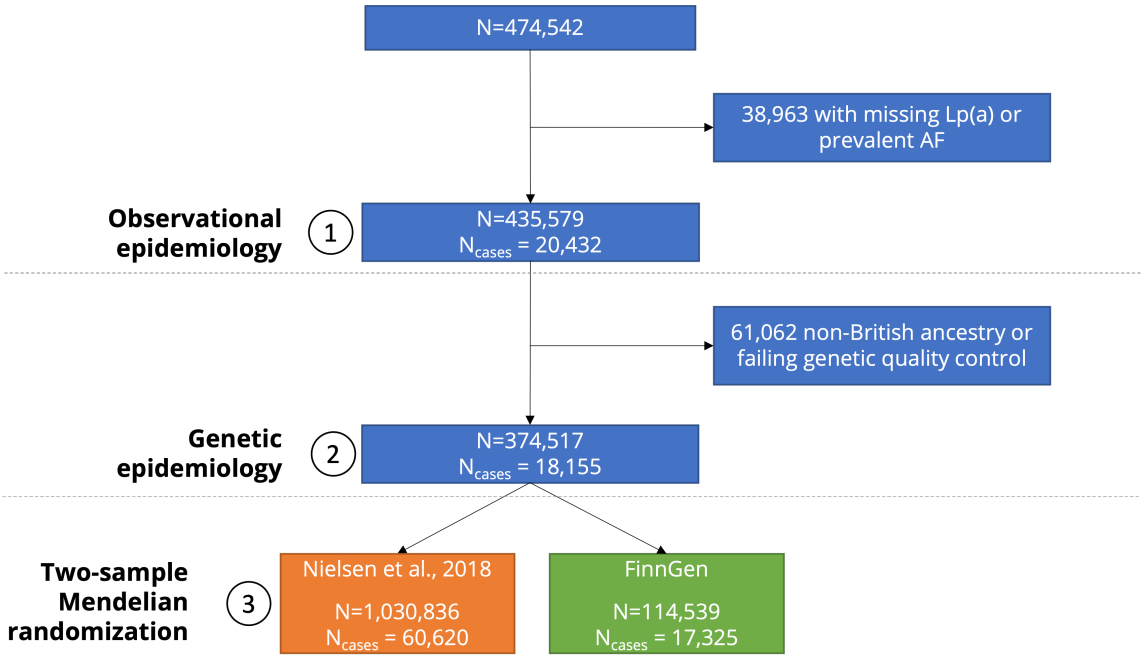


Figure 1. Flowchart depicting study design. *Study was testing Lp(a) as a causal risk factor for atrial fibrillation. In the UK Biobank (blue), participants were excluded if their Lp(a) was missing or if they had atrial fibrillation before their assessment centre visit. Genetic epidemiology was used to help infer the unconfounded relationship between Lp(a) and atrial fibrillation by leveraging random allocation of genetic variants at meiosis akin to the random assignment of participants to experimental groups in a randomized controlled trial. For these analyses, participants were further excluded if they were of non-British ancestry or missing genetic data. Finally, two-sample Mendelian randomization was performed to evaluate the effect of genetically predicted Lp(a) on atrial fibrillation in two independent cohorts from the Nielsen et al. and FinnGen studies. AF, atrial fibrillation; Lp(a), lipoprotein(a).*

Methods

Study Population

The UK Biobank is a population-scale longitudinal cohort study that recruited over 500,000 people between the ages of 37-73 across the United Kingdom from 2006-2010 (27). UK Biobank received ethical approval from the North West Multi-Centre Research Ethics Committee (REC reference: 11/NW/0382). This research was conducted using the UK Biobank under Application Number 15255. For this study, participant selection is depicted in Figure 1.

Measurement of Lipoprotein(a) in UK Biobank

Lp(a) was measured on a Randox AU5800 immunoturbidimetric assay that used a 5-point calibrator to reflect heterogeneity in Lp(a) isoform size. Analytical range for Lp(a) was from 5.76 to 189 nmol/L (2.68 to 87.91 mg/dL). Within-laboratory coefficient of variation for high, medium, and low concentration quality controls samples were 6.1, 4.4, and 3.8%. To monitor assay consistency, all samples were run with internal quality control samples between batches and operations used external quality assurance schemes according to the ISO 17025:2005 standard. Lp(a) values were detectable but flagged outside of the reportable range in 16.3% of samples. There were 45,352 and 31,775 participants with values below and above the reportable range, respectively (Supplemental Figure 1). Detectable values outside of analytical range were included in all analyses (Return 2321) (Supplemental Figure 1A). Sensitivity analyses were performed where any values that fell outside of the analytical range were winsorized to the lowest and highest reportable values, as appropriate (Supplemental Figure 1B). Lp(a) was converted from nmol/L to mg/dL by dividing by 2.15, as previously described (28).

Definitions of Atrial Fibrillation and Risk Factors in UK Biobank

Health information was collected from self-reported medical history and physical measures at recruitment, and ongoing developments through linked electronic health records. AF was defined based on the occurrence of one or more ICD-10 (I48.0 through I48.9) or OPCS-4 (K62.2, K62.3) codes in electronic health records from hospital inpatient admissions (field ID 41270) or death register (field ID 40001 and 40002). Sensitivity and specificity for ascertainment of AF using ICD-10 billing codes can vary between 42-97 and 88-100%, respectively (29, 30). Prevalent AF was defined as any event with a date of occurrence before the participant's first visit for recruitment into the study, while incident AF was defined as an event occurring after the first study visit. For subgroup analyses, prevalent disease (ischemic heart disease, type 2 diabetes, heart failure, or aortic valve stenosis) was defined as any event with a date of occurrence before the date of occurrence for AF. Definitions of risk factors for AF used in subgroup analyses and other relevant variables from the UK Biobank are provided in Supplemental Table 1.

Statistical Analysis

Survival Analysis in UK Biobank

For observational epidemiological analyses, we restricted participants to those with Lp(a) measured at recruitment and without prevalent AF (n= 435,579). Cox proportional hazards regression modelled the relationship between Lp(a) at recruitment and incident AF events in the overall cohort and each ancestry group (African, n=6,833; British, n=395,497; Non-British Caucasian, n=26,350; South Asian, n=6,899). Models were adjusted for age, sex, genetic ancestry, Townsend deprivation index, BMI, height, physical activity, diastolic blood pressure, systolic blood pressure, prevalent type 2 diabetes, smoking status, alcoholic

drinks per week, total cholesterol, HDL cholesterol, triglycerides, lipid-lowering medication, and anti-hypertensive medication use. The continuous relationship between Lp(a) and hazard ratios of incident AF was modelled using a restricted cubic spline adjusted for the aforementioned covariates. To identify effect modifiers of the Lp(a) and AF relationship, subgroup analyses were performed for common risk factors of AF selected *a priori* as specified in the Supplemental Methods. Given Lp(a) lowering therapies are being tested in participants with elevated Lp(a) (>70 mg/dL or 150 nmol/L), Lp(a) was also modelled as a categorical variable according to clinically accepted cutoffs for normal (<30 mg/dL or 75 nmol/L) and elevated (>70 mg/dL or 150 nmol/L) Lp(a) levels (31) [ClinicalTrials.gov Identifier: NCT04023552].

Genetic Risk Score Analysis in UK Biobank

For genetic epidemiological analyses, we further restricted the subset from observational epidemiological analyses to only participants with British ancestry and genetic data that passed quality control (n=374,516). Quality control of genetic data in the UK Biobank is described in Supplemental Methods. Least absolute shrinkage and selection operator (LASSO) regression was used to construct the genetic risk scores for Lp(a) using genetic variants within 500Kb of the *LPA* gene as described in the Supplemental Methods. Sensitivity analyses were performed using variants across the entire genome or within 50Kb of the LPA gene. In the UK Biobank, the genetic subset was divided into a training and testing set. Model training for the genetic risk score for Lp(a) was done using 10-fold cross-validation in the training set. Then, model performance was evaluated using the variance of Lp(a) explained (r^2) by the genetic score in an independent testing set.

Furthermore, we evaluated the association of the genetic risk score for Lp(a) with AF in the testing set. To maximize power to detect an association between genetically predicted Lp(a) and incident AF, the testing set consisted of all 18,155 cases of incident AF and 36,287 randomly selected age- (± 5 years) and sex-matched controls (n=54,442). The training set consisted of all remaining participants as it was exclusively used to generate the genetic risk score for Lp(a) (n=320,075). Logistic regression was used to model the relationship between the allele score for Lp(a) at recruitment and incident AF events in the overall cohort as a case-control analysis. Models were adjusted for age, sex, chip type (Affymetrix UK BiLEVE or UK Biobank Axiom arrays), assessment centre, and 40 principal components of ancestry. As a sensitivity analysis, we conducted a case-control analysis with only prevalent AF cases (n_{cases}=6,156) and combining prevalent with incident AF cases. Subgroup analyses were performed to identify effect modifiers as specified in the Supplemental Methods.

Mendelian randomization analysis of Lp(a) as a causal risk factor for atrial fibrillation

As described in the Supplemental Methods, 15 independent (r^2<0.001) genetic variants within 500Kb of *LPA* associated ($p<5x10^{-8}$) with Lp(a) were selected from the same training set from the UK Biobank. To replicate analyses in independent datasets and maximize statistical power, genetic associations with AF were derived from publicly available genome-wide association study (GWAS) summary statistics for two-sample Mendelian randomization analyses. The study by Nielsen *et al.* was the largest GWAS of AF to-date with 60,620 cases of AF and 970,216 controls (32). Since UK Biobank was a participating cohort in the study by Nielsen *et al*., there is partial overlap of samples. To

obtain a truly independent cohort for two-sample Mendelian randomization, summary statistics for AF were also downloaded from FinnGen freeze v4 release with 17,325 cases of AF and 97,214 controls (33).

Mendelian randomization compares the effect of each independent genetic variant on Lp(a) with its effect on AF using the inverse variance-weighted (IVW) method (34). To detect deviation from key assumptions underlying Mendelian randomization, standard sensitivity analyses were conducted: MR-Egger, weighted median, weighted mode, MR-RAPS (Robust Adjusted Profile Score), and MR-PRESSO (Mendelian Randomization Pleiotropy RESidual Sum and Outlier) (35–37). MR-PRESSO was performed using the *MR-PRESSO* package and all other Mendelian randomization analyses were implemented using the *TwoSampleMR* package (37, 38). Finally, Mendelian randomization was used to detect whether AF itself had a causal effect on Lp(a) by comparing effect of 112 independent genetic variants associated with AF ($p<5x10^{-8}$) in the study by Nielsen *et al.* Detailed methods are provided in the Supplemental Methods.

Evaluating potential mediation of lipoprotein(a) through atherosclerotic cardiovascular disease

Given the atherogenic properties of Lp(a) and role of atherosclerotic cardiovascular diseases as a risk factor for AF, we sought to assess whether part of the effect of Lp(a) on AF was mediated independently of atherosclerotic cardiovascular disease by comparing effect estimates from other atherogenic lipid fractions. As described for Lp(a), Cox proportional hazards regression was used to model the relationship between LDL and

triglycerides at recruitment and incident AF events in the UK Biobank cohort adjusted for aforementioned covariates in Lp(a) analyses. Furthermore, Mendelian randomization analysis was performed for LDL (OpenGWAS ID: ukb-d-30780_raw) and triglycerides (OpenGWAS ID: ukb-d-30870_raw) on AF using publicly available GWAS summary statistics in the UK Biobank from the Neale lab after excluding variants on chromosome 6 associated with Lp(a) ($p<0.05$) or within 500Kb of the *LPA* gene (39). To compare Lp(a) and LDL or triglyceride effect sizes, estimates were standardized with respect to equivalent effect of each lipid particle on CAD based on Mendelian randomization analysis using GWAS summary statistics from CARDIoGRAMplusC4D (n_{cases}=60,801) (40). The effect of LDL on AF is reported per 0.27 mmol/L of LDL and the effect of triglycerides on AF is reported per 0.69 mmol/L of triglycerides, which corresponds to the same effect on CAD as 50 nmol/L (23 mg/dL) increased Lp(a). Details for the calculations are provided in the Supplemental Methods. For each lipid fraction, IVW effect estimates from Nielsen *et al.* and FinnGen were meta-analyzed using a fixed-effects model. Cochran's Q statistic was used to test for heterogeneity between the effect estimates of Lp(a) and each lipid fraction on AF ($p_{hetero}<0.05$).

All individual-level UK Biobank data is available by application directly to the UK Biobank. Summary-level data for the studies by Nielsen *et al.* and FinnGen are publicly available for download.

All statistical analyses were performed in R version 3.6.0. A two-sided p-value less than 5.56×10^{-3} (0.05/9) was considered statistically significant for subgroup analyses. For

all other analyses, a two-sided p-value less than 0.05 was considered statistically significant.

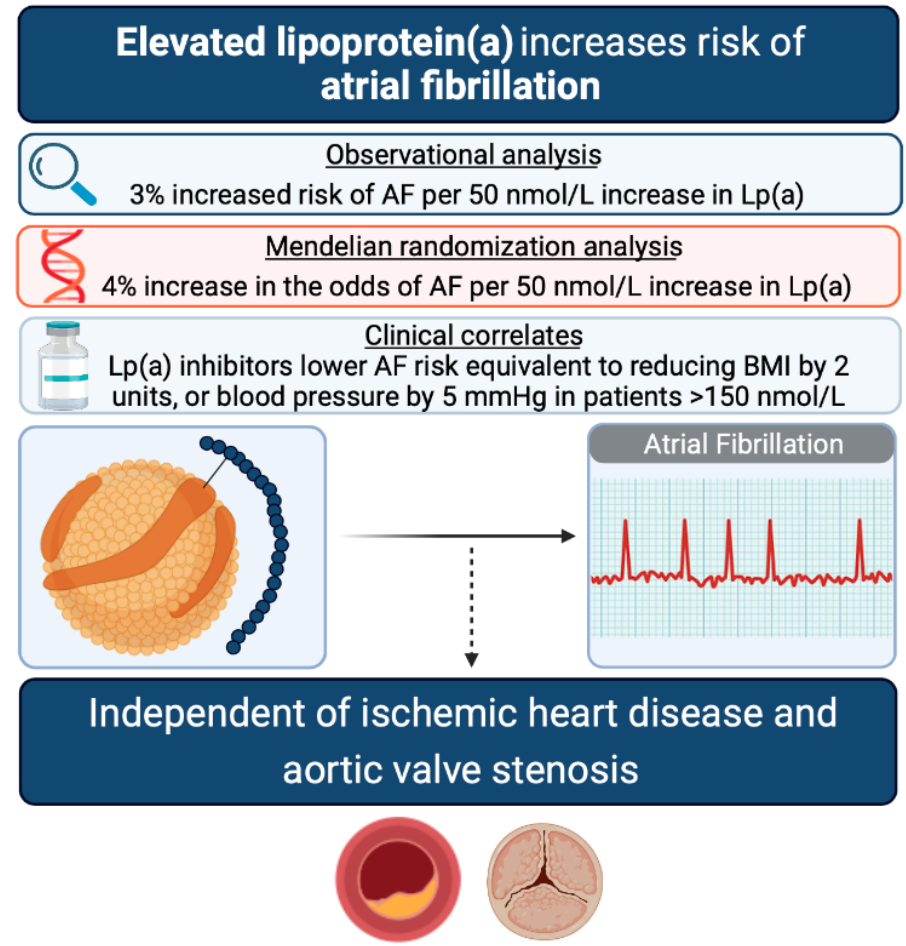
Results

Epidemiological association of lipoprotein(a) and atrial fibrillation in UK Biobank

After excluding participants with prevalent AF (1.4%, n=6,677) and missing Lp(a) measurement at recruitment (6.9%, n=32,748) in the UK Biobank, 435,579 participants were included (Supplemental Table 2). After a median 11 years of follow-up (interquartile range, 10.3 to 11.8 years), 20,432 participants developed incident AF corresponding to a rate of 4.37 events per 1000 person years.

After adjustment for common risk factors, Cox proportional hazards regression identified an increased incidence of AF by 3% per 50 nmol/L (23 mg/dL) increase in Lp(a) (HR=1.03 per 50 nmol/L Lp(a) increase; 95%CI=1.02 to 1.04; p= 1.65×10^{-8}) (Central Illustration). Sensitivity analysis winsorizing participants with Lp(a) levels outside of the reportable range showed consistent results (HR=1.03 per 50 nmol/L Lp(a) increase; 95%CI=1.02 to 1.05; p= 1.81×10^{-4}). Effect estimates were overlapping in non-British ancestries but less precise due to limited numbers of AF cases (Supplemental Figure 2 and Supplemental Table 3). Participants with Lp(a) above the cut-off to receive Lp(a)-lowering therapies (>150 nmol/L or >70 mg/dL) in a current phase 3 clinical trial had higher risk of AF (HR=1.10; 95%CI=1.05 to 1.15; p=1.72×10^{-5}) relative to participants with normal Lp(a) (<75 nmol/L or <30 mg/dL) (Figure 2A). In this group of at-risk participants, the prevalence of AF is 5.2% and assuming 80% decline in Lp(a), the prevalence of AF would fall to 4.9% representing 0.4% and 8% absolute and relative reduction, respectively, comparable to 2

units reduction in body mass index or 5 mmHg reduction in blood pressure (Supplemental Methods). In addition, subgroup analysis showed no evidence of effect modification or interaction according to common risk factors of AF (Supplemental Figure 4 and Supplemental Table 4).



Central Illustration. Lipoprotein(a) increases atrial fibrillation risk independent of atherosclerotic cardiovascular disease. *Lipoprotein(a) increased risk of atrial fibrillation using epidemiologic and genetic analyses across multiple independent cohorts and biobank studies. The effect is partially mediated independently of its effects on ischemic heart disease and aortic valve stenosis.*

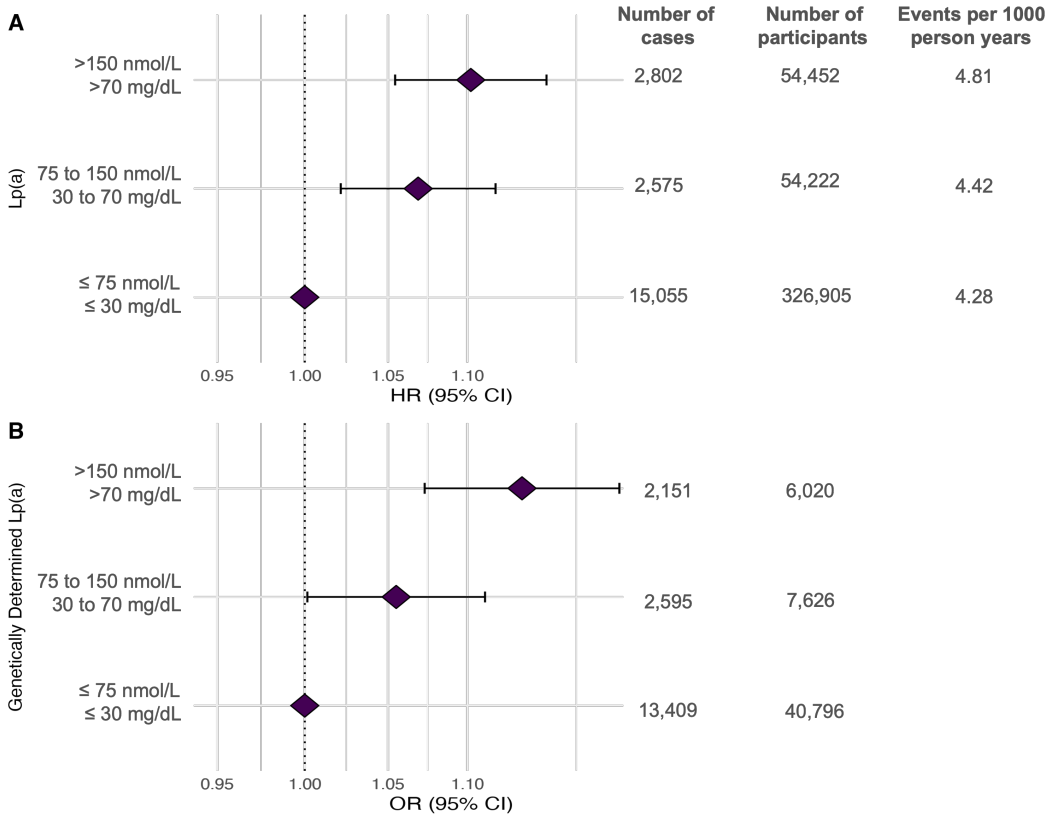


Figure 2. Greater risk of incident atrial fibrillation with increased Lp(a). *A) observed and B) genetically determined lipoprotein(a). Normal lipoprotein(a) levels are less than 75 nmol/L or 30 mg/dL. Participants with lipoprotein(a) levels above 150 nmol/L or 70 mg/dL are eligible for lipoprotein(a)-lowering therapies in a current clinical trial. Lp(a), lipoprotein(a)*

Association of genetically predicted lipoprotein(a) and atrial fibrillation in UK Biobank

In the UK Biobank, the genetic risk score alone explained 71.4% of the variance in Lp(a) (Figure 3). Moreover, each 50 nmol/L increase in genetically predicted Lp(a) was associated with an increased risk of incident AF (OR=1.03; 95%CI=1.02 to 1.05;

p=1.33x10^{-5}). Participants with genetically determined Lp(a) above the cut-off to receive Lp(a)-lowering therapies (>150 nmol/L or >70 mg/dL) had higher odds of AF (OR=1.14; 95%CI=1.07 to 1.20; p=1.22x10^{-5}) relative to participants with normal Lp(a) (<75 nmol/L or <30 mg/dL) (Figure 2B). An increased risk was consistently observed when looking at the association of genetically predicted Lp(a) with only prevalent AF cases (OR=1.03 per 50 nmol/L increased Lp(a); 95%CI= 1.01 to 1.06; p= 9.26x10^{-4}) and combined prevalent and incident AF cases (OR=1.03 per 50 nmol/L increased Lp(a); 95%CI= 1.02 to 1.04; p= 4.60x10^{-9}) (Supplemental Figure 5). Sensitivity analyses were performed using variants ±50Kb from the *LPA* gene and genome-wide variants that showed the majority of Lp(a) levels were explained by variants around the *LPA* gene (Supplemental Table 5) with a consistent risk-conferring effect of Lp(a) concentration on AF. In addition, there was no evidence of heterogeneous effects across *a priori* risk factor subgroups nor evidence of interaction with quantitative risk factors (all $p_{interaction}$>0.05): (Supplemental Figure 6 and Supplemental Table 6).

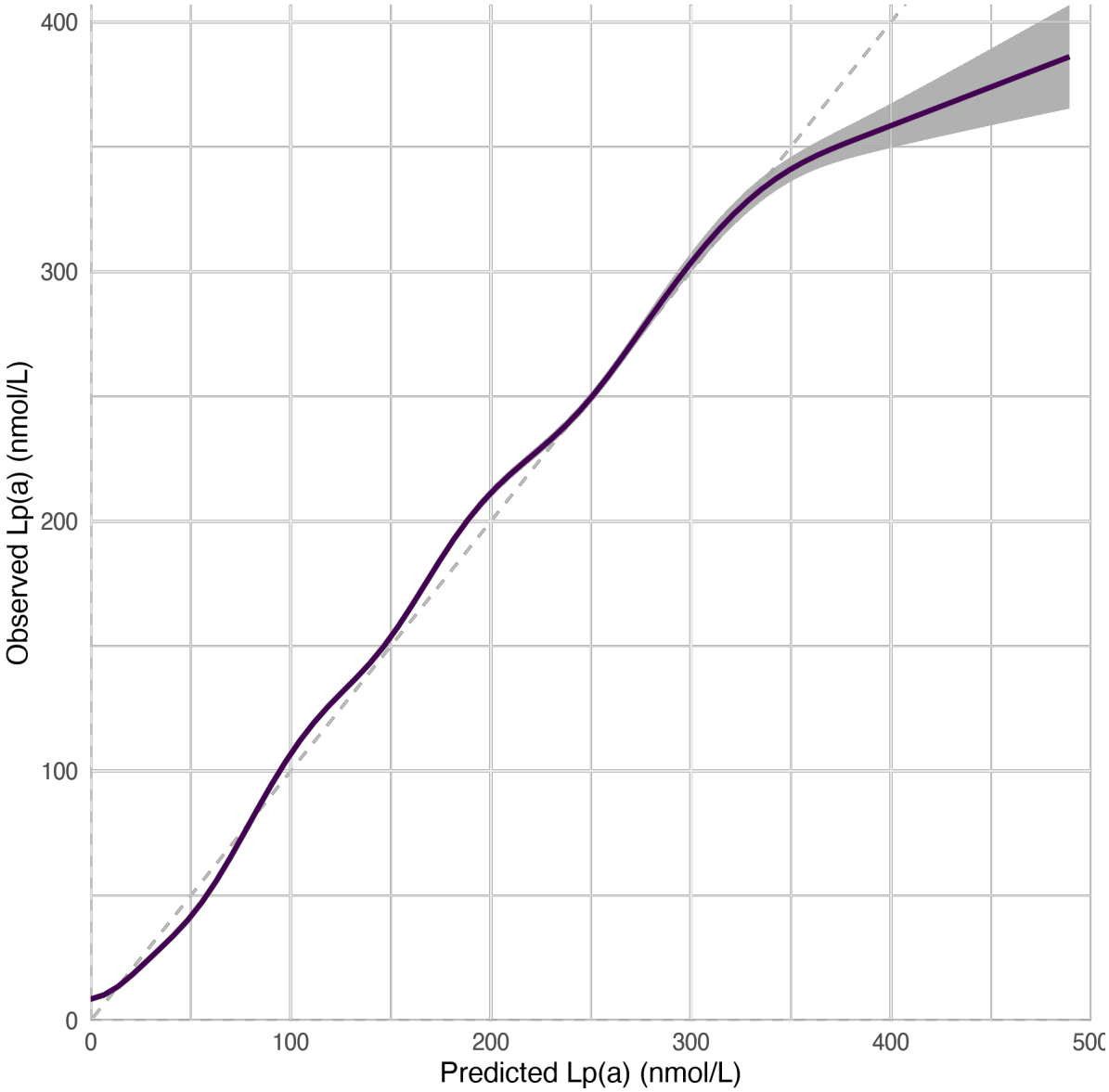


Figure 3. Genetically determined lipoprotein(a) levels are highly predictive of observed lipoprotein(a). *Genetic score for lipoprotein(a) explained 70% of the variance in observed lipoprotein(a) levels in British UK Biobank participants. Lp(a), lipoprotein(a)*

Mendelian randomization analysis of lipoprotein(a) and atrial fibrillation

To test for replication of the observed effect of Lp(a) on AF in independent datasets additional to the UK Biobank, we performed a two-sample Mendelian randomization analysis using publicly available GWAS summary statistics of AF from the study by Nielsen *et al.* and the FinnGen cohort. Using Mendelian randomization, 15 genetic variants within 500 Kb of the *LPA* gene associated with Lp(a) concentration in the UK Biobank were associated with increased risk of AF in both Nielsen *et al.* (OR=1.03 per 50 nmol/L Lp(a) increase; 95%CI=1.02 to 1.05; $p=9.93x10^{-8}$) and FinnGen (OR=1.08 per 50 nmol/L Lp(a) increase; 95%CI=1.04 to 1.12; $p=9.54x10^{-6}$) (Supplemental Table 7 and Figure 4). Sensitivity analyses were directionally consistent with no evidence of directional pleiotropy (Egger intercept p=0.42 and 0.99 for Nielsen *et al.* and FinnGen, respectively) or other apparent violation in the underlying assumptions for Mendelian randomization analyses (Supplemental Figure 7 and Supplemental Table 8). Moreover, Mendelian randomization analysis did not show an effect in the reverse direction for AF on Lp(a) levels itself (-0.195 nmol/L Lp(a) per 1 SD increase in log(OR) for AF; 95%CI=-0.587 to 0.197; p=0.33) (Supplemental Table 9). Finally, a Mendelian randomization analysis using only two variants associated with Lp(a) in previous studies showed a consistent effect of Lp(a) on AF in both Nielsen *et al.* (OR=1.03 per 50 nmol/L Lp(a) increase; 95%CI=1.02 to 1.04; $p=3.74x10^{-7}$) and FinnGen (OR=1.07 per 50 nmol/L Lp(a) increase; 95%CI=1.04 to 1.10; $p=7.58x10^{-6}$) (Supplemental Table 7) (19).

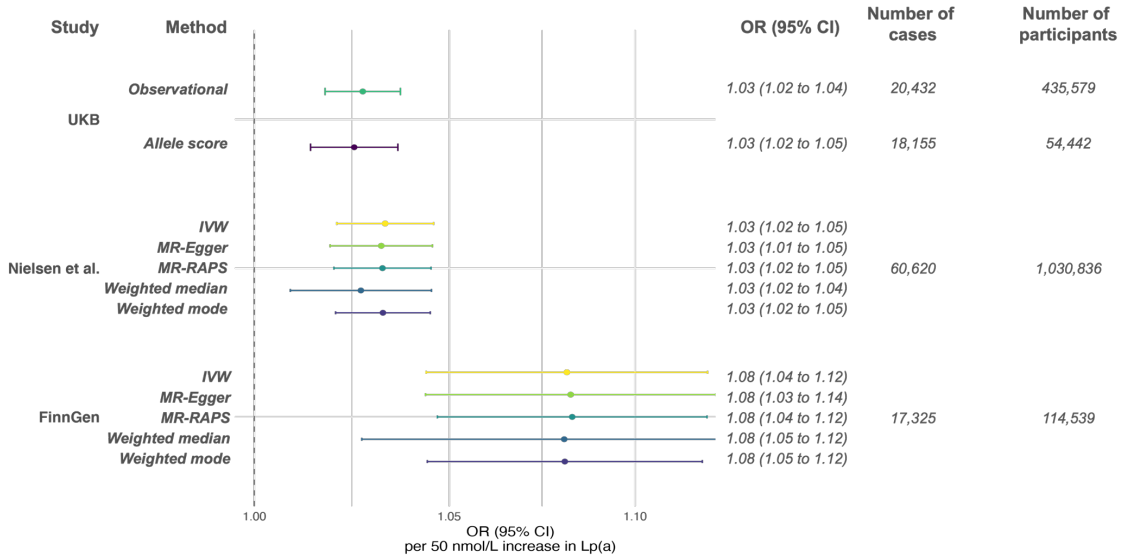


Figure 4. Lipoprotein(a) increases risk of atrial fibrillation in multiple independent studies. *Effect of lipoprotein(a) on atrial fibrillation is consistent using observed and genetically predicted levels in UK Biobank, and Mendelian randomization analysis using Nielsen et al. and FinnGen studies. CI, confidence interval; IVW, inverse variance weighted; Lp(a), lipoprotein (a); MR, Mendelian randomization; OR, odds ratio; RAPS, robust adjusted profile score.*

Evaluating potential mediation of lipoprotein(a) through atherosclerotic cardiovascular disease

Given the known relationship between Lp(a) and ischemic heart disease and aortic valve stenosis, we sought to establish whether the relationship of Lp(a) with AF was mediated through atherosclerotic cardiovascular disease or valvular disease. Subgroup analyses did not identify any modification of the effect of either observed or genetically predicted Lp(a) on AF according to prevalent ischemic heart disease (8.8%, n=38,543) or aortic valve stenosis (0.7%, n=3,085) status (Supplemental Figures 3 and 6). After

excluding participants with prevalent ischemic heart disease, the relationship between Lp(a) and AF remained unchanged (HR=1.02; 95%CI=1.00 to 1.04; $p=6.84x10^{-4}$) (Supplemental Figure 3). Mediation analysis determined that 62.2% (95%CI=48.9% to 80.4%) of the effect of observed Lp(a) and 39.2% (95%CI=26.7% to 73.3%) of the effect of genetically predicted Lp(a) on increased risk of atrial fibrillation was mediated through prevalent ischemic heart disease and aortic valve stenosis considered together (Supplemental Table 10). Furthermore, we tested whether the effect on AF was specific to Lp(a) or applicable to other lipid fractions. Unlike Lp(a), observed LDL (HR=0.89 per mmol/L; 95%CI= 0.87 to 0.91; $p=1.20x10^{-29}$) and triglyceride (HR=0.91 per mmol/L; 95%CI= 0.89 to 0.92; $p=1.22x10^{-28}$) levels at recruitment showed an inverse association with risk of incident AF in the UK Biobank. However, Mendelian randomization analyses showed no effect of LDL or triglycerides on AF (Supplemental Tables 11, 12, 13 and Supplemental Figure 8, 9). In fact, after standardizing changes in Lp(a) and LDL with respect to equivalent effects on CAD, there was a significant difference between effects reported for Lp(a) (OR=1.04 per 50 nmol/L Lp(a) increase; 95%CI=1.03 to 1.05; $p=9.23x10^{-10}$) and both LDL (OR=1.00 per 0.27 mmol/L LDL increase; 95%CI=0.99 to 1.02; p=0.62) ($p_{hetero} = 2.30x10^{-3}$) and triglycerides (OR=1.00 per 0.69 mmol/L triglyceride increase; 95%CI=0.97 to 1.03; p=0.99) (p_{hetero} = 0.04) on AF using Mendelian randomization.

Discussion

Using epidemiological and genetic analyses, our results implicate Lp(a) as a potentially causal mediator of AF. Higher Lp(a) levels were associated with greater risk of

incident AF independent of age, sex, BMI, blood pressure, smoking, and other risk factors in the UK Biobank. Genetic variants explaining 71% of variation in circulating Lp(a) levels were associated with a 3% increased risk of incident AF per 50 nmol/L increase in genetically predicted Lp(a). This result was replicated in two independent Mendelian randomization analyses using summary statistics from studies by Nielsen *et al.* and the FinnGen cohort. Ischemic heart disease and aortic valve stenosis are both known consequences of elevated Lp(a) and risk factors for AF, but they did not modify the effect of Lp(a) on incident AF (21, 22). Indeed, mediation analysis determined there was some residual effect of Lp(a) on AF independent from ischemic heart disease or aortic valve stenosis as these diseases mediated 39% of the total risk-conferring effect of genetically predicted Lp(a) on AF. Moreover, there was an inverse association in epidemiological analyses but no detectable effect in Mendelian randomization of genetically predicted LDL cholesterol or triglycerides on increased AF risk and a significant difference in effect size relative to Lp(a), making the effect specific to Lp(a) and unlikely to be mediated through atherosclerotic cardiovascular disease. Altogether, epidemiologic and genetic analyses support a role for Lp(a) on incident development of AF independent of atherosclerotic cardiovascular disease, demonstrating its effect on myocardial tissues extends beyond the aortic valve and coronary arteries that may point to new disease pathways and pharmacological targets.

Although the effect does not appear to be mediated solely through atherosclerotic cardiovascular disease, we can only speculate at the mechanism behind the risk-conferring effect of Lp(a) on AF. Lp(a) is a unique molecule with potential pro-atherogenic, pro-

thrombotic, and pro-inflammatory properties. In aortic valve stenosis, mechanical stress increases endothelial permeability allowing Lp(a) to infiltrate valvular tissue and induce gene expression that results in microcalcifications and local cell death (20, 41). A similar mechanism may extend to the atria resulting in cellular damage and electrical remodelling. Moreover, an *ex vivo* study showed that large reductions in Lp(a) induced anti-inflammatory gene expression and lower activation of circulating monocytes (42). Notably, reductions in Lp(a) with sufficient magnitude were only observed for antisense oligonucleotides against apolipoprotein(a) and were not observed with PCSK9 inhibitors.

A secondary analysis of the FIDELIO-DKD phase 3 trial recently found finerenone – a nonsteroidal mineralocorticoid receptor antagonist for treatment of diabetic kidney disease – reduced incidence of new-onset AF, and highlights the importance of therapies for AF comorbidities with ancillary benefits on AF (43). To this end, there is considerable interest regarding Lp(a) as an intervenable risk factor for cardiovascular diseases. Lp(a)HORIZON is an ongoing phase 3 clinical trial expected to complete in 2024 that is testing the efficacy of Lp(a) lowering for prevention of major adverse cardiovascular events [ClinicalTrials.gov Identifier: NCT04023552]. However, AF is not currently being assessed as an outcome. Our findings highlight the potential for Lp(a) to help fill the unmet need for preventative therapies of AF with complementary effects on cardiovascular comorbidities and suggest AF should be considered as an additional outcome in future studies.

A major strength of this study is the large-scale epidemiological analyses conferred by the UK Biobank cohort and public data from the study by Nielsen *et al.* and FinnGen

cohort. In comparison with previous observational studies of Lp(a) and AF, this study included 20-fold more incident AF cases improving statistical power (24, 25). A previous Mendelian randomization study explored the effects of seven lipids on risk of AF and identified a nominal association with Lp(a) that was not significant after correcting for multiple hypothesis testing. Our study provides more robust data to answer this question by triangulating evidence from observational and genetic analyses, increasing the sample size through the addition of UK Biobank and FinnGen cohorts, and exploring the mechanism mediating the effect between Lp(a) and AF (26). Accordingly, we demonstrate the beneficial mechanism of Lp(a) reduction on AF risk is partly independent of its known effects on atherosclerotic cardiovascular disease, which is relevant given the role of ischemic heart disease as a risk factor for AF. This finding is important for Lp(a) research as it suggests the role of Lp(a) in myocardial tissues extends beyond atherosclerotic cardiovascular disease. Furthermore, Mendelian randomization is particularly powerful for Lp(a) given the extent of variation in Lp(a) concentration explained by genetic variants and provides stronger evidence of a causal relationship between Lp(a) and AF. Notably, genetic determinants in the Lp(a) instrument lie within a narrow window around the *LPA* gene increasing confidence that their effects occur via Lp(a) and satisfy the pleiotropy assumptions underlying Mendelian randomization. Finally, we triangulated evidence across multiple types of analyses and independent studies to answer the central question. Lines of evidence from survival analyses, genetic risk scores for Lp(a), and Mendelian randomization with independent datasets all point to a risk-conferring effect of elevated Lp(a) in AF.

Study Limitations

Our study has limitations. Included population-scale cohorts ascertained cases of AF using electronic health records. An enrichment for disease that required hospitalization may limit generalizability to individuals with less severe disease, while undetected AF among controls might lead to an underestimate of the true effect size. Indeed, the magnitude of effect was larger in FinnGen, which may reflect cohort-specific differences such as more participant ascertainment in hospitals compared to the general population. Follow-up studies in cohorts designed to study AF are warranted. Relatedly, a limitation of the mediation analysis is that some atherosclerotic cardiovascular disease cases may be undetected among AF cases leading to an underestimate of the true proportion of mediation between Lp(a) and AF. Sample sizes for non-European ancestries were relatively smaller in the UK Biobank making it difficult to draw conclusions in other ancestries. In light of large ethnic differences in Lp(a) levels, it is particularly important to examine the generalizability of this relationship in other ancestries, although we did not find evidence of a large difference (15). The number of Kringle IV2 domain repeats within the *LPA* gene is the largest contributor to genetic variation in Lp(a) concentration and could not be directly measured in this investigation, but we were still able to explain over 70% of variation in Lp(a) using a predictor based on genotyping array data. Subgroup and interaction analyses were *ad hoc* and exploratory in nature, as such careful interpretation of results and follow-up studies are warranted. Finally, future research is required to identify the specific properties or subcomponents of Lp(a) that are responsible for the risk-

conferring effect, such as size versus concentration or oxidized phospholipid content relative to apolipoprotein(a).

Conclusions

Epidemiologic and genetic analyses implicate a potentially causal role for Lp(a) on risk of incident AF in population-scale cohorts that is independent of its effect on atherosclerotic cardiovascular disease. Given the role of Lp(a) as a risk factor for common comorbidities of AF – namely stroke and myocardial infarction – and the development of antisense oligonucleotides to lower Lp(a) to treat these comorbidities, these findings highlight an important beneficial pleiotropic effect these therapies may have on prevention or treatment of AF in high-risk groups with highly elevated Lp(a). Future experimental work and randomized controlled trials will be required to elucidate underlying physiological mechanisms and conclusively evaluate its efficacy as a preventative therapy for AF.

References

1. Chugh SS, Havmoeller R, Narayanan K, et al. Worldwide epidemiology of atrial fibrillation: A global burden of disease 2010 study. Circulation 2014;129:837–847.

2. Benjamin EJ, Muntner P, Alonso A, et al. Heart Disease and Stroke Statistics—2019 Update: A Report From the American Heart Association. Circulation 2019;139:e56–e528. Available at: https://www.ahajournals.org/doi/10.1161/CIR.0000000000000659.

3. Hindricks G, Potpara T, Dagres N, et al. 2020 ESC Guidelines for the diagnosis and management of atrial fibrillation developed in collaboration with the European Association for Cardio-Thoracic Surgery (EACTS). Eur. Heart J. 2021;42:373–498.

4. Conen D, Chae CU, Glynn RJ, et al. Risk of Death and Cardiovascular Events in Initially Healthy Women with New-Onset Atrial Fibrillation. JAMA - J. Am. Med. Assoc. 2011;305:2080–2087.

5. Chatterjee NA, Chae CU, Kim E, et al. Modifiable Risk Factors for Incident Heart Failure in Atrial Fibrillation. JACC Hear. Fail. 2017;5:552–560. Available at: https://linkinghub.elsevier.com/retrieve/pii/S2213177917302639.

6. Kalantarian S, Stern TA, Mansour M, Ruskin JN. Cognitive Impairment Associated With Atrial Fibrillation. Ann. Intern. Med. 2013;158:338. Available at: http://annals.org/article.aspx?doi=10.7326/0003-4819-158-5-201303050-00007.

7. Meyre P, Blum S, Berger S, et al. Risk of Hospital Admissions in Patients With Atrial Fibrillation: A Systematic Review and Meta-analysis. Can. J. Cardiol. 2019;35:1332–1343. Available at: https://doi.org/10.1016/j.cjca.2019.05.024.

8. Lau DH, Nattel S, Kalman JM, Sanders P. Modifiable Risk Factors and Atrial Fibrillation. Circulation 2017;136:583–596.

9. Wijffels MCEF, Kirchhof CJHJ, Dorland R, Allessie MA. Atrial Fibrillation Begets Atrial Fibrillation. Circulation 1995;92:1954–1968. Available at: https://www.ahajournals.org/doi/10.1161/01.CIR.92.7.1954.

10. Khatib R, Joseph P, Briel M, Yusuf S, Healey J. Blockade of the renin–angiotensin–aldosterone system (RAAS) for primary prevention of non-valvular atrial fibrillation: A systematic review and meta analysis of randomized controlled trials. Int. J. Cardiol. 2013;165:17–24. Available at: http://dx.doi.org/10.1016/j.ijcard.2012.02.009.

11. Schneider MP, Hua TA, Böhm M, Wachtell K, Kjeldsen SE, Schmieder RE.

Prevention of Atrial Fibrillation by Renin-Angiotensin System Inhibition. J. Am. Coll. Cardiol. 2010;55:2299–2307. Available at: https://linkinghub.elsevier.com/retrieve/pii/S0735109710010892.

12. Neefs J, van den Berg NWE, Limpens J, et al. Aldosterone Pathway Blockade to Prevent Atrial Fibrillation: A Systematic Review and Meta-Analysis. Int. J. Cardiol. 2017;231:155–161. Available at: http://dx.doi.org/10.1016/j.ijcard.2016.12.029.

13. Bang CN, Greve AM, Abdulla J, Køber L, Gislason GH, Wachtell K. The preventive effect of statin therapy on new-onset and recurrent atrial fibrillation in patients not undergoing invasive cardiac interventions: A systematic review and meta-analysis. Int. J. Cardiol. 2013;167:624–630. Available at: http://dx.doi.org/10.1016/j.ijcard.2012.08.056.

14. Bhatt DL, Steg PG, Miller M, et al. Cardiovascular Risk Reduction with Icosapent Ethyl for Hypertriglyceridemia. N. Engl. J. Med. 2019;380:11–22. Available at: http://www.nejm.org/doi/10.1056/NEJMoa1812792.

15. Paré G, Çaku A, McQueen M, et al. Lipoprotein(a) Levels and the Risk of Myocardial Infarction Among 7 Ethnic Groups. Circulation 2019;139:1472–1482. Available at: https://www.ahajournals.org/doi/10.1161/CIRCULATIONAHA.118.034311.

16. Zekavat SM, Ruotsalainen S, Handsaker RE, et al. Deep coverage whole genome sequences and plasma lipoprotein(a) in individuals of European and African ancestries. Nat. Commun. 2018;9:1–14.

17. Larsson SC, Gill D, Mason AM, et al. Lipoprotein(a) in Alzheimer, Atherosclerotic, Cerebrovascular, Thrombotic, and Valvular Disease. Circulation 2020;141:1826–1828. Available at:

https://www.ahajournals.org/doi/10.1161/CIRCULATIONAHA.120.045826.

18. Thanassoulis G, Campbell CY, Owens DS, et al. Genetic Associations with Valvular Calcification and Aortic Stenosis. N. Engl. J. Med. 2013;368:503–512. Available at: http://www.nejm.org/doi/10.1056/NEJMoa1109034.

19. Clarke R, Peden JF, Hopewell JC, et al. Genetic Variants Associated with Lp(a) Lipoprotein Level and Coronary Disease. N. Engl. J. Med. 2009;361:2518–2528. Available at: http://www.nejm.org/doi/abs/10.1056/NEJMoa0902604. Accessed June 22, 2017.

20. Zheng KH, Tsimikas S, Pawade T, et al. Lipoprotein(a) and Oxidized Phospholipids Promote Valve Calcification in Patients With Aortic Stenosis. J. Am. Coll. Cardiol. 2019;73:2150–2162. Available at: https://doi.org/10.1016/j.jacc.2019.01.070.

21. Benjamin EJ, Levy D, Vaziri SM, D'agostino RB, Belanger AJ, Wolf PA. Independent Risk Factors for Atrial Fibrillation in a Population-Based Cohort: The Framingham Heart Study. JAMA J. Am. Med. Assoc. 1994;271:840–844. Available at: http://jama.jamanetwork.com/article.aspx?doi=10.1001/jama.1994.03510350050036.

22. Greve AM, Gerdts E, Boman K, et al. Prognostic importance of atrial fibrillation in asymptomatic aortic stenosis: The Simvastatin and Ezetimibe in Aortic Stenosis study. Int. J. Cardiol. 2013;166:72–76. Available at: http://dx.doi.org/10.1016/j.ijcard.2011.09.064.

23. Igarashi Y, Yamaura M, Ito M, Inuzuka H, Ojima K, Aizawa Y. Elevated serum lipoprotein(a) is a risk factor for left atrial thrombus in patients with chronic atrial fibrillation: A transesophageal echocardiographic study. Am. Heart J. 1998;136:965–971.

Available at: https://linkinghub.elsevier.com/retrieve/pii/S0002870398701516.

24. Garg PK, Guan W, Karger AB, et al. Lp(a) (Lipoprotein [a]) and Risk for Incident Atrial Fibrillation. Circ. Arrhythmia Electrophysiol. 2020;13:477–479. Available at: https://www.ahajournals.org/doi/10.1161/CIRCEP.120.008401.

25. Aronis KN, Zhao D, Hoogeveen RC, et al. Associations of Lipoprotein(a) Levels With Incident Atrial Fibrillation and Ischemic Stroke: The ARIC (Atherosclerosis Risk in Communities) Study. J. Am. Heart Assoc. 2017;6:1–11. Available at: https://www.ahajournals.org/doi/10.1161/JAHA.117.007372.

26. Jiang Q, Qin D, Yang L, et al. Causal effects of plasma lipids on the risk of atrial fibrillation: A multivariable mendelian randomization study. Nutr. Metab. Cardiovasc. Dis. 2021;31:1569–1578. Available at: https://doi.org/10.1016/j.numecd.2021.02.011.

27. Sudlow C, Gallacher J, Allen N, et al. UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. PLOS Med. 2015;12:e1001779. Available at: http://journals.plos.org/plosmedicine/article/file?id=10.1371/journal.pmed.1001779&type=printable. Accessed April 29, 2018.

28. Khera A V., Everett BM, Caulfield MP, et al. Lipoprotein(a) Concentrations, Rosuvastatin Therapy, and Residual Vascular Risk. Circulation 2014;129:635–642. Available at: https://www.ahajournals.org/doi/10.1161/CIRCULATIONAHA.113.004406.

29. Shah RU, Mukherjee R, Zhang Y, et al. Impact of Different Electronic Cohort Definitions to Identify Patients With Atrial Fibrillation From the Electronic Medical

Record. J. Am. Heart Assoc. 2020;9:e014527.

30. Kokotailo RA, Hill MD. Coding of stroke and stroke risk factors using International Classification of Diseases, revisions 9 and 10. Stroke 2005;36:1776–1781.

31. Anderson TJ, Grégoire J, Pearson GJ, et al. 2016 Canadian Cardiovascular Society Guidelines for the Management of Dyslipidemia for the Prevention of Cardiovascular Disease in the Adult. Can. J. Cardiol. 2016;32:1263–1282.

32. Nielsen JB, Thorolfsdottir RB, Fritsche LG, et al. Biobank-driven genomic discovery yields new insight into atrial fibrillation biology. Nat. Genet. 2018;50:1234–1239. Available at: http://dx.doi.org/10.1038/s41588-018-0171-3.

33. Borodulin K, Tolonen H, Jousilahti P, et al. Cohort Profile: The National FINRISK Study. Int. J. Epidemiol. 2018;47:696-696i. Available at: https://academic.oup.com/ije/article/47/3/696/4641873.

34. Burgess S, Dudbridge F, Thompson SG. Combining information on multiple instrumental variables in Mendelian randomization: Comparison of allele score and summarized data methods. Stat. Med. 2016;35:1880–1906.

35. Bowden J, Smith GD, Burgess S. Mendelian randomization with invalid instruments: Effect estimation and bias detection through Egger regression. Int. J. Epidemiol. 2015;44:512–525.

36. Zhao Q, Wang J, Hemani G, Bowden J, Small DS. Statistical inference in two-sample summary-data Mendelian randomization using robust adjusted profile score. arXiv Prepr. 2018. Available at: https://arxiv.org/pdf/1801.09652.pdf. Accessed July 30, 2018.

37. Verbanck M, Chen C-Y, Neale B, Do R. Detection of widespread horizontal

pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. Nat. Genet. 2018;50:693–698. Available at: http://www.nature.com.libaccess.lib.mcmaster.ca/articles/s41588-018-0099-7.pdf. Accessed April 25, 2018.

38. Hemani G, Zheng J, Elsworth B, et al. The MR-Base platform supports systematic causal inference across the human phenome. Elife 2018;7:e34408. Available at: https://elifesciences.org/articles/34408. Accessed May 31, 2018.

39. Elsworth B, Lyon M, Alexander T, et al. The MRC IEU OpenGWAS data infrastructure. bioRxiv 2020.

40. Nikpay M, Goel A, Won H-HH, et al. A comprehensive 1000 Genomes–based genome-wide association meta-analysis of coronary artery disease. Nat. Genet. 2015;47:1121–1130. Available at: http://www.nature.com/articles/ng.3396. Accessed December 13, 2016.

41. Bouchareb R, Mahmut A, Nsaibia MJ, et al. Autotaxin derived from lipoprotein(a) and valve interstitial cells promotes inflammation and mineralization of the aortic valve. Circulation 2015;132:677–690.

42. Stiekema LCA, Prange KHM, Hoogeveen RM, et al. Potent lipoprotein(a) lowering following apolipoprotein(a) antisense treatment reduces the pro-inflammatory activation of circulating monocytes in patients with elevated lipoprotein(a). Eur. Heart J. 2020;41:2262–2271.

43. Filippatos G, Bakris GL, Pitt B, et al. Finerenone Reduces Onset of Atrial Fibrillation in Patients with Chronic Kidney Disease and Type 2 Diabetes. J. Am. Coll. Cardiol.

2021;184:107229. Available at: https://doi.org/10.1016/j.buildenv.2020.107229.

CHAPTER 4:

Effects of lifelong testosterone exposure on health and disease using Mendelian randomization

Published in *eLife*. 9:e58914 (2020)

RESEARCH ARTICLE

Effects of lifelong testosterone exposure on health and disease using Mendelian randomization

Pedrum Mohammadi-Shemirani[1,2,3], Michael Chong[1,2,4], Marie Pigeyre[1,5], Robert W Morton[6], Hertzel C Gerstein[1,5], Guillaume Paré[1,2,7,8]*

[1]Population Health Research Institute, David Braley Cardiac, Vascular and Stroke Research Institute, Hamilton, Canada; [2]Thrombosis and Atherosclerosis Research Institute, David Braley Cardiac, Vascular and Stroke Research Institute, Hamilton, Canada; [3]Department of Medical Sciences, McMaster University, Hamilton, Canada; [4]Department of Biochemistry and Biomedical Sciences, McMaster University, Hamilton, Canada; [5]Department of Medicine, McMaster University, Hamilton Health Sciences, Hamilton, Canada; [6]Department of Kinesiology, McMaster University, Hamilton, Canada; [7]Department of Pathology and Molecular Medicine, McMaster University, Michael G. DeGroote School of Medicine, Hamilton, Canada; [8]Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, Canada

Abstract Testosterone products are prescribed to males for a variety of possible health benefits, but causal effects are unclear. Evidence from randomized trials are difficult to obtain, particularly regarding effects on long-term or rare outcomes. Mendelian randomization analyses were performed to infer phenome-wide effects of free testosterone on 461 outcomes in 161,268 males from the UK Biobank study. Lifelong increased free testosterone had beneficial effects on increased bone mineral density, and decreased body fat; adverse effects on decreased HDL, and increased risks of prostate cancer, androgenic alopecia, spinal stenosis, and hypertension; and context-dependent effects on increased hematocrit and decreased C-reactive protein. No benefit was observed for type 2 diabetes, cardiovascular or cognitive outcomes. Mendelian randomization suggests benefits of long-term increased testosterone should be considered against adverse effects, notably increased prostate cancer and hypertension. Well-powered randomized trials are needed to conclusively address risks and benefits of testosterone treatment on these outcomes.

***For correspondence:** pareg@mcmaster.ca

Competing interest: See page 13

Funding: See page 13

Received: 14 May 2020
Accepted: 13 October 2020
Published: 16 October 2020

Reviewing editor: Dolores Shoback, University of California, San Francisco, United States

Introduction

In developed countries, rising rates of both serum testosterone level testing and therapy initiation have been observed among older male patients (***Handelsman, 2013***; ***Layton et al., 2014***). In the USA alone, it is estimated 1.5–1.7% of males are prescribed testosterone (***Baillargeon et al., 2018***; ***Jasuja et al., 2017***). Randomized clinical trials (RCT) have attempted to elucidate the benefits and risks of testosterone treatment (***Bhasin et al., 2018a***; ***Gagliano-Jucá and Basaria, 2019***). These studies identified short-term beneficial effects on bone mineral density (BMD), sexual function, body fat and muscle mass, and anaemia; potential adverse effects on venous thrombosis and coronary artery plaque; and no effects on cognitive function, fatigue, or hemoglobin $A1_c$ ($HbA1_c$) (***Bhasin et al., 2018a***; ***Gagliano-Jucá and Basaria, 2019***; ***Mohler et al., 2018***; ***Snyder et al., 2018***). However, given the logistic and financial challenges involved in a well-powered RCT with appropriate follow-up, there is unlikely to be satisfactory evidence regarding long-term effects and risks of

eLife digest Men experience a gradual decline in their testosterone levels as they grow older. However, the effects of testosterone and the consequences of supplementation on the human body have been unclear.

Scientists use so-called randomized controlled trials to establish cause-and-effect and to reduce bias. In these experiments, participants are randomly assigned to a either a treatment group (that receives the intervention being tested) or a control group (that either receives an alternative intervention, a dummy or placebo, or no intervention at all).

Randomization ensures that both groups are balanced, and any resulting differences can be attributed to the treatment. However, randomized controlled trials are time-consuming and expensive, so trials of testosterone have had relatively small numbers of participants and short follow-up periods. This makes it difficult to draw conclusions about any potential effects of testosterone administration on less common diseases in men.

Now, Paré et al. investigated the effects of naturally produced testosterone using Mendelian randomization, which mimics randomized trials by exploiting the fact that parents randomly pass on their unique genetic variants to their children at conception. This random assignment of genetic variants leads to its informal namesake, "nature's clinical trial", and provides the ability to study cause-and-effect for any genetically determined factors, such as testosterone levels.

Paré et al. studied the long-term effects of testosterone on 22 diseases previously explored in randomized controlled trials, and hundreds of other traits and diseases that have not been investigated in any randomized controlled trials yet.

The Mendelian randomization analysis made it possible to examine the effects of lifelong naturally elevated testosterone levels on 469 traits and diseases. Paré et al. found that testosterone increased the density of bone mineral and decreased body fat. However, it also increased the risks of prostate cancer, high blood pressure, baldness and a condition affecting the spine. It also increased the number of red blood cells and decreased a marker of inflammation, which may be beneficial or detrimental depending on the context.

This shows that genetic analyses can be powerful methods to prioritize the allocation of limited resources towards investigating the most pressing clinical questions. The results of this study may help inform physicians and patients about the effects of long-term testosterone use. Ultimately, large randomized controlled trials are needed to conclusively address the cause-and-effect on these diseases.

adverse outcomes, such as myocardial infarction (MI), stroke and cancer (***Gagliano-Jucá and Basaria, 2019***). Given the rates of testosterone prescription, efforts to resolve the causal effects of testosterone on health outcomes have important public health implications (***Bhasin et al., 2018a***).

Mendelian randomization (MR) is a technique for causal inference that leverages the random allocation of genetic variants to infer the unconfounded relationship between an exposure and outcome. Similar to the random assignment of participants to experimental groups in a RCT, genetic variants are randomly allocated at meiosis (***Davies et al., 2018***). For instance, if individuals genetically randomized to produce higher testosterone develop different rates of cardiovascular disease (CVD), then MR analysis supports a causal effect of testosterone on risk of CVD (***Figure 1—figure supplement 1***). Notably, this technique has previously replicated RCT findings, among others demonstrating causal roles for LDL cholesterol and dysglycemia on CVD risk (***Holmes et al., 2015***; ***Ross et al., 2015***). Earlier MR studies investigating the effects of testosterone have demonstrated harmful effects on lipid levels but inconsistent effects on CVD, and they were limited by the small number of genetic variants (***Schooling et al., 2018***; ***Zhao et al., 2014***). A recent MR study using the UK Biobank identified a large number of genetic variants associated with testosterone and found evidence for harmful effects on several types of cancers but sex-specific effects on type 2 diabetes (T2D) (***Ruth et al., 2020***). This study highlighted the importance of performing sex-specific analyses for testosterone, but it was focused on glycemic and oncologic traits (***Ruth et al., 2020***). Therefore, we sought to expand the scope of prior studies by performing a comprehensive scan of the effects of free testosterone on human disease in males.

We hypothesized that MR and genetic risk score (GRS) analyses would enable estimation of the causal effects of longstanding exposure to high levels of free testosterone on health outcomes in males. We first conducted a genome-wide association study (GWAS) for calculated free testosterone (CFT) in male participants of the UK Biobank (n = 161,268) cohort to identify genetic determinants of free testosterone levels. Then, using MR, we investigated the causal effects of lifelong genetically-elevated free testosterone levels on a priori health outcomes previously investigated in RCTs of testosterone treatment, encompassing: expected clinical benefits (physical activity, strength, fat-free body mass, body fat, BMD, dementia, depression) and potential adverse effects (androgenic alopecia, heematocrit, T2D, prostate cancer, benign prostate hyperplasia, blood pressure, CVD, heart failure, ischemic stroke) (***Figure 1***; ***Bhasin et al., 2018a***; ***Gagliano-Jucá and Basaria, 2019***; ***Mohler et al., 2018***; ***Snyder et al., 2018***). Finally, we used GRS to investigate the associations of lifelong genetically-elevated free testosterone levels on 439 health outcomes, encompassing diseases (n = 415) and biomarkers of health (n = 24) (***Figure 1***).

Results

Genetic determinants of CFT in males

To calculate free testosterone levels, 187,524 males in the white, British subset of the UK Biobank cohort were excluded if they had missing levels of total testosterone, SHBG and albumin, or self-reported taking androgen medications. After these exclusions, the study population consisted of 161,268 males with an average CFT of 0.210 nmol/L (***Supplementary file 1*** - Table 1 and ***Figure 1—figure supplement 2***).

There were 13,338 genetic variants associated with CFT that reached genome-wide significance ($p<5\times10^{-8}$). After removing genetic variants associated with natural-log-transformed SHBG, there were 7048 genetic variants that comprised 93 independent signals carried forward for subsequent genetic analyses (***Supplementary file 1*** - Table 2 and ***Figure 1—figure supplement 3***). Overall, chip-based heritability of CFT was estimated at 15% (95% CI = 14 to 16), while these 93 independent genetic variants associated with CFT explained 3.7% of the total variance of CFT levels in males from the UK Biobank.

Effect of genetically-predicted free testosterone on 22 a priori health outcomes

In males from the UK Biobank, sample size for the quantitative risk factors ranged from 30,439 to 156,403, while number of cases for dichotomous outcomes ranged from 1003 to 70,283 (***Table 1***). After adjusting for the 22 outcomes tested, one-sample MR analysis using IVW regression identified significant effects of CFT on hematocrit percentage, body fat-free percentage, body fat percentage, heel BMD, androgenic alopecia, and prostate cancer (***Table 1***). Each 0.1 nmol/L higher CFT had beneficial effects on increased heel BMD (0.40 SD; 95% CI = 0.25 to 0.54; $p=1.10\times10^{-7}$), increased body fat-free percentage (1.91%; 95% CI = 1.48 to 2.35; $p=9.06\times10^{-18}$), and decreased body fat percentage (−1.88%; 95% CI = −2.31 to −1.45; $p=1.65\times10^{-17}$), but deleterious effects on increased hematocrit percentage (1.37%; 95% CI = 1.12 to 1.62; $p=1.03\times10^{-27}$), risk of prostate cancer (OR = 1.51; 95% CI = 1.21 to 1.88; $p=2.1\times10^{-4}$), and risk of androgenic alopecia (OR = 1.49; 95% CI = 1.19 to 1.86; $p=5.28\times10^{-4}$) (***Figure 3—figure supplements 1–6***). Leave-one-out analyses did not identify any outlying individual genetic variants responsible for the observed effects on any significant outcomes.

Sensitivity analyses were performed to detect violations of MR assumptions. Egger regression did not detect evidence of directional pleiotropy for any outcomes ($p_{intercept}$ <0.05) (***Supplementary file 1*** - Table 3). Results using MR-RAPS were consistent with IVW regression method for all significant outcomes (***Supplementary file 1*** – Table 4). However, MR-PRESSO detected evidence of pleiotropic variants for hematocrit percentage, body fat-free percentage, body fat percentage, heel BMD, androgenic alopecia, whole body fat-free mass, hemoglobin A1C, glucose, handgrip strength, systolic blood pressure, diastolic blood pressure, T2D, and benign prostate hyperplasia (***Supplementary file 1*** - Table 5). However, removal of pleiotropic variants made no changes to the significance or interpretation of earlier results using IVW regression (***Supplementary file 1*** - Table 5).

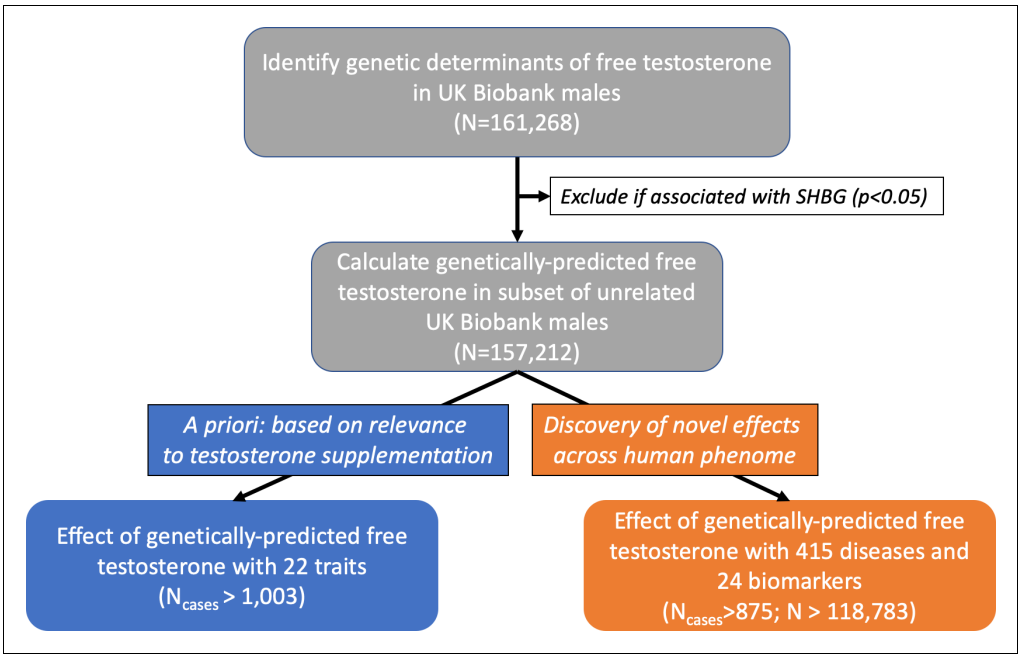


Figure 1. Flowchart depicting overall study design. Free testosterone levels were calculated in males from the UK Biobank cohort. Then, genetic variants were tested for association with levels of CFT and carried forward if: genome-wide significant ($p<5\times10^{-8}$) and unassociated with SHBG ($p<0.05$). In the subset of unrelated males, these genetic variants were used to investigate the effect of genetically-predicted CFT on: (1) 22 a priori outcomes relevant to suspected effects of testosterone treatment using Mendelian randomization, and (2) 439 outcomes in a hypothesis-free approach using a weighted genetic risk score. CFT, calculated free testosterone; MR, Mendelian randomization; SHBG, sex hormone-binding globulin.
The online version of this article includes the following figure supplement(s) for figure 1:

Figure supplement 1. Comparison of randomized controlled trial (RCT) and Mendelian randomization (MR) study designs demonstrating the common foundation behind interpretation of a causal effect of testosterone on cardiovascular disease (CVD).
Figure supplement 2. Distribution of free testosterone levels calculated using the Vermeulen equation in males from the UK Biobank cohort.
Figure supplement 3. Manhattan plot showing distribution of p-values from genome-wide association study of calculated free testosterone after exclusion of SHBG-associated variants based on chromosomal location.
Figure supplement 4. Distribution of sex hormone-binding globulin in males from the UK Biobank.
Figure supplement 5. Quantile-quantile plot for genome-wide association study of calculated free testosterone levels (before exclusion of SHBG-associated genetic variants).
Figure supplement 6. Distribution of total testosterone levels in males from the UK Biobank cohort.

Phenome-wide effects of genetically-predicted free testosterone

To discover novel effects of free testosterone, we tested for the association of a GRS for testosterone with 415 diseases and 24 biomarkers in the same subpopulation of unrelated males from the UK Biobank. Sample size for biomarkers ranged from 118,783 for lipoprotein(a) to 149,940 for total cholesterol, while number of cases for diseases ranged from 876 for 'localized superficial swelling, mass, or lump' to 40,960 for 'hypertension' (***Figure 2—source data 1***). After adjusting for the 439 outcomes tested, each 0.1 nmol/L increase in genetically-predicted CFT was significantly associated with beneficial effects on lowered C-reactive protein ($\beta = -0.085$ SD; 95% CI = −0.119 to −0.052; $p=6.15\times10^{-7}$) but adverse effects on increased creatinine ($\beta = 0.113$ SD; 95% CI = 0.079 to 0.146; $p=4.78\times10^{-11}$), lowered apolipoprotein A ($\beta = -0.018$ g/L; 95% CI = −0.026 to −0.01;

Table 1. Effect of calculated free testosterone on 22 health outcomes from the UK Biobank relevant to effects of testosterone treatment in males.

Outcome	Effect per 0.1 nmol/L increased CFT (95% CI)	P-value	Sample Size Cases/Controls
Outcomes with Expected Clinical Benefits			
Body fat-free percentage*	1.91% (1.48 to 2.35)	9.06E-18	154254
Body fat percentage*	−1.88% (−2.31 to −1.45)	1.65E-17	153772
Heel bone mineral density*	0.40 SD (0.25 to 0.54)	1.10E-07	90676
Depression	OR = 1.45 (1.1 to 1.91)	7.77E-03	4725/152485
Accelerometer-based physical activity	0.89 milligravity (−0.05 to 1.82)	0.06	30439
All fracture	OR = 0.89 (0.71 to 1.11)	0.30	9133/148077
Handgrip strength	0.29 kg (−0.31 to 0.89)	0.34	156400
All dementia	OR = 1.26 (0.67 to 2.34)	0.47	1003/156207
Outcomes with Potential Adverse Effects			
Hematocrit percentage*	1.37% (1.12 to 1.62)	1.03E-27	152872
Prostate cancer*	OR = 1.51 (1.21 to 1.88)	2.10E-04	7586/149624
Androgenic alopecia*	OR = 1.49 (1.19 to 1.86)	5.28E-04	70283/85756
Benign prostatic hyperplasia	OR = 1.36 (1.10 to 1.67)	3.80E-03	10894/146316
Myocardial infarction	OR = 1.23 (1 to 1.53)	0.05	9398/147812
Glucose	−0.06 mmol/L (−0.14 to 0.02)	0.12	138307
Hemoglobin A1c	−0.34 mmol/mol (−0.82 to 0.15)	0.17	149828
All stroke	OR = 1.18 (0.90 to 1.56)	0.23	4569/152641
Diastolic blood pressure	0.27 mmHg (−0.30 to 0.85)	0.35	148384
Ischemic stroke	OR = 0.92 (0.61 to 1.37)	0.67	2122/155088
Systolic blood pressure	−0.12 mmHg (−1.23 to 1.00)	0.84	148383
Type 2 diabetes	OR = 1.02 (0.81 to 1.28)	0.87	11079/146131
Venous thromboembolism	OR = 1.02 (0.74 to 1.4)	0.92	4127/153083
Heart failure	OR = 1.01 (0.76 to 1.34)	0.95	4288/152922

* Significant adjusting for Bonferroni correction of 22 outcomes ($p<2.27\times10^{-3}$) .
CFT, calculated free testosterone.

$p=1.55\times10^{-5}$), lowered HDL (β = −0.074 SD; 95% CI = −0.109 to −0.039; $p=3.62\times10^{-5}$), and increased risks of hypertension (OR = 1.17; 95% CI = 1.08 to 1.26; $p=2.83\times10^{-5}$), and spinal stenosis (OR = 2.03; 95% CI = 1.51 to 2.75; $p=3.82\times10^{-6}$) (***Table 2*** and ***Figure 2***).

As confirmation, we demonstrated the GRS was indeed not associated with natural log-transformed natural log-transformed SHBG levels in males (p=0.12). For all statistically significant outcomes, associations were directionally consistent after removing participants taking blood pressure medication (***Supplementary file 1*** - Table 6) or cholesterol-lowering medication (***Supplementary file 1*** - Table 7). Further sensitivity analyses were performed by repeating the one-sample MR analysis using 52 genetic variants associated with total testosterone in males from the UK Biobank (***Supplementary file 1*** - Table 8). For all statistically significant outcomes, effects observed using total testosterone genetic variants were directionally consistent with CFT, and results for all outcomes are presented in ***Supplementary file 1*** - Tables 9 and 10. Finally, most effect estimates for genetically-predicted testosterone in this stu dy were comparable in magnitude to effect sizes reported in RCTs except bone mineral density (***Figure 3***).

Discussion

We herein perform MR and GRS analyses of CFT to identify effects of endogenous free testosterone in males on 461 health outcomes. All effects are reported in terms of 0.1 nmol/L of CFT to

Table 2. Effects of calculated free testosterone on 439 health outcomes in males from the UK Biobank significant after adjusting for multiple hypothesis testing using Bonferroni correction ($p<1.14\times10^{-4}$).

Outcome	Effect per 0.1 nmol/L increased CFT (95% CI)	P-value	Sample Size Cases/Controls
Creatinine	0.113 SD (0.079 to 0.146)	4.78×10^{-11}	149849
C-reactive protein	−0.085 SD (−0.119 to −0.052)	6.15×10^{-7}	149547
Spinal stenosis	OR = 2.03 (1.51 to 2.75)	3.82×10^{-6}	1917/150919
Apolipoprotein A	−0.018 g/L (−0.026 to −0.01)	1.55×10^{-5}	138185
HDL cholesterol	−0.074 SD (−0.109 to −0.039)	3.62×10^{-5}	138394
Essential hypertension	OR = 1.17 (1.08 to 1.27)	7.53×10^{-5}	40809/115957
Hypertension	OR = 1.17 (1.08 to 1.26)	1.05×10^{-4}	40960/115957

CFT, calculated free testosterone; HDL, high density lipoprotein; GRS, genetic risk score.

approximate expected effect sizes after initiation of testosterone treatment (***Bhasin et al., 2018b***). Among 22 a priori outcomes with suspected effects based on RCTs of testosterone treatment, MR analyses demonstrated that each 0.1 nmol/L increase in CFT was associated with adverse effects on

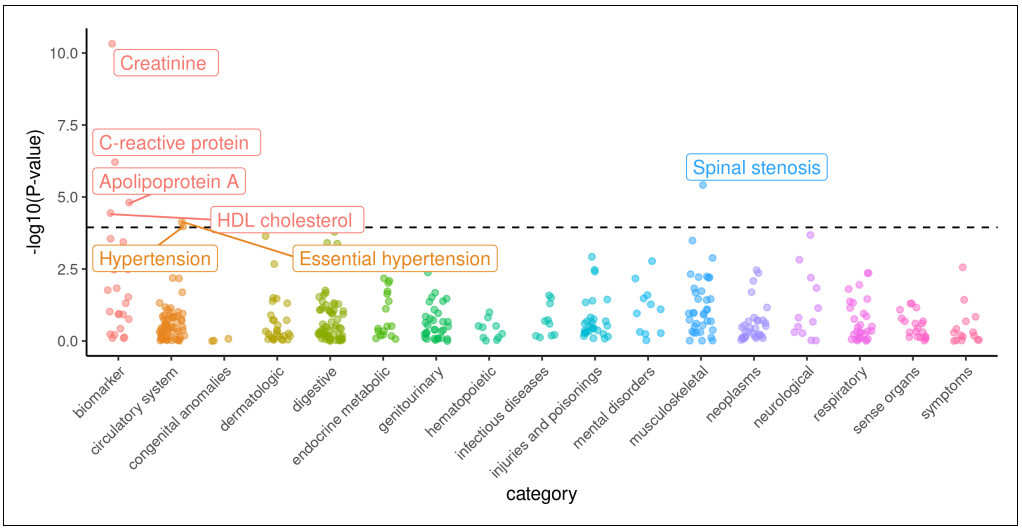


Figure 2. Phenome-wide survey of effects of genetically-predicted calculated free testosterone on 439 health outcomes in males from the UK Biobank. Logistic or linear regression was used to assess the association of the genetic score for free testosterone against each dichotomous or quantitative outcome, respectively. $-\log_{10}$(p-values) for the association of each outcome on the y-axis are stratified into subcategories on the x-axis. Labelled outcomes were statistically significant adjusting for multiple hypothesis testing ($p<1.14\times10^{-4}$).
The online version of this article includes the following source data for figure 2:

Source data 1. Associations of genetically-predicted calculated free testosterone for 439 health outcomes across the human phenome.

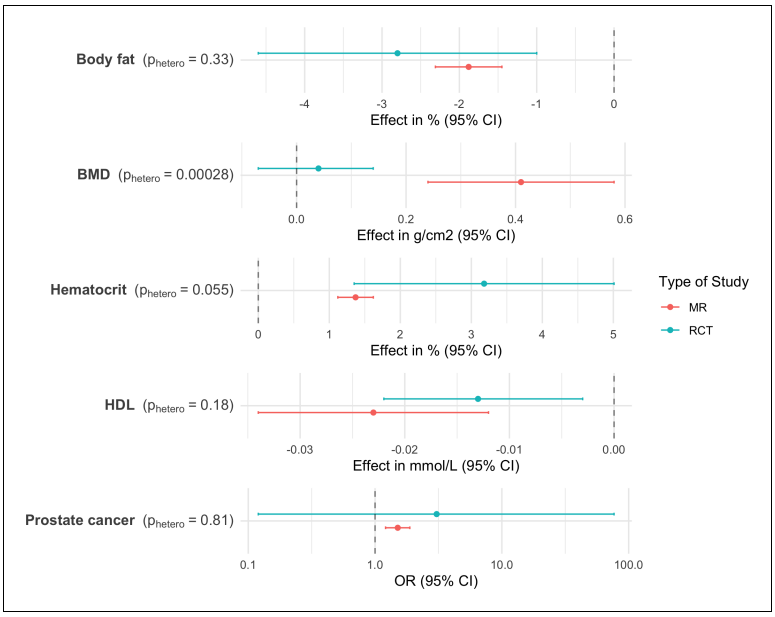


Figure 3. Comparison of effect sizes reported in randomized controlled trials and Mendelian randomization analyses. Error bars indicate 95% confidence intervals around the effect estimate. MR effect estimates are reported in terms of 0.1 nmol/L of CFT to approximate expected effect sizes after initiation of testosterone treatment (***Bhasin et al., 2018b***).

The online version of this article includes the following figure supplement(s) for figure 3:

Figure supplement 1. Comparison of effect of calculated free testosterone on hematocrit percentage using Mendelian randomization with IVW and Egger regression methods.

Figure supplement 2. Comparison of effect of calculated free testosterone on body fat-free percentage using Mendelian randomization with IVW and Egger regression methods.

Figure supplement 3. Comparison of effect of calculated free testosterone on body fat percentage using Mendelian randomization with IVW and Egger regression methods.

Figure supplement 4. Comparison of effect of calculated free testosterone on heel bone mineral density using Mendelian randomization with IVW and Egger regression methods.

Figure supplement 5. Comparison of effect of calculated free testosterone on prostate cancer using Mendelian randomization with IVW and Egger regression methods.

Figure supplement 6. Comparison of effect of calculated free testosterone on androgenic alopecia using Mendelian randomization with IVW and Egger regression methods.

increased risk of prostate cancer, risk of androgenic alopecia, and hematocrit percentage, but beneficial effects on increased heel BMD, increased body fat-free percentage and decreased body fat percentage. Findings on body composition, hematocrit, and BMD are consistent with short-term effects in randomized trials of testosterone treatment (***Bhasin et al., 2018a***). Although testosterone treatment has not been conclusively shown to increase risk of prostate cancer and androgenic alopecia in RCTs, androgen suppression therapies, such as of 5α-reductase inhibitors, are used as treatment for androgenic alopecia and prostate cancer (***Adil and Godwin, 2017***; ***Andriole et al., 2010***). The increased risk of prostate cancer replicates effects of testosterone observed in a previous MR analysis using independent data from the PRACTICAL consortium, and further supports the role of testosterone in development of these outcomes. As the leading cause of cancer among men, the predicted 1.5-fold increased risk as a result of changes in testosterone observed after initiation of

testosterone treatment warrants further investigation in clinical trials and greater scrutiny in at-risk patient populations (**American Cancer Society, 2019**; **Bhasin et al., 2018b**). Furthermore, these results cast doubt on cardiovascular, cognitive, or metabolic benefit for increased testosterone, as we do not find evidence of a beneficial effect of CFT on hard endpoints, such as dementia, MI, stroke, fractures, or T2D (**Aukrust et al., 2009**). Most of the estimates from MR analyses were comparable with effect sizes from RCTs (**Figure 3**). There was only significant heterogeneity between the effects on BMD for MR and RCT, but it is difficult to make direct comparisons due to variable change in testosterone levels after administration of testosterone in each RCT, different methods and anatomical sites of BMD estimation, and differences between short-term effects in RCTs relative to lifelong effects in MR.

Among the remaining outcomes without well-established effects from RCTs, we identified evidence of novel associations between an increased GRS for CFT with adverse effects on creatinine, HDL, apolipoprotein A, hypertension, and spinal stenosis, but beneficial effects on C-reactive protein. Higher genetically-predicted free testosterone was associated with increased creatinine (β = 0.113 SD; 95% CI = 0.079 to 0.146; $p=4.78\times10^{-11}$). Mechanistically, effects of testosterone on renal function are unclear, but this effect may be mediated through the known effect of testosterone on increased muscle mass which is tightly related to serum creatinine (**Carrero et al., 2009**; **Filler et al., 2016**; **Schutte et al., 1981**). HDL cholesterol (β = −0.074 SD; 95% CI = −0.109 to −0.039; $p=3.62\times10^{-5}$) and its main protein component, apolipoprotein A (β = −0.018 g/L; 95% CI = −0.026 to −0.01; $p=1.55\times10^{-5}$), were both decreased with higher genetically-predicted free testosterone. Likewise, the Testosterone Trials found male participants over 65 years of age randomized to testosterone experienced mildly lowered HDL cholesterol levels after 12 months (**Mohler et al., 2018**; **Snyder et al., 2018**). Higher free testosterone was associated with decreased C-reactive protein (CRP) (β = −0.085 SD; 95% CI = −0.119 to −0.052; $p=6.15\times10^{-7}$). Although the Testosterone Trials did not find any change in CRP in its testosterone arm, testosterone is widely-believed to have suppressive effects on the immune system which may extend to markers of inflammation such as CRP (**Trigunaite et al., 2015**). Furthermore, despite no effect on SBP or DBP, our analyses suggest 0.1 mol/L higher free testosterone is associated with increased risk of hypertension (OR = 1.17; 95% CI = 1.08 to 1.27; $p=1.05\times10^{-4}$). Given the multifactorial nature of this disease, the apparent discrepancy between blood pressure and hypertension may be explained by an effect on other risk factors that develop into hypertension. Moreover, both human and animal studies suggest a role of testosterone on hypertension. A randomized controlled trial found testosterone administration increased levels of NT-proBNP, and studies of both transgender men and anabolic steroid users have found testosterone increased arterial stiffness and blood pressure (**Bachmann et al., 2019**; **Hartgens and Kuipers, 2004**; **Velho et al., 2017**). Meanwhile, animal models have shown testosterone may aggravate hypertension and exacerbate increased production of reactive oxygen species specifically in hypertensive but not normotensive rat vascular endothelial tissue (**Chignalia et al., 2012**; **Reckelhoff et al., 1998**). Testosterone is widely-believed to have anti-inflammatory and osteogenic effects, but our analyses showed an association with higher risk of spinal stenosis (OR = 2.03; 95% CI = 1.51 to 2.75; $p=3.82\times10^{-6}$). However, the literature shows some evidence that higher testosterone is associated with greater loss of cartilage in healthy older males, and evidence from mouse models suggest testosterone has a sex-specific role in worsening osteoarthritis, a common risk factor for spinal stenosis (**Hanna, 2005**; **Hl et al., 2007**).

In comparison to previous MR studies, our results broaden the scope of the existing literature by comprehensively assessing the effects of testosterone on 461 health outcomes including hard endpoints and intermediate biomarkers. Moreover, a key strength of this study was the stringent attempt to control for pleiotropic effects of SHBG on free testosterone by conservatively removing any genetic variants in the GRS that were associated with SHBG ($p<0.05$). The apparent difference between protective effects of testosterone observed in a previous MR analysis of testosterone and lack of protective effect in our study might be a result of less stringent control for pleiotropic effects of SHBG in the previous study. Given studies have identified associations between SHBG and risk of T2D independent of testosterone and a direct role of SHBG in mediating signalling on target cells, insufficient controls for SHBG may lead to residual pleiotropic effects (**Lakshman et al., 2010**; **Rosner et al., 2010**; **Vikan et al., 2010**). Other reasons may include genetic variants explaining less variation in testosterone levels in our study, fewer cases of T2D leading to inadequate statistical

power to detect weaker effects in our study, or other differences between the populations of the UK Biobank in our study and DIAGRAM consortium used by ***Ruth et al., 2020***.

There are several limitations of this study. First, an assumption of the MR analysis is that the effect of the genetic variant on the outcome occurs only through free testosterone levels, such that there are no pleiotropic effects through other proteins or mechanisms (***Davies et al., 2018***). This concern was minimized by the use of multiple genetic variants, which limited the likelihood of a common alternative pathway confounding our observation. Moreover, we performed several sensitivity analyses and excluded genetic variants associated with SHBG levels, which is a potential source of pleiotropy through its effects on other hormones. Although a stringent p-value threshold was selected for genetic variants, the winner's curse phenomenon may still bias genetic effect sizes due to the same sample being used to select genetic variants and estimate effect sizes on testosterone. Additionally, one-sample MR may be susceptible to bias towards the confounded estimate if the genetic variants are 'weak instruments', which can occur if the genetic variants don't explain enough of the variance in free testosterone levels (***Davies et al., 2018***). To address this concern, we confirmed the selected genetic variants were strong instruments using a common threshold in MR literature (F-statistic >10) (***Davies et al., 2018***). Next, the UK Biobank is generally healthier and higher socioeconomic status than the general population, so there are insufficient cases to detect effects on certain rarer outcomes, such as Alzheimer's disease, and inadequate power to identify weaker effects of free testosterone on common outcomes. Relatedly, an inherent limitation for outcomes ascertained using linked electronic medical records is a lack of adjudication and consistent application of codes in clinical practice. In the UK Biobank, CFT levels were below the reference ranges for young healthy individuals, which may be attributable to the older age of the cohort and inherent inaccuracy of immunoassays at lower levels of total testosterone. Total testosterone levels are similarly low relative to reference ranges and comparable to previous studies in the UK Biobank (***Peila et al., 2020***; ***Petermann-Rocha et al., 2020***). Additional sources of variability introduced into the total testosterone measurements include differences in fasting times, diets, and time of day at which blood was drawn from participants. Nevertheless, genetic variants associated with testosterone consistently replicated known effects of testosterone on established outcomes, such as body fat, body fat-free mass, and hematocrit (***Table 1***). Furthermore, although the free hormone hypothesis is still debated by experts, we found largely consistent effects on outcomes using genetically-predicted free testosterone and total testosterone (***Handelsman, 2017***). The only significant outcomes from MR analyses with free testosterone that showed no significant effect with total testosterone across all MR methods were HDL (p=0.55) and apolipoprotein A (p=0.45). Finally, these results represent lifelong effects of endogenous free testosterone and may not necessarily reflect effects of exogenous testosterone treatment, which can vary in duration, age of initiation, and dosage.

Taken altogether, the decision to initiate long-term testosterone use warrants careful consideration of benefits and risk. Beneficial effects on body composition, sexual function, hematocrit, and BMD should be weighed against detrimental effects on androgenic alopecia, prostate cancer, hypertension and spinal stenosis, and no detectable beneficial effects on other major clinical endpoints. Ultimately, well-designed and appropriately powered RCTs, such as the ongoing TRAVERSE trials (clinicaltrials.gov, NCT03518034), are necessary to conclusively address questions of safety and effectiveness of testosterone treatment. However, as demonstrated in this study, genetically-informed analyses can be powerful tools to aid health professionals in prioritizing allocation of limited resources towards investigating the most pressing questions.

Materials and methods

Study population - UK Biobank

The UK Biobank is a large-scale longitudinal cohort study that recruited over 500,000 people between the ages of 37–73 across the United Kingdom from 2006 to 2010 (***Sudlow et al., 2015***) (RRID:SCR_012815). UK Biobank received ethical approval from the North West Multi-Centre Research Ethics Committee (REC reference: 11/NW/0382). This research was conducted using the UK Biobank under Application Number 15255. For this study, UK Biobank participants were included if white British ancestry, and no self-reported androgen medication at recruitment based on field ID 20003.

Measurement of testosterone and sex hormone-binding globulin in UK Biobank

In the UK Biobank, total testosterone and sex hormone-binding globulin (SHBG) were measured on a Beckman Coulter Unicel DXI 800 using a one-step competitive analysis and two-step sandwich immunoassay, respectively. Analytical range for the immunoassays of total testosterone and SHBG were 0.35 to 55.52 and 0.33 to (226-242) nmol/L, respectively. For total testosterone, within-laboratory CV for high, medium, and low concentration quality control samples were 4.15, 3.66, and 8.34%. For SHBG, within-laboratory CV for high, medium, and low concentration quality control samples were 5.22, 5.25, and 5.67%. For each blood sample drawn at recruitment, testosterone, SHBG, and albumin were each measured only once. Testosterone and SHBG measurements were flagged if they fell outside the manufacturer's observed reportable range, or samples reported high levels of bilirubin, hemoglobin or lipids/turbidity that might interfere with the assay. Testosterone measurements were flagged if levels of total protein (<55 or>85 g/L) or triglycerides (>20 mmol/L) could interfere with the assay measurements. To monitor assay consistency, all samples were run with internal quality control samples between batches and operations used external quality assurance schemes against the ISO 17025:2005 standard.

Genome-wide association study of CFT

Individual-level genetic data was available for 488,317 participants that consented to blood collection and genotyping. Genotyping was performed with the Applied Biosystems UK Biobank Lung Exome Variant Evaluation (UK BiLEVE) and UK Biobank Axiom arrays (Affymetrix Research Services Laboratory, Santa Clara, California, USA). Description of quality control has been previously described in detail (***Bycroft et al., 2017***). Genetic variants located in the human leukocyte antigen gene complex were excluded due to extensive pleiotropic effects.

For genome-wide association testing, samples were restricted to a subset of 161,268 males with white British ancestry, no androgen medication (n = 2,137), and no missing values of testosterone, SHBG, or albumin at recruitment. Free testosterone at recruitment was calculated using the Vermeulen equation (***Vermeulen et al., 1999***). CFT levels were winsorized such that outlying values greater or less than four standard deviations (SD) away from the mean in males were set to 4 SD.

This study was restricted to genetic variants from 'v3' release of the UK Biobank data including those present in the Haplotype Reference Consortium and 1000 Genomes panels with imputation imputation quality greater than 0.7, no deviation from Hardy-Weinberg equilibrium ($p>1\times10^{-10}$) and minor allele frequency greater than 1% (***McCarthy et al., 2016***). To allow for genetic relatedness between participants, linear mixed models in BOLT-LMM were used to test for associations of genetic variants (***Loh et al., 2015***). The model was adjusted for age, age^2, chip type, assessment center, and the first 20 genetic principal components. Genetic variants near the *SHBG* gene may alter binding affinity for testosterone thereby violating assumptions of the Vermeulen equation, or risk having pleiotropic effects through binding of other sex hormones (***Ohlsson et al., 2011***). Therefore, any genetic variants associated with CFT reaching genome-wide significance ($p\leq5\times10^{-8}$) were excluded if associated with natural log-transformed SHBG levels at a stringent threshold ($p<0.05$) in the same subset of the UK Biobank (***Figure 1—figure supplement 4***). To arrive at an independent set of genetic variants, variants associated with CFT but not SHBG were pruned based on linkage disequilibrium (LD) at a threshold of r^2 <0.01 using Europeans from 1000 Genomes phase three as reference panel (***Abecasis et al., 2012***) (RRID:SCR_006828).

Genomic inflation factor (λ) was 1.2 and calculated as the ratio of the median test statistic from the GWAS relative to the expected median test statistic under a null model (***Figure 1—figure supplement 5***). To distinguish between an inflated λ due to population stratification or polygenic inheritance of the trait, the intercept of an LD score regression line was determined to be 1.03 indicating the observed inflation could be attributed to polygenicity rather than uncontrolled population stratification. LD score regression was performed and intercept was calculated with LDSC software (***Bulik-Sullivan et al., 2015***) using 1000 Genomes Europeans phase three data as the LD reference panel (***Abecasis et al., 2012***).

Definition of health-related UK Biobank outcomes

For MR analyses, 22 health outcomes were selected a priori based on relevance with known or suspected effects of testosterone treatment and categorized based on expected beneficial or adverse effects from RCT data. Outcomes with expected beneficial effects were fractures at any site, heel BMD, body fat percentage, body fat-free percentage, dementia, depression, handgrip strength, and physical activity level measured by wrist-worn accelerometer. Outcomes with potential adverse effects were stroke, androgenic alopecia, benign prostate hyperplasia (BPH), blood pressure, glucose, hematocrit percentage, hemoglobin A1c, heart failure, prostate cancer, MI, type 2 diabetes (T2D), and venous thromboembolism. Depression was coded using a 'broad' definition as previously described, which included self-reported depressive symptoms with associated impairment, or having sought help for 'nerves, anxiety, tensions or depression' (***Howard et al., 2018***). Androgenic alopecia was defined based on participants' responses to the question, 'Which of the following best describes your hair/balding pattern?' (field ID 2395). Available options were four pictures of hair patterns (***Supplementary file 1*** – ***Figure 1***). Individuals with pattern 3 or four were cases, pattern 1 and 2 were controls, and 'do not know' or 'prefer not to answer' responses were excluded. Physical activity was assessed using the overall acceleration average from wrist-worn accelerometer devices over the course of approximately 7 days. Following UK Biobank recommendations, individuals were excluded from the analysis based on poorly calibrated data (field ID: 90016) or having worn the device for insufficient time to get a stable measure of physical activity (field ID: 90015) (***Doherty et al., 2017***). Blood pressure measures were coded as the average of two automated measurements of blood pressure taken a few moments apart by a registered nurse using an Omron 705 IT electronic blood pressure monitor. Body fat percentage and whole body fat-free mass were estimated based on impedance measurements from a Tanita BC418MA body composition analyser. Heel BMD was estimated as a T-score based on quantitative ultrasound index through the calcaneus relative to that expected in someone of the same sex. Handgrip strength was calculated as the average of right and left hands measured using a Jamar J00105 hydraulic hand dynamometer. hemoglobin $A1_C$ was measured using high performance liquid chromatography analysis on a Bio-Rad VARIANT II Turbo. Glucose was measured using hexokinase analysis on a Beckman Coulter AU5800. Hematocrit percentage was measured using a Coulter LH750 and calculated as the relative volume of packed erythrocytes to whole blood, computed by the formula: $\frac{red\ blood\ cells * mean\ corpuscular\ volume}{10}$. Detailed descriptions of all 22 outcomes are shown in ***Supplementary file 1*** – Table 11.

For hypothesis-free GRS analyses, we included 24 blood biomarkers measured at recruitment and 415 diseases derived from linked electronic medical records (***Supplementary file 1*** - Table 12; ***Brion et al., 2013***; ***Denny et al., 2013***; ***Wu et al., 2019***). Disease outcomes were defined using the previously published 'PheCode' scheme to aggregate ICD-10 codes from hospital episodes (field ID 41270), death registry (field ID 40001 and 40002), and cancer registry (field ID 40006) records (***Denny et al., 2013***; ***Wu et al., 2019***). Given the small number of cases for many disease outcomes, any outcomes with detectable odds ratios less than 0.5 or greater than 2 per 0.1 nmol/L at 80% power were excluded ($n_{cases} < 871$) based on approximate changes in response to testosterone supplementation (***Bhasin et al., 2018b***; ***Brion et al., 2013***; ***Traustadóttir et al., 2018***). After these exclusions, there were 415 diseases that remained for subsequent analyses in this study. Furthermore, all blood biomarkers measured by the UK Biobank at recruitment were included except estradiol and rheumatoid factor, which were complicated by majority missing values below the limit of detection of the assay ($n_{biomarkers} = 24$). Detailed descriptions of all 439 outcomes (415 diseases and 24 biomarkers) are shown in ***Supplementary file 1*** – Table 12.

Mendelian randomization analysis

In a subset of unrelated males with White British ancestry, the association of all independent genetic variants associated with CFT were determined for each of the 22 a priori outcomes using additive genetic models in BGENIE v1.2 and adjusted for the same covariates as the model for CFT (***Bycroft et al., 2017***). For each of the 22 outcomes, one-sample MR analysis was used to combine the effect of each independent genetic variant on CFT with its effect on the outcome using the inverse variance-weighted (IVW) method (***Burgess et al., 2016***). Effect estimates were reported per 0.1 nmol/L increase in CFT levels based on approximate changes in response to testosterone treatment (***Bhasin et al., 2018b***). For dichotomous outcomes, odds ratios were approximated as

previously described (**Adams et al., 2018**) by converting linear effect estimates from BGENIE to log-odds scale using:

$\log(OR) = \frac{}{k(1-k)}$, where k is the proportion of cases for the given outcome.

Given the polygenic nature of testosterone and potential for pleiotropy, for outcomes with statistically significant effects using the IVW method, standard sensitivity analyses were conducted to correct for pleiotropic effects, such as MR-Egger, MR-RAPS, and MR-PRESSO (**Bowden et al., 2015**; **Verbanck et al., 2018**). To investigate and correct for directional pleiotropy on each outcome, we performed Egger regression. For outcomes with y-intercept of the regression line significantly different from 0 ($p<0.05$), there was evidence of directional pleiotropy and the causal estimate from MR Egger was reported to attempt to control for pleiotropic effects (**Bowden et al., 2015**). As a sensitivity analysis robust to idiosyncratic pleiotropy and weak instrument bias, MR-RAPS (Robust Adjusted Profile Score) was conducted using overdispersion and Tukey's loss function (**Zhao et al., 2018**). To detect and correct for potential bias from invalid variants with pleiotropic effects, we performed the MR-PRESSO (Mendelian Randomization Pleiotropy RESidual Sum and Outlier) test with 10,000 simulations (**Verbanck et al., 2018**). The global test p-value evaluated whether there was any overall horizontal pleiotropy among all genetic variants. For outcomes with significant p-values ($p<0.05$), outlying genetic variants with predicted pleiotropic effects were removed and MR analysis repeated to correct for horizontal pleiotropy. The distortion test evaluated whether removal of the pleiotropic variants resulted in a significantly different causal estimate ($p<0.05$). Leave-one-out analysis was performed such that the IVW MR analysis was repeated after each genetic variant was excluded to identify effects on an outcome that are driven by a single outlying genetic variant. Furthermore, the set of genetic variants used in MR analysis were assessed for 'weak instrument bias', which can result in biased estimates if genetic variants don't explain enough variance in exposure (e.g., CFT) levels (**Pierce et al., 2011**). Lastly, as a sensitivity analysis, all MR and GRS analyses were repeated using genetic variants associated with total testosterone. Finally, for significant outcomes, we compared estimated effect sizes from this MR study with reported effect sizes from random controlled trials of testosterone therapy, where possible, in **Figure 3** (**Cui et al., 2014**; **Fernández-Balsells et al., 2010**; **Ng Tang Fui et al., 2016**; **Zhang et al., 2020**).

In consideration of 'weak instrument bias', the F-statistic was 66 for the genetic variants associated with CFT, which was considered a strong instrument based on the recommended threshold of greater than 10 (**Davies et al., 2018**). MR-PRESSO was performed using the *MR-PRESSO* package and all other MR analyses were implemented using the *TwoSampleMR* package (**Hemani et al., 2018**; **Verbanck et al., 2018**) (RRID:SCR_019010).

Genetic risk score analysis

A genetically-predicted value of CFT was determined for each individual by constructing weighted GRS in the unrelated White British subset of UK Biobank males (n = 157,252). Weighted GRS were calculated by multiplying the effect of each CFT-associated genetic variant by the number of effect-corresponding alleles and summing this value for each individual. The GRS was tested for association with outcomes using logistic or linear regression models for case-control or quantitative outcomes, respectively, and adjusted for the same covariates as the GWAS for CFT. Effect estimates were reported per 0.1 nmol/L increase in CFT levels based on approximate changes in response to testosterone treatment (**Bhasin et al., 2018b**). As sensitivity analyses, we repeated GRS analyses after excluding males that self-reported taking blood pressure (n = 38,676) or cholesterol medication (n = 35,737) at recruitment based on field ID 6177.

Genetic determinants and effects of total testosterone in males

As a set of sensitivity checks, we repeated all GWAS, MR, and GRS analyses using total testosterone. In the White British subset of the UK Biobank, there were 175,421 males with total testosterone measured with an average 11.9 nmol/L (**Figure 1—figure supplement 6**). In this population, a genome-wide association study was conducted for total testosterone as described herein for CFT. After removing genetic variants associated with natural-log-transformed SHBG and LD pruning for independent SNPs (r^2 <0.01), there were 52 independent genetic variants associated ($p<5\times10^{-8}$) with total testosterone in males from the UK Biobank (**Supplementary file 1** – Table 8).

All statistical analyses were performed under R version 3.6.0, unless otherwise specified (RRID: SCR_001905). A two-sided p-value less than 5×10^{-8} for GWAS, 2.27×10^{-3} (0.05/22 outcomes) for a priori MR analyses, and 1.14×10^{-4} (0.05/439 outcomes) for hypothesis-free GRS analyses was considered statistically significant.

Acknowledgements

The authors are thankful for all the participants that contributed to the UK Biobank study.

Additional information

Competing interests

Hertzel C Gerstein: HCG reports research grants from Eli Lilly, AstraZeneca, Merck, Novo Nordisk, and Sanofi; honoraria for speaking from AstraZeneca, Boehringer Ingelheim, Eli Lilly, Novo Nordisk, and Sanofi; and consulting fees from Abbott, AstraZeneca, Boehringer Ingelheim, Eli Lilly, Merck, Novo Nordisk, Janssen, Sanofi, Kowa, and Cirius. The other authors declare that no competing interests exist.

Funding

Funder	Grant reference number	Author
Canadian Institutes of Health Research	Frederick Banting and Charles Best Canada Graduate Scholarships Doctoral Award	Michael Chong
Canadian Institutes of Health Research	Post-Doctoral Fellowship	Robert W Morton
McMaster University	E.J. Moran Campbell Internal Career Research Award	Marie Pigeyre
McMaster University	McMaster-Sanofi Population Health Institute Chair in Diabetes Research and Care	Hertzel C Gerstein
Cisco Systems	Professorship in Integrated Health Biosystems	Guillaume Paré
Canada Research Chairs	Canada Research Chair in Genetic and Molecular Epidemiology	Guillaume Paré

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Author contributions

Pedrum Mohammadi-Shemirani, Data curation, Software, Formal analysis, Investigation, Visualization, Writing - original draft; Michael Chong, Data curation, Software, Formal analysis, Writing - review and editing; Marie Pigeyre, Hertzel C Gerstein, Conceptualization, Writing - review and editing, Analysis and interpretation of data; Robert W Morton, Writing - review and editing, Analysis and interpretation of data; Guillaume Paré, Conceptualization, Data curation, Supervision, Funding acquisition, Methodology, Project administration, Writing - review and editing

Author ORCIDs

Pedrum Mohammadi-Shemirani https://orcid.org/0000-0001-6740-7858
Robert W Morton http://orcid.org/0000-0003-0099-4167
Guillaume Paré https://orcid.org/0000-0002-6795-4760

Ethics

Human subjects: UK Biobank received ethical approval from the North West Multi-Centre Research Ethics Committee (REC reference: 11/NW/0382). This research was conducted using the UK Biobank under Application Number 15255.

Decision letter and Author response

Decision letter https://doi.org/10.7554/eLife.58914.sa1
Author response https://doi.org/10.7554/eLife.58914.sa2

Additional files

Supplementary files

- Supplementary file 1. Supplementary Tables. Table 1. Characteristics at recruitment for study population of males from UK Biobank cohort study Table 2. Independent genetic variants associated with calculated free testosterone (CFT) at genome-wide significance ($p<5\times10\text{-}8$) and not associated with sex hormone-binding globulin in males Table 3. Results of Mendelian randomization analysis using Egger regression for 22 a priori outcomes relevant to testosterone treatment Table 4. Results of Mendelian randomization analysis using MR-RAPS for effect of CFT on 22 a priori outcomes relevant to testosterone treatment Table 5. Results of Mendelian randomization analysis using MR-PRESSO for effect of CFT on 22 a priori outcomes relevant to testosterone treatment Table 6. Associations of genetically-predicted CFT for 439 health outcomes across the human phenome excluding individuals on antihypertensive medication Table 7. Associations of genetically-predicted CFT for 439 health outcomes across the human phenome excluding individuals on cholesterol-lowering medication Table 8. Independent genetic variants associated with total testosterone at genome-wide significance ($p<5\times10\text{-}8$) and not associated with sex hormone-binding globulin in 175,421 males from UK Biobank Table 9. All Mendelian randomization analyses of total testosterone on 22 a priori outcomes Table 10. Associations of genetically-predicted total testosterone for 439 health outcomes across the human phenome. Table 11. Definitions for 22 health outcomes with suspected relevance with testosterone treatment Table 12. Definitions for 439 phenome-wide health outcomes ***Figure 1***. Screenshot of options shown to male UK Biobank participants for selection of hair/baldness pattern.
- Transparent reporting form

Data availability

Individual-level data cannot be provided, but it is available to all researchers by application to the UK Biobank. Summary-level GWAS data will be returned to the UK Biobank Access Team for use by other researchers. All MR results and genome-wide significant SNPs have been provided in Supplementary Tables 4 to 12 in Supplementary file 1.

References

Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, Kang HM, Marth GT, McVean GA, 1000 Genomes Project Consortium. 2012. An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**:56–65. DOI: https://doi.org/10.1038/nature11632, PMID: 23128226

Adams M, Hill WD, Howard DM, Davis KAS, Deary IJ, Hotopf M, McIntosh AM. 2018. Factors associated with sharing email information and mental health survey participation in two large population cohorts. *bioRxiv*. DOI: https://doi.org/10.1101/471433

Adil A, Godwin M. 2017. The effectiveness of treatments for androgenetic alopecia: a systematic review and meta-analysis. *Journal of the American Academy of Dermatology* **77**:136–141. DOI: https://doi.org/10.1016/j.jaad.2017.02.054

American Cancer Society. 2019. *Cancer Facts & Figures 2019*.

Andriole GL, Bostwick DG, Brawley OW, Gomella LG, Marberger M, Montorsi F, Pettaway CA, Tammela TL, Teloken C, Tindall DJ, Somerville MC, Wilson TH, Fowler IL, Rittmaster RS REDUCE Study Group. 2010. Effect of dutasteride on the risk of prostate Cancer. *New England Journal of Medicine* **362**:1192–1202. DOI: https://doi.org/10.1056/NEJMoa0908127, PMID: 20357281

Aukrust P, Ueland T, Gullestad L, Yndestad A. 2009. Testosterone: a novel therapeutic approach in chronic heart failure? *Journal of the American College of Cardiology* **54**:928–929. DOI: https://doi.org/10.1016/j.jacc.2009.05.039, PMID: 19712803

Bachmann KN, Huang S, Lee H, Dichtel LE, Gupta DK, Burnett JC, Miller KK, Wang TJ, Finkelstein JS. 2019. Effect of testosterone on natriuretic Peptide Levels. *Journal of the American College of Cardiology* **73**:1288–1296. DOI: https://doi.org/10.1016/j.jacc.2018.12.062

Baillargeon J, Kuo Y-F, Westra JR, Urban RJ, Goodwin JS. 2018. Testosterone prescribing in the united states, 2002-2016. *JAMA* **320**:200–202. DOI: https://doi.org/10.1001/jama.2018.7999

Bhasin S, Brito JP, Cunningham GR, Hayes FJ, Hodis HN, Matsumoto AM, Snyder PJ, Swerdloff RS, Wu FC, Yialamas MA. 2018a. Testosterone therapy in men with hypogonadism: an endocrine society* clinical practice guideline. *The Journal of Clinical Endocrinology & Metabolism* **103**:1715–1744. DOI: https://doi.org/10.1210/jc.2018-00229

Bhasin S, Ellenberg SS, Storer TW, Basaria S, Pahor M, Stephens-Shields AJ, Cauley JA, Ensrud KE, Farrar JT, Cella D, Matsumoto AM, Cunningham GR, Swerdloff RS, Wang C, Lewis CE, Molitch ME, Barrett-Connor E, Crandall JP, Hou X, Preston P, et al. 2018b. Effect of testosterone replacement on measures of mobility in older men with mobility limitation and low testosterone concentrations: secondary analyses of the testosterone trials. *The Lancet Diabetes & Endocrinology* **6**:879–890. DOI: https://doi.org/10.1016/S2213-8587(18)30171-2

Bowden J, Davey Smith G, Burgess S. 2015. Mendelian randomization with invalid instruments: effect estimation and Bias detection through egger regression. *International Journal of Epidemiology* **44**:512–525. DOI: https://doi.org/10.1093/ije/dyv080

Brion M-JA, Shakhbazov K, Visscher PM. 2013. Calculating statistical power in mendelian randomization studies. *International Journal of Epidemiology* **42**:1497–1501. DOI: https://doi.org/10.1093/ije/dyt179

Bulik-Sullivan BK, Loh P-R, Finucane HK, Ripke S, Yang J, Patterson N, Daly MJ, Price AL, Neale BM. 2015. LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature Genetics* **47**:291–295. DOI: https://doi.org/10.1038/ng.3211

Burgess S, Dudbridge F, Thompson SG. 2016. Combining information on multiple instrumental variables in mendelian randomization: comparison of allele score and summarized data methods. *Statistics in Medicine* **35**:1880–1906. DOI: https://doi.org/10.1002/sim.6835

Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, Motyer A, Vukcevic D, Delaneau O, O'Connell J, Cortes A, Welsh S, McVean G, Leslie S, Donnelly P, Marchini J. 2017. Genome-wide genetic data on ~500,000 UK biobank participants. *bioRxiv*. DOI: https://doi.org/10.1101/166298

Carrero JJ, Qureshi AR, Parini P, Arver S, Lindholm B, Bárány P, Heimbürger O, Stenvinkel P. 2009. Low serum testosterone increases mortality risk among male Dialysis patients. *Journal of the American Society of Nephrology* **20**:613–620. DOI: https://doi.org/10.1681/ASN.2008060664

Chignalia AZ, Schuldt EZ, Camargo LL, Montezano AC, Callera GE, Laurindo FR, Lopes LR, Avellar MCW, Carvalho MHC, Fortes ZB, Touyz RM, Tostes RC. 2012. Testosterone induces vascular smooth muscle cell migration by NADPH oxidase and c-Src–Dependent Pathways. *Hypertension* **59**:1263–1271. DOI: https://doi.org/10.1161/HYPERTENSIONAHA.111.180620

Cui Y, Zong H, Yan H, Zhang Y. 2014. The effect of testosterone replacement therapy on prostate Cancer: a systematic review and meta-analysis. *Prostate Cancer and Prostatic Diseases* **17**:132–143. DOI: https://doi.org/10.1038/pcan.2013.60

Davies NM, Holmes MV, Davey Smith G. 2018. Reading mendelian randomisation studies: a guide, glossary, and checklist for clinicians. *BMJ* **362**:k601. DOI: https://doi.org/10.1136/bmj.k601

Denny JC, Bastarache L, Ritchie MD, Carroll RJ, Zink R, Mosley JD, Field JR, Pulley JM, Ramirez AH, Bowton E, Basford MA, Carrell DS, Peissig PL, Kho AN, Pacheco JA, Rasmussen LV, Crosslin DR, Crane PK, Pathak J, Bielinski SJ, et al. 2013. Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data. *Nature Biotechnology* **31**:1102–1111. DOI: https://doi.org/10.1038/nbt.2749

Doherty A, Jackson D, Hammerla N, Plötz T, Olivier P, Granat MH, White T, van Hees VT, Trenell MI, Owen CG, Preece SJ, Gillions R, Sheard S, Peakman T, Brage S, Wareham NJ. 2017. Large scale population assessment of physical activity using wrist worn accelerometers: the UK biobank study. *PLOS ONE* **12**:e0169649. DOI: https://doi.org/10.1371/journal.pone.0169649

Fernández-Balsells MM, Murad MH, Lane M, Lampropulos JF, Albuquerque F, Mullan RJ, Agrwal N, Elamin MB, Gallegos-Orozco JF, Wang AT, Erwin PJ, Bhasin S, Montori VM. 2010. Adverse effects of testosterone therapy in adult men: a systematic review and Meta-Analysis. *The Journal of Clinical Endocrinology & Metabolism* **95**:2560–2575. DOI: https://doi.org/10.1210/jc.2009-2575

Filler G, Ramsaroop A, Stein R, Grant C, Marants R, So A, McIntyre C. 2016. Is testosterone detrimental to renal function? *Kidney International Reports* **1**:306–310. DOI: https://doi.org/10.1016/j.ekir.2016.07.004

Gagliano-Jucá T, Basaria S. 2019. Testosterone replacement therapy and cardiovascular risk. *Nature Reviews Cardiology* **16**:555–574. DOI: https://doi.org/10.1038/s41569-019-0211-4

Handelsman DJ. 2013. Global trends in testosterone prescribing, 2000–2011: expanding the spectrum of prescription drug misuse. *Medical Journal of Australia* **199**:548–551. DOI: https://doi.org/10.5694/mja13.10111

Handelsman DJ. 2017. Free testosterone: pumping up the tires or ending the free ride? *Endocrine Reviews* **38**:297–301. DOI: https://doi.org/10.1210/er.2017-00171

Hanna F. 2005. Factors influencing longitudinal change in knee cartilage volume measured from magnetic resonance imaging in healthy men. *Annals of the Rheumatic Diseases* **64**:1038–1042. DOI: https://doi.org/10.1136/ard.2004.029355

Hartgens F, Kuipers H. 2004. Effects of Androgenic-Anabolic steroids in Athletes. *Sports Medicine* **34**:513–554. DOI: https://doi.org/10.2165/00007256-200434080-00003

Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, Laurin C, Burgess S, Bowden J, Langdon R, Tan VY, Yarmolinsky J, Shihab HA, Timpson NJ, Evans DM, Relton C, Martin RM, Davey Smith G, Gaunt TR, Haycock PC. 2018. The MR-Base platform supports systematic causal inference across the human phenome. *eLife* **7**:e34408. DOI: https://doi.org/10.7554/eLife.34408

Hl M, Blanchet TJ, Peluso D, Hopkins B, Morris EA, Glasson SS. 2007. Osteoarthritis severity is sex dependent in a surgical mouse model. *Osteoarthr Cartil* **15**:695–700. DOI: https://doi.org/10.1016/j.joca.2006.11.005

Holmes MV, Asselbergs FW, Palmer TM, Drenos F, Lanktree MB, Nelson CP, Dale CE, Padmanabhan S, Finan C, Swerdlow DI, Tragante V, van Iperen EPA, Sivapalaratnam S, Shah S, Elbers CC, Shah T, Engmann J, Giambartolomei C, White J, Zabaneh D, et al. 2015. Mendelian randomization of blood lipids for coronary heart disease. *European Heart Journal* **36**:539–550. DOI: https://doi.org/10.1093/eurheartj/eht571

Howard DM, Adams MJ, Shirali M, Clarke T-K, Marioni RE, Davies G, Coleman JRI, Alloza C, Shen X, Barbu MC, Wigmore EM, Gibson J, Hagenaars SP, Lewis CM, Ward J, Smith DJ, Sullivan PF, Haley CS, Breen G, Deary IJ, et al. 2018. Genome-wide association study of depression phenotypes in UK biobank identifies variants in excitatory synaptic pathways. *Nature Communications* **9**:1470. DOI: https://doi.org/10.1038/s41467-018-03819-3

Jasuja GK, Bhasin S, Rose AJ. 2017. Patterns of testosterone prescription overuse. *Current Opinion in Endocrinology & Diabetes and Obesity* **24**:240–245. DOI: https://doi.org/10.1097/MED.0000000000000336

Lakshman KM, Bhasin S, Araujo AB. 2010. Sex Hormone-Binding globulin as an independent predictor of incident type 2 diabetes mellitus in men. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences* **65A**:503–509. DOI: https://doi.org/10.1093/gerona/glq002

Layton JB, Li D, Meier CR, Sharpless JL, Stürmer T, Jick SS, Brookhart MA. 2014. Testosterone lab testing and initiation in the united kingdom and the united states, 2000 to 2011. *The Journal of Clinical Endocrinology & Metabolism* **99**:835–842. DOI: https://doi.org/10.1210/jc.2013-3570

Loh P-R, Tucker G, Bulik-Sullivan BK, Vilhjálmsson BJ, Finucane HK, Salem RM, Chasman DI, Ridker PM, Neale BM, Berger B, Patterson N, Price AL. 2015. Efficient bayesian mixed-model analysis increases association power in large cohorts. *Nature Genetics* **47**:284–290. DOI: https://doi.org/10.1038/ng.3190

McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood AR, Teumer A, Kang HM, Fuchsberger C, Danecek P, Sharp K, Luo Y, Sidore C, Kwong A, Timpson N, Koskinen S, Vrieze S, Scott LJ, Zhang H, Mahajan A, Veldink J, et al. 2016. A reference panel of 64,976 haplotypes for genotype imputation. *Nature Genetics* **48**:1279–1283. DOI: https://doi.org/10.1038/ng.3643, PMID: 27548312

Mohler ER, Ellenberg SS, Lewis CE, Wenger NK, Budoff MJ, Lewis MR, Barrett-Connor E, Swerdloff RS, Stephens-Shields A, Bhasin S, Cauley JA, Crandall JP, Cunningham GR, Ensrud KE, Gill TM, Matsumoto AM, Molitch ME, Pahor M, Preston PE, Hou X, et al. 2018. The effect of testosterone on cardiovascular biomarkers in the testosterone trials. *The Journal of Clinical Endocrinology & Metabolism* **103**:681–688. DOI: https://doi.org/10.1210/jc.2017-02243, PMID: 29253154

Ng Tang Fui M, Prendergast LA, Dupuis P, Raval M, Strauss BJ, Zajac JD, Grossmann M. 2016. Effects of testosterone treatment on body fat and lean mass in obese men on a hypocaloric diet: a randomised controlled trial. *BMC Medicine* **14**:153. DOI: https://doi.org/10.1186/s12916-016-0700-9, PMID: 27716209

Ohlsson C, Wallaschofski H, Lunetta KL, Stolk L, Perry JR, Koster A, Petersen AK, Eriksson J, Lehtimäki T, Huhtaniemi IT, Hammond GL, Maggio M, Coviello AD, Ferrucci L, Heier M, Hofman A, Holliday KL, Jansson JO, Kähönen M, Karasik D, et al. 2011. Genetic determinants of serum testosterone concentrations in men. *PLOS Genetics* **7**:e1002313. DOI: https://doi.org/10.1371/journal.pgen.1002313, PMID: 21998597

Peila R, Arthur RS, Rohan TE. 2020. Association of sex hormones with risk of cancers of the pancreas, kidney, and brain in the UK biobank cohort study. *Cancer Epidemiology Biomarkers & Prevention* **29**::1832–1836. DOI: https://doi.org/10.1158/1055-9965.EPI-20-0246

Petermann-Rocha F, Gray SR, Pell JP, Celis-Morales C, Ho FK. 2020. Biomarkers profile of people with Sarcopenia: a Cross-sectional analysis from UK biobank. *Journal of the American Medical Directors Association* **5**:e005. DOI: https://doi.org/10.1016/j.jamda.2020.05.005

Pierce BL, Ahsan H, Vanderweele TJ. 2011. Power and instrument strength requirements for mendelian randomization studies using multiple genetic variants. *International Journal of Epidemiology* **40**:740–752. DOI: https://doi.org/10.1093/ije/dyq151, PMID: 20813862

Reckelhoff JF, Zhang H, Granger JP. 1998. Testosterone exacerbates hypertension and reduces pressure-natriuresis in male spontaneously hypertensive rats. *Hypertension* **31**:435–439. DOI: https://doi.org/10.1161/01.HYP.31.1.435, PMID: 9453341

Rosner W, Hryb DJ, Kahn SM, Nakhla AM, Romas NA. 2010. Interactions of sex hormone-binding globulin with target cells. *Molecular and Cellular Endocrinology* **316**:79–85. DOI: https://doi.org/10.1016/j.mce.2009.08.009, PMID: 19698759

Ross S, Gerstein HC, Eikelboom J, Anand SS, Yusuf S, Paré G. 2015. Mendelian randomization analysis supports the causal role of dysglycaemia and diabetes in the risk of coronary artery disease. *European Heart Journal* **36**:1454–1462. DOI: https://doi.org/10.1093/eurheartj/ehv083, PMID: 25825043

Ruth KS, Day FR, Tyrrell J, Thompson DJ, Wood AR, Mahajan A, Beaumont RN, Wittemans L, Martin S, Busch AS, Erzurumluoglu AM, Hollis B, O'Mara TA, McCarthy MI, Langenberg C, Easton DF, Wareham NJ, Burgess S, Murray A, Ong KK, et al. 2020. Using human genetics to understand the disease impacts of testosterone in men and women. *Nature Medicine* **26**:252–258. DOI: https://doi.org/10.1038/s41591-020-0751-5, PMID: 32042192

Schooling CM, Luo S, Au Yeung SL, Thompson DJ, Karthikeyan S, Bolton TR, Mason AM, Ingelsson E, Burgess S. 2018. Genetic predictors of testosterone and their associations with cardiovascular disease and risk factors: a mendelian randomization investigation. *International Journal of Cardiology* **267**:171–176. DOI: https://doi.org/10.1016/j.ijcard.2018.05.051, PMID: 29804699

Schutte JE, Longhurst JC, Gaffney FA, Bastian BC, Blomqvist CG. 1981. Total plasma creatinine: an accurate measure of total striated muscle mass. *Journal of Applied Physiology* **51**:762–766. DOI: https://doi.org/10.1152/jappl.1981.51.3.762, PMID: 7327978

Snyder PJ, Bhasin S, Cunningham GR, Matsumoto AM, Stephens-Shields AJ, Cauley JA, Gill TM, Barrett-Connor E, Swerdloff RS, Wang C, Ensrud KE, Lewis CE, Farrar JT, Cella D, Rosen RC, Pahor M, Crandall JP, Molitch ME, Resnick SM, Budoff M, et al. 2018. Lessons from the testosterone trials. *Endocrine Reviews* **39**:369–386. DOI: https://doi.org/10.1210/er.2017-00234, PMID: 29522088

Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, Downey P, Elliott P, Green J, Landray M, Liu B, Matthews P, Ong G, Pell J, Silman A, Young A, Sprosen T, Peakman T, Collins R. 2015. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLOS Medicine* **12**:e1001779. DOI: https://doi.org/10.1371/journal.pmed.1001779, PMID: 25826379

Traustadóttir T, Harman SM, Tsitouras P, Pencina KM, Li Z, Travison TG, Eder R, Miciek R, McKinnon J, Woodbury E, Basaria S, Bhasin S, Storer TW. 2018. Long-Term testosterone supplementation in older men attenuates Age-Related decline in aerobic capacity. *The Journal of Clinical Endocrinology & Metabolism* **103**:2861–2869. DOI: https://doi.org/10.1210/jc.2017-01902, PMID: 29846604

Trigunaite A, Dimo J, Jørgensen TN. 2015. Suppressive effects of androgens on the immune system. *Cellular Immunology* **294**:87–94. DOI: https://doi.org/10.1016/j.cellimm.2015.02.004, PMID: 25708485

Velho I, Fighera TM, Ziegelmann PK, Spritzer PM. 2017. Effects of testosterone therapy on BMI, blood pressure, and laboratory profile of transgender men: a systematic review. *Andrology* **5**:881–888. DOI: https://doi.org/10.1111/andr.12382, PMID: 28709177

Verbanck M, Chen CY, Neale B, Do R. 2018. Detection of widespread horizontal pleiotropy in causal relationships inferred from mendelian randomization between complex traits and diseases. *Nature Genetics* **50**:693–698. DOI: https://doi.org/10.1038/s41588-018-0099-7, PMID: 29686387

Vermeulen A, Verdonck L, Kaufman JM. 1999. A critical evaluation of simple methods for the estimation of free testosterone in serum. *The Journal of Clinical Endocrinology & Metabolism* **84**:3666–3672. DOI: https://doi.org/10.1210/jcem.84.10.6079, PMID: 10523012

Vikan T, Schirmer H, Njølstad I, Svartberg J. 2010. Low testosterone and sex hormone-binding globulin levels and high estradiol levels are independent predictors of type 2 diabetes in men. *European Journal of Endocrinology* **162**:747–754. DOI: https://doi.org/10.1530/EJE-09-0943, PMID: 20061333

Wu P, Gifford A, Meng X, Li X, Campbell H, Varley T, Zhao J, Carroll R, Bastarache L, Denny JC, Theodoratou E, Wei W-Q. 2019. Mapping ICD-10 and ICD-10-CM codes to phecodes: workflow development and initial evaluation. *JMIR Medical Informatics* **7**:e14325. DOI: https://doi.org/10.2196/14325

Zhang Z, Kang D, Li H. 2020. The effects of testosterone on bone health in males with testosterone deficiency: a systematic review and meta-analysis. *BMC Endocrine Disorders* **20**:1–12. DOI: https://doi.org/10.1186/s12902-020-0509-6

Zhao J, Jiang C, Lam TH, Liu B, Cheng KK, Xu L, Au Yeung SL, Zhang W, Leung GM, Schooling CM. 2014. Genetically predicted testosterone and cardiovascular risk factors in men: a mendelian randomization analysis in the guangzhou biobank cohort study. *International Journal of Epidemiology* **43**:140–148. DOI: https://doi.org/10.1093/ije/dyt239, PMID: 24302542

Zhao Q, Wang J, Hemani G, Bowden J, Small DS. 2018. Statistical inference in two-sample summary-data mendelian randomization using robust adjusted profile score. *arXiv*. https://arxiv.org/abs/1801.09652.

CHAPTER 5:

A Mendelian randomization-based approach to identify early and sensitive diagnostic biomarkers of disease

Published in *Clinical Chemistry*. 65(3):427-436 (2019)

Clinical Chemistry 65:3
427–436 (2019)

A Mendelian Randomization-Based Approach to Identify Early and Sensitive Diagnostic Biomarkers of Disease

Pedrum Mohammadi-Shemirani,[1,2,3] Jennifer Sjaarda,[1,2,3] Hertzel C. Gerstein,[1,4] Darin J. Treleaven,[5] Michael Walsh,[5] Johannes F. Mann,[6] Matthew J. McQueen,[7] Sibylle Hess,[8] and Guillaume Paré[1,2,7,9*]

BACKGROUND: Identifying markers of chronic kidney disease (CKD) that occur early in the disease process and are specific to loss of kidney function rather than other underlying causes of disease may allow earlier, more accurate identification of patients who will develop CKD. We therefore sought to identify diagnostic blood markers of early CKD that are caused by loss of kidney function by using an innovative "reverse Mendelian randomization" (MR) approach.

METHODS: We applied this technique to genetic and biomarker data from 4147 participants in the Outcome Reduction with Initial Glargine Intervention (ORIGIN) trial, all with known type 2 diabetes, impaired fasting glucose, or impaired glucose tolerance. Two-sample MR was conducted using variants associated with creatinine-based eGFR ($eGFR_{crea}$) from the CKDGen Consortium (n = 133 814) to estimate the effect of genetically decreased $eGFR_{crea}$ on 238 serum biomarkers.

RESULTS: With reverse MR, trefoil factor 3 (TFF3) was identified as a protein that is increased owing to decreased $eGFR_{crea}$ ($\beta = 1.86$ SD per SD decrease $eGFR_{crea}$; 95% CI, 0.95–2.76; $P = 8.0 \times 10^{-5}$). Reverse MR findings were consistent with epidemiological associations for incident CKD in ORIGIN (OR = 1.28 per SD increase in TFF3; 95% CI, 1.18–1.38; $P = 4.58 \times 10^{-10}$). Addition of TFF3 significantly improved discrimination for incident CKD relative to $eGFR_{crea}$ alone (net reclassification improvement = 0.211; $P = 9.56 \times 10^{-12}$) and in models including additional risk factors.

CONCLUSIONS: Our results suggest TFF3 is a valuable diagnostic marker for early CKD in dysglycemic populations and acts as a proof of concept for the application of this novel MR technique to identify diagnostic biomarkers for other chronic diseases.

CLINICALTRIALS.GOV IDENTIFIER: NCT00069784

Diabetic renal disease is characterized by a progressive increase in albumin excretion and gradual decline in glomerular filtration rate (GFR)[10] *(1)*. Chronic kidney disease (CKD) increases the likelihood of early mortality, cardiovascular disease, end-stage renal disease, and a host of further serious chronic conditions *(2–4)*. Current guidelines recommend the use of albuminuria and estimated GFR based on the serum creatinine ($eGFR_{crea}$) to gauge the severity of diabetic renal disease. However, these tests have significant limitations *(5)*. Albuminuria may be transient in the early stages of CKD *(6)*, may be a late manifestation of kidney disease *(7)*, and may be discordant from GFR *(4)*. $eGFR_{crea}$ is insensitive to early kidney damage *(8)* and is affected by such factors as age, sex, ethnicity, muscle mass, diet, and intraglomerular hemodynamics, which may confound the relationship between serum creatinine and GFR in diabetics *(9, 10)*. The limitations in these biomarkers of renal disease subsequently impair clinicians' abilities to accurately identify individuals at early risk of CKD when treatments may mitigate their future risks as recommended in guidelines *(5)*. It is ultimately hoped that early, focused interven-

[1] Population Health Research Institute, McMaster University, Hamilton Health Sciences, Hamilton, Ontario, Canada; [2] Thrombosis and Atherosclerosis Research Institute, McMaster University, Hamilton Health Sciences, Hamilton, Ontario, Canada; [3] Department of Medical Sciences, McMaster University, Hamilton Ontario, Canada; [4] Department of Medicine, McMaster University, Hamilton Health Sciences, Hamilton, Ontario, Canada; [5] Division of Nephrology, Department of Medicine, McMaster University, Hamilton, Ontario, Canada; [6] KfH Kidney Center, München, Germany; [7] Department of Pathology and Molecular Medicine, McMaster University, Michael G. DeGroote School of Medicine, Hamilton, Ontario, Canada; [8] Sanofi Aventis Deutschland GmbH, Research and Development Division, Translational Medicine and Early Development, Biomarkers and Clinical Bioanalyses, Frankfurt, Germany; [9] Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, Ontario, Canada.

* Address correspondence to this author at: McMaster University, Population Health Research Institute, David Braley Cardiac, Vascular and Stroke Institute, 237 Barton St. East – C4 126. Fax +905-297-3789; e-mail pareg@mcmaster.ca.

Received June 19, 2018; accepted September 5, 2018.
Previously published online at DOI: 10.1373/clinchem.2018.291104

[10] Nonstandard abbreviations: GFR, glomerular filtration rate; CKD, chronic kidney disease; $eGFR_{crea}$, estimated GFR based on serum creatinine; MR, Mendelian randomization; ORIGIN, Outcome Reduction with Initial Glargine Intervention; ACR, albumin-to-creatinine ratio; SNP, single nucleotide polymorphism; $eGFR_{cys}$, estimated GFR based on cystatin C; NRI, net reclassification improvement index; AUC, area under the curve.

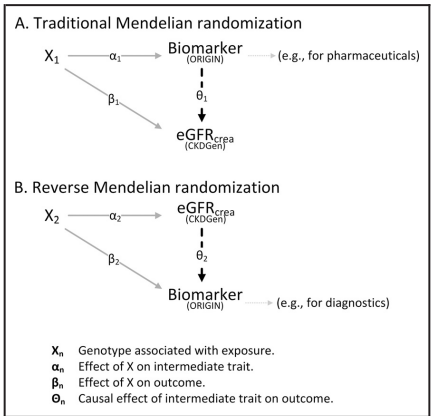


Fig. 1. Schematic representation of (A) traditional MR compared to (B) reverse MR studies.

In both (A) traditional MR and (B) reverse MR analyses, we are identifying biomarkers related to disease by using genotypes. In traditional MR analysis (A), we are interested in identifying biomarkers that cause disease development (θ_1), which are expected to be promising pharmaceutical targets. Therefore, we would infer concentrations of a biomarker on the basis of genotypes (X_1) associated with the biomarker from ORIGIN (α_1) and find the genetic associations of these genotypes with $eGFR_{crea}$ from the CKDGen consortium (β_1). However, in reverse MR (B), we are interested in identifying biomarkers that are themselves affected by a disease (θ_2), which are expected to be promising diagnostic markers. Therefore, we would identify genotypes (X_2) that are associated with $eGFR_{crea}$, as determined in the CKDGen consortium (α_2), and find their associations with a given biomarker from ORIGIN (β_2).

tion will reduce the contribution of CKD to the global burden disease *(11)*.

An early effect of CKD is the accumulation of proteins in the blood due to changes in the glomerular barrier of the kidneys. Thus, the serum of affected individuals may harbor markers of the antecedent causes of CKD, confounders commonly linked to either dysglycemia or CKD, and CKD itself *(5, 12–14)* (see Fig. 1 in the Data Supplement that accompanies the online version of this article at http://www.clinchem.org/content/vol65/issue3). We propose that identification of these biomarkers related to CKD itself can be achieved by using a variation of a powerful genetic technique called Mendelian randomization (MR) (Fig. 1A) *(15)*. Genetic variants have the unique property of being randomly and independently inherited from one another *(15)*. Thus, nature effectively randomizes individuals to make high or low levels of particular proteins. If people born with a genetic propensity to make higher (or lower) levels of a particular protein also develop some chronic disease more commonly than people born with a propensity to make lower (or higher) levels of this protein, an MR analysis may support the conclusion that this protein has a causal effect on the disease *(16, 17)* (Fig. 1A). Using similar reasoning, we can identify serum proteins that change concentration as a result of a chronic disease rather than simply being associated with the cause of the chronic disease. In the case of diabetic renal disease, this means we would identify serum proteins that change because of kidney damage rather than because of dysglycemia. We label this novel approach "reverse Mendelian randomization." In this study, we demonstrate its utility as a tool in the context of identifying sensitive and early diagnostic biomarkers of kidney damage specifically (Fig. 1B). However, it has broader applications in identifying these biomarkers for any chronic disease with a set of robustly associated genetic variants.

We therefore sought to identify new biomarkers of early renal disease in patients with dysglycemia. We identified genetic variants linked with $eGFR_{crea}$ in the publicly available CKDGen consortium study *(18)*, found which of the 238 biomarkers measured in the Outcome Reduction with Initial Glargine Intervention (ORIGIN) (NCT00069784) biomarker study were statistically associated with these variants *(19)*, and then tested the identified biomarkers for their association with incident CKD in models adjusting for other risk factors of a reduced eGFR (Fig. 2).

Methods

STUDY POPULATION–ORIGIN

The ORIGIN trial has been described previously *(20)*. Briefly, from 2003 to 2005, 12 537 participants with cardiovascular risk factors and impaired fasting glucose, impaired glucose tolerance, or type 2 diabetes were randomized in a 2 × 2 factorial design to either subcutaneous insulin glargine vs usual care and ω-3 fatty acid supplementation vs placebo and followed for 5–7 years (median 6.2 years). At randomization, all participants were insulin naive.

The ORIGIN biomarker substudy consisted of 8401 participants of European, Native Latin, and African ethnicity. As previously described *(19, 21)*, 238 biomarkers related to metabolism, inflammation, and cardiovascular disease were selected to be measured. Of those, 237 serum biomarkers were measured with a customized human discovery multianalyte platform on the Luminex 100/200 instrument (Myriad RBM), and serum troponin I was assayed with the Architect Stat high-sensitivity assay on the Abbott Architect System *(21)*. Because the primary aim of this study was to identify

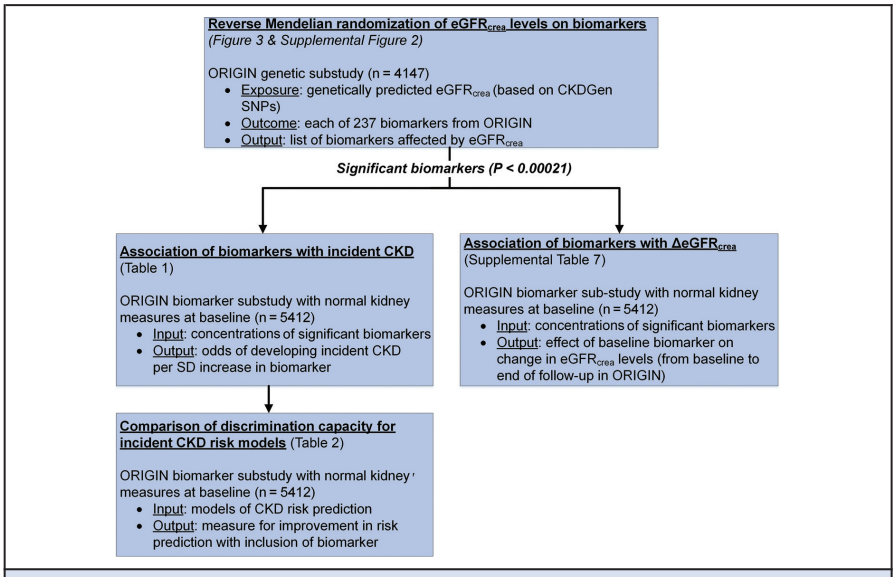


Fig. 2. Simplified diagram of the overall study design.

First, reverse MR analysis was used to identify biomarkers in the ORIGIN trial biomarker panel that were affected by genetically predicted $eGFR_{crea}$ (on the basis of SNPs from CKDGen consortium) shown in Fig. 3 and Fig. 2 in the online Data Supplement. Next, significant biomarkers from the reverse MR were subject to follow-up analyses in individuals with normal urinary ACR and $eGFR_{crea}$ at baseline in the ORIGIN trial. These involved testing for epidemiological associations of baseline biomarkers with incident CKD and change in $eGFR_{crea}$ from baseline to the end of follow-up in the ORIGIN trial, shown in Table 1 and Table 7 in the online Data Supplement, respectively. Finally, as a follow-up to the analysis in Table 1, improvement in discrimination ability for incident CKD was assessed after addition of biomarkers to models with established risk factors in Table 2.

early blood biomarkers of incident CKD, our primary patient population was a subset of the ORIGIN biomarker substudy participants with $eGFR_{crea}$ and urinary albumin-to-creatinine ratio (ACR) at baseline within clinically normal ranges, defined as >60 mL/min/1.73m^2 and <30 mg/g (<3.4 mg/mmol), respectively.

In total, 5078 biomarker substudy participants consented to genetic analysis. Genotyping was performed on Illumina's HumanCore Exome chip and filtered with standard quality-control measures in PLINK *(22)* and GCTA *(23)*. After quality control, 4147 participants and 284 024 single nucleotide polymorphisms (SNPs) from 2 ethnic groups, Native Latin and Europeans, remained. Imputation was done after quality control with IMPUTE2 *(24)* and the 1000 Genomes Project *(25)* as a reference panel. SNPs with low imputation certainty, defined as INFO <0.7 by IMPUTE2, were then removed (detailed information in the Methods in the online Data Supplement).

STUDY OUTCOMES–ORIGIN

The primary renal outcome was incident CKD in the ORIGIN biomarker substudy with normal $eGFR_{crea}$ and urinary ACR at baseline. It was defined as ≥1 of the following: incident worsening of albuminuria category, doubling of serum creatinine, end-stage renal disease, or development of $eGFR_{crea}$ <60 mL/min per 1.73 m^2 by the end of follow-up. Detailed definitions of renal outcomes are described in the Methods in the online Data Supplement. Renal function during follow-up was assessed at year 2 and at the end of study on the basis of serum creatinine and urinary ACR. Both renal markers were measured centrally *(20)*. The secondary renal outcome was change in $eGFR_{crea}$ from baseline to end of study time point.

CKDGen CONSORTIUM DATA

Genome-wide significant variants for $eGFR_{crea}$ were obtained from the CKDGen Consortium *(18)*. In brief, 49

genome-wide association studies totaling 133814 individuals of European descent were meta-analyzed to identify 53 independent loci at a genome-wide significant threshold ($P < 5 \times 10^{-8}$) associated with $eGFR_{crea}$, which we selected for further analysis. ORIGIN did not contribute samples to the consortium.

ETHICS STATEMENT

The ORIGIN clinical trial and CKDGen consortium both received approvals from the local ethics committees at all participating sites, and all participants provided written informed consent.

"REVERSE" MR ANALYSIS OF eGFRcrea ON 238 BIOMARKERS

Of the 53 $eGFR_{crea}$ SNPs, 50 were successfully imputed in the ORIGIN genetic substudy (see Table 1 in the online Data Supplement). First, each of the 50 SNPs was tested for association with each of the 238 biomarkers in ORIGIN by linear regression. Models were constructed with biomarker concentration as the dependent variable under an additive genetic model in each ethnic group separately (European and Latin), adjusting for age, sex, and the first 5 genetic principal components of each ethnicity by SNPtest v2.5 (26). Estimates from the 2 ethnic-specific models were then meta-analyzed together with a fixed-effects model in META v1.7 *(26)*.

Next, a 2-sample reverse MR analysis *(15)* was performed for each of the 238 serum biomarkers b use of the 50 SNPs associated with $eGFR_{crea}$ in the CKDGen Consortium. SNP– biomarker effect estimates (calculated in ORIGIN) for each of the 238 biomarkers, and SNP–$eGFR_{crea}$ effect estimates (from the CKDGen consortium) were used as input variables for the reverse MR analysis (Fig. 1B). The inverse-variance weighted method was used to obtain MR associations by regressing the genetic effect estimates for each biomarker against genetic effect estimates for $eGFR_{crea}$ *(27)*. Bootstrapping *(28)* was done to calculate significance of the regression model for each biomarker. Two-tailed *P* values are reported for 100000 random simulations, with effect sizes sampled assuming a normal distribution of genetic effects according to mean and standard deviations taken from the CKDGen data or ORIGIN calculations. Significance threshold was adjusted for multiple hypothesis testing with Bonferroni correction for 238 biomarkers ($P < 0.05/238$). Sensitivity analyses included "leave-one-out" validation and repetition of reverse MR using genetic variants associated with related traits and risk factors of decreased $eGFR_{crea}$ (see Methods in the online Data Supplement).

To determine whether the identified biomarkers themselves were causally affecting CKD, we used a traditional MR approach (Fig. 1A) to identify the effect of significant biomarkers on $eGFR_{crea}$. As previously described *(17)*, MR analysis was restricted to nearby SNPs (within 300 Kb of the gene encoding the biomarker) significantly associated with the biomarker ($P < 0.01$) and pruned for independence on the basis of linkage disequilibrium (see Methods in the online Data Supplement for more details).

ASSOCIATION OF SIGNIFICANT BIOMARKERS WITH BASELINE $eGFR_{crea}$, INCIDENT CKD, AND CHANGE IN $eGFR_{crea}$ IN ORIGIN

We assessed whether the significant biomarkers predicted incident CKD and change in $eGFR_{crea}$ in those ORIGIN biomarker substudy participants whose $eGFR_{crea}$ and urinary ACR were both normal at baseline. We used logistic regression models to test the association of each biomarker (independent variable) with incident CKD (dependent variable). Three separate multivariate models were constructed. The basic model was adjusted for age, sex, and ethnicity. The laboratory model additionally included baseline natural log-transformed urinary ACR and $eGFR_{crea}$. The full model included all variables in the basic and laboratory model and added fasting plasma glucose, systolic blood pressure, body mass index, prior diabetes, prior cardiovascular disease, antihypertensive drug use, and smoking status. As sensitivity analyses, we also tested each component of the composite CKD end point, namely, worsening of albuminuria category, doubling of serum creatinine, and $eGFR_{crea}$ <60 mL/min per 1.73 m^2. The significant biomarker concentrations were categorized as quartiles within this population. Kaplan–Meier survival probabilities from incident CKD were calculated for each quartile adjusting for age, sex, and ethnicity at each of the primary follow-up time points. Logistic regression models were created for incident CKD against biomarker quartiles and adjusted for the aforementioned base, laboratory, and full models. We determined *P* values for trend across the quartiles by modeling the biomarker quartiles as an ordered linear term. Similarly, linear regression models were used to assess the relationship between changes in $eGFR_{crea}$ from baseline to the end of usual follow-up time point, adjusting for the same models. Sensitivity analyses were performed by further adjusting models with randomization status for insulin glargine and ω-3 fatty acids.

The ability of the significant biomarkers to discriminate between people who do and do not develop incident stage 3 CKD was assessed by estimating the category-free net reclassification improvement (NRI) index after inclusion of the biomarker to several risk reference models containing known CKD risk factors with use of the Hmisc *(29)* R package. As a sensitivity measure, we also calculated ROC curves and the area under the curve (AUC) for the aforementioned risk models with and without the biomarker, using the pROC *(30)* R package.

All statistical analyses were performed on R version 3.0.1, unless otherwise specified.

DATA AVAILABILITY

Patient consent forms for the ORIGIN trial state that analysis of individual-level data must be approved by the principal investigator (Hertzel Gerstein, gerstein@mcmaster.ca). All relevant summary-level data are included within the report and its supporting information. CKDGen consortium data are available at http://ckdgen.imbi.uni-freiburg.de/.

Results

ORIGIN STUDY PARTICIPANT CHARACTERISTICS

The primary study population consisted of 5300 individuals from the ORIGIN trial who had normal $eGFR_{crea}$ and no albuminuria at baseline. In this subset, mean $eGFR_{crea}$, and urinary ACR at baseline were 82.0 mL/min per 1.73 m^2 and 6.2 mg/g (0.7 mg/mmol), respectively, and 1353 of 5300 participants developed CKD (980 albuminuria progression, 66 doubling of serum creatinine, and 515 $eGFR_{crea}$ <60 mL/min per 1.73 m^2) by study end. End-stage renal disease developed in only 2 of 5300 participants.

Key clinical characteristics of the study population are shown in Table 2 in the online Data Supplement. Those participants with prior renal disease were excluded from the present analysis, if not stated otherwise.

IDENTIFICATION OF SERUM BIOMARKERS AFFECTED BY $eGFR_{crea}$ BY USE OF REVERSE MR

With use of reverse MR on 238 serum biomarkers, trefoil factor 3 (TFF3)[11] and uromodulin (UMOD) were identified as being significantly affected by $eGFR_{crea}$ after adjustment for multiple hypothesis testing ($P < 0.05/238$) (see Table 3 in the online Data Supplement). The reverse MR analysis for TFF3 suggests an inverse relationship between $eGFR_{crea}$ and serum TFF3 levels ($\beta = -1.86$ SD of TFF3 per 1 unit increase in log-transformed $eGFR_{crea}$; 95% CI, −2.76 to −0.95; $P = 8 \times 10^{-5}$), or in other words increased TFF3 levels are associated with decreased kidney function (Fig. 3). While an inverse relationship between $eGFR_{crea}$ and UMOD was observed, a leave-one-out analysis suggested a single SNP (at the *UMOD* locus) was responsible for this association (see Fig. 2 in the online Data Supplement). Since the association with TFF3 remained unchanged upon leave-one-out analysis, we focused further analyses on this relationship. Consistent with the $eGFR_{crea}$ analysis, TFF3 was also associated with CKD ($P = 0.028$) but not associated with cystatin C-based eGFR ($eGFR_{cys}$), an alternative measure of GFR, in reverse MR analysis (see Table 4 in the online Data Supplement). To ensure that the $eGFR_{crea}$–TFF3 relationship was not due to some factor associated with $eGFR_{crea}$, we conducted reverse MR analysis for known risk factors of reduced GFR (see Methods in the online Data Supplement). TFF3 was not associated with any of these traits ($P > 0.05$; see Table 4 in the online Data Supplement). To investigate the potential for a causal effect of TFF3 on $eGFR_{crea}$, we determined whether the genetic variants near *TFF3* are also linked to $eGFR_{crea}$ and found no such relationship ($\beta = 0.01$; 95% CI, 0.00–0.02; $P = 0.24$).

ASSOCIATION OF TFF3 WITH PREVALENT AND INCIDENT CKD IN ORIGIN

We sought to test the ability of TFF3 to identify individuals at risk of CKD despite normal $eGFR_{crea}$ and no albuminuria. In the 5300 ORIGIN participants with a normal $eGFR_{crea}$ and no albuminuria at baseline, TFF3 was associated with a higher incidence of incident CKD (OR = 1.28 per SD; 95% CI, 1.18–1.38; $P = 4.58 \times 10^{-10}$). The relationship with incident CKD remained significant after adjusting for $eGFR_{crea,}$ albuminuria, other known CKD risk factors (Table 1), and treatment allocation in the ORIGIN trial (data not shown). The associations were also significant with different definitions of CKD (Table 1). After adjusting for known CKD risk factors, participants in the highest (OR = 1.74; 95% CI, 1.43–2.12; $P = 3.58 \times 10^{-8}$) and upper-middle TFF3 quartiles (OR = 1.41; 95% CI, 1.16–1.70; $P = 4.50 \times 10^{-4}$) had a significantly increased risk of incident CKD events relative to the lowest quartile (P for trend = 3.04×10^{-9}) (Fig. 4; see Table 5 in the online Data Supplement). Of patients in the highest quartile of serum TFF3, 65.7% were free from incident CKD, whereas those with upper-middle, lower-middle, and lowest quartile of serum TFF3 had 72.0%, 77.6%, and 80.2% freedom from incident CKD by study end, respectively (see Table 6 in the online Data Supplement). Moreover, higher levels of TFF3 predicted a greater decline in $eGFR_{crea}$ (−4.33 mL/min per 1.73 m^2 $eGFR_{crea}$ per 1 SD increase in TFF3; 95% CI, −5.13 to −3.53; $P = 7.92 \times 10^{-26}$) from baseline to the end of study before and after adjusting for baseline $eGFR_{crea}$ and urinary ACR (see Table 7 in the online Data Supplement).

To determine the improvement in discrimination ability for incident CKD after addition of TFF3 to models with established risk factors, we constructed ROC curves and calculated net reclassification improvement (NRI) scores. In models adjusting for age, sex, and ethnicity, addition of TFF3 moderately but significantly improved risk prediction relative to $eGFR_{crea}$ alone (AUC with TFF3 and $eGFR_{crea}$ = 0.60; AUC with $eGFR_{crea}$ alone = 0.59; NRI = 0.211; $P = 9.56 \times 10^{-12}$) (Table 2). Addition of TFF3 to models with clinical risk factors, urinary ACR, and other risk factors consistently showed significant improvements in NRI (Table 2). TFF3 alone

[11] Human Genes: *TFF3,* trefoil factor 3; *UMOD,* uromodulin.

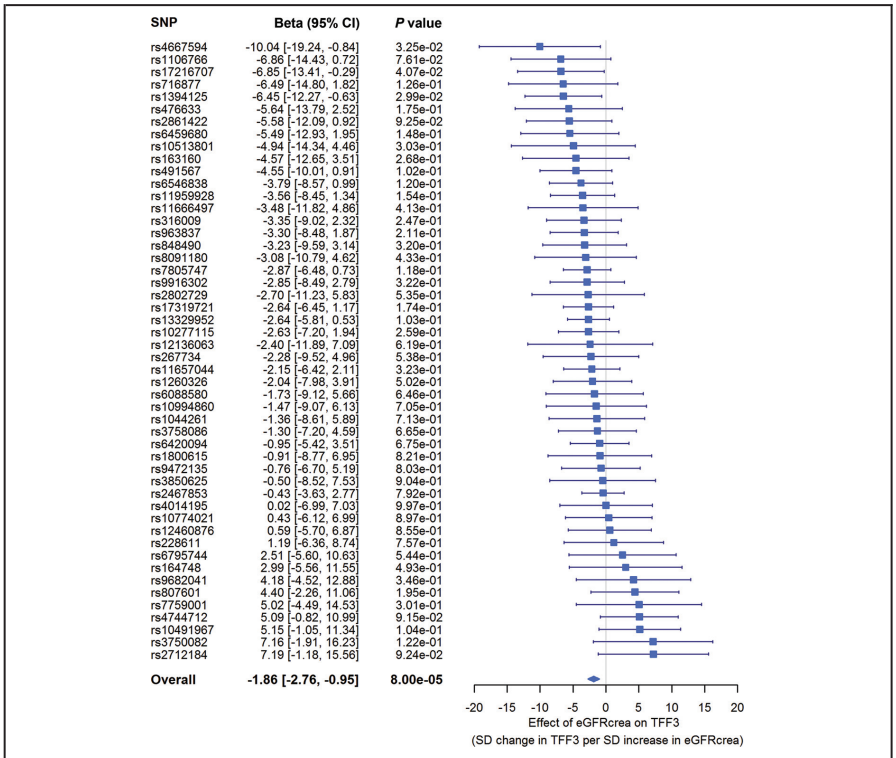


Fig. 3. Effect of $eGFR_{crea}$ on trefoil factor 3 with reverse MR.

Forest plots depict a summary of the reverse MR results for TFF3. Single SNP MR was conducted for each SNP used in the combined MR analysis. Betas represent the effect of $eGFR_{crea}$ on TFF3 as explained by that single SNP. Betas were determined by the Wald method *(39)* by regressing the effect estimates from the biomarker association (from ORIGIN trial) on the $eGFR_{crea}$ association (from CKDGen consortium). A two-tailed *P* value was calculated with a *z* test from 100 000 random simulations.

was also still significantly better at discriminating incident CKD than $eGFR_{crea}$ alone (AUC for $eGFR_{crea}$ = 0.59; AUC for TFF3 = 0.60; NRI = 0.112; $P = 1.86 \times 10^{-4}$).

Sensitivity analyses investigating the relationship between TFF3 and baseline renal characteristics were consistent with the overall analysis (see Results in the online Data Supplement).

Discussion

We herein demonstrated a novel approach to diagnostic biomarker discovery called "reverse MR." As a proof of concept, we applied it in the context of diabetic renal disease. TFF3 was identified as a serum biomarker that is causally increased by reduced $eGFR_{crea}$ in patients with dysglycemia. The observed effect on TFF3 was specific to $eGFR_{crea}$ and CKD, and TFF3 itself was not found to affect CKD. The reverse MR finding was subsequently validated with statistically significant and directionally consistent epidemiological associations, and we demonstrated TFF3 concentration is a significant independent predictor of incident CKD in a dysglycemic population.

TFF3 belongs to a family of 3 mammalian trefoil factors and is a 6.6-kDa protein that contains 1 trefoil

Table 1. Adjusted odds ratios for incident CKD in subset of ORIGIN with normal kidney measures.

	Age-, sex-, and ethnicity-adjusted		Laboratory model[a]		Full model[b]	
Clinical endpoint	OR[c,d] (95% CI)	*P* value	OR[c] (95% CI)	*P* value	OR[c] (95% CI)	*P* value
CKD	1.28 (1.18–1.38)	4.58E−10	1.21 (1.12–1.31)	2.22E−06	1.24 (1.1–1.34)	3.41E−07
Worsening in albuminuria category	1.17 (1.07–1.27)	3.01E−04	1.12 (1.02–1.22)	1.26E−02	1.13 (1.04–1.24)	6.38E−03
Doubling of serum creatinine	1.33 (1.01–1.76)	0.045	1.62 (1.22–2.16)	8.60E−04	1.66 (1.23–2.22)	8.11E−04
$eGFR_{crea}$ <60 mL/min per 1.73m^2 at EUF[d]	1.57 (1.41–1.75)	6.82E−16	1.44 (1.28–1.61)	2.58E−10	1.46 (1.30–1.64)	1.33E−10

[a] Adjusted for age, sex, ethnicity, baseline $eGFR_{crea}$, and natural log-transformed urinary albumin-to-creatinine ratio.
[b] Adjusted for age, sex, ethnicity, baseline $eGFR_{crea}$, natural log-transformed urinary albumin-to-creatinine ratio, fasting plasma glucose, systolic blood pressure, body mass index, prior diabetes, prior cardiovascular disease, antihypertensive drug use, and smoking status.
[c] OR per 1 SD increase in serum trefoil factor 3.
[d] OR, odds ratio; EUF, end of usual study follow-up.

domain, a characteristic of the trefoil protein family, and exists in monomeric or dimeric forms *(31, 32)*. The biological role of TFF3 is unclear, but it is believed to be involved in cellular repair and restitution. It is produced in the tubular cells of the renal cortex *(32)*, and both increased urinary *(33)* and serum TFF3 concentrations have been associated with onset *(34)* and severity of CKD *(32)*. With SOMAscan assay technology (SomaLogic), it was shown that TFF3 measured in plasma showed a negative correlation with measured GFR in a Swedish cohort with 389 participants *(35)*. Moreover, our results also suggest that TFF3 concentrations predict declines in $eGFR_{crea}$ and incident CKD in dysglycemic individuals with normal albuminuria and renal function. Indeed, in such individuals, TFF3 improved CKD risk discrimination compared to $eGFR_{crea}$ and models that included other risk factors. Those individuals with TFF3 in the highest quartile had 1.74 times greater odds of

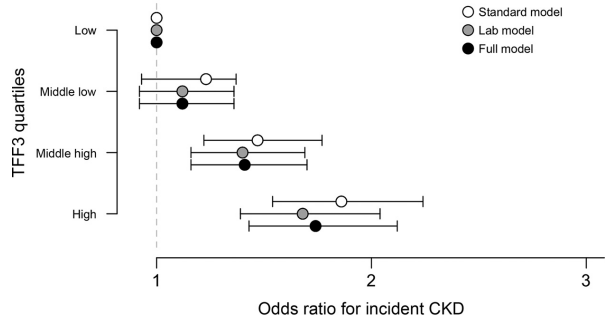


Fig. 4. Odds ratios of incident CKD according to TFF3 quartiles in subset of ORIGIN with normal kidney measures.
Figure represents the odds ratio of developing incident CKD in the given quartile of TFF3 relative to the lowest ratio, with adjustment for each set of model covariates. TFF3 concentrations were transformed into quartiles in a subset of the ORIGIN trial population with normal albuminuria and $eGFR_{crea}$. Standard model covariates included age, sex, and ethnicity. Laboratory model covariates included standard model as well as baseline $eGFR_{crea}$ and natural log-transformed urinary ACR. Full model covariates included laboratory model as well as fasting plasma glucose, systolic blood pressure, body mass index, prior diabetes, prior cardiovascular disease, antihypertensive drug use, and smoking status. Incident CKD was defined as worsening of albuminuria category, doubling of serum creatinine, end-stage renal disease, or $eGFR_{crea}$ <60 mL/min per 1.73m^2 at the end of follow-up.

Table 2. Comparison of discrimination capacity for incident CKD after addition of TFF3 to risk models.

Reference model	AUC for reference model (95% CI)	Model of interest	AUC for model of interest (95% CI)	NRI[a] (95% CI)	*P* value
$eGFR_{crea}$	0.586 (0.569–0.604)	TFF3	0.601 (0.584–0.618)	0.112 (0.050–0.174)	1.86E−04
		$eGFR_{crea}$ + TFF3	0.602 (0.585–0.610)	0.211 (0.149– 0.272)	9.56E−12
$eGFR_{crea}$ + clinical risk factors[b]	0.614 (0.596– 0.631)	$eGFR_{crea}$ + clinical risk factors[b] + TFF3	0.626 (0.609–0.643)	0.206 (0.145–0.268)	2.55E−11
$eGFR_{crea}$ + clinical risk factors[b] + ln(ACR)	0.664 (0.647–0.680)	$eGFR_{crea}$ + clinical risk factors[b] + ln(ACR) + TFF3	0.670 (0.653–0.687)	0.174 (0.112–0.235)	1.65E−08
Full model[c]	0.667 (0.650–0.683)	Full model[c] + TFF3	0.673 (0.657–0.690)	0.193 (0.131–0.254)	3.96E−10

[a] NRI index comparing model of interest to reference model.
[b] Prior cardiovascular disease + prior diabetes + systolic blood pressure + smoking status + anti-hypertensive drug use.
[c] Baseline $eGFR_{crea}$ + natural log-transformed baseline urinary albumin-to-creatinine ratio + prior cardiovascular disease + prior diabetes + systolic blood pressure + smoking status + anti-hypertensive drug use + body mass index + fasting plasma glucose.

developing CKD than the lowest after adjusting for known risk factors, including eGFR.

The dysglycemic population in ORIGIN is well suited for identification of novel markers because $eGFR_{crea}$ is temporally and paradoxically increased in the early stages of diabetic nephropathy *(9, 10)*. However, there are several limitations of our study. First, our stringent statistical criteria may have excluded other markers causally affected by renal disease. For example, 2 other established renal biomarkers, β_2 microglobulin *(14)* and cystatin C *(5)* (see Table 3 in the online Data Supplement), failed to meet the stringent Bonferroni-corrected significance threshold of 0.00021, although they were nominally significant (P, 3.4×10^{-4} and 1.5×10^{-3}, respectively) and cannot be ruled out as potential biomarkers. Second, MR relies on the assumption that the genetic variants have no pleiotropic effects *(15)*, meaning the genetic variants have an effect on the biomarker only through $eGFR_{crea}$. This concern was minimized by the use of multiple genetic variants, which limited the likelihood of an alternative pathway confounding our observation, and sensitivity analyses, like leave-one-out, to reduce the possibility of any individual variants driving results. For instance, we identified that the association of UMOD in the reverse MR analysis was driven by a single variant located in the *UMOD* gene, which is also related to hypertension *(36)* and strongly associated with $eGFR_{crea}$ *(18)*. In this case, UMOD is a known causal mediator of CKD, as leave-one-out analysis of a cis SNP resulting in a null association likely corresponds to a significant traditional MR result for that biomarker *(17)*. Third, we used genetic variants associated with $eGFR_{crea}$, a specific estimate of the kidney's abilities to function. The CKDGen Consortium genome-wide association studies for $eGFR_{cys}$ was a smaller sample size (n = 33152) and yielded weaker instruments ($F = 11.8$) than $eGFR_{crea}$ (n = 133814), which had an overall stronger set of instruments ($F = 67.6$). Consequently, the lack of a significant association in the reverse MR between $eGFR_{cys}$, an alternative measure of kidney function, and TFF3 could be due to the limited number of variants strongly associated with $eGFR_{cys}$, which limits our power to detect an association. Moreover, Pattaro et al. reported that the majority of $eGFR_{crea}$ SNPs in CKDGen were associated with a consistent direction of effect for $eGFR_{cys}$—albeit with less significance given the smaller sample size; this association further supported the expectation that the $eGFR_{crea}$ reverse MR findings are not reflecting pathways specific to creatinine metabolism but rather true kidney function overall *(18)*. Indeed, the reverse MR results themselves imply that the majority of variants are reflecting kidney function, since a nominal association was observed for cystatin C ($\beta = -1.39$ SD of cystatin C per 1 unit increase in log-transformed $eGFR_{crea}$; 95% CI, −2.30 to −0.48; $P = 2.8 \times 10^{-3}$), which is a known marker of CKD and itself corresponds to an increase in $eGFR_{cys}$.

Disease prediction does not necessitate causal biomarkers, and often better accuracy can be achieved by adding markers simply associated from epidemiology. However, finding the biomarkers causally affected by disease is particularly promising for diagnostic and population-screening applications. Because these markers are less prone to confounding, they would be expected to be more specific, sensitive, and altered early in the disease pathology. Improvements in diagnostic methods for early detection of CKD are important for strategies to slow progression of kidney disease and its complications, particularly in high-risk populations such as patients with dysglycemia *(5)*. Growing evidence for protective effects

of SGLT2 inhibitors and GLP1-receptor agonists *(37)* for diabetic nephropathy further highlights the timely nature and importance of early markers *(38)*. Future research should be directed toward better elucidating the biological role of TFF3 and trefoil family proteins in the kidneys and further evaluating the clinical utility of TFF3 as an early diagnostic tool for CKD in broader populations. Importantly, the validation of our findings with epidemiological models of CKD risk also suggests that reverse MR could be a novel method used to identify sensitive and early diagnostic biomarkers for a variety of other diseases. The growing number of publicly available genome-wide association studies data sets will clearly facilitate such analyses.

Author Contributions: *All authors confirmed they have contributed to the intellectual content of this paper and have met the following 4 requirements: (a) significant contributions to the conception and design, acquisition of data, or analysis and interpretation of data; (b) drafting or revising the article for intellectual content; (c) final approval of the published article; and (d) agreement to be accountable for all aspects of the article thus ensuring that questions related to the accuracy or integrity of any part of the article are appropriately investigated and resolved.*

P. Mohammadi-Shemirani, statistical analysis; J. Sjaarda, statistical analysis; G. Pare, financial support, administrative support, provision of study material or patients.

Authors' Disclosures or Potential Conflicts of Interest: *Upon manuscript submission, all authors completed the author disclosure form. Disclosures and/or potential conflicts of interest:*

Employment or Leadership: S. Hess, Sanofi.
Consultant or Advisory Role: H.C. Gerstein, Sanofi, Lilly, Merck, Novo Nordisk, AstraZeneca, Abbott, Boehringer Ingelheim, Janssen.
Stock Ownership: S. Hess, Sanofi.
Honoraria: H.C. Gerstein, Sanofi, Lilly, Merck, Novo Nordisk, Astra Zeneca, Boehringer Ingelheim.
Research Funding: The ORIGIN trial and biomarker project were financially supported by Sanofi and the Canadian Institute of Health Research (CIHR) (FRN 125794).
Expert Testimony: None declared.
Patents: None declared.

Role of Sponsor: The funding organizations played no role in the design of study, choice of enrolled patients, or review and interpretation of data. Sanofi provided comments on the manuscript.

Acknowledgment: The authors are thankful to all the participants having agreed to contribute to this project and to all the consortia for making their data publicly available. In particular, the authors wish to thank the ORIGIN investigators at the Population Health Research Institute for having led the biomarker project and the CKDGen consortium for providing free access to their data.

References

1. Molitch ME, Adler AI, Flyvbjerg A, Nelson RG, So W-Y, Wanner C, et al. Diabetic kidney disease: a clinical update from Kidney Disease: Improving Global Outcomes. Kidney Int 2015;87:20–30.
2. Matsushita K, van der Velde M, Astor BC, Woodward M, Levey AS, de Jong PE, et al. Association of estimated glomerular filtration rate and albuminuria with all-cause and cardiovascular mortality in general population cohorts: a collaborative meta-analysis. Lancet 2010;375:2073–81.
3. Hemmelgarn BR, Manns BJ, Lloyd A, James MT, Klarenbach S, Quinn RR, et al. Relation between kidney function, proteinuria, and adverse outcomes. JAMA 2010; 303:423–9.
4. Retnakaran R, Cull CA, Thorne KI, Adler AI, Holman RR. Risk factors for renal dysfunction in type 2 diabetes: U.K. Prospective Diabetes Study 74. Diabetes 2006;55: 1832–9.
5. Group KDIGO (KDIGO) CW. KDIGO clinical practice guideline for the evaluation and management of chronic kidney disease. Kidney Int Suppl 2013;3: 1–150.
6. Johnson DW, Jones GRD, Mathew TH, Ludlow MJ, Chadban SJ, Usherwood T, et al. Chronic kidney disease and measurement of albuminuria or proteinuria: a position statement. Med J Aust 2012;197:224–5.
7. Nelson RG, Bennett PH, Beck GJ, Tan M, Knowler WC, Mitch WE, et al. Development and progression of renal disease in Pima Indians with non-insulin-dependent diabetes mellitus. N Engl J Med 1996;335:1636–42.
8. Shemesh O, Golbetz H, Kriss JP, Myers BD. Limitations of creatinine as a filtration marker in glomerulopathic patients. Kidney Int 1985;28:830–8.
9. Bank N. Mechanisms of diabetic hyperfiltration. Kidney Int 1991;40:792–807.
10. Jin Y, Moriya T, Tanaka K, Matsubara M, Fujita Y. Glomerular hyperfiltration in non-proteinuric and non-hypertensive Japanese type 2 diabetic patients. Diabetes Res Clin Pract 2006;71:264–71.
11. GBD 2016 Causes of Death Collaborators, Naghavi M, Abajobir AA, Abbafati C, Abbas KM, Abd-Allah F, et al. Global, regional, and national age-sex specific mortality for 264 causes of death, 1980–2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet 2017;390:1151–210.
12. Fliser D, Kollerits B, Neyer U, Ankerst DP, Lhotta K, Lingenhel A, et al. Fibroblast growth factor 23 (FGF23) predicts progression of chronic kidney disease: the Mild to Moderate Kidney Disease (MMKD) Study. J Am Soc Nephrol 2007;18:2600–8.
13. Bhavsar NA, Köttgen A, Coresh J, Astor BC. Neutrophil gelatinase-associated lipocalin (NGAL) and kidney injury molecule 1 (KIM-1) as predictors of incident CKD stage 3: the Atherosclerosis Risk in Communities (ARIC) Study. Am J Kidney Dis 2012;60:233–40.
14. Foster MC, Coresh J, Hsu C, Xie D, Levey AS, Nelson RG, et al. Serum β-trace protein and β2-microglobulin as predictors of ESRD, mortality, and cardiovascular disease in adults with CKD in the Chronic Renal Insufficiency Cohort (CRIC) Study. Am J Kidney Dis 2016;68: 68–76.
15. Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. Hum Mol Genet 2014;23:R89–98.
16. Clarke R, Peden JF, Hopewell JC, Kyriakou T, Goel A, Heath SC, et al. Genetic variants associated with Lp(a) lipoprotein level and coronary disease. N Engl J Med 2009;361:2518–28.
17. Sjaarda J, Gerstein HC, Yusuf S, Treleaven D, Walsh M, Mann JFE, et al. Blood HER2 and uromodulin as causal mediators of CKD. J Am Soc Nephrol 2018; 29:1326–35.
18. Pattaro C, Teumer A, Gorski M, Chu AY, Li M, Mijatovic V, et al. Genetic associations at 53 loci highlight cell types and biological pathways relevant for kidney function. Nat Commun 2016;7:10023.
19. Gerstein HC, Paré G, McQueen MJ, Haenel H, Lee SF, Pogue J, et al. Identifying novel biomarkers for cardiovascular events or death in people with dysglycemia. Circulation 2015;132:2297–304.
20. Gerstein HC, Yusuf S, Riddle MC, Ryden L, Bosch J. Rationale, design, and baseline characteristics for a large international trial of cardiovascular disease prevention in people with dysglycemia: the ORIGIN trial (Outcome Reduction with an Initial Glargine Intervention). Am Heart J 2008;155:26–32, 32.e1–6.
21. Gerstein HC, Paré G, McQueen MJ, Lee SF, Hess S. Validation of the ORIGIN cardiovascular biomarker panel and the value of adding troponin I in dysglycemic people. J Clin Endocrinol Metab 2017;102:2251–7.
22. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007;81:559–75.
23. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet 2011;88:76–82.
24. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet 2009;5:e1000529.
25. Gibbs RA, Boerwinkle E, Doddapaneni H, Han Y, Korchina V, Kovar C, et al. A global reference for human genetic variation. Nature 2015;526:68–74.
26. Marchini J, Howie B. Genotype imputation for genome-wide association studies. Nat Rev Genet 2010;11:499–511.
27. Burgess S, Butterworth A, Thompson SG. Mendelian

randomization analysis with multiple genetic variants using summarized data. Genet Epidemiol 2013;37:658–65.
28. Efron B. Bootstrap methods: another look at the jackknife. Ann Stat 1979;7:1–26.
29. Harrell F. Hmisc: A package of miscellaneous R functions. 2014.
30. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J-C, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics 2011;12:1–8.
31. Aihara E, Engevik KA, Montrose MH. Trefoil factor peptides and gastrointestinal function. Annu Rev Physiol 2017;79:357–80.
32. Du T, Luo H, Qin H, Wang F, Wang Q, Xiang Y, et al. Circulating serum trefoil factor 3 (TFF3) is dramatically increased in chronic kidney disease. PLoS One 2013;8:e80271.
33. Astor BC, Köttgen A, Hwang S-J, Bhavsar N, Fox CS, Coresh J. Trefoil factor 3 predicts incident chronic kidney disease: a case-control study nested within the Atherosclerosis Risk in Communities (ARIC) study. Am J Nephrol 2011;34:291–7.
34. Lebherz-Eichinger D, Tudor B, Ankersmit HJ, Reiter T, Haas M, Roth-Walter F, et al. Trefoil factor 1 excretion is increased in early stages of chronic kidney disease. PLoS One 2015;10:e0138312.
35. Christensson A, Ash JA, DeLisle RK, Gaspar FW, Ostroff R, Grubb A, et al. The impact of the glomerular filtration rate on the human plasma proteome. Proteomics Clin Appl 2018;12:1700067.
36. Padmanabhan S, Melander O, Johnson T, Di Blasio AM, Lee WK, Gentilini D, et al. Genome-wide association study of blood pressure extremes identifies variant near UMOD associated with hypertension. PLoS Genet 2010;6:e1001177.
37. Mann JFE, Ørsted DD, Brown-Frandsen K, Marso SP, Poulter NR, Rasmussen S, et al. Liraglutide and renal outcomes in type 2 diabetes. N Engl J Med 2017;377:839–48.
38. Fioretto P, Zambon A, Rossato M, Busetto L, Vettor R. SGLT2 inhibitors and the diabetic kidney. Diabetes Care 2016;39:S165–71.
39. Wald A. The fitting of straight lines if both variables are subject to error. Ann Math Stat 1940;11:284–300.

CHAPTER 6:

DISCUSSION

CHAPTER 6: DISCUSSION

6.1 GENERAL OVERVIEW

Precision medicine aims to stratify patients and provide tailored treatments, as such it is expected that better understanding of individual biological characteristics and their relationship to disease will be instrumental in advancing such initiatives (Correa Rojo et al., 2021). More specifically, studies of peripheral blood biomarkers have applications to the discovery of novel therapeutic targets, prediction of drug efficacy and safety, and improvements in disease screening. However, as demonstrated in this body of work, establishing causality in epidemiological studies of biomarkers is critical for these applications and Mendelian randomization analysis can act as a complementary tool to pinpoint such biomarkers in the context of cardiometabolic diseases. This section serves as a summary of the key findings presented in this thesis as well as their implications, limitations, and future directions of research.

6.2 CHAPTER SUMMARIES

6.2.1 CHAPTER 3 SUMMARY

Lipoprotein(a) is a lipid particle consisting of an LDL core attached to an apolipoprotein(a) chain and serves as a well-established risk factor for atherosclerotic cardiovascular disease but its role in atrial fibrillation is unclear. In this chapter, higher lipoprotein(a) was demonstrated to be associated with increased risk of atrial fibrillation in participants from the UK Biobank. Moreover, these results were replicated through Mendelian randomization analysis using independent data from two of the largest genome-wide association studies of atrial fibrillation. Notably, no effect of other atherogenic lipids

and some residual effect from lipoprotein(a) that was not mediated through atherosclerotic cardiovascular disease suggests an independent mechanism that underlies the role of lipoprotein(a) in atrial fibrillation pathology.

6.2.2 CHAPTER 4 SUMMARY

In this chapter, Mendelian randomization analysis was applied to test causal effects of lifelong higher free testosterone in males. Among 22 outcomes relevant to testosterone supplementation, we identified adverse effects on increased risk of prostate cancer, risk of androgenic alopecia, and hematocrit percentage, but beneficial effects on increased heel bone mineral density, increased body fat-free percentage and decreased body fat percentage. In a hypothesis-free survey of 439 diseases and traits, we found lesser-appreciated adverse effects on creatinine, HDL, hypertension, and spinal stenosis, but beneficial effects on C-reactive protein.

6.2.3 CHAPTER 5 SUMMARY

Traditionally, Mendelian randomization analysis has been focused on identifying biomarkers that cause disease. Although non-causal biomarkers, such as those associated with confounders or antecedent causes of disease, can be informative for some applications, biomarkers that are specifically caused by disease have advantageous properties. We proposed that Mendelian randomization can be used to identify biomarkers that are consequences of disease thereby representing early and sensitive markers for disease. Among 238 serum biomarkers measured in the ORIGIN trial, trefoil factor 3 was associated with genetic predisposition to reduced renal function as measured by $eGFR_{crea}$. As hypothesized, trefoil factor 3 showed promise for early diagnosis as it predicted incident

chronic kidney disease in participants without risk factors and improved discriminative ability of models for early stages of chronic kidney disease even after adjusting for known risk factors.

6.3 SIGNIFICANCE OF FINDINGS AND IMPLICATIONS

6.3.1 LIPOPROTEIN(A)

Lipoprotein(a) is an established risk factor for coronary artery disease, ischemic stroke, and aortic valve stenosis (Clarke et al., 2009; Larsson et al., 2020; Paré et al., 2019; Thanassoulis et al., 2013). Its role in atrial fibrillation is unclear as smaller studies investigating the relationship had inconclusive findings (Aronis et al., 2017; Garg et al., 2020) (Jiang et al., 2021). Our work was able to demonstrate a causal effect of lipoprotein(a) on atrial fibrillation by integrating evidence from observational and genetic epidemiology across multiple independent cohorts representing a 20-fold increase in atrial fibrillation cases over previous observational studies. The lack of preventative therapies for atrial fibrillation with ancillary benefits on cardiovascular comorbidities makes lipoprotein(a) an intriguing candidate worthy of further investigation (Bang et al., 2013; Bhatt et al., 2019; Khatib et al., 2013; Neefs et al., 2017; Schneider et al., 2010). Lp(a)HORIZON is an ongoing phase III trial testing pelacarsen, an antisense oligonucleotide against lipoprotein(a), for prevention of major adverse cardiovascular events. In light of our results, secondary analyses or future trials should consider atrial fibrillation as an additional outcome.

Furthermore, we demonstrated the effect of lipoprotein(a) on atrial fibrillation is partly mediated independent of known effects on atherosclerotic cardiovascular diseases.

From a structural and biochemical perspective, lipoprotein(a) has an LDL-like core covalently attached to an apolipoprotein(a) chain – a homologue of plasminogen with an inactive protease site – and acts as a carrier of proinflammatory oxidized phospholipids in circulation (Tsimikas et al., 2020) (Berglund & Ramakrishnan, 2004). Therefore, lipoprotein(a) is a unique molecule with potential atherogenic, thrombogenic, and inflammatory properties that provide alternative mechanisms that could mediate an effect on atrial fibrillation. Indeed, both thrombosis and inflammation have been associated with atrial fibrillation (Spronk et al., 2017) (Pretorius et al., 2007) (Conen et al., 2010). Colchicine – an anti-inflammatory drug – is being explored as a secondary prevention to lower rates of post-operative atrial fibrillation (Imazio et al., 2011). To this end, an *ex vivo* study showed that large reductions in lipoprotein(a) levels induced anti-inflammatory gene expression in circulating monocytes (Stiekema et al., 2020).

More broadly, this study provides further evidence that deleterious effects of lipoprotein(a) are not limited to arterial tissue. Previous studies in aortic valve stenosis have shown mechanical stresses allow lipoprotein(a) to infiltrate valvular tissue and induce gene expression resulting in calcification and cell death (Bouchareb et al., 2015; Zheng et al., 2019). Our results suggest lipoprotein(a) effects may extend to atrial tissue and raises questions regarding the potential for effects of lipoprotein(a) on other diseases.

6.3.2 TESTOSTERONE

Testosterone is a biomarker of great clinical interest given age-related declines in its levels, frequency of prescriptions for testosterone supplementation, and controversy surrounding its effects on cardiometabolic disease (Bhasin et al., 2018). As might be

expected for an endogenous hormone, we found testosterone exhibited both beneficial and deleterious effects in males. Many effects replicated well-established effects based on observations from trials of short-term testosterone administration, such as lower body fat, higher fat-free mass, higher hematocrit, and higher bone density (Bhasin et al., 2018) (Snyder et al., 2018). Other observed effects on risk of prostate cancer and androgenic alopecia were supported by randomized controlled trials of androgen suppression therapy, such as 5α-reductase inhibitors, and this work raises the possibility of higher endogenous testosterone levels as a risk factor (Adil & Godwin, 2017) (Andriole et al., 2010).

When considering cardiometabolic outcomes, we did not find evidence of a beneficial effect on any endpoints, casting doubt on claims of benefit for cardiovascular or metabolic health. However, it should be noted that recent Mendelian randomization studies have been divided on the role of testosterone in type 2 diabetes. *Ruth et al.* identified a protective effect for testosterone on type 2 diabetes in males (Ruth et al., 2020), while a recent preprint by *Leinonen et al.* using data from the FinnGen cohort found no effect after adjustment for sex hormone-binding globulin (SHBG), which matched our findings (Leinonen et al., 2021). Explanations for the discrepancies could include differences in the amount of control for pleiotropy, differences in power due to numbers of type 2 diabetes cases, or other differences between study populations. In particular, isolating testosterone is challenging and there is molecular and epidemiological evidence that SHBG may have physiological effects independent from regulating sex hormone concentration (Lakshman et al., 2010; Rosner et al., 2010; Vikan et al., 2010). Once completed, the TRAVERSE trial (clinicaltrials.gov, NCT03518034) will be able to shed more light on the effects of

testosterone supplementation in hypogonadal men for cardiovascular and prostate safety as well as anemia, bone density, depression, diabetes, and sexual activity.

Furthermore, in a hypothesis-free scan of human disease and biomarkers, we detected effects of testosterone on higher risk of hypertension, higher risk of spinal stenosis, and decrease C-reactive protein. Testosterone is believed to have immunosuppressive properties which may explain its effect on C- reactive protein, a common marker of inflammation (Trigunaite et al., 2015). Scientific literature shows some evidence that higher testosterone is associated with greater loss of cartilage in healthy older males, and evidence from mouse models suggest testosterone has a sex-specific role in worsening osteoarthritis, a common risk factor for spinal stenosis (Hanna et al., 2005) (Ma et al., 2007). With regards to hypertension, since it is a multifactorial disease, a lack of association with systolic or diastolic blood pressure does not preclude a real effect of testosterone on hypertension. Indeed, small randomized controlled trials have observed testosterone administration lowered NT-proBNP levels, a protective factor against hypertension, in healthy men (Bachmann et al., 2019). Likewise, animal models show testosterone may exacerbate early stages of hypertension by increasing production of reactive oxygen species in hypertensive rats (Hartgens & Kuipers, 2004) (Reckelhoff et al., 1998).

Since the publication of this work, additional Mendelian randomization studies of testosterone have been published. A study of bioavailable testosterone in the UK Biobank found no effects on social and economic outcomes (Harrison et al., 2021). This study reported no effect of testosterone on BMI in males, which could reflect opposing effects of testosterone on decreased body fat and increased body fat-free mass as reported in our study

and others. In addition, there is a preprint for a phenome-wide analysis that builds on our work by including participants from the FinnGen cohort, though it includes comparatively fewer outcomes (Leinonen et al., 2021). Although outcomes are not directly comparable, these independent results are largely concordant with our study.

6.3.3 "REVERSE" MR AND TREFOIL FACTOR 3

Our results highlighted trefoil factor 3 as a promising diagnostic and prognostic biomarker that predicted worsening kidney function at an early stage of disease. Trefoil factor 3 is not well-known for its role in kidney so there is still a scarcity of literature. Since our study was published some independent researchers using broad panels of biomarkers have similarly reported associations of trefoil factor 3 with kidney function in various populations (Marcovecchio et al., 2020) (Grams et al., 2021) (Ascher et al., 2021). Ultimately, the growing appreciation of this biomarker in kidney disease driven by hypothesis-free studies reinforces the importance of unbiased exploratory analyses to enable the discovery of new biomarkers for disease.

Furthermore, the reverse Mendelian randomization method has started to gain traction as it has been applied to other traits and diseases. Two studies surveyed proteins on the SOMAScan assay to identify associations with genetic predisposition to BMI (Goudswaard et al., 2021) (Zaghlool et al., 2021). *Goundwaard et al.* identified 8 differentially regulated proteins, and *Zaghlool et al.* identified 24 differentially regulated proteins, of which 21 had a BMI-to-protein causal relationship. Importantly, it is not necessary to restrict reverse Mendelian randomization to identifying proteins that are consequences of disease and other researchers have broadened the scope to include other

types of molecules. A preprint by *Gobeil et al.* explored blood metabolites that were associated with genetic predisposition to non-alcoholic fatty liver disease, and identified elevated tyrosine levels as a plausible metabolite causally affected by liver disease (Gobeil et al., 2021). Another study surveyed 12 clinically relevant traits and diseases to develop a comprehensive map of differential gene expression caused by each disease (Porcu et al., 2021). As the rate of adoption of multiplex quantitative proteomic technologies increases, it will lead to additional studies that discover novel diagnostic biomarkers. Finally, unlike therapeutic targets, causality does not need to be established for diagnostic biomarkers. However, as our work demonstrates, biomarkers that are causal consequences of disease are likely to provide unique properties as early markers that can be ideal to screen or monitor disease (Califf, 2018) (Atkinson et al., 2001).

6.4 LIMITATIONS

As previously discussed, Mendelian randomization has several limitations. First, genetic variants are assumed to act on the outcome exclusively through the biomarker (i.e., no horizontal pleiotropy). Second, estimates from Mendelian randomization represent lifelong effects of the endogenous biomarker on each outcome, which may not reflect short-term interventions with exogenous drugs. This is particularly important when extrapolating findings to predict effects of short-term interventions, such as administration of exogenous testosterone in randomized controlled trials. Third, estimates from Mendelian randomization represent on-target effects. Again, this is relevant when extrapolating findings to predicting efficacy and safety of therapeutics, as different therapeutic agents and modalities can have off-target effects. Similarly, unless tissue-specific genetic variants

associated with biomarker levels are incorporated, these analyses would not predict tissue-specific effects. The latter point can be an important consideration as therapies might have different effects across tissues, and other modalities, such as oligonucleotide-based therapies, are currently limited to delivery in specific tissues (Roberts et al., 2020) (Zhao et al., 2020).

With respect to individual chapters, there are some additional limitations. First, independent replication is necessary for novel findings, such as the association of trefoil factor 3 with incident chronic kidney disease. At the time, we were limited by the availability of public data with the necessary measurements, but wider adoption of proteomics in other epidemiological cohorts have now addressed these concerns providing independent replication for this association (Grams et al., 2021). Second, electronic health records are commonly utilized and invaluable datasets for conducting exploratory phenome-wide in large-scale biobank studies. However, there are inherent biases in this type of data due to ascertainment and selection biases in hospital settings. In chapter 3, we noted that atrial fibrillation cases might reflect more severe disease that required hospitalization while controls might include undetected atrial fibrillation. Similarly, outcomes in the phenome-wide analysis from chapter 4 could suffer from these biases as they were derived from electronic health records in the UK Biobank. Third, genetic studies were limited to Europeans to control for confounding by genetic ancestry, but this limits generalizability to other populations. In some cases, such as lipoprotein(a), there are large ethnic differences in biomarker levels and potentially differences in risk, which make trans-ethnic analysis particularly relevant (Paré et al., 2019). Often non-European ancestries had

smaller sample sizes that decreased statistical power and limited the ability to draw firm conclusions, which underscores the need for increased diversity (Martin et al., 2019). On a related note, our Mendelian randomization study of testosterone was limited to males. The decision was motivated by the sex-specific genetic architecture of testosterone, which can result in opposing sex-specific effects that would otherwise be masked, but this limited generalizability to females (Leinonen et al., 2021) (Ruth et al., 2020).

6.5 FUTURE DIRECTIONS

Naturally, there are numerous outstanding questions that remain unanswered for each chapter in this thesis. Although there is considerable excitement around lipoprotein(a) inhibitors, the Lp(a)HORIZON trial is still ongoing to see if pelacarsen will prevent cardiovascular disease. Until then, there is still uncertainty regarding the exact mechanisms that mediate the effect of lipoprotein(a) and atrial fibrillation indirect from atherosclerotic cardiovascular disease. Targeted cohorts focused on studying atrial fibrillation have better phenotyping that could help address limitations related to use of electronic healthcare records (Conen et al., 2017). Additionally, experimental models or multivariable analyses can help disentangle the specific components of lipoprotein(a) that mediate the effect on atrial fibrillation, such as size, concentration, or oxidized phospholipid content. Furthermore, there have been phenome-wide Mendelian randomization studies of lipoprotein(a), but with updates to the UK Biobank and release of other biobank studies, a refresh may be warranted eventually to discover novel effects of lipoprotein(a) (Chong et al., 2019) (W. Zhou et al., 2021).

Although it is tempting to think about biomarkers in binary terms, most endogenous molecules are inherently essential in human bodies at certain levels. Therefore, a limitation of Mendelian randomization studies for testosterone has been the assumption of a linear relationship with disease. Instead, it is more likely that there is a range of healthy values and excessively high or low levels both confer deleterious effects. Non-linear Mendelian randomization allows modelling of non-linear causal relationship and it could be used to estimate optimal levels of testosterone for overall health (Staley & Burgess, 2017). Indeed, other studies using this technique have found evidence for non-linear relationships, such as vitamin D – another endogenous hormone – and all-cause mortality, which was previously missed by traditional Mendelian randomization (A. Zhou et al., 2021) (Sofianopoulou et al., 2021) (Sun et al., 2019).

With the release of newer datasets from the CKDGen consortium, the reverse Mendelian randomization study for diagnostic biomarkers of kidney function may warrant an update that includes a larger panel of biomarkers (Stanzick et al., 2021) (Ferkingstad et al., 2021). However, independent validation is important to assess diagnostic accuracy, which can present a challenge as the availability of individual-level data from studies with the same biomarker panel is often limited. Alternatively, it might be more feasible to design targeted panels based on biomarkers identified from Mendelian randomization analyses that could then validate their performance in longitudinal cohorts. Furthermore, there have been methodological developments related to Mendelian randomization that could be worth exploring. Novel Mendelian randomization methods, such as the contamination mixture method, group genetic variants according to their individual causal estimates and provide

valid estimates if the largest group of genetic variants contributing to causal effect follow Mendelian randomization assumptions (Burgess et al., 2020). As such, these methods are more robust and identify instances where a cluster of genetic variants might have a causal effect on a biomarker. The contaminant mixture method can further disentangle instances where a complex disease may have multiple effects on an outcome through different mechanisms by inspecting the modality of the distribution of likely causal effect estimates (Burgess et al., 2020). These methods can be applied more broadly as sample sizes in genome-wide association studies increase, leading to the discovery of additional genetic variants that better capture the multifactorial nature of complex traits. For instance, genetic variants associated with BMI reflect diverse biological pathways, such as appetite control, growth, insulin secretion, and thermogenesis (Loos & Yeo, 2021). Therefore, it is plausible that the contaminant mixture method may enable an unbiased identification of biomarkers that are caused by specific biological pathways underlying complex disease and enable discovery of novel disease subtypes.

Finally, proteomics tools are beginning to emerge that provide single-molecule protein sequencing, such as Quantum-Si (https://www.quantum-si.com/), Nautilus Bio (https://www.nautilus.bio/), and others. Although this technology is still very early, it could enable better characterization of human proteoforms and perhaps avoid limitations of existing quantitative proteomic methods that rely on epitope recognition for protein detection (Pietzner et al., 2021).

6.6 CONCLUSION

In this thesis, we demonstrated that Mendelian randomization is a valuable tool to identify causal relationship between biomarker and cardiometabolic disease. It is expected that such biomarkers can inform efficacy and safety of novel therapeutic targets, uncover drug repurposing opportunities, and discover diagnostic biomarkers. In chapter 3, we applied Mendelian randomization to a pre-specified hypothesis revealing lipoprotein(a) is a risk factor for atrial fibrillation independent of atherosclerotic cardiovascular disease. In chapter 4, we broadened the scope by conducting a phenome-wide Mendelian randomization and identifying both beneficial and adverse effects for higher testosterone levels. In chapter 5, we test whether Mendelian randomization can identify biomarkers caused by disease and discover trefoil factor 3 as a novel diagnostic biomarker for early-stage chronic kidney disease. As we stand on the shoulders of giants, we hope the work outlined in this thesis will form the foundation for future discoveries that will improve patient lives.

6.7 REFERENCES

Adil, A., & Godwin, M. (2017). The effectiveness of treatments for androgenetic alopecia: A systematic review and meta-analysis. *Journal of the American Academy of Dermatology*, *77*(1), 136-141.e5. https://doi.org/10.1016/j.jaad.2017.02.054

Andriole, G. L., Bostwick, D. G., Brawley, O. W., Gomella, L. G., Marberger, M., Montorsi, F., Pettaway, C. A., Tammela, T. L., Teloken, C., Tindall, D. J., Somerville, M. C., Wilson, T. H., Fowler, I. L., & Rittmaster, R. S. (2010). Effect of Dutasteride on the Risk of Prostate Cancer. *New England Journal of Medicine*,

362(13), 1192–1202. https://doi.org/10.1056/NEJMoa0908127

Aronis, K. N., Zhao, D., Hoogeveen, R. C., Alonso, A., Ballantyne, C. M., Guallar, E., Jones, S. R., Martin, S. S., Nazarian, S., Steffen, B. T., Virani, S. S., & Michos, E. D. (2017). Associations of Lipoprotein(a) Levels With Incident Atrial Fibrillation and Ischemic Stroke: The ARIC (Atherosclerosis Risk in Communities) Study. *Journal of the American Heart Association*, *6*(12), 1–11. https://doi.org/10.1161/JAHA.117.007372

Ascher, S. B., Scherzer, R., Estrella, M. M., Muiru, A. N., Jotwani, V. K., Grunfeld, C., Shigenaga, J., Spaulding, K. A., Ng, D. K., Gustafson, D., Spence, A. B., Sharma, A., Cohen, M. H., Parikh, C. R., Ix, J. H., & Shlipak, M. G. (2021). Kidney tubule health scores and their associations with incident CKD in women living with HIV. *HIV Medicine*, *22*(7), 527–537. https://doi.org/10.1111/hiv.13081

Atkinson, A. J., Colburn, W. A., DeGruttola, V. G., DeMets, D. L., Downing, G. J., Hoth, D. F., Oates, J. A., Peck, C. C., Schooley, R. T., Spilker, B. A., Woodcock, J., & Zeger, S. L. (2001). Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework. *Clinical Pharmacology & Therapeutics*, *69*(3), 89–95. https://doi.org/10.1067/mcp.2001.113989

Bachmann, K. N., Huang, S., Lee, H., Dichtel, L. E., Gupta, D. K., Burnett, J. C., Miller, K. K., Wang, T. J., & Finkelstein, J. S. (2019). Effect of Testosterone on Natriuretic Peptide Levels. *Journal of the American College of Cardiology*, *73*(11), 1288–1296. https://doi.org/10.1016/j.jacc.2018.12.062

Bang, C. N., Greve, A. M., Abdulla, J., Køber, L., Gislason, G. H., & Wachtell, K.

(2013). The preventive effect of statin therapy on new-onset and recurrent atrial fibrillation in patients not undergoing invasive cardiac interventions: A systematic review and meta-analysis. *International Journal of Cardiology*, *167*(3), 624–630. https://doi.org/10.1016/j.ijcard.2012.08.056

Berglund, L., & Ramakrishnan, R. (2004). Lipoprotein(a): an elusive cardiovascular risk factor. *Arteriosclerosis, Thrombosis, and Vascular Biology*, *24*(12), 2219–2226. https://doi.org/10.1161/01.ATV.0000144010.55563.63

Bhasin, S., Brito, J. P., Cunningham, G. R., Hayes, F. J., Hodis, H. N., Matsumoto, A. M., Snyder, P. J., Swerdloff, R. S., Wu, F. C., & Yialamas, M. A. (2018). Testosterone Therapy in Men With Hypogonadism: An Endocrine Society* Clinical Practice Guideline. *The Journal of Clinical Endocrinology & Metabolism*, *103*(5), 1715–1744. https://doi.org/10.1210/jc.2018-00229

Bhatt, D. L., Steg, P. G., Miller, M., Brinton, E. A., Jacobson, T. A., Ketchum, S. B., Doyle, R. T., Juliano, R. A., Jiao, L., Granowitz, C., Tardif, J.-C., & Ballantyne, C. M. (2019). Cardiovascular Risk Reduction with Icosapent Ethyl for Hypertriglyceridemia. *New England Journal of Medicine*, *380*(1), 11–22. https://doi.org/10.1056/NEJMoa1812792

Bouchareb, R., Mahmut, A., Nsaibia, M. J., Boulanger, M. C., Dahou, A., Lépine, J. L., Laflamme, M. H., Hadji, F., Couture, C., Trahan, S., Pagé, S., Bossé, Y., Pibarot, P., Scipione, C. A., Romagnuolo, R., Koschinsky, M. L., Arsenault, B. J., Marette, A., & Mathieu, P. (2015). Autotaxin derived from lipoprotein(a) and valve interstitial cells promotes inflammation and mineralization of the aortic valve. *Circulation*,

132(8), 677–690. https://doi.org/10.1161/CIRCULATIONAHA.115.016757

Burgess, S., Foley, C. N., Allara, E., Staley, J. R., & Howson, J. M. M. (2020). A robust and efficient method for Mendelian randomization with hundreds of genetic variants. *Nature Communications*, *11*(1). https://doi.org/10.1038/s41467-019-14156-4

Califf, R. M. (2018). Biomarker definitions and their applications. *Experimental Biology and Medicine*, *243*(3), 213–221. https://doi.org/10.1177/1535370217750088

Chong, M., Sjaarda, J., Pigeyre, M., Mohammadi-Shemirani, P., Lali, R., Shoamanesh, A., Gerstein, H. C., & Paré, G. (2019). Novel Drug Targets for Ischemic Stroke Identified Through Mendelian Randomization Analysis of the Blood Proteome. *Circulation*, *140*(10), 819–830. https://doi.org/10.1161/circulationaha.119.040180

Clarke, R., Peden, J. F., Hopewell, J. C., Kyriakou, T., Goel, A., Heath, S. C., Parish, S., Barlera, S., Franzosi, M. G., Rust, S., Bennett, D., Silveira, A., Malarstig, A., Green, F. R., Lathrop, M., Gigante, B., Leander, K., de Faire, U., Seedorf, U., … Farrall, M. (2009). Genetic Variants Associated with Lp(a) Lipoprotein Level and Coronary Disease. *New England Journal of Medicine*, *361*(26), 2518–2528. https://doi.org/10.1056/NEJMoa0902604

Conen, D., Ridker, P. M., Everett, B. M., Tedrow, U. B., Rose, L., Cook, N. R., Buring, J. E., & Albert, C. M. (2010). A multimarker approach to assess the influence of inflammation on the incidence of atrial fibrillation in women. *European Heart Journal*, *31*(14), 1730–1736. https://doi.org/10.1093/eurheartj/ehq146

Conen, D., Rodondi, N., Müller, A., Beer, J. H., Auricchio, A., Ammann, P., Hayoz, D., Kobza, R., Moschovitis, G., Shah, D., Schläpfer, J., Novak, J., Di Valentino, M.,

Erne, P., Sticherling, C., Bonati, L. H., Ehret, G., Roten, L., Fischer, U., … Osswald, S. (2017). Design of the Swiss Atrial Fibrillation Cohort Study (Swiss-AF): Structural brain damage and cognitive decline among patients with atrial fibrillation. *Swiss Medical Weekly*, *147*(July), 1–9. https://doi.org/10.4414/smw.2017.14467

Correa Rojo, A., Heylen, D., Aerts, J., Thas, O., Hooyberghs, J., Ertaylan, G., & Valkenborg, D. (2021). Towards Building a Quantitative Proteomics Toolbox in Precision Medicine: A Mini-Review. *Frontiers in Physiology*, *12*(August). https://doi.org/10.3389/fphys.2021.723510

Ferkingstad, E., Sulem, P., Atlason, B. A., Sveinbjornsson, G., Magnusson, M. I., Styrmisdottir, E. L., Gunnarsdottir, K., Helgason, A., Oddsson, A., Halldorsson, B. V., Jensson, B. O., Zink, F., Halldorsson, G. H., Masson, G., Arnadottir, G. A., Katrinardottir, H., Juliusson, K., Magnusson, M. K., Magnusson, O. T., … Stefansson, K. (2021). Large-scale integration of the plasma proteome with genetics and disease. *Nature Genetics*, *53*(12), 1712–1721. https://doi.org/10.1038/s41588-021-00978-w

Garg, P. K., Guan, W., Karger, A. B., Steffen, B. T., O'Neal, W., Heckbert, S. R., Michos, E. D., & Tsai, M. Y. (2020). Lp(a) (Lipoprotein [a]) and Risk for Incident Atrial Fibrillation. *Circulation: Arrhythmia and Electrophysiology*, *13*(5), 477–479. https://doi.org/10.1161/CIRCEP.120.008401

Gobeil, É., Maltais-payette, I., Taba, N., Brière, F., Ghodsian, N., Abner, E., Bourgault, J., Gagnon, É., Manikpurage, H. D., Couture, C., Mitchell, P. L., Julien, F., Corbeil, J., & Vohl, M. (2021). Mendelian randomization analysis identifies blood tyrosine

levels as a biomarker of non-alcoholic fatty liver disease. *MedRxiv*. https://doi.org/10.1101/2021.11.26.21266879

Goudswaard, L. J., Bell, J. A., Hughes, D. A., Corbin, L. J., Walter, K., Davey Smith, G., Soranzo, N., Danesh, J., Di Angelantonio, E., Ouwehand, W. H., Watkins, N. A., Roberts, D. J., Butterworth, A. S., Hers, I., & Timpson, N. J. (2021). Effects of adiposity on the human plasma proteome: observational and Mendelian randomisation estimates. *International Journal of Obesity*, *45*(10), 2221–2229. https://doi.org/10.1038/s41366-021-00896-1

Grams, M. E., Surapaneni, A., Chen, J., Zhou, L., Yu, Z., Dutta, D., Welling, P. A., Chatterjee, N., Zhang, J., Arking, D. E., Chen, T. K., Rebholz, C. M., Yu, B., Schlosser, P., Rhee, E. P., Ballantyne, C. M., Boerwinkle, E., Lutsey, P. L., Mosley, T., … Coresh, J. (2021). Proteins associated with risk of kidney function decline in the general population. *Journal of the American Society of Nephrology*, *32*(9), 2291–2302. https://doi.org/10.1681/ASN.2020111607

Hanna, F., Ebeling, P. R., Wang, Y., O'Sullivan, R., Davis, S., Wluka, A. E., & Cicuttini, F. M. (2005). Factors influencing longitudinal change in knee cartilage volume measured from magnetic resonance imaging in healthy men. *Annals of the Rheumatic Diseases*, *64*(7), 1038–1042. https://doi.org/10.1136/ard.2004.029355

Harrison, S., Davies, N. M., Howe, L. D., & Hughes, A. (2021). Testosterone and socioeconomic position: Mendelian randomization in 306,248 men and women in UK Biobank. *Science Advances*, *7*(31), 1–16. https://doi.org/10.1126/sciadv.abf8257

Hartgens, F., & Kuipers, H. (2004). Effects of androgenic-anabolic steroids in athletes.

Sports Medicine, *34*(8), 513–554. https://doi.org/10.2165/00007256-200434080-00003

Imazio, M., Brucato, A., Ferrazzi, P., Rovere, M. E., Gandino, A., Cemin, R., Ferrua, S., Belli, R., Maestroni, S., Simon, C., Zingarelli, E., Barosi, A., Sansone, F., Patrini, D., Vitali, E., Trinchero, R., Spodick, D. H., & Adler, Y. (2011). Colchicine reduces postoperative atrial fibrillation: Results of the Colchicine for the Prevention of the Postpericardiotomy Syndrome (COPPS) atrial fibrillation substudy. *Circulation*, *124*(21), 2290–2295. https://doi.org/10.1161/CIRCULATIONAHA.111.026153

Jiang, Q., Qin, D., Yang, L., Lin, Y., Zhai, L., Zhang, Y., Yang, G., Wang, K., Tong, D., Li, X., Chen, Z., Huang, K., Yu, T., Xiang, X., Cui, C., Cai, C., Shi, J., Li, M., & Chen, M. (2021). Causal effects of plasma lipids on the risk of atrial fibrillation: A multivariable mendelian randomization study. *Nutrition, Metabolism and Cardiovascular Diseases*, *31*(5), 1569–1578. https://doi.org/10.1016/j.numecd.2021.02.011

Khatib, R., Joseph, P., Briel, M., Yusuf, S., & Healey, J. (2013). Blockade of the renin–angiotensin–aldosterone system (RAAS) for primary prevention of non-valvular atrial fibrillation: A systematic review and meta analysis of randomized controlled trials. *International Journal of Cardiology*, *165*(1), 17–24. https://doi.org/10.1016/j.ijcard.2012.02.009

Lakshman, K. M., Bhasin, S., & Araujo, A. B. (2010). Sex hormone-binding globulin as an independent predictor of incident type 2 diabetes mellitus in men. *Journals of Gerontology - Series A Biological Sciences and Medical Sciences*, *65 A*(5), 503–509.

https://doi.org/10.1093/gerona/glq002

Larsson, S. C., Gill, D., Mason, A. M., Jiang, T., Bäck, M., Butterworth, A. S., & Burgess, S. (2020). Lipoprotein(a) in Alzheimer, Atherosclerotic, Cerebrovascular, Thrombotic, and Valvular Disease. *Circulation*, *141*(22), 1826–1828. https://doi.org/10.1161/CIRCULATIONAHA.120.045826

Leinonen, J. T., Mars, N., Lehtonen, L. E., Ahola-Olli, A., Ruotsalainen, S., Lehtimäki, T., Kähönen, M., Raitakari, O., Gen, F., Daly, M., Tuomi, T., Ripatti, S., Pirinen, M., & Tukiainen, T. (2021). Genetic analyses on the health impacts of testosterone highlight effects on female-specific diseases and sex differences. In *medRxiv*.

Loos, R. J. F., & Yeo, G. S. H. (2021). The genetics of obesity: from discovery to biology. *Nature Reviews Genetics*, *0123456789*. https://doi.org/10.1038/s41576-021-00414-z

Ma, H. L., Blanchet, T. J., Peluso, D., Hopkins, B., Morris, E. A., & Glasson, S. S. (2007). Osteoarthritis severity is sex dependent in a surgical mouse model. *Osteoarthritis and Cartilage*, *15*(6), 695–700. https://doi.org/10.1016/j.joca.2006.11.005

Marcovecchio, M. L., Colombo, M., Dalton, R. N., McKeigue, P. M., Benitez-Aguirre, P., Cameron, F. J., Chiesa, S. T., Couper, J. J., Craig, M. E., Daneman, D., Davis, E. A., Deanfield, J. E., Donaghue, K. C., Jones, T. W., Mahmud, F. H., Marshall, S. M., Neil, A., Colhoun, H. M., & Dunger, D. B. (2020). Biomarkers associated with early stages of kidney disease in adolescents with type 1 diabetes. *Pediatric Diabetes*, *May*, 1322–1332. https://doi.org/10.1111/pedi.13095

Martin, A. R., Kanai, M., Kamatani, Y., Okada, Y., Neale, B. M., & Daly, M. J. (2019). Clinical use of current polygenic risk scores may exacerbate health disparities. *Nature Genetics*, *51*(4), 584–591. https://doi.org/10.1038/s41588-019-0379-x

Neefs, J., van den Berg, N. W. E., Limpens, J., Berger, W. R., Boekholdt, S. M., Sanders, P., & de Groot, J. R. (2017). Aldosterone Pathway Blockade to Prevent Atrial Fibrillation: A Systematic Review and Meta-Analysis. *International Journal of Cardiology*, *231*, 155–161. https://doi.org/10.1016/j.ijcard.2016.12.029

Paré, G., Çaku, A., McQueen, M., Anand, S. S., Enas, E., Clarke, R., Boffa, M. B., Koschinsky, M., Wang, X., & Yusuf, S. (2019). Lipoprotein(a) Levels and the Risk of Myocardial Infarction Among 7 Ethnic Groups. *Circulation*, *139*(12), 1472–1482. https://doi.org/10.1161/CIRCULATIONAHA.118.034311

Pietzner, M., Wheeler, E., Carrasco-Zanini, J., Kerrison, N. D., Oerton, E., Koprulu, M., Luan, J., Hingorani, A. D., Williams, S. A., Wareham, N. J., & Langenberg, C. (2021). Synergistic insights into human health from aptamer- and antibody-based proteomic profiling. *Nature Communications*, *12*(1). https://doi.org/10.1038/s41467-021-27164-0

Porcu, E., Sadler, M. C., Lepik, K., Auwerx, C., Wood, A. R., Weihs, A., Sleiman, M. S. B., Ribeiro, D. M., Bandinelli, S., Tanaka, T., Nauck, M., Völker, U., Delaneau, O., Metspalu, A., Teumer, A., Frayling, T., Santoni, F. A., Reymond, A., & Kutalik, Z. (2021). Differentially expressed genes reflect disease-induced rather than disease-causing changes in the transcriptome. *Nature Communications*, *12*(1), 1–9. https://doi.org/10.1038/s41467-021-25805-y

Pretorius, M., Donahue, B. S., Yu, C., Greelish, J. P., Roden, D. M., & Brown, N. J. (2007). Plasminogen activator inhibitor-1 as a predictor of postoperative atrial fibrillation after cardiopulmonary bypass. *Circulation*, *116*(11 SUPPL. 1), 1–7. https://doi.org/10.1161/CIRCULATIONAHA.106.677906

Reckelhoff, J. F., Zhang, H., & Granger, J. P. (1998). Testosterone Exacerbates Hypertension and Reduces Pressure-Natriuresis in Male Spontaneously Hypertensive Rats. *Hypertension*, *31*(1), 435–439. https://doi.org/10.1161/01.HYP.31.1.435

Roberts, T. C., Langer, R., & Wood, M. J. A. (2020). Advances in oligonucleotide drug delivery. *Nature Reviews Drug Discovery*, *19*(10), 673–694. https://doi.org/10.1038/s41573-020-0075-7

Rosner, W., Hryb, D. J., Kahn, S. M., Nakhla, A. M., & Romas, N. A. (2010). Interactions of sex hormone-binding globulin with target cells. *Molecular and Cellular Endocrinology*, *316*(1), 79–85. https://doi.org/10.1016/j.mce.2009.08.009

Ruth, K. S., Day, F. R., Tyrrell, J., Thompson, D. J., Wood, A. R., Mahajan, A., Beaumont, R. N., Wittemans, L., Martin, S., Busch, A. S., Erzurumluoglu, A. M., Hollis, B., O'Mara, T. A., McCarthy, M. I., Langenberg, C., Easton, D. F., Wareham, N. J., Burgess, S., Murray, A., … Perry, J. R. B. (2020). Using human genetics to understand the disease impacts of testosterone in men and women. *Nature Medicine*, *26*(2), 252–258. https://doi.org/10.1038/s41591-020-0751-5

Schneider, M. P., Hua, T. A., Böhm, M., Wachtell, K., Kjeldsen, S. E., & Schmieder, R. E. (2010). Prevention of Atrial Fibrillation by Renin-Angiotensin System Inhibition.

Journal of the American College of Cardiology, *55*(21), 2299–2307. https://doi.org/10.1016/j.jacc.2010.01.043

Snyder, P. J., Bhasin, S., Cunningham, G. R., Matsumoto, A. M., Stephens-Shields, A. J., Cauley, J. A., Gill, T. M., Barrett-Connor, E., Swerdloff, R. S., Wang, C., Ensrud, K. E., Lewis, C. E., Farrar, J. T., Cella, D., Rosen, R. C., Pahor, M., Crandall, J. P., Molitch, M. E., Resnick, S. M., … Ellenberg, S. S. (2018). Lessons from the Testosterone Trials. *Endocrine Reviews*, *39*(3), 369–386. https://doi.org/10.1210/er.2017-00234

Sofianopoulou, E., Kaptoge, S. K., Afzal, S., Jiang, T., Gill, D., Gundersen, T. E., Bolton, T. R., Allara, E., Arnold, M. G., Mason, A. M., Chung, R., Pennells, L. A. M., Shi, F., Sun, L., Willeit, P., Forouhi, N. G., Langenberg, C., Sharp, S. J., Panico, S., … Burgess, S. (2021). Estimating dose-response relationships for vitamin D with coronary heart disease, stroke, and all-cause mortality: observational and Mendelian randomisation analyses. *The Lancet Diabetes and Endocrinology*, *9*(12), 837–846. https://doi.org/10.1016/S2213-8587(21)00263-1

Spronk, H. M. H., De Jong, A. M., Verheule, S., De Boer, H. C., Maass, A. H., Lau, D. H., Rienstra, M., Van Hunnik, A., Kuiper, M., Lumeij, S., Zeemering, S., Linz, D., Kamphuisen, P. W., Ten Cate, H., Crijns, H. J., Van Gelder, I. C., Van Zonneveld, A. J., & Schotten, U. (2017). Hypercoagulability causes atrial fibrosis and promotes atrial fibrillation. *European Heart Journal*, *38*(1), 38–50. https://doi.org/10.1093/eurheartj/ehw119

Staley, J. R., & Burgess, S. (2017). Semiparametric methods for estimation of a nonlinear

exposure-outcome relationship using instrumental variables with application to Mendelian randomization. *Genetic Epidemiology*, *41*(4), 341–352. https://doi.org/10.1002/gepi.22041

Stanzick, K. J., Li, Y., Schlosser, P., Gorski, M., Wuttke, M., Thomas, L. F., Rasheed, H., Rowan, B. X., Graham, S. E., Vanderweff, B. R., Patil, S. B., Robinson-Cohen, C., Gaziano, J. M., O'Donnell, C. J., Willer, C. J., Hallan, S., Åsvold, B. O., Gessner, A., Hung, A. M., … Winkler, T. W. (2021). Discovery and prioritization of variants and genes for kidney function in >1.2 million individuals. *Nature Communications*, *12*(1), 1–17. https://doi.org/10.1038/s41467-021-24491-0

Stiekema, L. C. A., Prange, K. H. M., Hoogeveen, R. M., Verweij, S. L., Kroon, J., Schnitzler, J. G., Dzobo, K. E., Cupido, A. J., Tsimikas, S., Stroes, E. S. G., de Winther, M. P. J., & Bahjat, M. (2020). Potent lipoprotein(a) lowering following apolipoprotein(a) antisense treatment reduces the pro-inflammatory activation of circulating monocytes in patients with elevated lipoprotein(a). *European Heart Journal*, *41*(24), 2262–2271. https://doi.org/10.1093/eurheartj/ehaa171

Sun, Y.-Q., Burgess, S., Staley, J. R., Wood, A. M., Bell, S., Kaptoge, S. K., Guo, Q., Bolton, T. R., Mason, A. M., Butterworth, A. S., Di Angelantonio, E., Vie, G. Å., Bjørngaard, J. H., Kinge, J. M., Chen, Y., & Mai, X.-M. (2019). Body mass index and all cause mortality in HUNT and UK Biobank studies: linear and non-linear mendelian randomisation analyses. *BMJ*, *364*, l1042. https://doi.org/10.1136/bmj.l1042

Thanassoulis, G., Campbell, C. Y., Owens, D. S., Smith, J. G., Smith, A. V., Peloso, G.

M., Kerr, K. F., Pechlivanis, S., Budoff, M. J., Harris, T. B., Malhotra, R., O'Brien, K. D., Kamstrup, P. R., Nordestgaard, B. G., Tybjaerg-Hansen, A., Allison, M. A., Aspelund, T., Criqui, M. H., Heckbert, S. R., … Post, W. S. (2013). Genetic Associations with Valvular Calcification and Aortic Stenosis. *New England Journal of Medicine*, *368*(6), 503–512. https://doi.org/10.1056/NEJMoa1109034

Trigunaite, A., Dimo, J., & Jørgensen, T. N. (2015). Suppressive effects of androgens on the immune system. *Cellular Immunology*, *294*(2), 87–94. https://doi.org/10.1016/j.cellimm.2015.02.004

Tsimikas, S., Karwatowska-Prokopczuk, E., Gouni-Berthold, I., Tardif, J.-C., Baum, S. J., Steinhagen-Thiessen, E., Shapiro, M. D., Stroes, E. S., Moriarty, P. M., Nordestgaard, B. G., Xia, S., Guerriero, J., Viney, N. J., O'Dea, L., & Witztum, J. L. (2020). Lipoprotein(a) Reduction in Persons with Cardiovascular Disease. *New England Journal of Medicine*, *382*(3), 244–255. https://doi.org/10.1056/nejmoa1905239

Vikan, T., Schirmer, H., Njølstad, I., & Svartberg, J. (2010). Low testosterone and sex hormone-binding globulin levels and high estradiol levels are independent predictors of type 2 diabetes in men. *European Journal of Endocrinology*, *162*(4), 747–754. https://doi.org/10.1530/EJE-09-0943

Zaghlool, S. B., Sharma, S., Molnar, M., Matías-García, P. R., Elhadad, M. A., Waldenberger, M., Peters, A., Rathmann, W., Graumann, J., Gieger, C., Grallert, H., & Suhre, K. (2021). Revealing the role of the human blood plasma proteome in obesity using genetic drivers. *Nature Communications*, *12*(1).

https://doi.org/10.1038/s41467-021-21542-4

Zhao, Z., Ukidve, A., Kim, J., & Mitragotri, S. (2020). Targeting Strategies for Tissue-Specific Drug Delivery. *Cell*, *181*(1), 151–167. https://doi.org/10.1016/j.cell.2020.02.001

Zheng, K. H., Tsimikas, S., Pawade, T., Kroon, J., Jenkins, W. S. A., Doris, M. K., White, A. C., Timmers, N. K. L. M., Hjortnaes, J., Rogers, M. A., Aikawa, E., Arsenault, B. J., Witztum, J. L., Newby, D. E., Koschinsky, M. L., Fayad, Z. A., Stroes, E. S. G., Boekholdt, S. M., & Dweck, M. R. (2019). Lipoprotein(a) and Oxidized Phospholipids Promote Valve Calcification in Patients With Aortic Stenosis. *Journal of the American College of Cardiology*, *73*(17), 2150–2162. https://doi.org/10.1016/j.jacc.2019.01.070

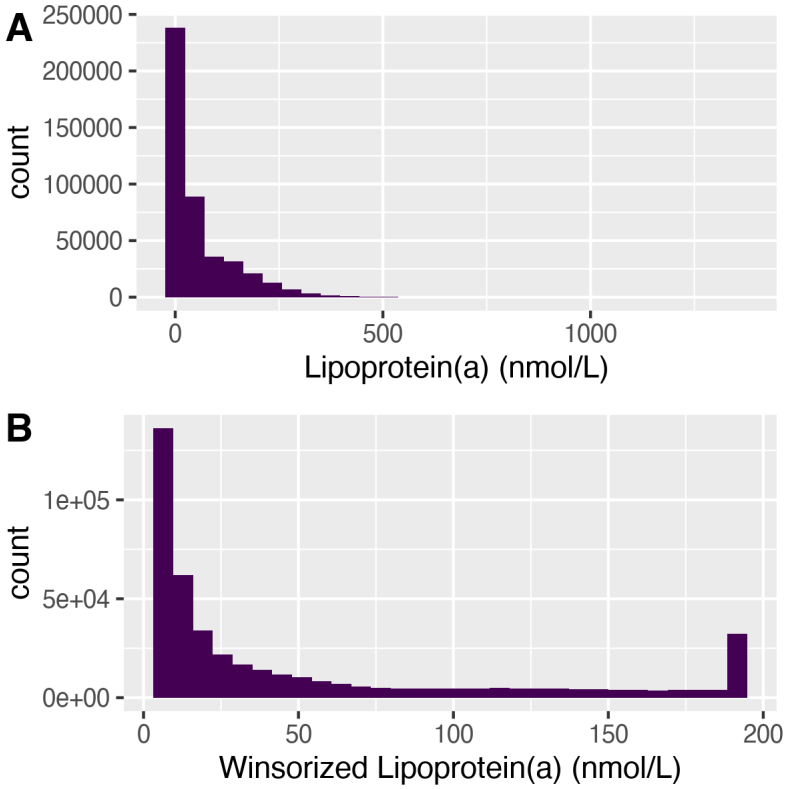
Zhou, A., Selvanayagam, J. B., & Hyppönen, E. (2021). Non-linear Mendelian randomization analyses support a role for vitamin D deficiency in cardiovascular disease risk. *European Heart Journal*, 1–10. https://doi.org/10.1093/eurheartj/ehab809

Zhou, W., Kanai, M., Wu, K. H., Humaira, R., Tsuo, K., & Hirbo, J. B. (2021). Global Biobank Meta-analysis Initiative: powering genetic discovery across human diseases. *MedRxiv*. https://doi.org/10.1101/2021.11.19.21266436

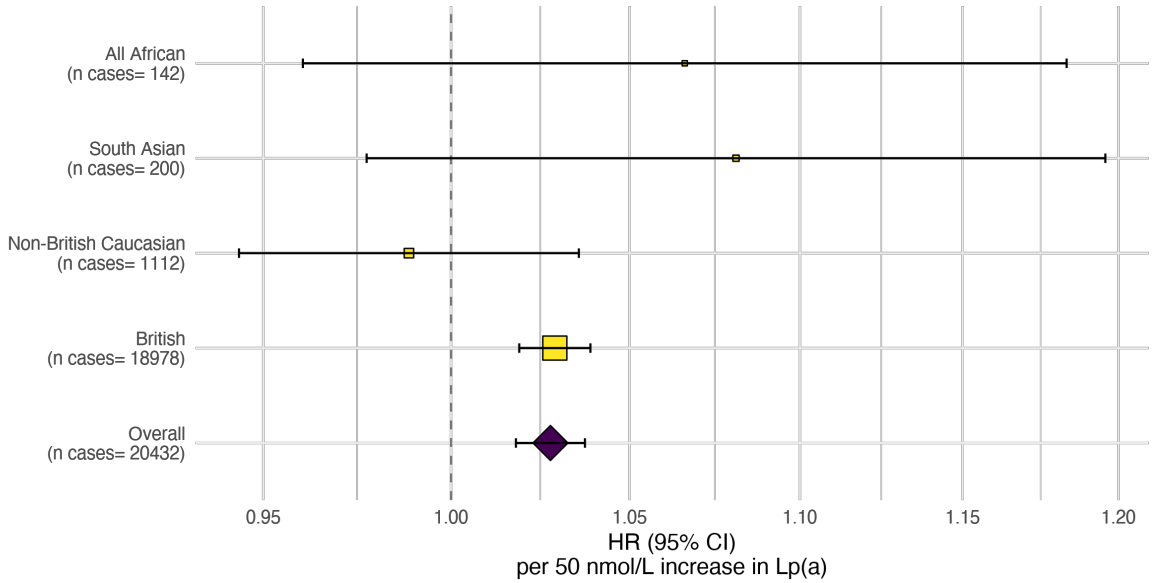
APPENDIX A:

SUPPLEMENTARY DATA FOR CHAPTER 3

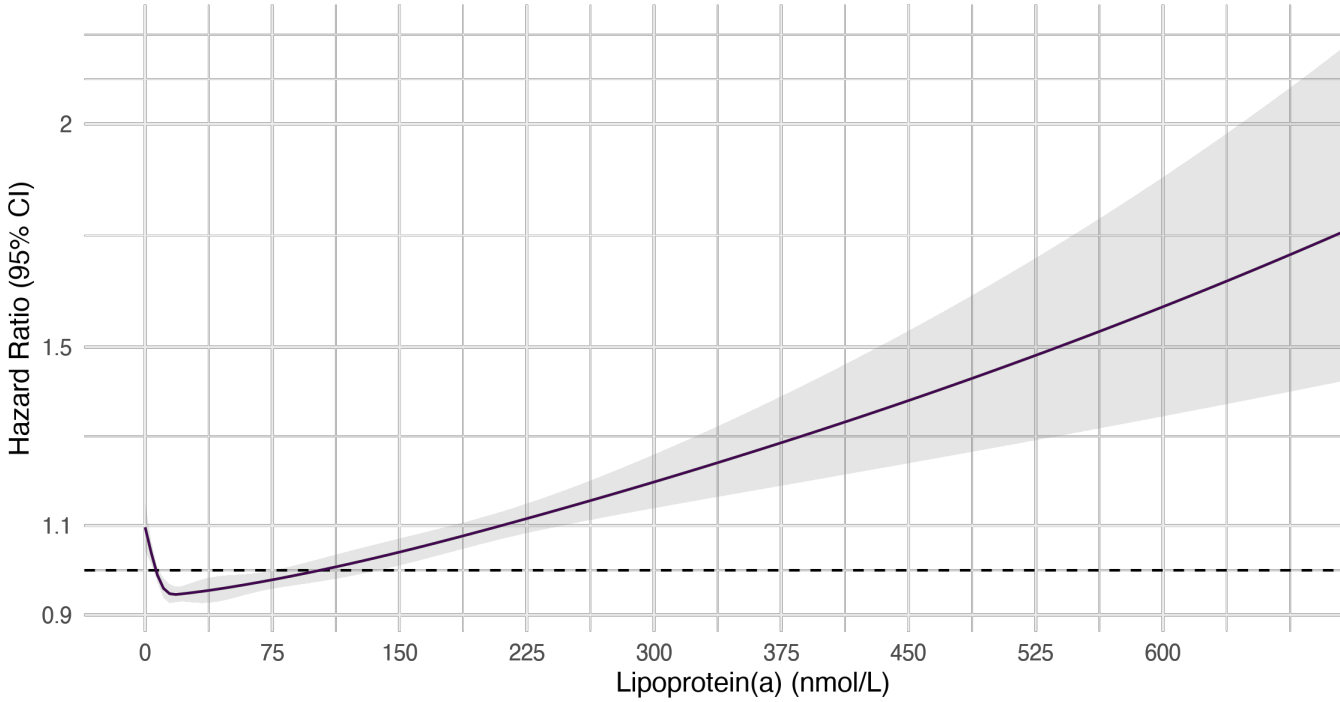
Supplemental Figures



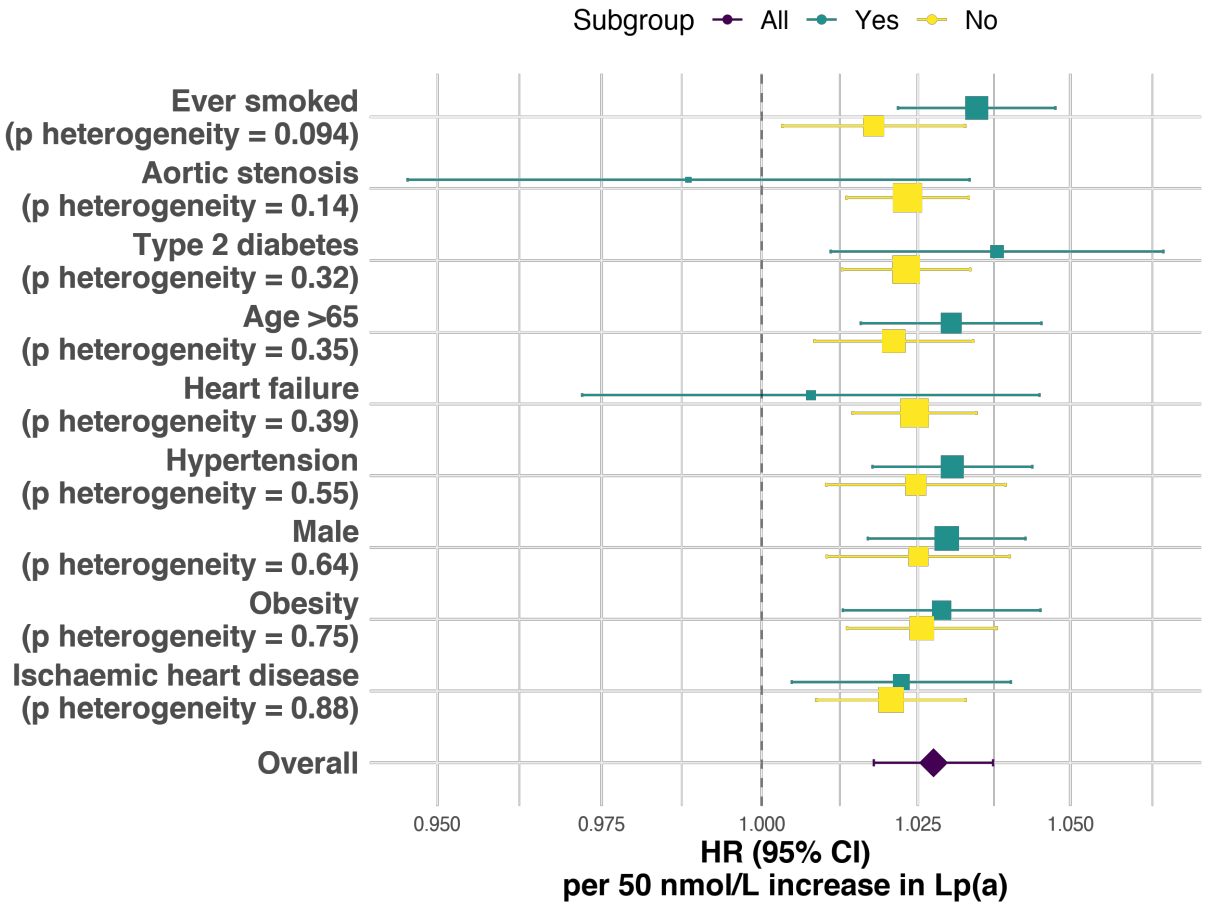
Supplemental Figure 1. Distribution of lipoprotein(a) levels from the UK Biobank cohort. *A) Raw lipoprotein(a) levels B) After winsorizing detectable values outside reportable range*



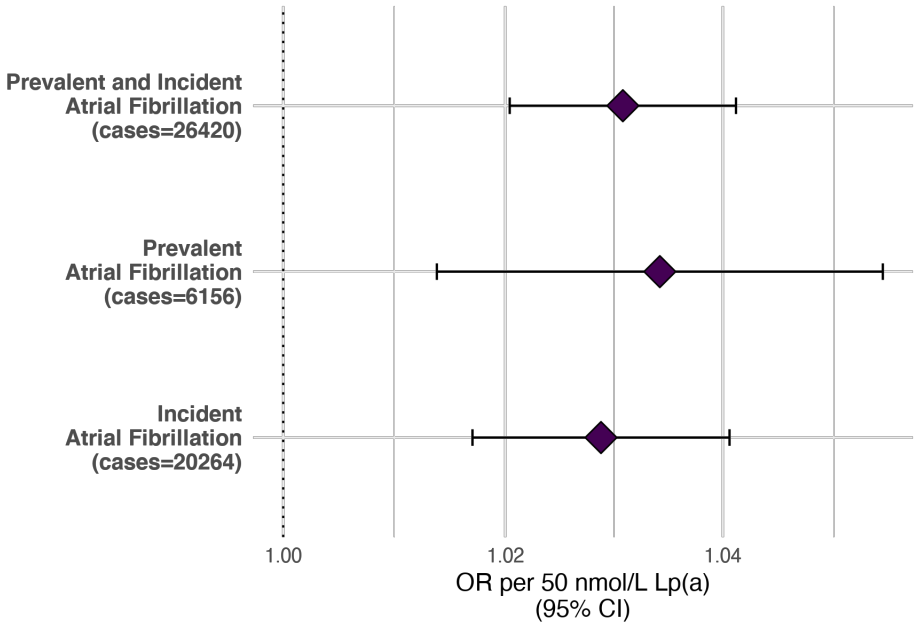
Supplemental Figure 2. Association between Lp(a) concentration at recruitment and incident atrial fibrillation events in UK Biobank participants. *CI, confidence interval; HR, hazard ratio; Lp(a), lipoprotein(a)*



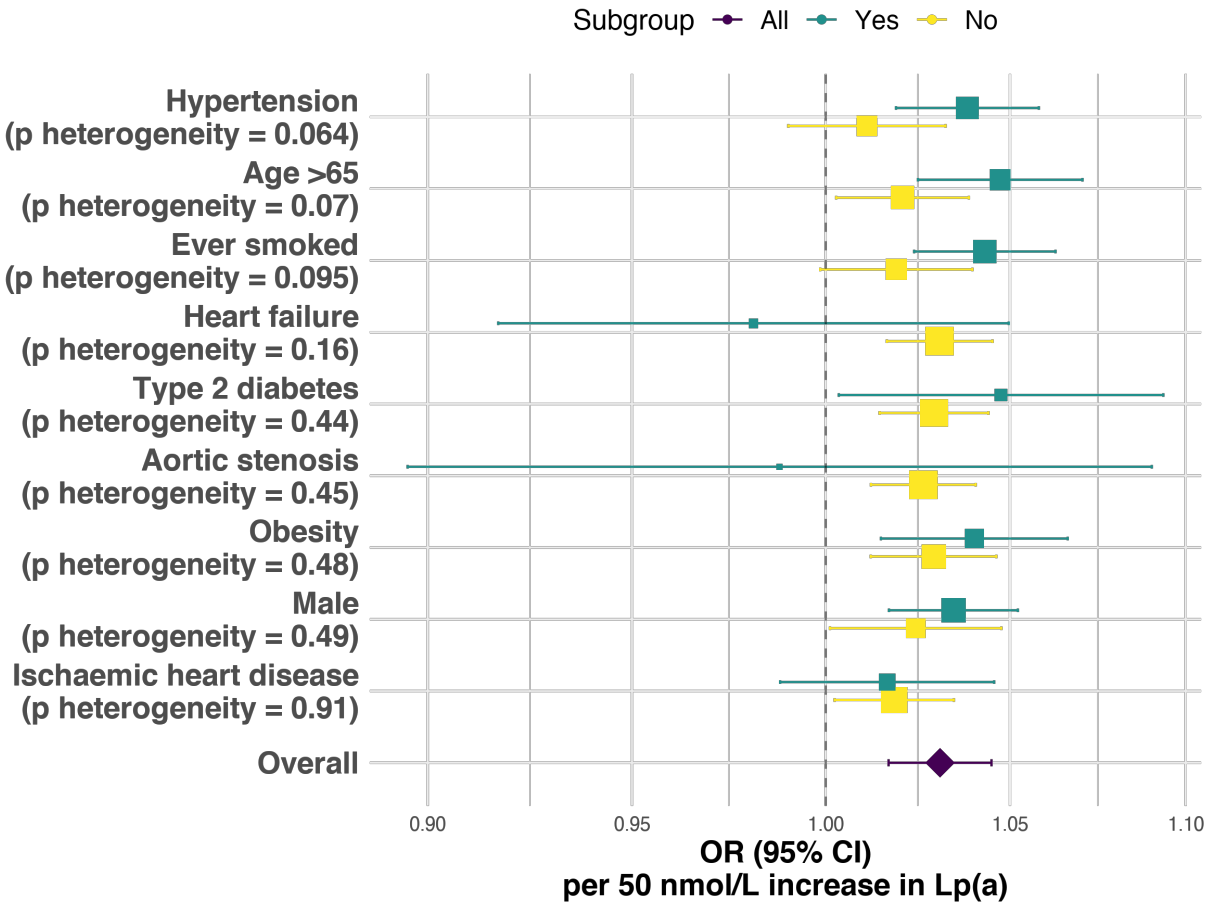
Supplemental Figure 3. Dose-response relationship between lipoprotein(a) and risk of incident atrial fibrillation in UK Biobank.



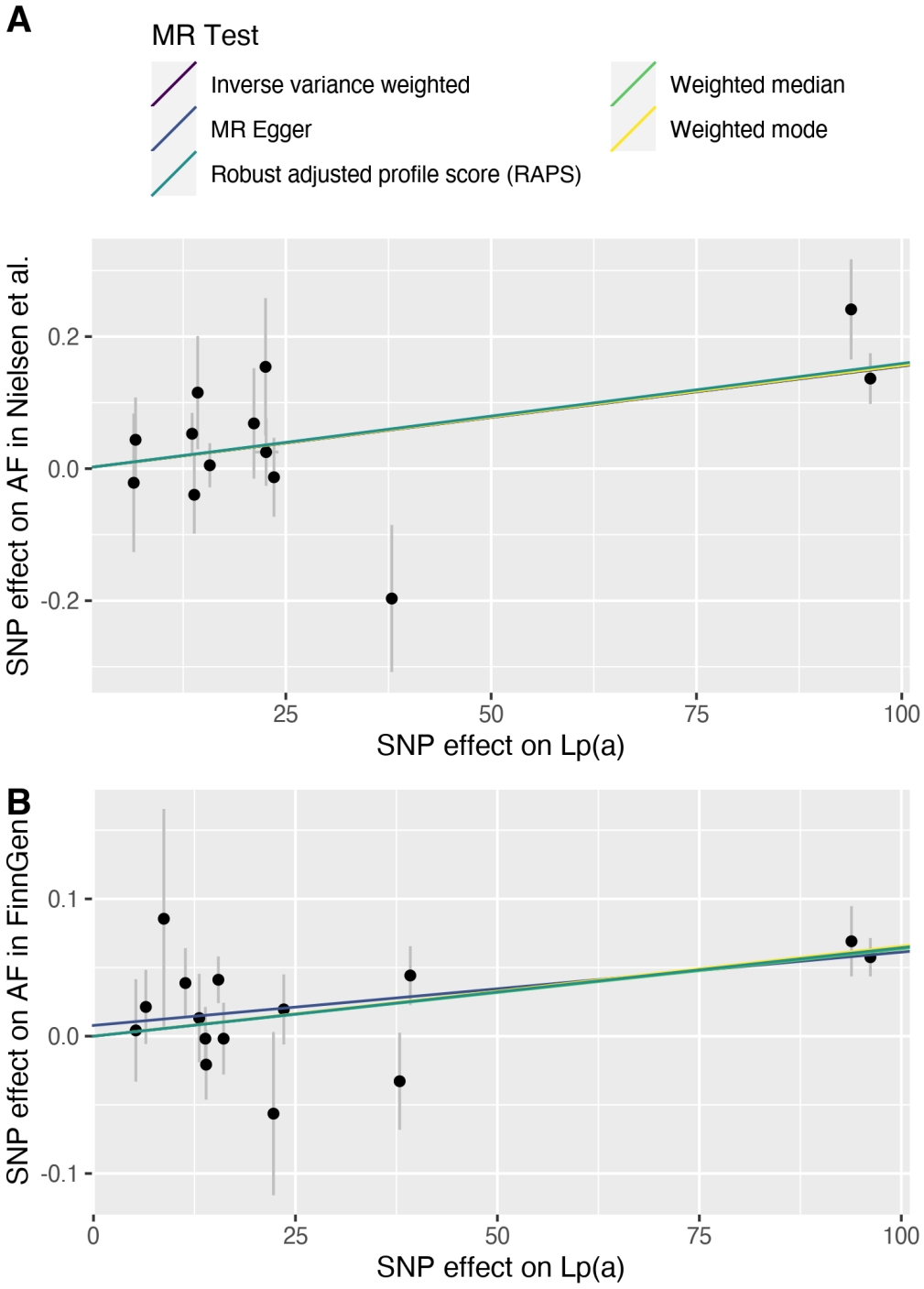
Supplemental Figure 4. No evidence of effect modification of lipoprotein(a) and atrial fibrillation in UK Biobank stratified according to common risk factors.



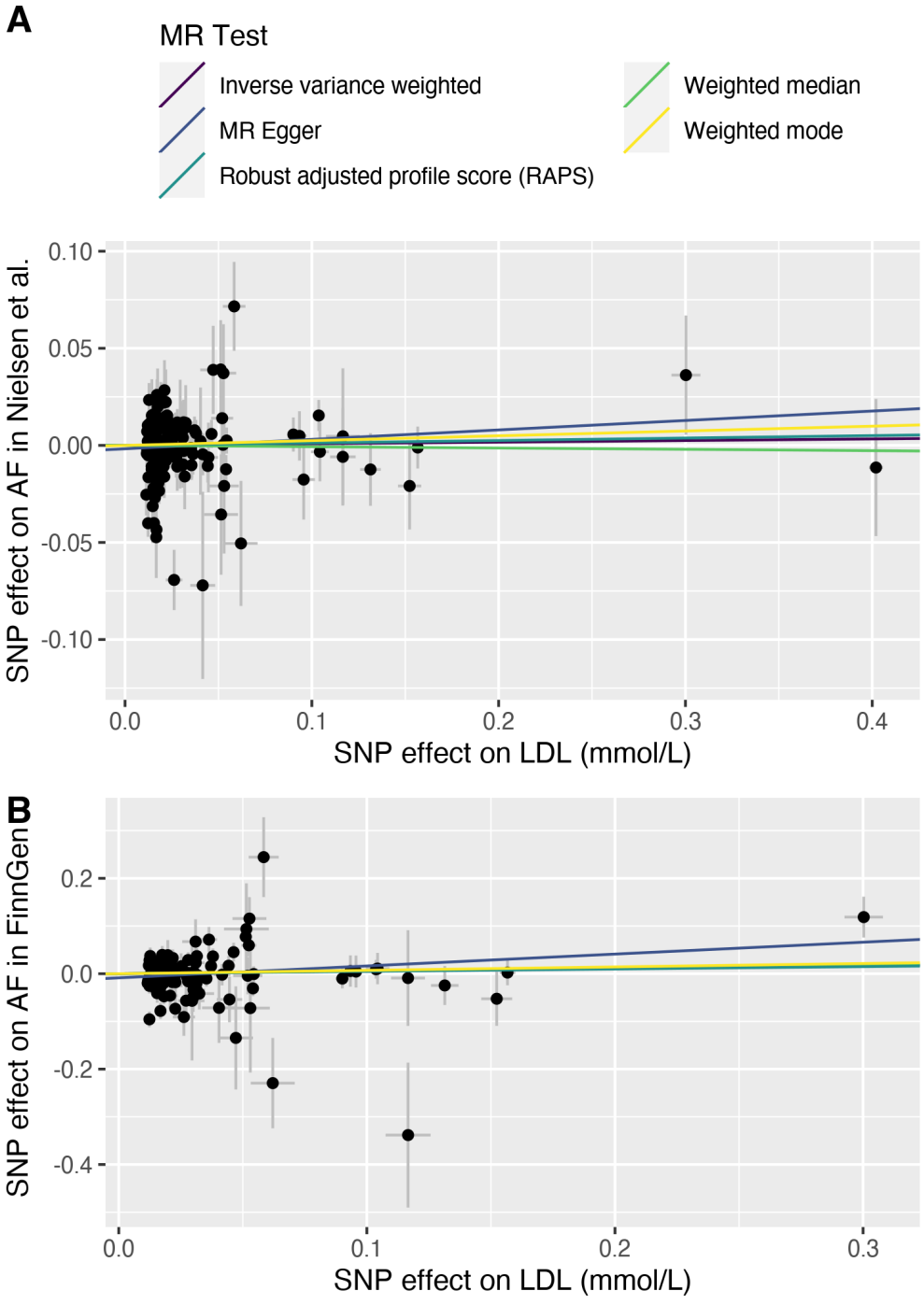
Supplemental Figure 5. Consistent risk-conferring effect of genetically predicted lipoprotein(a) on prevalent or incident atrial fibrillation in UK Biobank.



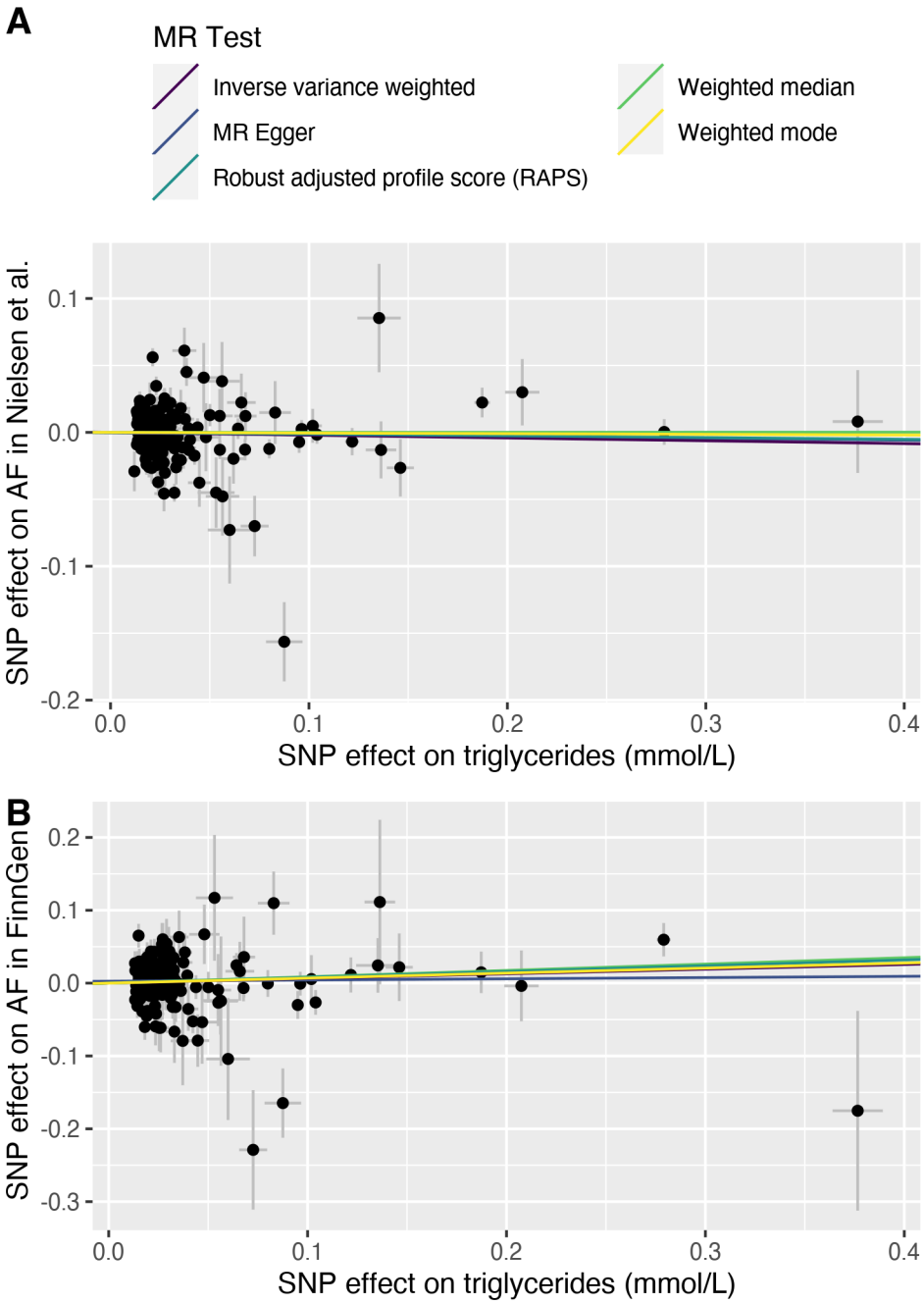
Supplemental Figure 6. No evidence of effect modification of genetically predicted lipoprotein(a) and atrial fibrillation in UK Biobank stratified according to common risk factors.



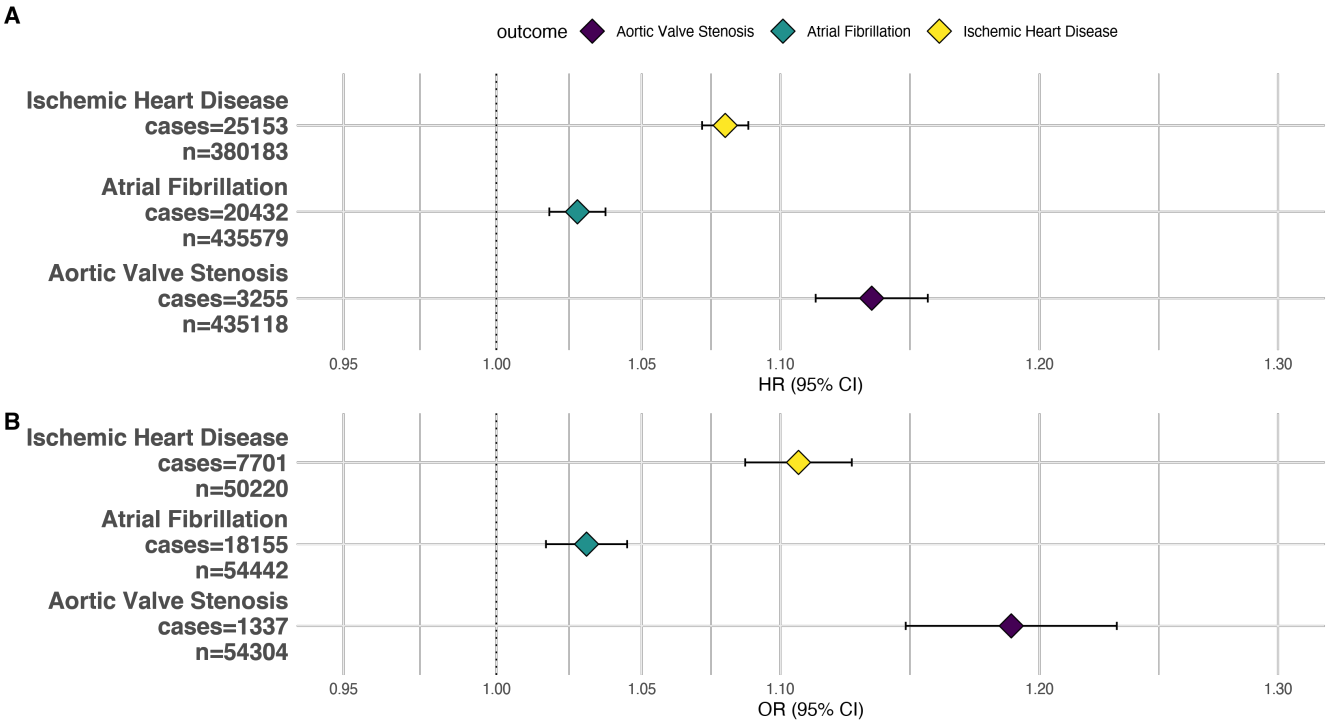
Supplemental Figure 7. Comparison of effect of lipoprotein(a) on atrial fibrillation using Mendelian randomization with various Mendelian randomization methods. *A) Using Nielsen et al. GWAS summary statistics for atrial fibrillation. B) Using FinnGen consortium GWAS summary statistics for atrial fibrillation.*



Supplemental Figure 8. Comparison of effect of LDL cholesterol on atrial fibrillation using Mendelian randomization with various Mendelian randomization methods. *A) Using Nielsen et al. GWAS summary statistics for atrial fibrillation. B) Using FinnGen consortium GWAS summary statistics for atrial fibrillation.*



Supplemental Figure 9. Comparison of effect of triglycerides on atrial fibrillation using Mendelian randomization with various Mendelian randomization methods. *A) Using Nielsen et al. GWAS summary statistics for atrial fibrillation. B) Using FinnGen consortium GWAS summary statistics for atrial fibrillation.*



Supplemental Figure 10. Comparison of risk estimates for lipoprotein(a) with incident aortic valve stenosis, ischemic heart disease, and atrial fibrillation in UK Biobank.

Supplemental Tables

Supplemental Table 1. Definitions for risk factors and other variables used in analyses.

Trait	Field ID	ICD10 codes	Notes
Lipoprotein(a)	30790		Participants above or below reportable range were defined by field ID 30796
Ischaemic heart disease	41270; 40001; 40002	I20- to I25-	
Aortic valve stenosis	41270; 40001; 40002	I35.0; I35.2	
Type 2 diabetes	41270; 40001; 40002	E11-	
Heart Failure	41270; 40001; 40002	I50-	
Age at recruitment	21022		
Sex	31		
Townsend deprivation index	189		
Income	738		"Do not know" and "prefer not to answer" were coded as missing
BMI	21001		Obesity was defined as BMI >30
Height	50		
Physical activity	22040		MET minutes for all activities in week
Systolic blood pressure	4080		Average of two consecutive measures for blood pressure; Hypertension was defined as SBP > 140 mmHg
Diastolic blood pressure	4079		Average of two consecutive measures for blood pressure
Smoking Status	20116		Ever smokers were defined as any participants that reported "current" or "previous" smoking status
Alcoholic drinks	1568; 1578; 1588; 1598; 1608; 5364		Number of alcoholic drinks per week were calculated as sum of all drinks per week across red wine, champagne plus white wine, beer plus cider, spirits, fortified wine, and other alcohols
Total cholesterol	30690		
HDL cholesterol	30760		
LDL cholesterol	30780		
Triglycerides	30870		
Lipid-lowering medication	6153; 6157		
Anti-hypertensive medication	6153; 6157		

Supplemental Table 2 - Characteristics at recruitment for population of participants with Lp(a) and without prevalent AF from UK Biobank cohort study.	
UK Biobank (N=435579)	**Mean (± SD) or N (%)**
Age at recruitment (years)	57 (8)
Sex (male)	198223 (45.5%)
Ethnicity	
African	*6833 (2%)*
British	*395497 (91%)*
Non-British Caucasian	*26350 (6%)*
South Asian	*6899 (2%)*
Lipoprotein(a) (nmol/L)	55.93 (76.14)
Lipoprotein(a) winsorized (nmol/L)	*50.76 (60.33)*
Incident atrial fibrillation	20432 (4.7%)
Prevalent ischemic heart disease	16834 (3.9%)
Prevalent type 2 diabetes	8146 (1.9%)
Body mass index (kg/m^2)	27.4 (4.8)
Systolic blood pressure (mmHg)	138 (19)
Diastolic blood pressure (mmHg)	82 (10)
Ever smoked	196496 (45.1%)
Townsend Deprivation Index	-1.37 (3.05)
Alcohol consumption (# drinks per week)	2.2 (2.0)

Supplemental Table 3 - Association of Lp(a) at recruitment with incident AF events in UK Biobank.						
Ethnicity	**OR per 50 nmol/L increased Lp(a)**	**95% CI**	**P-value**	**cases**	**controls**	**n**
Overall	1.03	(1.02 to 1.04)	1.65E-08	20432	415147	435579
British	1.03	(1.02 to 1.04)	1.07E-08	18978	376519	395497
Non-British Caucasian	0.99	(0.94 to 1.04)	0.63	1112	25238	26350
African	1.07	(0.96 to 1.18)	0.23	142	6691	6833
South Asian	1.08	(0.98 to 1.20)	0.13	200	6699	6899

Supplemental Table 4 - Association of interaction terms for quantitative risk factors and Lp(a) with incident AF in UK Biobank.	
Interaction Term	**P-value for interaction**
Age	0.424
BMI	0.569
Diastolic blood pressure	0.197
Systolic blood pressure	0.367
Weekly alcohol consumption	0.476

Supplemental Table 5 - Association of each Lp(a) genetic score with observed Lp(a) and incident AF in testing set from UK Biobank (n=54062).					
cis window	**R2 for observed Lp(a) using genetic score**	**OR* for incident AF (95% CI)**	**p-value**	**cases**	**controls**
50Kb	0.680	1.03 (1.02 to 1.04)	5.36E-05	18155	36287
500Kb	0.714	1.03 (1.02 to 1.05)	1.33E-05	18155	36287
whole genome	0.728	1.03 (1.02 to 1.04)	2.21E-04	18155	36287

* per 50 nmol/L increase in Lp(a) genetic score

Supplemental Table 6 - Association of interaction terms for quantitative risk factors and Lp(a) genetic score with incident AF in UK Biobank.	
Interaction Term	**P-value for interaction**
Age	0.302
BMI	0.505
Diastolic blood pressure	0.507
Systolic blood pressure	0.529
Weekly alcohol consumption	0.697

Supplemental Table 7 - Association of independent genetic variants with lipoprotein(a) in UK Biobank training set within 500Kb of *LPA* gene.								
Chr	**Position (hg19)**	**SNP**	**EA**	**OA**	**EAF**	**Beta (nmol/L)**	**Standard Error**	**-log10(p-value)**
6	160578069	rs146534110	G	T	0.987	-37.90	0.665	708.157
6	160628128	rs117648937	C	T	0.988	13.09	0.722	72.7517
6	160644552	rs494554	C	G	0.972	13.85	0.474	187.279
6	160762506	rs149210101	C	A	0.979	13.94	0.549	141.632
6	160779981	rs76596562	G	A	0.974	11.38	0.472	127.447
6	160922870	rs117733303	A	G	0.981	-93.83	0.562	6053.64
6	160939982	rs71565772	T	C	0.955	15.46	0.366	389.482
6	161010118	rs10455872	A	G	0.919	-96.17	0.278	25933.8
6	161017363	rs73596816	G	A	0.967	-39.22	0.423	1871.61
6	161062118	rs569944069	G	A	0.984	22.28	0.615	286.824
6	161080457	rs117857195	G	T	0.975	23.56	0.527	435.14
6	161104060	rs565310232	C	T	0.984	8.71	0.669	38.0536
6	161290860	rs143368848	A	G	0.980	16.10	0.571	174.153
6	161314975	rs143292133	G	C	0.981	-6.48	0.565	29.777
6	161447465	rs113304208	T	C	0.992	5.24	0.834	9.4866

Chr, chromosome; EA, effect allele; EAF, effect allele frequency; OA, other allele; SNP, single nucleotide polymorphism

Supplemental Table 8 - Effect of lipoprotein(a) on atrial fibrillation using Mendelian randomization in Nielsen et al. and FinnGen cohorts.

				MR-PRESSO			Heterogeneity		MR Egger	
method	**nsnp**	**OR (95%CI)**	**P-value**	**Global P-value**	**Distortion P-value**	**Outliers**	**Cochran's Q**	**P-value**	**Intercept (SE)**	**P-value**
50Kb - Nielsen et al.										
IVW	6	1.03 (1.02 to 1.04)	2.31E-06	NA			1.13	0.95	NA	
MR Egger	6	1.04 (1.01 to 1.06)	3.36E-02	NA			NA		-0.0152 (0.0177)	0.44
Weighted median	6	1.03 (1.02 to 1.04)	8.89E-06	NA			NA		NA	
Weighted mode	6	1.03 (1.02 to 1.04)	5.44E-03	NA			NA		NA	
RAPS	6	1.03 (1.02 to 1.04)	4.15E-06	NA			NA		NA	
MR PRESSO	6	NA	NA	0.96	NA	NA	NA		NA	
50Kb - FinnGen										
IVW	8	1.08 (1.05 to 1.12)	2.66E-06	NA			4.73	0.69	NA	
MR Egger	8	1.1 (1.04 to 1.16)	1.20E-02	NA			NA		-0.0258 (0.0313)	0.44
Weighted median	8	1.08 (1.04 to 1.12)	5.46E-05	NA			NA		NA	
Weighted mode	8	1.08 (1.04 to 1.12)	4.57E-03	NA			NA		NA	
RAPS	8	1.08	4.73E-06	NA			NA		NA	

		(1.05 to 1.12)								
MR PRESSO	8	NA	NA	0.63	NA	NA	NA		8	
500Kb - Nielsen et al.										
IVW	15	1.03 (1.02 to 1.05)	9.93E-08	NA			13.16	0.51	NA	
MR Egger	15	1.03 (1.01 to 1.05)	1.10E-02	NA			NA		0.0079 (0.0095)	0.42
Weighted median	15	1.03 (1.02 to 1.05)	1.16E-06	NA			NA		NA	
Weighted mode	15	1.03 (1.02 to 1.04)	4.25E-05	NA			NA		NA	
RAPS	15	1.03 (1.02 to 1.05)	2.07E-07	NA			NA		NA	
MR PRESSO	15	NA	NA	0.44	NA	NA	NA		NA	
500Kb - FinnGen										
IVW	13	1.08 (1.04 to 1.12)	9.54E-06	NA			13.14	0.36	NA	
MR Egger	13	1.08 (1.03 to 1.14)	1.21E-02	NA			NA		0.0002 (0.023)	0.99
Weighted median	13	1.08 (1.04 to 1.12)	1.03E-05	NA			NA		NA	
Weighted mode	13	1.08 (1.05 to 1.12)	6.23E-04	NA			NA		NA	
RAPS	13	1.08 (1.05 to 1.12)	3.71E-06	NA			NA		NA	

MR PRESSO	13	NA	NA	0.44	NA	NA	NA		13	
Whole genome - Nielsen et al.										
IVW	40	1.03 (1.02 to 1.05)	3.65E-06	NA			51.24	0.09	NA	
MR Egger	40	1.03 (1.02 to 1.05)	8.22E-05	NA			NA		-0.0001 (0.0021)	0.98
Weighted median	40	1.03 (1.02 to 1.05)	1.02E-06	NA			NA		NA	
Weighted mode	40	1.03 (1.02 to 1.05)	1.62E-05	NA			NA		NA	
RAPS	40	1.03 (1.02 to 1.05)	5.89E-07	NA			NA		NA	
MR PRESSO	40	NA	NA	0.17	NA	NA	NA		NA	
Whole genome - FinnGen										
IVW	44	1.2 (1.11 to 1.3)	1.42E-05	NA			52.22	0.08	NA	
MR Egger	40	1.08 (1.04 to 1.12)	3.90E-05	NA			NA		0.0017 (0.005)	0.74
Weighted median	40	1.08 (1.04 to 1.12)	4.99E-04	NA			NA		NA	
Weighted mode	40	1.08 (1.04 to 1.12)	1.56E-04	NA			NA		NA	
RAPS	40	1.07 (1.03 to 1.11)	3.97E-04	NA			NA		NA	
MR PRESSO	40	NA	NA	0.13	NA	NA	NA		NA	

Clarke et al. – Nielsen et al.						
IVW	2	1.03 (1.02 to 1.04)	3.74E-07	NA	0.22	NA
RAPS	2	1.03 (1.02 to 1.04)	7.33E-07	NA	NA	NA
Clarke et al. – FinnGen						
IVW	2	1.07 (1.04 to 1.10)	4.36E-06	NA	0.27	NA
RAPS	2	1.07 (1.04 to 1.10)	7.58E-06	NA	NA	NA

IVW, inverse variance weighted; nsnp, number of single nucleotide polymorphisms; RAPS, robust adjusted profile score

Supplemental Table 9 - Effect of atrial fibrillation on lipoprotein(a) using Mendelian randomization.

				MR-PRESSO			Heterogeneity		Egger	
method	**nsnp**	**Effect (50 mol/L) (95%CI)**	**P-value**	**Global P-value**	**Distortion P-value**	**Outliers**	**Cochran's Q**	**P-value**	**Intercept (SE)**	**P-value**
IVW	112	-0.195 (-0.587 to 0.197)	0.33	NA			152.26	3.95E-03	NA	NA
MR Egger	112	0.108 (-0.747 to 0.963)	0.80	NA			NA	NA	-0.0249 (0.0319)	0.44
Weighted median	112	-0.049 (-0.591 to 0.493)	0.86	NA			NA	NA	NA	NA
Weighted mode	112	-0.078 (-0.715 to 0.559)	0.81	NA			NA	NA	NA	NA
RAPS	112	-0.103 (-0.499 to 0.292)	0.61	NA			NA	NA	NA	NA
MR PRESSO	112	-0.100 (-0.467 to 0.267)	0.59	0.0022	0.32	rs775498; rs1278493	NA	NA	NA	NA

IVW, inverse variance weighted; nsnp, number of single nucleotide polymorphisms; RAPS, robust adjusted profile score

Supplemental Table 10 - Effect of lipoprotein(a) on atrial fibrillation mediated through atherosclerotic cardiovascular disease.

Mediator	Direct Effect	Indirect Effect	Proportion of Total Effect Mediated (%)
Observed Lp(a)			
Ischemic Heart Disease	0.0019 (0.0001 to 0.0038)	0.0025 (0.0023 to 0.0027)	59.4 (39.7 to 97.8)
Aortic Valve Stenosis	0.0035 (0.0021 to 0.0049)	0.0086 (0.0007 to 0.0010)	19.8 (13.8 to 27.7)
Ischemic Heart Disease and Aortic Valve Stenosis	0.0017 (0.0006 to 0.0029)	0.0026 (0.0024 to 0.0029)	62.2 (48.9 to 80.4)
Genetically Predicted Lp(a)			
Ischemic Heart Disease	0.0182 (0.0059 to 0.0334)	0.0100 (0.0083 to 0.0116)	35.6 (23.5 to 63.4)
Aortic Valve Stenosis	0.0213 (0.0115 to 0.0318)	0.0045 (0.0029 to 0.0057)	17.3 (10.4 to 31.6)
Ischemic Heart Disease and Aortic Valve Stenosis	0.0168 (0.0044 to 0.0322)	0.0110 (0.0088 to 0.0132)	39.2 (26.7 to 73.3)

Supplemental Table 11 - Association of independent genetic variants with LDL in UK Biobank after excluding LPA genetic variants.

Chromosome	Position (hg19)	SNP	Effect Allele	Other Allele	Effect Allele Frequency	Beta (mmol/L)	Standard Error	p-value
18	47158186	rs10438978	C	T	0.8178	0.015164	0.0026591	1.18E-08
9	139320069	rs10448340	G	T	0.32184	-0.014238	0.0021903	8.02E-11
6	29937795	rs1061537	A	G	0.46166	0.016162	0.0020502	3.20E-15
19	45413233	rs1065853	T	G	0.080275	-0.40201	0.0037043	1.00E-200
10	124686656	rs10794579	C	T	0.5763	0.014373	0.0020696	3.80E-12
19	45412955	rs1081105	C	A	0.027878	0.15234	0.0062156	1.56E-132
1	234734956	rs10910476	T	C	0.55699	0.012392	0.002071	2.18E-09
1	174898715	rs10912854	C	G	0.27215	0.012683	0.0023075	3.88E-08
7	100216773	rs10953298	T	C	0.23569	-0.017618	0.0024225	3.54E-13
11	18639167	rs11024735	T	C	0.7397	0.013354	0.0023322	1.03E-08
12	9061971	rs11047939	A	G	0.22676	0.015695	0.0024554	1.64E-10
16	83980965	rs11149612	T	C	0.46149	-0.015554	0.0020656	5.09E-14
3	132217703	rs113177823	A	G	0.054411	-0.031935	0.0045654	2.66E-12
9	136149711	rs115478735	T	A	0.18375	0.051192	0.0026462	2.47E-83
11	103870640	rs115739682	A	T	0.19105	-0.016709	0.0026169	1.72E-10
1	55505647	rs11591147	T	G	0.017695	-0.30026	0.007733	1.00E-200
11	5701074	rs11601507	A	C	0.068465	0.030005	0.0039966	6.03E-14
14	24871926	rs11621792	T	C	0.45362	0.016436	0.0020675	1.87E-15
5	52095024	rs116734477	T	C	0.041372	-0.044622	0.0051499	4.55E-18
12	121426478	rs1169292	T	C	0.3075	0.021572	0.002222	2.80E-22
3	142648844	rs11709868	T	G	0.29778	-0.014707	0.002246	5.84E-11
8	29024943	rs117139027	A	G	0.017217	-0.05303	0.0078561	1.48E-11

8	74899141	rs11775193	T	C	0.31395	-0.013049	0.0022102	3.55E-09
9	107647019	rs11789603	T	C	0.10762	0.020987	0.0033079	2.23E-10
19	11197261	rs12151108	A	G	0.11874	-0.15668	0.0031533	1.00E-200
4	100239319	rs1229984	C	T	0.977714	0.047196	0.0069313	9.84E-12
15	75083912	rs12442901	G	A	0.73778	0.01672	0.0023292	7.07E-13
16	11706100	rs12445804	A	G	0.074481	0.02304	0.0039348	4.76E-09
1	109817590	rs12740374	T	G	0.22176	-0.1037	0.0024557	1.00E-200
10	18495885	rs1277763	C	T	0.7981	0.014239	0.0025517	2.41E-08
5	74656539	rs12916	C	T	0.39977	0.054077	0.0020866	6.08E-148
2	44327463	rs13020929	A	G	0.45694	0.012833	0.0020589	4.58E-10
3	58398215	rs13066351	T	C	0.082304	-0.027956	0.0037291	6.57E-14
3	12239931	rs13098031	T	G	0.26877	-0.017143	0.002311	1.19E-13
4	3443931	rs13108218	G	A	0.61829	-0.016763	0.0021242	3.00E-15
2	165528876	rs13389219	T	C	0.39347	-0.012123	0.0020938	7.05E-09
8	6563868	rs1365041	T	G	0.68341	0.012144	0.0022067	3.73E-08
22	41272143	rs138354	C	T	0.53403	-0.012176	0.0020498	2.85E-09
2	21149771	rs140798831	C	T	0.65258	-0.031359	0.0021558	6.38E-48
14	94768196	rs145730801	C	T	0.045445	0.029537	0.0050078	3.68E-09
4	81164723	rs1458038	T	C	0.29289	-0.016814	0.0022561	9.16E-14
19	45337918	rs147711004	A	G	0.037007	0.13139	0.0055391	2.81E-124
19	45346666	rs148601586	G	C	0.013544	0.11661	0.0090611	6.82E-38
2	234522619	rs149247216	C	A	0.068811	0.022847	0.0040443	1.61E-08
17	64228995	rs149394327	C	G	0.029739	0.058407	0.0060424	4.23E-22
8	18272881	rs1495741	A	G	0.77908	-0.0171	0.0024638	3.92E-12
2	118845121	rs150474434	A	G	0.10187	-0.031382	0.0034066	3.22E-20
15	58683366	rs1532085	G	A	0.61337	-0.016135	0.0021007	1.58E-14

10	71093392	rs16926246	T	C	0.13069	-0.019763	0.0030332	7.25E-11
6	100620931	rs17185536	T	C	0.24437	-0.01548	0.0023897	9.32E-11
11	61603510	rs174576	A	C	0.3527	-0.027986	0.0021426	5.58E-39
7	44581986	rs17725246	C	T	0.18254	0.031461	0.0026413	1.05E-32
20	43042364	rs1800961	T	C	0.031399	-0.051996	0.0058643	7.58E-19
20	39179822	rs1883711	C	G	0.031082	0.095621	0.006005	4.56E-57
9	107586753	rs2066714	C	T	0.12709	0.02071	0.0030664	1.44E-11
12	50650057	rs2160994	C	T	0.64706	0.016101	0.0021422	5.65E-14
22	35678256	rs2179050	G	A	0.69051	-0.012521	0.0022321	2.03E-08
13	32959199	rs2238162	T	C	0.52456	-0.015568	0.0020472	2.87E-14
7	75614777	rs2302429	A	G	0.18325	0.015681	0.0026428	2.97E-09
16	56990716	rs247617	A	C	0.32478	-0.030169	0.0021858	2.55E-43
19	11228745	rs2569550	C	T	0.59296	0.037956	0.0020859	5.96E-74
17	45664861	rs2611867	G	A	0.51292	-0.021804	0.0020526	2.37E-26
15	58727325	rs261332	G	A	0.78938	-0.020206	0.0025123	8.79E-16
20	17844684	rs2618566	T	G	0.66141	-0.021268	0.0021603	7.25E-23
1	220970028	rs2642438	G	A	0.70247	0.02176	0.002235	2.13E-22
2	63248968	rs2710644	C	A	0.69906	0.017912	0.0022333	1.06E-15
5	131638817	rs272838	T	C	0.16495	-0.016877	0.0027849	1.36E-09
8	116667539	rs2737263	T	G	0.28411	-0.01916	0.0022696	3.13E-17
9	107661742	rs2740488	C	A	0.26548	-0.020655	0.0023224	5.93E-19
8	126500031	rs28601761	G	C	0.41935	-0.054389	0.002096	2.64E-148
8	55451193	rs28615248	C	T	0.19749	0.021387	0.0025804	1.15E-16
1	235109214	rs28631087	C	T	0.21218	-0.01472	0.00251	4.51E-09
19	45173951	rs28807203	C	A	0.048867	-0.10432	0.0047597	2.08E-106
3	119813282	rs334558	G	A	0.33584	0.015183	0.0021769	3.07E-12

12	618790	rs34019521	C	G	0.26177	0.012866	0.0023335	3.52E-08
1	10798489	rs34071855	G	C	0.34069	-0.012353	0.002167	1.20E-08
19	49155255	rs34488585	T	C	0.085278	-0.022526	0.0037133	1.31E-09
6	31514448	rs34568880	T	C	0.013032	0.051441	0.0090229	1.19E-08
19	58662235	rs35081008	T	C	0.14661	-0.030705	0.0029086	4.77E-26
2	20363666	rs35135293	T	C	0.52048	-0.013042	0.0020543	2.18E-10
9	2640759	rs3780181	G	A	0.067305	-0.025128	0.0041137	1.01E-09
16	72097827	rs3794695	T	C	0.18827	0.044279	0.002618	3.83E-64
6	116316882	rs3822855	T	G	0.40059	0.016405	0.002089	4.07E-15
2	44072576	rs4299376	T	G	0.6765	-0.046289	0.0021856	1.74E-99
11	126244955	rs4307732	A	G	0.10532	0.041646	0.00334	1.13E-35
19	11285390	rs440677	A	G	0.62303	-0.015172	0.0021278	1.00E-12
1	55521313	rs472495	T	G	0.65079	0.037196	0.0021467	3.12E-67
8	59393273	rs4738684	G	A	0.66476	-0.0267	0.0021666	6.90E-35
21	40709171	rs4818025	G	A	0.57159	0.011403	0.0020731	3.79E-08
6	30696762	rs4947288	G	A	0.22085	0.015906	0.0024715	1.23E-10
10	113933009	rs5024318	A	T	0.24753	0.017448	0.0023726	1.93E-13
1	109696333	rs542049	C	T	0.33009	-0.013454	0.0021893	8.00E-10
17	7080316	rs55714927	T	C	0.19216	-0.025364	0.002596	1.52E-22
8	145031968	rs55831924	T	C	0.36104	0.016018	0.0021414	7.42E-14
10	94822686	rs55843714	T	C	0.59271	0.017969	0.0020808	5.85E-18
3	69810294	rs55921103	T	G	0.65099	0.011846	0.0021615	4.24E-08
7	21598753	rs56130071	C	G	0.2191	0.027901	0.0024814	2.50E-29
2	21289432	rs581411	A	G	0.82133	0.090106	0.0026677	1.00E-200
5	156399039	rs58198139	T	C	0.63579	0.030182	0.002124	8.20E-46
19	19379549	rs58542926	T	C	0.07546	-0.093367	0.0038733	2.81E-128

12	111973358	rs597808	G	A	0.51664	0.022783	0.002053	1.31E-28
1	55492083	rs59784135	G	A	0.69766	-0.021101	0.0023064	5.76E-20
7	98874892	rs60612724	G	A	0.037656	0.030964	0.005384	8.88E-09
20	62909520	rs6090101	A	G	0.19903	0.016205	0.0025922	4.07E-10
20	39780932	rs6093446	A	G	0.28692	0.017525	0.0022636	9.80E-15
20	17804068	rs61433703	A	G	0.16344	0.015733	0.0027919	1.75E-08
19	45617263	rs62118464	A	G	0.1183	0.027128	0.003243	6.03E-17
19	10622478	rs62131897	C	G	0.026592	-0.041576	0.0066524	4.11E-10
2	169832276	rs62171034	T	A	0.6039	-0.017242	0.0021054	2.64E-16
9	22081850	rs6475606	T	C	0.48109	-0.018219	0.0020477	5.74E-19
9	78730766	rs6560499	A	G	0.57694	-0.011853	0.0020816	1.24E-08
13	114551993	rs6602909	C	T	0.32526	0.019873	0.0021866	1.01E-19
1	198994619	rs6667939	T	C	0.7197	0.014352	0.0022932	3.89E-10
2	44080324	rs6709904	G	A	0.11102	-0.036362	0.0032664	8.84E-29
8	9173209	rs7012637	A	G	0.47499	0.022562	0.0020626	7.63E-28
14	70846954	rs7157399	C	T	0.86123	0.01694	0.002962	1.07E-08
5	71849807	rs71628040	C	T	0.055714	-0.026255	0.0044683	4.21E-09
16	71642946	rs7186717	T	C	0.37504	0.016359	0.0021125	9.69E-15
16	72217113	rs7202323	G	T	0.22869	-0.022612	0.0024353	1.62E-20
17	67191270	rs72631343	G	C	0.12732	-0.027414	0.003066	3.87E-19
10	115786233	rs72823013	A	G	0.12807	-0.019459	0.0030686	2.28E-10
1	62999675	rs7534572	G	C	0.6456	0.035578	0.0021382	3.84E-62
1	27180088	rs75460349	C	A	0.02368	0.052673	0.0068347	1.29E-14
2	203527979	rs7569317	C	T	0.53196	0.016745	0.0020487	3.01E-16
2	135597764	rs7608700	A	C	0.32311	0.014803	0.0022579	5.53E-11
5	74472939	rs7707394	A	G	0.35469	0.035352	0.0021349	1.45E-61

5	122848876	rs7734476	A	G	0.54853	0.018928	0.0020544	3.18E-20
6	16126934	rs7746081	A	G	0.30134	-0.0219	0.0022339	1.10E-22
17	67081278	rs77542162	G	A	0.022999	0.11658	0.0068278	2.45E-65
18	47109955	rs77960347	G	A	0.013481	0.062069	0.0088598	2.46E-12
2	27741237	rs780094	C	T	0.61824	-0.030117	0.002104	1.84E-46
18	19640350	rs79120103	G	A	0.031287	-0.032465	0.0059419	4.67E-08
10	52373245	rs80276949	A	G	0.0231	0.04041	0.0068255	3.21E-09
7	6440437	rs836550	G	A	0.4079	0.011642	0.0020829	2.28E-08
7	25934357	rs896311	A	G	0.70494	0.014438	0.0022578	1.61E-10
6	127097775	rs9398815	C	T	0.45289	-0.012378	0.0020633	1.99E-09
11	116648917	rs964184	C	G	0.86787	-0.052434	0.0030169	1.25E-67
3	32535382	rs9834932	G	A	0.088902	-0.031014	0.0035927	6.03E-18
4	69373407	rs9884390	C	T	0.2365	0.021633	0.0024368	6.86E-19
17	7571080	rs9894946	G	A	0.84264	-0.015836	0.0028565	2.96E-08

Supplemental Table 12 - Effect of LDL on atrial fibrillation using Mendelian randomization.

				Unscaled		Scaled		
Exposure	**Outcome GWAS**	**N SNP**	**Method**	**OR**	**95%CI**	**OR***	**95%CI**	**P-value**
LDL	Nielsen et al.	140	Inverse variance weighted	1.01	(0.94 to 1.08)	1.01	(0.96 to 1.05)	0.82
LDL	Nielsen et al.	140	MR Egger	1.05	(0.94 to 1.17)	1.03	(0.96 to 1.11)	0.38
LDL	Nielsen et al.	140	Weighted median	0.99	(0.92 to 1.07)	1	(0.95 to 1.04)	0.85
LDL	Nielsen et al.	140	Weighted mode	1.03	(0.96 to 1.09)	1.02	(0.98 to 1.06)	0.43
LDL	Nielsen et al.	140	Robust adjusted profile score (RAPS)	1.01	(0.95 to 1.08)	1.01	(0.97 to 1.05)	0.69
LDL	FinnGen	128	Inverse variance weighted	1.06	(0.91 to 1.24)	1.04	(0.94 to 1.15)	0.48
LDL	FinnGen	128	MR Egger	1.28	(1.02 to 1.61)	1.18	(1.01 to 1.36)	0.035
LDL	FinnGen	128	Weighted median	1.06	(0.88 to 1.27)	1.04	(0.92 to 1.17)	0.54
LDL	FinnGen	128	Weighted mode	1.07	(0.92 to 1.25)	1.05	(0.95 to 1.16)	0.36
LDL	FinnGen	128	Robust adjusted profile score (RAPS)	1.05	(0.91 to 1.22)	1.03	(0.94 to 1.14)	0.51

*Effect is scaled to the equivalent effect on CAD from Lp(a) (i.e., the effect on atrial fibrillation per unit of LDL that corresponds to the same effect on CAD as 50 nmol/L increase in Lp(a))

Supplemental Table 13 - Effect of triglycerides on atrial fibrillation using Mendelian randomization.

				Unscaled		Scaled		
Exposure	**Outcome GWAS**	**N SNP**	**Method**	**OR**	**95%CI**	**OR***	**95%CI**	**P-value**
Triglycerides	Nielsen et al.	178	Inverse variance weighted	0.98	(0.93 to 1.04)	0.99	(0.91 to 1.07)	0.46
Triglycerides	Nielsen et al.	178	MR Egger	0.99	(0.91 to 1.07)	0.99	(0.88 to 1.11)	0.74
Triglycerides	Nielsen et al.	178	Weighted median	1.00	(0.94 to 1.06)	1.00	(0.92 to 1.09)	0.99
Triglycerides	Nielsen et al.	178	Weighted mode	1.00	(0.95 to 1.04)	1.00	(0.94 to 1.06)	0.82
Triglycerides	Nielsen et al.	178	Robust adjusted profile score (RAPS)	0.99	(0.93 to 1.05)	0.99	(0.91 to 1.08)	0.68
Triglycerides	FinnGen	166	Inverse variance weighted	1.07	(0.97 to 1.18)	1.05	(0.91 to 1.21)	0.20
Triglycerides	FinnGen	166	MR Egger	1.02	(0.88 to 1.17)	1.01	(0.82 to 1.24)	0.82
Triglycerides	FinnGen	166	Weighted median	1.09	(0.96 to 1.24)	1.06	(0.88 to 1.28)	0.19
Triglycerides	FinnGen	166	Weighted mode	1.07	(0.96 to 1.20)	1.05	(0.89 to 1.23)	0.22
Triglycerides	FinnGen	166	Robust adjusted profile score (RAPS)	1.08	(0.98 to 1.20)	1.06	(0.91 to 1.22)	0.12

*Effect is scaled to the equivalent effect on CAD from Lp(a) (i.e., the effect on atrial fibrillation per unit of triglycerides that corresponds to the same effect on CAD as 50 nmol/L increase in Lp(a))

Supplemental Methods

Subgroup analyses in UK Biobank

In order to identify effect modifiers of the Lp(a) and AF relationship, subgroup analyses were performed for common risk factors of AF selected *a priori* including age (*above versus below 65 years of age*), sex, obesity, smoking status (*current and former versus never smokers*), hypertension, prevalent ischemic heart disease, prevalent aortic valve stenosis, prevalent heart failure, and prevalent type 2 diabetes. Moreover, risk factors of AF modeled as quantitative risk factors for interaction analyses included age, BMI, systolic blood pressure, diastolic blood pressure, or alcohol intake.

Estimating effect of Lp(a)-lowering therapies in relation to other clinical risk factors

To provide an estimate of reduction in AF expected with Lp(a)-lowering therapies, we restricted the UK Biobank sample to participants with Lp(a) above 150 nmol/L that would be eligible for current Lp(a)-lowering trials. For each participant, the probability of AF was predicted using a logistic regression model adjusted for the same covariates. By modifying Lp(a) levels in the model, the predicted prevalence of AF was compared before and after reducing Lp(a) levels by 80%, based on the median reduction in Lp(a) observed in phase 2 trials of Lp(a)-lowering therapies (1). Risk estimates for reduction of blood pressure and BMI were obtained from previous literature (2, 3).

Genotyping quality control in UK Biobank

In the UK Biobank, genotyping was performed with the Applied Biosystems UK Biobank Lung Exome Variant Evaluation (UK BiLEVE) and UK Biobank Axiom arrays (Affymetrix Research Services Laboratory, Santa Clara, California, USA). Quality control has been previously described (4). In brief, UK Biobank centrally excluded poor quality markers or samples based on standard metrics, such as batch effects, plate effects, Hardy-Weinberg equilibrium, sex effects, array effects, missing rate, and heterozygosity. For our analyses, data was derived from almost 17 million genetic variants from 'v3' release of the UK Biobank genetic data including those present in the Haplotype Reference Consortium and 1000 Genomes panels (4). Genetic variants located in the human leukocyte antigen gene complex were excluded due to extensive pleiotropic effects.

Genetic risk score analyses in UK Biobank

After excluding rare genetic variants (MAF<0.1%) or variants with low imputation quality (INFO<0.7), three scores were constructed using variants across the entire genome ($p<5x10^{-6}$), within 500Kb of the *LPA* gene, and within 50Kb of the *LPA* gene. After regularization, the allele score was calculated by multiplying the effect of each selected genetic variant by the number of effect alleles and summing the value for each individual resulting in a genetically predicted Lp(a) level.

$$GRS = \sum \beta_1 g_1 + \beta_2 g_2 + \cdots + \beta_i g_i,$$

where β is the effect of a genetic variant on Lp(a), and g is the number of effect alleles (0, 1, or 2)

Genome-wide association study for Lp(a) in UK Biobank

A genome-wide association study (GWAS) for Lp(a) levels was conducted in the same training set from the UK Biobank. To allow for genetic relatedness between participants, REGENIE was used to test for associations after excluding rare genetic variants (MAF<0.1%) (5). The model was adjusted for age, sex, genotyping chip type, assessment centre, and 40 principal components of ancestry. To arrive at an independent set, genetic variants associated at genome-wide significance ($p<5x10^{-8}$) were pruned based on linkage disequilibrium (LD) at a threshold of $r^2<0.001$ using Europeans from 1000 Genomes phase 3 as reference panel (6).

Mendelian randomization sensitivity analyses

An Egger intercept significantly different from 0 ($p<0.05$) was considered evidence of directional pleiotropy and the causal estimate from MR Egger was reported to attempt to control for pleiotropic effects(7). As a sensitivity analysis robust to idiosyncratic pleiotropy and weak instrument bias, MR-RAPS (Robust Adjusted Profile Score) was conducted using overdispersion and Tukey's loss function(8). To detect and correct for potential bias from invalid variants with pleiotropic effects, we performed the MR-PRESSO (Mendelian Randomization Pleiotropy RESidual Sum and Outlier) test with 10,000 simulations(9). The global test p-value evaluated whether there was any overall horizontal pleiotropy among all genetic variants. If global test p-value was significant ($p<0.05$), outlying genetic variants with predicted pleiotropic effects were removed and MR analysis repeated to correct for horizontal pleiotropy. The distortion test evaluated whether removal of the pleiotropic variants resulted in a significantly different causal estimate ($p<0.05$). As another sensitivity analyses, Mendelian randomization was repeated using two genetic variants (rs10455872 and rs3798220) associated with Lp(a) that were previously reported by Clarke *et al.* (10)

Standardizing effect estimates for Lp(a), LDL, and triglycerides

In order to facilitate comparisons, the effects derived from Mendelian randomization analysis of Lp(a), LDL, and triglycerides on risk of AF were standardized. The effect of each lipid particle on CAD using GWAS summary statistics from CARDIoGRAMplusC4D (n_{cases}=60,801) was used to standardize effects, such that effects of LDL and triglycerides on CAD are equivalent to the effect of 50 nmol/L change in Lp(a) on CAD. In detail, for each lipid, the effect of 50 nmol/L change in Lp(a) on CAD was divided by the effect of that lipid fraction on CAD to determine the units of the lipid fraction that correspond to the same effect as 50 nmol/L change in Lp(a), then the effect of the lipid fraction on AF was adjusted by multiplying by this number of units. As an example, the formula for obtaining the effect of LDL on AF standardized with respect to the effect on CAD caused by 50 nmol/L change in Lp(a) is shown below.

$$\text{Standardized log } (OR)_{AF \sim LDL} = \frac{\log (OR)_{CAD \sim LPA}}{\log (OR)_{CAD \sim LDL}} \log (OR)_{AF \sim LDL}$$

$$= \frac{0.167}{0.612} * 0.0083 = 0.0023$$

Mediation analysis

To formally quantify the proportion increased risk of atrial fibrillation due to increased Lp(a) that is mediated by prevalent ischemic heart disease and aortic valve stenosis. Mediation analysis was performed using the *mediation* package in R with 1,000 simulations and quasi-Bayesian approximation of confidence intervals (11). As recommended, Lp(a) "treatment" and "control" values were selected by dichotomizing through a natural cut point. In this case, participants were categorized based on Lp(a) levels above or below 150 nmol/L.

Supplemental References

1. Tsimikas S, Karwatowska-Prokopczuk E, Gouni-Berthold I, et al. Lipoprotein(a) Reduction in Persons with Cardiovascular Disease. N. Engl. J. Med. 2020;382:244–255.
2. Wong CX, Sullivan T, Sun MT, et al. Obesity and the risk of incident, post-operative, and post-ablation atrial fibrillation: A meta-analysis of 626,603 individuals in 51 studies. JACC Clin. Electrophysiol. 2015;1:139–152.
3. Hyman MC, Levin MG, Gill D, et al. Genetically Predicted Blood Pressure and Risk of Atrial Fibrillation. Hypertens. (Dallas, Tex. 1979) 2021;77:376–382.
4. Bycroft C, Freeman C, Petkova D, et al. Genome-wide genetic data on ~500,000 UK Biobank participants. bioRxiv 2017. Available at: http://dx.doi.org/10.1101/166298. Accessed July 23, 2018.
5. Mbatchou J, Barnard L, Backman J, et al. Computationally efficient whole genome regression for quantitative and binary traits. bioRxiv 2020:1–88.
6. Altshuler DM, Durbin RM, Abecasis GR, et al. An integrated map of genetic variation from 1,092 human genomes. Nature 2012;491:56–65. Available at: https://www.nature.com/articles/nature11632.pdf. Accessed April 20, 2018.
7. Bowden J, Smith GD, Burgess S. Mendelian randomization with invalid instruments: Effect estimation and bias detection through Egger regression. Int. J. Epidemiol. 2015;44:512–525.
8. Zhao Q, Wang J, Hemani G, Bowden J, Small DS. Statistical inference in two-sample summary-data Mendelian randomization using robust adjusted profile score. arXiv Prepr. 2018. Available at: https://arxiv.org/pdf/1801.09652.pdf. Accessed July 30, 2018.
9. Verbanck M, Chen C-Y, Neale B, Do R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. Nat. Genet. 2018;50:693–698. Available at: http://www.nature.com.libaccess.lib.mcmaster.ca/articles/s41588-018-0099-7.pdf. Accessed April 25, 2018.
10. Clarke R, Peden JF, Hopewell JC, et al. Genetic Variants Associated with Lp(a) Lipoprotein Level and Coronary Disease. N. Engl. J. Med. 2009;361:2518–2528. Available at: http://www.nejm.org/doi/abs/10.1056/NEJMoa0902604.
11. Imai K, Keele L, Tingley D. A general approach to causal mediation analysis. Psychol. Methods 2010;15:309–334. Available at:

http://doi.apa.org/getdoi.cfm?doi=10.1037/a0020761.

APPENDIX B:

SUPPLEMENTARY DATA FOR CHAPTER 4

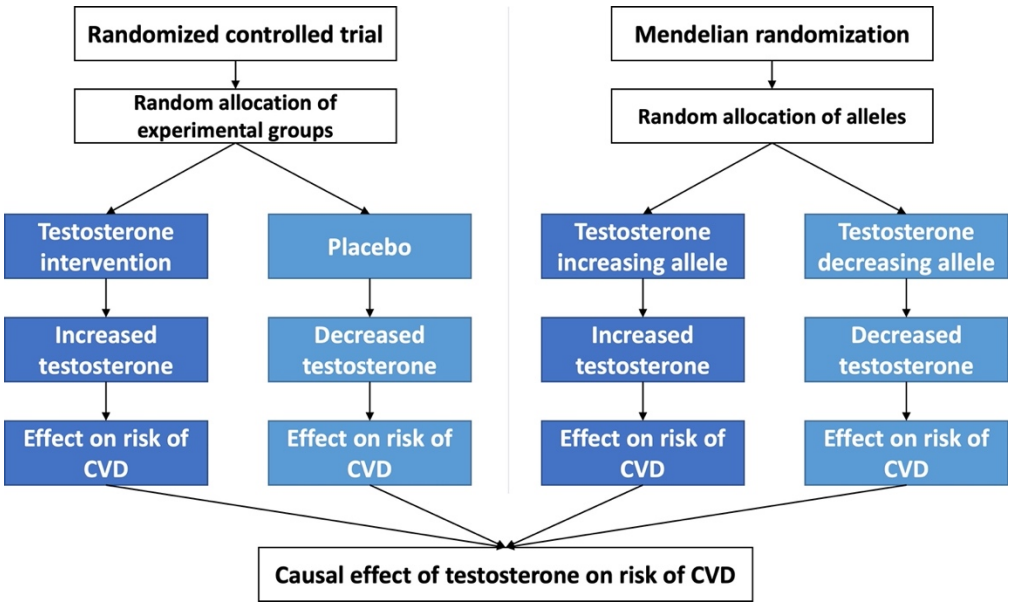


Figure 1 – Figure Supplement 1. Comparison of randomized controlled trial (RCT) and Mendelian randomization (MR) study designs demonstrating the common foundation behind interpretation of a causal effect of testosterone on cardiovascular disease (CVD). In accordance with Mendel's second law, random and independent inheritance of alleles can be thought of akin to random allocation of treatment vs. placebo in RCT. Therefore, by the same reasoning, if MR finds genetic variants affecting testosterone are associated with a difference in CVD risk, it provides evidence that testosterone causally affects CVD.

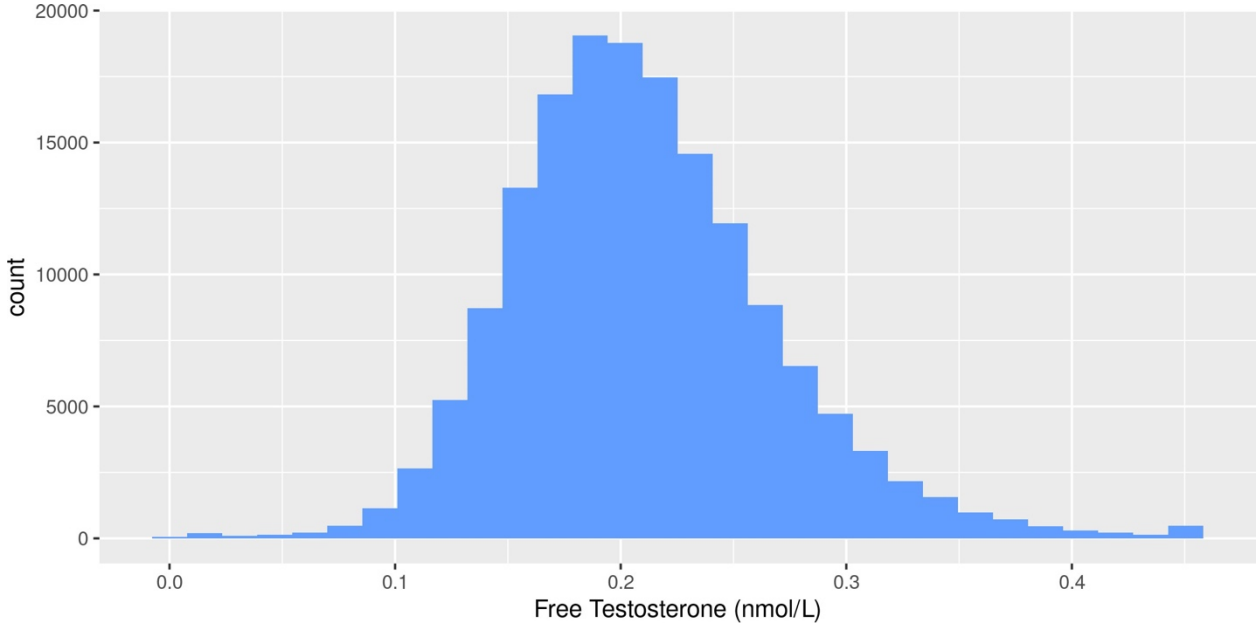


Figure 1 – Figure Supplement 2. Distribution of free testosterone levels calculated using the Vermeulen equation in males from the UK Biobank cohort.

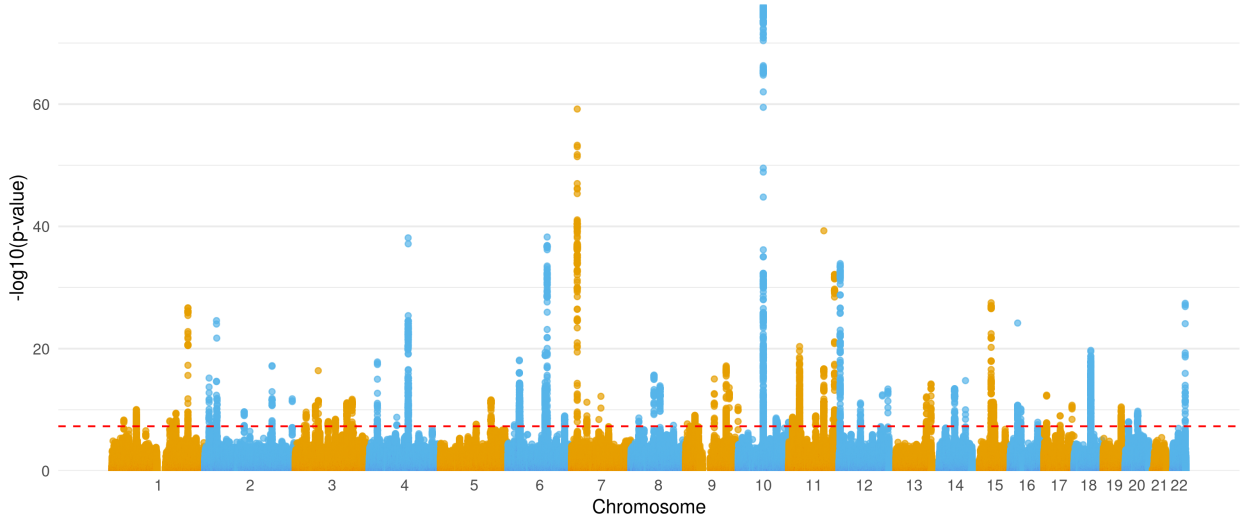


Figure 1 – Figure Supplement 3. Manhattan plot showing distribution of p-values from genome-wide association study of calculated free testosterone after exclusion of SHBG-associated variants based on chromosomal location.

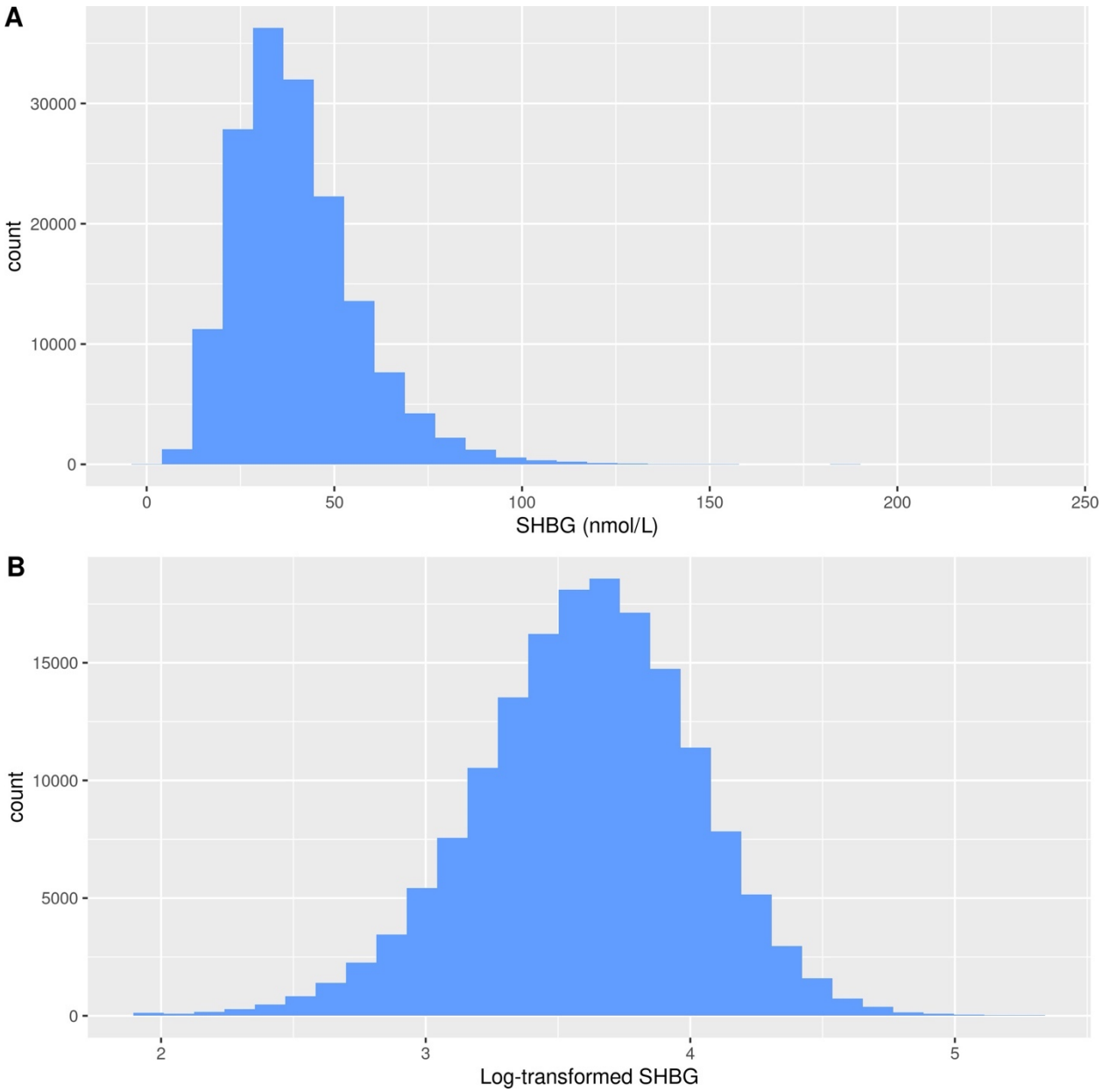


Figure 1 – Figure Supplement 4. Distribution of sex hormone-binding globulin in males from the UK Biobank. (A) Distribution of raw sex hormone-binding globulin levels in males from the UK Biobank cohort (B) Distribution of natural log-transformed sex hormone-binding globulin levels in males from the UK Biobank cohort

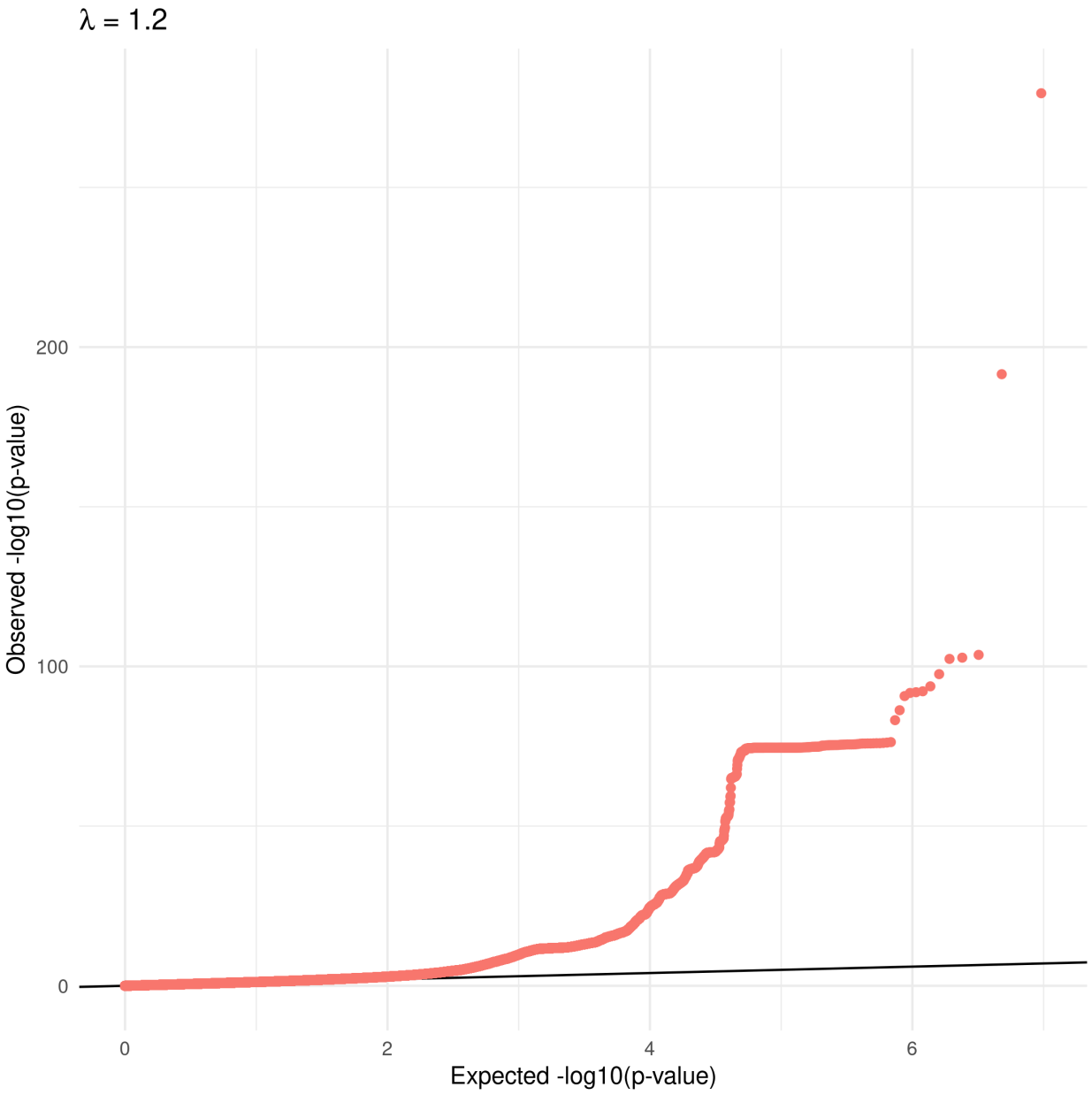


Figure 1 – Figure Supplement 5. Quantile-quantile plot for genome-wide association study of calculated free testosterone levels (before exclusion of SHBG-associated genetic variants). Plot shows observed test statistics (y-axis) relative to expected test statistics under a null model (x-axis), and lambda (λ) represents genomic inflation factor.

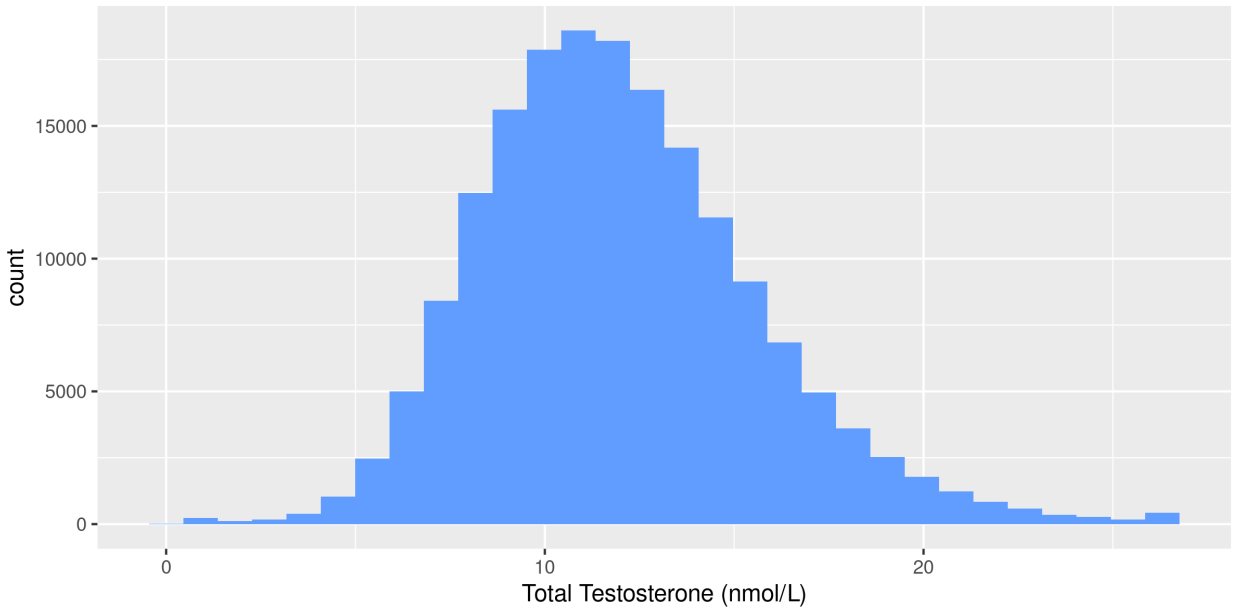


Figure 1 – Figure Supplement 6. Distribution of total testosterone levels in males from the UK Biobank cohort.

Figure 2 – Source Data 1. Associations of genetically predicted calculated free testosterone for 439 health outcomes across the human phenome.

Trait	Effect per 0.1 nmol/L increase CFT (95% CI)	P-value	FDR-adjusted p-value	Sample Size	Number of Cases	Number of Controls	Phecode	Category
Creatinine	**0.113 (0.079 to 0.146)**	**4.78E-11**	**2.11E-08**	**149849**	**NA**	**NA**	**NA**	**biomarker**
C-reactive protein	**-0.085 (-0.119 to -0.052)**	**6.15E-07**	**1.35E-04**	**149547**	**NA**	**NA**	**NA**	**biomarker**
Spinal stenosis	**OR=2.03 (1.51 to 2.75)**	**3.82E-06**	**5.61E-04**	**152836**	**1917**	**150919**	**phecode 720**	**musculoskeletal**
Apolipoprotein A	**-0.018 (-0.026 to -0.01)**	**1.55E-05**	**1.71E-03**	**138185**	**NA**	**NA**	**NA**	**biomarker**
HDL cholesterol	**-0.074 (-0.109 to -0.039)**	**3.62E-05**	**3.19E-03**	**138394**	**NA**	**NA**	**NA**	**biomarker**
Essential hypertension	**OR=1.17 (1.08 to 1.27)**	**7.53E-05**	**5.54E-03**	**156766**	**40809**	**115957**	**phecode 401.1**	**circulatory system**
Hypertension	**OR=1.17 (1.08 to 1.26)**	**1.05E-04**	**6.63E-03**	**156917**	**40960**	**115957**	**phecode 401**	**circulatory system**
Umbilical hernia	OR=1.64 (1.27 to 2.13)	1.64E-04	9.01E-03	132863	2624	130239	phecode 550.4	digestive
Sleep apnea	OR=1.56 (1.23 to 1.98)	2.09E-04	0.010	156734	3124	153610	phecode 327.3	neurological
Degenerative skin conditions and other dermatoses	OR=0.61 (0.47 to 0.79)	2.27E-04	0.010	155651	2584	153067	phecode 702	dermatologic
IGF1	0.333 (0.153 to 0.513)	2.81E-04	0.011	149151	NA	NA	NA	biomarker
Unspecified monoarthritis	OR=1.29 (1.12 to 1.48)	3.25E-04	0.012	147634	9832	137802	phecode 716.2	musculoskeletal
Phosphate	-0.01 (-0.016 to -0.005)	3.64E-04	0.012	138207	NA	NA	NA	biomarker
Inguinal hernia	OR=1.24 (1.1 to 1.39)	3.94E-04	0.012	144145	13906	130239	phecode 550.1	digestive
Abdominal hernia	OR=1.17 (1.07 to 1.28)	4.18E-04	0.012	157211	26972	130239	phecode 550	digestive
Total protein	-0.235 (-0.376 to -0.095)	1.05E-03	0.027	138299	NA	NA	NA	biomarker
Inflammatory diseases of prostate	OR=1.56 (1.19 to 2.03)	1.08E-03	0.027	44293	2695	41598	phecode 601	genitourinary
Intestinal obstruction without mention of hernia	OR=1.66 (1.22 to 2.25)	1.17E-03	0.027	127446	1884	125562	phecode 560	digestive
Allergyoradverse effect of penicillin	OR=1.34 (1.12 to 1.6)	1.19E-03	0.027	152298	5686	146612	phecode 960.2	injuries and poisonings
GERD	OR=1.24 (1.09 to 1.42)	1.26E-03	0.027	147762	10877	136885	phecode 530.11	digestive
Spondylosis and allied disorders	OR=1.46 (1.16 to 1.84)	1.30E-03	0.027	154215	3296	150919	phecode 721	musculoskeletal
Sleep disorders	OR=1.43 (1.15 to 1.78)	1.50E-03	0.030	157211	3601	153610	phecode 327	neurological
Symptoms involving head and neck	OR=1.67 (1.21 to 2.29)	1.68E-03	0.032	157211	1709	155502	phecode 293	mental disorders
Total bilirubin	0.054 (0.02 to 0.088)	1.74E-03	0.032	149300	NA	NA	NA	biomarker
Alkaline phosphatase	-0.054 (-0.087 to -0.02)	1.87E-03	0.033	149937	NA	NA	NA	biomarker
Seborrheic keratosis	OR=0.58 (0.41 to 0.82)	2.12E-03	0.036	156046	1419	154627	phecode 702.2	dermatologic
Other symptoms involving abdomen and pelvis	OR=1.69 (1.2 to 2.38)	2.51E-03	0.041	147838	1496	146342	phecode 579	digestive
Back pain	OR=1.31 (1.1 to 1.56)	2.76E-03	0.044	157211	5797	151414	phecode 760	symptoms
Albumin	-0.133 (-0.221 to -0.044)	3.26E-03	0.048	138474	NA	NA	NA	biomarker
Calcium	-0.005 (-0.008 to -0.002)	3.37E-03	0.048	138426	NA	NA	NA	biomarker
Internal derangement of knee	OR=1.25 (1.08 to 1.46)	3.39E-03	0.048	156183	7941	148242	phecode 835	injuries and poisonings
Lipoma	OR=1.42 (1.12 to 1.79)	3.46E-03	0.048	156819	3221	153598	phecode 214	neoplasms
Poisoning by antibiotics	OR=1.29 (1.08 to 1.52)	3.87E-03	0.052	152812	6200	146612	phecode 960	injuries and poisonings
Prostatitis	OR=1.68 (1.18 to 2.4)	4.13E-03	0.053	43064	1466	41598	phecode 601.1	genitourinary
Obstructive chronic bronchitis	OR=1.32 (1.09 to 1.6)	4.34E-03	0.053	147942	5003	142939	phecode 496.21	respiratory
Cancer; suspected or other	OR=1.19 (1.06 to 1.34)	4.37E-03	0.053	151649	13842	137807	phecode 195	neoplasms
Chronic bronchitis	OR=1.32 (1.09 to 1.59)	4.44E-03	0.053	148033	5094	142939	phecode 496.2	respiratory
Arthropathy NOS	OR=1.17 (1.05 to 1.3)	4.75E-03	0.055	155282	17480	137802	phecode 716.9	musculoskeletal
Contracture of palmar fascia Dupuytrens disease	OR=0.69 (0.53 to 0.9)	5.75E-03	0.063	145632	2606	143026	phecode 728.71	musculoskeletal
Fasciitis	OR=0.7 (0.54 to 0.9)	6.02E-03	0.063	145804	2778	143026	phecode 728.7	musculoskeletal
Other arthropathies	OR=1.16 (1.04 to 1.29)	6.19E-03	0.063	155325	17523	137802	phecode 716	musculoskeletal
Other peripheral nerve disorders	OR=1.35 (1.09 to 1.68)	6.31E-03	0.063	154804	3798	151006	phecode 351	neurological
Osteoarthrosis	OR=1.19 (1.05 to 1.36)	6.36E-03	0.063	157211	11721	145490	phecode 740	musculoskeletal
Nonrheumatic mitral valve disorders	OR=0.65 (0.48 to 0.89)	6.49E-03	0.063	150859	1859	149000	phecode 395.1	circulatory system
Hypercholesterolemia	OR=1.15 (1.04 to 1.28)	6.69E-03	0.063	155658	19758	135900	phecode 272.11	endocrine metabolic
Mitral valve disease	OR=0.66 (0.48 to 0.89)	6.73E-03	0.063	150902	1902	149000	phecode 394.2	circulatory system
Delirium dementia and amnestic and other cognitive disorders	OR=1.57 (1.13 to 2.17)	6.77E-03	0.063	154523	1652	152871	phecode 290	mental disorders
Malignant neoplasm; other	OR=1.18 (1.04 to 1.33)	8.18E-03	0.074	151312	13505	137807	phecode 195.1	neoplasms
Gout	OR=1.38 (1.09 to 1.75)	8.20E-03	0.074	156915	3131	153784	phecode 274.1	endocrine metabolic
Gout and other crystal arthropathies	OR=1.35 (1.08 to 1.69)	9.59E-03	0.085	157211	3427	153784	phecode 274	endocrine metabolic
Chronic airway obstruction	OR=1.25 (1.05 to 1.49)	0.011	0.097	148957	6018	142939	phecode 496	respiratory
Convulsions	OR=1.6 (1.1 to 2.33)	0.014	0.121	148322	1219	147103	phecode 345.3	neurological
Direct bilirubin	0.043 (0.009 to 0.078)	0.015	0.121	139800	NA	NA	NA	biomarker
Spondylosis without myelopathy	OR=1.43 (1.07 to 1.91)	0.015	0.121	153008	2089	150919	phecode 721.1	musculoskeletal
Respiratory abnormalities	OR=1.27 (1.05 to 1.54)	0.016	0.126	157211	4784	152427	phecode 513	respiratory
Cholesterol	-0.046 (-0.083 to -0.008)	0.017	0.135	149940	NA	NA	NA	biomarker
Gastrointestinal hemorrhage	OR=1.18 (1.03 to 1.36)	0.018	0.136	156012	9670	146342	phecode 578	digestive
Osteoarthritis; localized	OR=1.21 (1.03 to 1.41)	0.018	0.141	152917	7427	145490	phecode 740.1	musculoskeletal
Hyperlipidemia	OR=1.13 (1.02 to 1.24)	0.019	0.141	157148	21248	135900	phecode 272.1	endocrine metabolic
Lipoma of skin and subcutaneous tissue	OR=1.4 (1.05 to 1.87)	0.020	0.147	155732	2134	153598	phecode 214.1	neoplasms
Other acute and subacute forms of ischemic heart disease	OR=1.56 (1.07 to 2.28)	0.020	0.147	137463	1221	136242	phecode 411.9	circulatory system
Disorders of muscle; ligament; and fascia	OR=0.75 (0.59 to 0.96)	0.021	0.147	146118	3092	143026	phecode 728	musculoskeletal
Urinary tract infection	OR=1.24 (1.03 to 1.49)	0.021	0.147	149462	5313	144149	phecode 591	genitourinary
Personal history of diseases of digestive system	OR=1.19 (1.03 to 1.39)	0.021	0.147	133795	8233	125562	phecode 564.9	digestive
Disorders of lipoid metabolism	OR=1.12 (1.02 to 1.24)	0.023	0.156	157211	21311	135900	phecode 272	endocrine metabolic
Other intestinal obstruction	OR=1.48 (1.05 to 2.08)	0.025	0.167	127050	1488	125562	phecode 560.4	digestive
Anxiety disorder	OR=1.34 (1.04 to 1.72)	0.025	0.167	138279	2725	135554	phecode 300.1	mental disorders
Postoperative infection	OR=1.39 (1.04 to 1.87)	0.026	0.170	156723	2028	154695	phecode 080	infectious diseases
Esophagitis; GERD and related diseases	OR=1.13 (1.01 to 1.27)	0.029	0.187	152090	15205	136885	phecode 530.1	digestive
Gamma glutamyltransferase	0.037 (0.004 to 0.071)	0.030	0.187	149840	NA	NA	NA	biomarker
Other symptomsordisorders or the urinary system	OR=1.15 (1.01 to 1.3)	0.030	0.187	157211	12410	144801	phecode 599	genitourinary
Intestinal infection	OR=1.24 (1.02 to 1.52)	0.031	0.191	157211	4483	152728	phecode 008	infectious diseases
Actinic keratosis	OR=0.67 (0.46 to 0.97)	0.032	0.193	155924	1297	154627	phecode 702.1	dermatologic
Anxiety disorders	OR=1.3 (1.02 to 1.65)	0.033	0.195	138570	3016	135554	phecode 300	mental disorders
Chronic dermatitis due to solar radiation	OR=0.68 (0.47 to 0.97)	0.034	0.195	155541	1353	154188	phecode 938.2	dermatologic
Other disorders of prostate	OR=1.53 (1.03 to 2.27)	0.034	0.195	42816	1218	41598	phecode 602	genitourinary
Osteoporosis	OR=0.64 (0.42 to 0.97)	0.035	0.195	156733	1004	155729	phecode 743.1	musculoskeletal
Osteoporosis NOS	OR=0.64 (0.42 to 0.97)	0.035	0.195	156733	1004	155729	phecode 743.11	musculoskeletal
Asthma	OR=1.16 (1.01 to 1.33)	0.035	0.195	152957	10018	142939	phecode 495	respiratory
Fracture of lower limb	OR=0.76 (0.59 to 0.98)	0.036	0.201	151628	2688	148940	phecode 800	injuries and poisonings
Sciatica	OR=1.58 (1.03 to 2.43)	0.037	0.202	156502	934	155568	phecode 764	symptoms

Other inflammatory spondylopathies	OR=1.55 (1.03 to 2.33)	0.038	0.202	157211	1022	156189	phecode 715	musculoskeletal
Poisoning by primarily systemic agents	OR=1.53 (1.02 to 2.29)	0.040	0.211	147670	1058	146612	phecode 963	injuries and poisonings
Dermatitis due to solar radiation	OR=0.69 (0.49 to 0.99)	0.041	0.211	155587	1399	154188	phecode 938	dermatologic
Urinary obstruction	OR=1.4 (1.01 to 1.95)	0.041	0.211	146437	1636	144801	phecode 599.1	genitourinary
Peritoneal adhesions postoperative postinfection	OR=1.48 (1.01 to 2.15)	0.042	0.211	143200	1234	141966	phecode 568.1	digestive
Hypothyroidism NOS	OR=1.31 (1.01 to 1.7)	0.042	0.211	156567	2588	153979	phecode 244.4	endocrine metabolic
Other diseases of respiratory system; not elsewhere classified	OR=1.21 (1.01 to 1.46)	0.042	0.212	157211	5112	152099	phecode 519	respiratory
Other disorders of urethra and urinary tract	OR=1.35 (1.01 to 1.8)	0.043	0.215	151230	2091	149139	phecode 597	genitourinary
Fracture of unspecified part of femur	OR=0.64 (0.41 to 0.99)	0.046	0.224	149838	898	148940	phecode 800.2	injuries and poisonings
Septal DeviationsorTurbinate Hypertrophy	OR=1.31 (1 to 1.7)	0.047	0.225	150593	2536	148057	phecode 470	respiratory
Right bundle branch block	OR=0.68 (0.47 to 1)	0.048	0.225	141247	1236	140011	phecode 426.31	circulatory system
Functional digestive disorders	OR=1.13 (1 to 1.27)	0.048	0.225	139238	13676	125562	phecode 564	digestive
Erythematous conditions	OR=1.56 (1 to 2.43)	0.049	0.225	155296	881	154415	phecode 695	dermatologic
Alanine aminotransferase	0.033 (0 to 0.067)	0.049	0.225	149830	NA	NA	NA	biomarker
E coli	OR=1.46 (1 to 2.14)	0.050	0.225	150693	1204	149489	phecode 041.4	infectious diseases
Degeneration of macula and posterior pole of retina	OR=1.51 (1 to 2.28)	0.050	0.225	94875	1053	93822	phecode 362.2	sense organs
Macular degeneration senile of retina NOS	OR=1.51 (1 to 2.28)	0.050	0.225	94875	1053	93822	phecode 362.29	sense organs
Other disorders of peritoneum	OR=1.43 (1 to 2.06)	0.052	0.233	143289	1323	141966	phecode 568	digestive
Altered mental status	OR=1.4 (1 to 1.97)	0.053	0.235	154362	1491	152871	phecode 292.4	mental disorders
Diseases of the oral soft tissues; excluding lesions specific for gingiva and tong	OR=1.36 (0.99 to 1.87)	0.055	0.239	156503	1744	154759	phecode 528	digestive
Diseases of esophagus	OR=1.11 (1 to 1.24)	0.057	0.245	153704	16819	136885	phecode 530	digestive
Peripheral vascular disease; unspecified	OR=1.33 (0.98 to 1.8)	0.067	0.286	153464	1889	151575	phecode 443.9	circulatory system
Secondary malignancy of bone	OR=1.4 (0.97 to 2.01)	0.069	0.291	139152	1345	137807	phecode 198.6	neoplasms
Other disorders of eyelids	OR=1.29 (0.98 to 1.7)	0.069	0.291	155466	2335	153131	phecode 374	sense organs
Diaphragmatic hernia	OR=1.13 (0.99 to 1.28)	0.071	0.291	141891	11652	130239	phecode 550.2	digestive
Other diseases of respiratory system; NEC	OR=1.19 (0.98 to 1.45)	0.071	0.291	156921	4822	152099	phecode 519.8	respiratory
Nonspecific chest pain	OR=1.11 (0.99 to 1.25)	0.071	0.291	157211	14178	143033	phecode 418	circulatory system
Hemiplegia	OR=0.67 (0.43 to 1.04)	0.072	0.291	148019	916	147103	phecode 342	neurological
Hypothyroidism	OR=1.26 (0.98 to 1.62)	0.076	0.305	156694	2715	153979	phecode 244	endocrine metabolic
Other mental disorder	OR=1.1 (0.99 to 1.22)	0.079	0.310	154074	18520	135554	phecode 306	mental disorders
Synovitis and tenosynovitis	OR=1.3 (0.97 to 1.73)	0.079	0.310	145109	2083	143026	phecode 727.1	musculoskeletal
Hemorrhoids	OR=1.13 (0.99 to 1.29)	0.081	0.310	151566	10293	141273	phecode 455	circulatory system
Hydronephrosis	OR=1.39 (0.96 to 2.02)	0.081	0.310	153509	1254	152255	phecode 595	genitourinary
Anal and rectal conditions	OR=1.15 (0.98 to 1.33)	0.081	0.310	149752	7786	141966	phecode 565	digestive
Otitis media and Eustachian tube disorders	OR=1.46 (0.95 to 2.23)	0.082	0.310	156494	954	155540	phecode 381	sense organs
Osteoarthrosis; localized; primary	OR=1.21 (0.98 to 1.51)	0.083	0.310	149212	3722	145490	phecode 740.11	musculoskeletal
Noninfectious gastroenteritis	OR=1.18 (0.98 to 1.41)	0.083	0.310	130942	5380	125562	phecode 558	digestive
Diseases of pancreas	OR=1.37 (0.96 to 1.96)	0.084	0.311	157211	1360	155851	phecode 577	digestive
Other biliary tract disease	OR=1.35 (0.96 to 1.88)	0.085	0.311	153390	1530	151860	phecode 575	digestive
Symptoms involving digestive system	OR=1.17 (0.98 to 1.39)	0.091	0.332	131291	5729	125562	phecode 561	digestive
Other disorders of intestine	OR=1.13 (0.98 to 1.31)	0.094	0.337	150634	8668	141966	phecode 569	digestive
Hemorrhage of rectum and anus	OR=1.17 (0.97 to 1.41)	0.094	0.337	151592	5250	146342	phecode 578.8	digestive
Vitamin D	0.029 (-0.005 to 0.063)	0.096	0.340	145566	NA	NA	NA	biomarker
Heart valve disorders	OR=0.83 (0.67 to 1.03)	0.097	0.343	152826	3826	149000	phecode 395	circulatory system
Atrial fibrillation and flutter	OR=1.12 (0.98 to 1.28)	0.098	0.344	150383	10372	140011	phecode 427.2	circulatory system
Other diseases of blood and blood forming organs	OR=1.27 (0.95 to 1.68)	0.101	0.350	155768	2180	153588	phecode 289	hematopoietic
Peripheral enthesopathies and allied syndromes	OR=1.15 (0.97 to 1.36)	0.106	0.362	149433	6407	143026	phecode 726	musculoskeletal
Intervertebral disc disorders	OR=1.22 (0.96 to 1.56)	0.106	0.362	153832	2913	150919	phecode 722	musculoskeletal
Hemorrhage of gastrointestinal tract	OR=1.22 (0.96 to 1.55)	0.108	0.362	149364	3022	146342	phecode 578.9	digestive
Urethral stricture not specified as infectious	OR=1.29 (0.95 to 1.75)	0.108	0.362	150977	1838	149139	phecode 597.1	genitourinary
Tobacco use disorder	OR=1.12 (0.98 to 1.28)	0.110	0.362	151119	10356	140763	phecode 318	mental disorders
Osteoarthrosis NOS	OR=1.17 (0.97 to 1.41)	0.111	0.362	150594	5104	145490	phecode 740.9	musculoskeletal
Cough	OR=1.34 (0.94 to 1.92)	0.111	0.362	152453	1344	151109	phecode 512.8	respiratory
Unstable angina intermediate coronary syndrome	OR=1.22 (0.95 to 1.55)	0.112	0.362	139239	2997	136242	phecode 411.1	circulatory system
Other disorders of synovium; tendon; and bursa	OR=1.19 (0.96 to 1.47)	0.113	0.362	146899	3873	143026	phecode 727	musculoskeletal
Esophageal bleeding varicesorhemorrhage	OR=0.71 (0.47 to 1.08)	0.113	0.362	137865	980	136885	phecode 530.2	digestive
LDL direct	-0.023 (-0.052 to 0.006)	0.113	0.362	149626	NA	NA	NA	biomarker
Gastric ulcer	OR=0.79 (0.59 to 1.06)	0.117	0.371	155429	2060	153369	phecode 531.2	digestive
Cystatin C	0.026 (-0.006 to 0.057)	0.118	0.371	149927	NA	NA	NA	biomarker
Other and unspecified disc disorder	OR=1.25 (0.94 to 1.67)	0.121	0.376	153055	2136	150919	phecode 722.9	musculoskeletal
Hypotension	OR=0.84 (0.67 to 1.05)	0.126	0.389	148804	3444	145360	phecode 458	circulatory system
Other local infections of skin and subcutaneous tissue	OR=0.81 (0.61 to 1.06)	0.127	0.390	153426	2334	151092	phecode 686	dermatologic
Acute pulmonary heart disease	OR=1.24 (0.94 to 1.64)	0.134	0.404	155042	2226	152816	phecode 415.1	circulatory system
Pulmonary embolism and infarction; acute	OR=1.24 (0.94 to 1.64)	0.134	0.404	155042	2226	152816	phecode 415.11	circulatory system
Cardiac pacemakerordevice in situ	OR=1.26 (0.93 to 1.71)	0.139	0.415	141903	1892	140011	phecode 426.9	circulatory system
Abnormal findings examination of lungs	OR=1.31 (0.92 to 1.86)	0.139	0.415	157211	1391	155820	phecode 514	respiratory
Renal failure	OR=1.13 (0.96 to 1.34)	0.142	0.421	155360	6446	148914	phecode 585	genitourinary
Malaise and fatigue	OR=1.3 (0.92 to 1.83)	0.144	0.423	157211	1440	155771	phecode 798	symptoms
Other specified peripheral vascular diseases	OR=1.23 (0.93 to 1.64)	0.145	0.423	153785	2210	151575	phecode 443.8	circulatory system
Transient cerebral ischemia	OR=0.76 (0.52 to 1.1)	0.146	0.423	152337	1222	151115	phecode 433.31	circulatory system
Poisoning by analgesics; antipyretics; and antirheumatics	OR=1.22 (0.93 to 1.61)	0.151	0.434	148899	2287	146612	phecode 965	injuries and poisonings
Diseases of white blood cells	OR=1.27 (0.92 to 1.75)	0.152	0.434	155271	1683	153588	phecode 288	hematopoietic
Non Hodgkins lymphoma	OR=1.32 (0.9 to 1.95)	0.153	0.436	155517	1170	154347	phecode 202.2	neoplasms
Epilepsy; recurrent seizures; convulsions	OR=1.21 (0.93 to 1.57)	0.155	0.438	149628	2525	147103	phecode 345	neurological
Hypotension NOS	OR=0.84 (0.66 to 1.07)	0.158	0.444	148355	2995	145360	phecode 458.9	circulatory system
Dizziness and giddiness Light headedness and vertigo	OR=1.22 (0.92 to 1.62)	0.164	0.457	156589	2216	154373	phecode 386.9	sense organs
Other disorders of bladder	OR=1.13 (0.95 to 1.35)	0.167	0.462	154968	5829	149139	phecode 596	genitourinary
Cardiac pacemaker in situ	OR=1.25 (0.91 to 1.72)	0.170	0.469	141740	1729	140011	phecode 426.91	circulatory system
Abnormal sputum	OR=1.2 (0.92 to 1.56)	0.173	0.470	157211	2581	154630	phecode 516	respiratory
Effects radiation NOS	OR=1.24 (0.91 to 1.68)	0.174	0.470	155439	1852	153587	phecode 990	injuries and poisonings
Acute pancreatitis	OR=1.34 (0.88 to 2.06)	0.175	0.472	156801	950	155851	phecode 577.1	digestive
Rheumatic disease of the heart valves	OR=0.85 (0.68 to 1.08)	0.177	0.472	152296	3296	149000	phecode 394	circulatory system
Heart valve replaced	OR=0.75 (0.5 to 1.14)	0.178	0.472	150046	1046	149000	phecode 395.6	circulatory system
Enthesopathy	OR=1.15 (0.93 to 1.42)	0.184	0.483	147072	4046	143026	phecode 726.1	musculoskeletal

Other hypertrophic and atrophic conditions of skin	OR=1.24 (0.9 to 1.69)	0.186	0.483	157153	1777	155376	phecode 701	dermatologic
Bacterial infection NOS	OR=1.13 (0.94 to 1.34)	0.186	0.483	155365	5876	149489	phecode 041	infectious diseases
Varicose veins	OR=1.16 (0.93 to 1.44)	0.187	0.483	144964	3691	141273	phecode 454	circulatory system
Cancer of other lymphoid; histiocytic tissue	OR=1.26 (0.89 to 1.79)	0.190	0.483	155784	1437	154347	phecode 202	neoplasms
Colon cancer	OR=1.22 (0.91 to 1.64)	0.191	0.483	118823	1997	116826	phecode 153.2	neoplasms
Atopicorcontact dermatitis due to other or unspecified	OR=1.31 (0.87 to 1.98)	0.191	0.483	155217	1029	154188	phecode 939	dermatologic
Chronic renal failure CKD	OR=1.17 (0.92 to 1.48)	0.191	0.483	152156	3242	148914	phecode 585.3	genitourinary
Chronic ulcer of skin	OR=1.26 (0.89 to 1.79)	0.192	0.483	157211	1420	155791	phecode 707	dermatologic
Pulmonary heart disease	OR=1.19 (0.91 to 1.55)	0.198	0.494	155334	2518	152816	phecode 415	circulatory system
Benign neoplasm of colon	OR=1.09 (0.96 to 1.24)	0.200	0.494	128626	11711	116915	phecode 208	neoplasms
Other derangement of joint	OR=1.3 (0.87 to 1.95)	0.202	0.494	155391	1053	154338	phecode 742.9	musculoskeletal
Bacterial enteritis	OR=1.28 (0.87 to 1.89)	0.203	0.494	153897	1169	152728	phecode 008.5	infectious diseases
Complications of transplants and reattached limbs	OR=1.1 (0.95 to 1.27)	0.205	0.494	154168	8858	145310	phecode 851	injuries and poisonings
Secondary malignancy of lymph nodes	OR=1.21 (0.9 to 1.61)	0.206	0.494	139898	2091	137807	phecode 198.1	neoplasms
Retinal detachments and defects	OR=0.82 (0.6 to 1.12)	0.206	0.494	152471	1845	150626	phecode 361	sense organs
Retinal detachment with retinal defect	OR=0.82 (0.6 to 1.12)	0.206	0.494	152471	1845	150626	phecode 361.1	sense organs
Abdominal pain	OR=1.08 (0.96 to 1.22)	0.207	0.494	157211	13297	143914	phecode 785	symptoms
Osteoporosis; osteopenia and pathological fracture	OR=0.8 (0.57 to 1.13)	0.208	0.494	157211	1482	155729	phecode 743	musculoskeletal
Fracture of upper limb	OR=0.86 (0.68 to 1.09)	0.208	0.494	152093	3153	148940	phecode 803	injuries and poisonings
Redundant prepuce and phimosisorBXO	OR=1.18 (0.91 to 1.54)	0.213	0.502	141251	2493	138758	phecode 604.1	genitourinary
Nerve root and plexus disorders	OR=1.21 (0.9 to 1.62)	0.214	0.502	153006	2000	151006	phecode 353	neurological
Other disorders of the kidney and ureters	OR=1.2 (0.9 to 1.61)	0.217	0.505	150972	2058	148914	phecode 586	genitourinary
Duodenitis	OR=1.14 (0.93 to 1.4)	0.219	0.509	146811	4217	142594	phecode 535.6	digestive
Left bundle branch block	OR=1.26 (0.86 to 1.85)	0.228	0.527	141209	1198	140011	phecode 426.32	circulatory system
Fracture of tibia and fibula	OR=0.77 (0.5 to 1.18)	0.230	0.527	149884	944	148940	phecode 800.3	injuries and poisonings
Renal colic	OR=0.79 (0.53 to 1.17)	0.233	0.532	153392	1137	152255	phecode 594.8	genitourinary
Malignant neoplasm of rectum; rectosigmoid junction; and anus	OR=1.23 (0.87 to 1.73)	0.235	0.535	118320	1494	116826	phecode 153.3	neoplasms
Atherosclerosis	OR=1.28 (0.85 to 1.92)	0.242	0.547	152626	1051	151575	phecode 440	circulatory system
Cataract	OR=1.09 (0.94 to 1.25)	0.246	0.551	157211	9843	147368	phecode 366	sense organs
Fracture of vertebral column without mention of spinal cord injury	OR=0.77 (0.5 to 1.2)	0.247	0.551	149858	918	148940	phecode 805	injuries and poisonings
Other retinal disorders	OR=1.17 (0.9 to 1.52)	0.247	0.551	96487	2665	93822	phecode 362	sense organs
Other disorders of soft tissues	OR=0.87 (0.69 to 1.1)	0.250	0.554	146197	3171	143026	phecode 729	musculoskeletal
Candidiasis	OR=1.26 (0.84 to 1.89)	0.254	0.560	156902	1064	155838	phecode 112	infectious diseases
Coronary atherosclerosis	OR=1.07 (0.95 to 1.2)	0.257	0.563	151051	14809	136242	phecode 411.4	circulatory system
Reflux esophagitis	OR=1.11 (0.92 to 1.34)	0.260	0.568	142036	5151	136885	phecode 530.14	digestive
Hemoptysis	OR=1.16 (0.89 to 1.51)	0.275	0.598	157105	2475	154630	phecode 516.1	respiratory
Tachycardia NOS	OR=0.81 (0.55 to 1.19)	0.278	0.601	141157	1146	140011	phecode 427.7	circulatory system
Frequency of urination and polyuria	OR=1.18 (0.87 to 1.59)	0.283	0.605	146763	1962	144801	phecode 599.5	genitourinary
Disturbance of skin sensation	OR=1.23 (0.84 to 1.79)	0.283	0.605	155152	1220	153932	phecode 687.4	dermatologic
Pain in joint	OR=1.14 (0.9 to 1.44)	0.285	0.605	157211	3215	153996	phecode 745	musculoskeletal
Secondary malignant neoplasm	OR=1.12 (0.91 to 1.37)	0.285	0.605	142238	4431	137807	phecode 198	neoplasms
Angina pectoris	OR=1.07 (0.94 to 1.23)	0.291	0.614	147250	11008	136242	phecode 411.3	circulatory system
Dysphagia	OR=0.88 (0.68 to 1.12)	0.293	0.614	139768	2883	136885	phecode 532	digestive
Superficial injury without mention of infection	OR=1.14 (0.89 to 1.47)	0.294	0.614	156951	2791	154160	phecode 915	injuries and poisonings
Decreased white blood cell count	OR=1.21 (0.85 to 1.72)	0.301	0.623	154969	1381	153588	phecode 288.1	hematopoietic
Neutropenia	OR=1.21 (0.85 to 1.72)	0.301	0.623	154969	1381	153588	phecode 288.11	hematopoietic
Colorectal cancer	OR=1.14 (0.89 to 1.45)	0.304	0.625	119810	2984	116826	phecode 153	neoplasms
Diabetes mellitus	OR=1.07 (0.94 to 1.21)	0.305	0.625	157211	12038	145173	phecode 250	endocrine metabolic
Fracture of radius and ulna	OR=0.84 (0.6 to 1.18)	0.306	0.625	150452	1512	148940	phecode 803.2	injuries and poisonings
Diabetic retinopathy	OR=1.25 (0.81 to 1.95)	0.311	0.628	94735	913	93822	phecode 250.7	endocrine metabolic
Calculus of kidney	OR=1.16 (0.87 to 1.56)	0.312	0.628	154296	2041	152255	phecode 594.1	genitourinary
Diseases of the larynx and vocal cords	OR=1.22 (0.83 to 1.8)	0.312	0.628	149221	1164	148057	phecode 473	respiratory
Peripheral vascular disease	OR=1.15 (0.88 to 1.51)	0.314	0.628	153963	2388	151575	phecode 443	circulatory system
Derangement of joint; non traumatic	OR=1.21 (0.84 to 1.75)	0.315	0.628	155611	1273	154338	phecode 742	musculoskeletal
Degenerative disease of the spinal cord	OR=1.23 (0.82 to 1.84)	0.316	0.628	148165	1062	147103	phecode 334	neurological
Other diseases of the teeth and supporting structures	OR=1.22 (0.83 to 1.79)	0.318	0.628	154150	1161	152989	phecode 525	digestive
Pericarditis	OR=0.8 (0.51 to 1.25)	0.322	0.634	155919	879	155040	phecode 420.2	circulatory system
Abnormal heart sounds	OR=0.9 (0.74 to 1.1)	0.325	0.637	153532	4532	149000	phecode 396	circulatory system
Cerebral ischemia	OR=1.14 (0.88 to 1.48)	0.329	0.641	153702	2587	151115	phecode 433.3	circulatory system
Other anemias	OR=0.91 (0.76 to 1.1)	0.333	0.647	154546	5218	149328	phecode 285	hematopoietic
Cancer of bronchus; lung	OR=0.84 (0.6 to 1.19)	0.337	0.649	156871	1458	155413	phecode 165.1	neoplasms
Other symptoms of respiratory system	OR=1.09 (0.92 to 1.29)	0.337	0.649	157211	6102	151109	phecode 512	respiratory
Ileostomy status	OR=1.23 (0.8 to 1.91)	0.345	0.661	126476	914	125562	phecode 559	digestive
Benign neoplasm of other parts of digestive system	OR=1.14 (0.86 to 1.51)	0.350	0.661	157211	2238	154973	phecode 211	neoplasms
Precordial pain	OR=0.86 (0.63 to 1.18)	0.351	0.661	144876	1843	143033	phecode 418.1	circulatory system
Hemorrhage or hematoma complicating a procedure	OR=1.14 (0.87 to 1.5)	0.352	0.661	147619	2309	145310	phecode 850	injuries and poisonings
Hypertensive heart andoror renal disease	OR=1.21 (0.81 to 1.81)	0.355	0.661	117021	1064	115957	phecode 401.2	circulatory system
Blood in stool	OR=1.18 (0.83 to 1.66)	0.355	0.661	147799	1457	146342	phecode 578.2	digestive
Cholelithiasis with other cholecystitis	OR=1.17 (0.84 to 1.64)	0.355	0.661	153385	1525	151860	phecode 574.12	digestive
Senile cataract	OR=1.1 (0.9 to 1.35)	0.355	0.661	151653	4285	147368	phecode 366.2	sense organs
Chemotherapy	OR=1.08 (0.92 to 1.26)	0.362	0.669	144944	7137	137807	phecode 197	neoplasms
Fracture of ankle and foot	OR=0.81 (0.52 to 1.27)	0.364	0.669	149820	880	148940	phecode 801	injuries and poisonings
Dislocation	OR=1.2 (0.81 to 1.79)	0.364	0.669	149324	1082	148242	phecode 830	injuries and poisonings
Hypovolemia	OR=1.16 (0.84 to 1.58)	0.369	0.676	154337	1763	152574	phecode 276.5	endocrine metabolic
Urate	1.102 (-1.318 to 3.523)	0.372	0.678	149755	NA	NA	NA	biomarker
Ulcer of esophagus	OR=1.12 (0.87 to 1.43)	0.375	0.680	139816	2931	136885	phecode 530.12	digestive
Swelling of limb	OR=0.88 (0.66 to 1.17)	0.380	0.686	156888	2127	154761	phecode 771.1	symptoms
Cardiac dysrhythmias	OR=1.05 (0.94 to 1.18)	0.388	0.698	154977	14966	140011	phecode 427	circulatory system
Appendiceal conditions	OR=0.86 (0.62 to 1.21)	0.401	0.718	157211	1516	155695	phecode 540	digestive
Urinary incontinence	OR=1.17 (0.81 to 1.71)	0.402	0.718	146048	1247	144801	phecode 599.4	genitourinary
Cardiac arrest and ventricular fibrillation	OR=1.2 (0.78 to 1.85)	0.404	0.719	140944	933	140011	phecode 427.4	circulatory system
Disorders of penis	OR=1.11 (0.87 to 1.41)	0.409	0.724	141790	3032	138758	phecode 604	genitourinary
Acquired foot deformities	OR=0.86 (0.61 to 1.23)	0.418	0.737	155349	1384	153965	phecode 735	musculoskeletal
Acute renal failure	OR=1.09 (0.88 to 1.35)	0.420	0.737	152907	3993	148914	phecode 585.1	genitourinary

Aortic aneurysm	OR=1.16 (0.81 to 1.67)	0.423	0.737	152901	1326	151575	phecode 442.1	circulatory system
Calculus of bile duct	OR=0.84 (0.54 to 1.3)	0.424	0.737	152759	899	151860	phecode 574.2	digestive
Complication due to other implant and internal device	OR=1.14 (0.83 to 1.58)	0.426	0.737	146977	1667	145310	phecode 859	injuries and poisonings
Diseases of hair and hair follicles	OR=0.9 (0.68 to 1.18)	0.428	0.737	156720	2347	154373	phecode 704	dermatologic
Skull and face fracture and other intercranial injury	OR=1.13 (0.84 to 1.52)	0.429	0.737	157211	1949	155262	phecode 819	injuries and poisonings
Ischemic Heart Disease	OR=1.04 (0.94 to 1.15)	0.429	0.737	156847	20605	136242	phecode 411	circulatory system
Localized superficial swelling; mass; or lump	OR=1.19 (0.76 to 1.86)	0.439	0.751	154808	876	153932	phecode 687.2	dermatologic
Other chronic ischemic heart disease; unspecified	OR=1.04 (0.94 to 1.16)	0.447	0.759	154252	18010	136242	phecode 411.8	circulatory system
Epistaxis or throat hemorrhage	OR=1.15 (0.8 to 1.64)	0.447	0.759	149441	1384	148057	phecode 477	respiratory
Disorders of fluid; electrolyte; and acid base balance	OR=1.08 (0.89 to 1.31)	0.450	0.759	157211	4637	152574	phecode 276	endocrine metabolic
Electrolyte imbalance	OR=1.1 (0.85 to 1.43)	0.454	0.759	155178	2604	152574	phecode 276.1	endocrine metabolic
Open wounds of extremities	OR=1.1 (0.86 to 1.41)	0.454	0.759	154725	2867	151858	phecode 871	injuries and poisonings
Bacterial pneumonia	OR=0.92 (0.75 to 1.14)	0.455	0.759	154608	4193	150415	phecode 480.1	respiratory
Cystitis and urethritis	OR=1.16 (0.78 to 1.72)	0.457	0.759	145262	1113	144149	phecode 592	genitourinary
Cerebral artery occlusion; with cerebral infarction	OR=0.87 (0.6 to 1.26)	0.460	0.759	152417	1302	151115	phecode 433.21	circulatory system
Carbuncle and furuncle	OR=0.87 (0.59 to 1.27)	0.461	0.759	152286	1194	151092	phecode 686.1	dermatologic
Paroxysmal supraventricular tachycardia	OR=0.86 (0.58 to 1.28)	0.463	0.759	141140	1129	140011	phecode 427.11	circulatory system
Disorders of refraction and accommodation; blindness and low vision	OR=1.14 (0.8 to 1.62)	0.465	0.759	157211	1411	155800	phecode 367	sense organs
Symptoms and disorders of the joints	OR=1.13 (0.82 to 1.55)	0.467	0.759	156044	1706	154338	phecode 741	musculoskeletal
Varicose veins of lower extremity	OR=1.09 (0.87 to 1.37)	0.467	0.759	144622	3349	141273	phecode 454.1	circulatory system
Appendicitis	OR=0.88 (0.62 to 1.24)	0.470	0.761	157158	1463	155695	phecode 540.1	digestive
Shortness of breath	OR=1.1 (0.85 to 1.4)	0.472	0.762	153980	2871	151109	phecode 512.7	respiratory
Open wounds of head; neck; and trunk	OR=1.11 (0.84 to 1.46)	0.474	0.762	154137	2279	151858	phecode 870	injuries and poisonings
Malignant neoplasm of other and ill defined sites within the digestive organs a	OR=1.09 (0.86 to 1.38)	0.480	0.768	118550	3285	115265	phecode 159	neoplasms
Neurological disorders	OR=1.09 (0.85 to 1.39)	0.482	0.768	155814	2943	152871	phecode 292	mental disorders
Peptic ulcer excl esophageal	OR=0.93 (0.75 to 1.15)	0.483	0.768	157211	3842	153369	phecode 531	digestive
Cancer of bladder	OR=0.9 (0.66 to 1.22)	0.487	0.773	156022	1868	154154	phecode 189.2	neoplasms
Other acquired deformities of limbs	OR=1.16 (0.75 to 1.8)	0.492	0.774	154886	921	153965	phecode 736	musculoskeletal
Musculoskeletal symptoms referable to limbs	OR=0.91 (0.7 to 1.19)	0.493	0.774	157211	2450	154761	phecode 771	symptoms
Abnormal movement	OR=1.11 (0.82 to 1.5)	0.495	0.774	157211	1915	155296	phecode 350	neurological
Emphysema	OR=1.15 (0.78 to 1.69)	0.495	0.774	144084	1145	142939	phecode 496.1	respiratory
Hereditary retinal dystrophies	OR=1.16 (0.75 to 1.79)	0.497	0.774	94759	937	93822	phecode 362.7	sense organs
Vertiginous syndromes and other disorders of vestibular system	OR=1.09 (0.85 to 1.4)	0.498	0.774	157211	2838	154373	phecode 386	sense organs
Pain in limb	OR=1.1 (0.84 to 1.44)	0.508	0.787	157211	2384	154827	phecode 773	symptoms
Renal failure NOS	OR=1.15 (0.75 to 1.77)	0.510	0.787	149863	949	148914	phecode 585.2	genitourinary
Acquired toe deformities	OR=0.87 (0.56 to 1.34)	0.517	0.790	154873	908	153965	phecode 735.2	musculoskeletal
Sepsis	OR=1.1 (0.82 to 1.49)	0.518	0.790	157211	1954	155257	phecode 994.2	injuries and poisonings
Sepsis and SIRS	OR=1.1 (0.82 to 1.49)	0.518	0.790	157211	1954	155257	phecode 994	injuries and poisonings
Acute appendicitis	OR=0.89 (0.61 to 1.28)	0.520	0.790	156962	1267	155695	phecode 540.11	digestive
Ganglion and cyst of synovium; tendon; and bursa	OR=1.15 (0.75 to 1.76)	0.523	0.793	143973	947	143026	phecode 727.4	musculoskeletal
Hematuria	OR=1.05 (0.9 to 1.22)	0.527	0.796	152438	8289	144149	phecode 593	genitourinary
Alcohol related disorders	OR=1.05 (0.9 to 1.22)	0.532	0.797	149004	8241	140763	phecode 317	mental disorders
Inflammatory and toxic neuropathy	OR=0.88 (0.6 to 1.31)	0.532	0.797	156804	1127	155677	phecode 357	neurological
Pleurisy; pleural effusion	OR=1.07 (0.87 to 1.32)	0.534	0.797	153858	4080	149778	phecode 507	respiratory
Other specified cardiac dysrhythmias	OR=0.91 (0.69 to 1.21)	0.535	0.797	142197	2186	140011	phecode 427.3	circulatory system
Fracture of clavicle or scapula	OR=0.88 (0.59 to 1.31)	0.537	0.798	150048	1108	148940	phecode 803.3	injuries and poisonings
Ulcerative colitis	OR=0.9 (0.65 to 1.26)	0.544	0.804	127122	1560	125562	phecode 555.2	digestive
Hematemesis	OR=1.13 (0.76 to 1.69)	0.548	0.804	147411	1069	146342	phecode 578.1	digestive
Superficial cellulitis and abscess	OR=1.06 (0.87 to 1.31)	0.549	0.804	155296	4204	151092	phecode 681	dermatologic
Nasal polyps	OR=1.1 (0.81 to 1.48)	0.549	0.804	150012	1955	148057	phecode 471	respiratory
Other disorders of arteries and arterioles	OR=1.13 (0.75 to 1.72)	0.553	0.807	152579	1004	151575	phecode 447	circulatory system
Symptoms affecting skin	OR=1.07 (0.85 to 1.35)	0.557	0.808	157211	3279	153932	phecode 687	dermatologic
Neoplasm of uncertain behavior	OR=0.9 (0.64 to 1.27)	0.558	0.808	139321	1514	137807	phecode 199	neoplasms
Palpitations	OR=0.9 (0.62 to 1.29)	0.560	0.808	141308	1297	140011	phecode 427.9	circulatory system
Thrombocytopenia	OR=1.14 (0.74 to 1.76)	0.562	0.808	156674	918	155756	phecode 287.3	hematopoietic
Unspecified diffuse connective tissue disease	OR=0.89 (0.59 to 1.34)	0.564	0.808	145036	1019	144017	phecode 709.7	dermatologic
Fracture of ribs	OR=0.89 (0.59 to 1.33)	0.565	0.808	149986	1046	148940	phecode 807	injuries and poisonings
Retention of urine	OR=1.05 (0.88 to 1.26)	0.568	0.808	150344	5543	144801	phecode 599.2	genitourinary
Substance addiction and disorders	OR=1.06 (0.86 to 1.32)	0.569	0.808	144599	3836	140763	phecode 316	mental disorders
Atrioventricular AV block	OR=1.1 (0.8 to 1.51)	0.570	0.808	141731	1720	140011	phecode 426.2	circulatory system
Apolipoprotein B	-0.002 (-0.01 to 0.006)	0.582	0.823	148956	NA	NA	NA	biomarker
Other disorders of stomach and duodenum	OR=1.07 (0.85 to 1.34)	0.588	0.828	145936	3342	142594	phecode 537	digestive
Secondary malignancy of respiratory organs	OR=0.9 (0.62 to 1.31)	0.592	0.831	139057	1250	137807	phecode 198.2	neoplasms
Pneumococcal pneumonia	OR=0.94 (0.76 to 1.17)	0.595	0.832	154129	3714	150415	phecode 480.11	respiratory
Hyposmolality andoror hyponatremia	OR=1.11 (0.74 to 1.66)	0.605	0.844	153659	1085	152574	phecode 276.12	endocrine metabolic
Constipation	OR=1.05 (0.86 to 1.29)	0.608	0.846	130028	4466	125562	phecode 563	digestive
Aspartate aminotransferase	0.009 (-0.025 to 0.043)	0.611	0.847	149354	NA	NA	NA	biomarker
Septicemia	OR=1.07 (0.83 to 1.38)	0.615	0.850	152151	2662	149489	phecode 038	infectious diseases
Ventral hernia	OR=1.08 (0.79 to 1.49)	0.620	0.854	131993	1754	130239	phecode 550.5	digestive
Other non epithelial cancer of skin	OR=1.04 (0.9 to 1.19)	0.623	0.855	155540	10010	145530	phecode 172.2	neoplasms
Other abnormal blood chemistry	OR=1.05 (0.87 to 1.26)	0.634	0.868	157098	5126	151972	phecode 790.6	symptoms
Overweight; obesity and other hyperalimentation	OR=1.04 (0.87 to 1.24)	0.645	0.880	157211	5734	151477	phecode 278	endocrine metabolic
Other aneurysm	OR=1.08 (0.77 to 1.51)	0.649	0.880	153158	1583	151575	phecode 442	circulatory system
Streptococcus infection	OR=1.11 (0.72 to 1.71)	0.650	0.880	150400	911	149489	phecode 041.2	infectious diseases
Cholecystitis without cholelithiasis	OR=0.91 (0.6 to 1.38)	0.651	0.880	152857	997	151860	phecode 574.3	digestive
Nonspecific findings on examination of blood	OR=1.04 (0.87 to 1.26)	0.654	0.880	157211	5239	151972	phecode 790	symptoms
Cancer of urinary organs incl kidney and bladder	OR=1.06 (0.83 to 1.34)	0.655	0.880	157211	3057	154154	phecode 189	neoplasms
Viral infection	OR=1.1 (0.73 to 1.64)	0.660	0.881	155858	1051	154807	phecode 079	infectious diseases
Diseases of sebaceous glands	OR=0.95 (0.76 to 1.19)	0.661	0.881	156848	3588	153260	phecode 706	dermatologic
Diffuse diseases of connective tissue	OR=0.92 (0.62 to 1.36)	0.663	0.881	145150	1133	144017	phecode 709	dermatologic
Cholelithiasis	OR=1.05 (0.85 to 1.29)	0.664	0.881	156001	4141	151860	phecode 574.1	digestive
Abnormal results of function study of liver	OR=1.07 (0.78 to 1.46)	0.667	0.881	154132	1811	152321	phecode 573.7	digestive
Obesity	OR=1.04 (0.87 to 1.24)	0.668	0.881	157184	5707	151477	phecode 278.1	endocrine metabolic
Fracture of hand or wrist	OR=0.93 (0.67 to 1.29)	0.673	0.881	150596	1656	148940	phecode 804	injuries and poisonings

Diverticulosis	OR=0.97 (0.86 to 1.1)	0.673	0.881	139014	13452	125562	phecode 562.1	digestive
Diverticulosis and diverticulitis	OR=0.97 (0.86 to 1.1)	0.673	0.881	139014	13452	125562	phecode 562	digestive
Cardiac conduction disorders	OR=1.04 (0.86 to 1.26)	0.683	0.891	145047	5036	140011	phecode 426	circulatory system
Sebaceous cyst	OR=0.96 (0.76 to 1.19)	0.687	0.891	156823	3563	153260	phecode 706.2	dermatologic
Inflammation of eyelids	OR=1.09 (0.73 to 1.62)	0.687	0.891	154216	1085	153131	phecode 371.3	sense organs
Complication of internal orthopedic device	OR=1.06 (0.8 to 1.41)	0.701	0.905	147475	2165	145310	phecode 858	injuries and poisonings
Skin cancer	OR=1.03 (0.9 to 1.17)	0.704	0.905	157141	11611	145530	phecode 172	neoplasms
Leukemia	OR=0.93 (0.64 to 1.36)	0.706	0.905	155553	1206	154347	phecode 204	neoplasms
Nausea and vomiting	OR=0.96 (0.78 to 1.18)	0.706	0.905	157211	4168	153043	phecode 789	symptoms
Inflammation of the eye	OR=1.07 (0.75 to 1.53)	0.710	0.907	154480	1349	153131	phecode 371	sense organs
Psoriasis vulgaris	OR=0.93 (0.61 to 1.41)	0.724	0.921	143312	981	142331	phecode 696.41	dermatologic
Paroxysmal tachycardia; unspecified	OR=0.94 (0.69 to 1.29)	0.724	0.921	141775	1764	140011	phecode 427.1	circulatory system
Malignant neoplasm of bladder	OR=0.94 (0.68 to 1.31)	0.728	0.923	155780	1626	154154	phecode 189.21	neoplasms
Hearing loss	OR=0.96 (0.74 to 1.24)	0.730	0.923	157185	2573	154612	phecode 389	sense organs
Urea	0.008 (-0.037 to 0.053)	0.742	0.935	149832	NA	NA	NA	biomarker
Respiratory failure	OR=1.06 (0.75 to 1.49)	0.747	0.938	151257	1479	149778	phecode 509.1	respiratory
Cancer within the respiratory system	OR=0.95 (0.69 to 1.3)	0.752	0.938	157193	1780	155413	phecode 165	neoplasms
Aphakia and other disorders of lens	OR=0.93 (0.61 to 1.43)	0.753	0.938	154756	969	153787	phecode 379.3	sense organs
Staphylococcus infections	OR=1.05 (0.77 to 1.43)	0.753	0.938	151291	1802	149489	phecode 041.1	infectious diseases
Glaucoma	OR=1.04 (0.8 to 1.36)	0.756	0.939	153095	2469	150626	phecode 365	sense organs
Iron deficiency anemias	OR=0.97 (0.77 to 1.21)	0.763	0.940	152701	3373	149328	phecode 280	hematopoietic
Iron deficiency anemias; unspecified or not due to blood loss	OR=0.97 (0.77 to 1.21)	0.763	0.940	152701	3373	149328	phecode 280.1	hematopoietic
Calculus of ureter	OR=1.05 (0.76 to 1.46)	0.765	0.940	153872	1617	152255	phecode 594.3	genitourinary
Degeneration of intervertebral disc	OR=1.06 (0.72 to 1.57)	0.765	0.940	152067	1148	150919	phecode 722.6	musculoskeletal
Chronic liver disease and cirrhosis	OR=0.96 (0.7 to 1.3)	0.772	0.945	154165	1844	152321	phecode 571	digestive
Other forms of chronic heart disease	OR=1.05 (0.75 to 1.46)	0.781	0.950	137837	1595	136242	phecode 414	circulatory system
Occlusion and stenosis of precerebral arteries	OR=1.06 (0.69 to 1.64)	0.782	0.950	152030	915	151115	phecode 433.1	circulatory system
Bundle branch block	OR=0.96 (0.74 to 1.26)	0.789	0.950	142464	2453	140011	phecode 426.3	circulatory system
Lipoprotein A	-0.005 (-0.043 to 0.033)	0.789	0.950	118783	NA	NA	NA	biomarker
Benign neoplasm of skin	OR=1.03 (0.81 to 1.32)	0.790	0.950	156940	2940	154000	phecode 216	neoplasms
Melanomas of skin	OR=1.04 (0.76 to 1.42)	0.792	0.950	147340	1810	145530	phecode 172.11	neoplasms
Melanomas of skin; dx or hx	OR=1.04 (0.76 to 1.42)	0.792	0.950	147340	1810	145530	phecode 172.1	neoplasms
Triglycerides	0.005 (-0.029 to 0.038)	0.793	0.950	149776	NA	NA	NA	biomarker
Urinary calculus	OR=1.03 (0.84 to 1.26)	0.796	0.950	156430	4175	152255	phecode 594	genitourinary
Acute upper respiratory infections of multiple or unspecified sites	OR=1.06 (0.69 to 1.62)	0.798	0.950	157150	942	156208	phecode 465	respiratory
Other open wound of head and face	OR=1.04 (0.77 to 1.4)	0.801	0.950	153838	1980	151858	phecode 870.3	injuries and poisonings
Other abnormality of urination	OR=1.05 (0.71 to 1.56)	0.803	0.950	145926	1125	144801	phecode 599.9	genitourinary
Diseases of hard tissues of teeth	OR=1.04 (0.74 to 1.47)	0.803	0.950	154484	1495	152989	phecode 521	digestive
Type 1 diabetes	OR=1.04 (0.74 to 1.46)	0.810	0.953	146702	1529	145173	phecode 250.1	endocrine metabolic
Phlebitis and thrombophlebitis	OR=0.97 (0.73 to 1.28)	0.811	0.953	143446	2173	141273	phecode 451	circulatory system
Dental caries	OR=1.04 (0.74 to 1.47)	0.813	0.953	154465	1476	152989	phecode 521.1	digestive
Rash and other nonspecific skin eruption	OR=0.95 (0.63 to 1.44)	0.817	0.955	154935	1003	153932	phecode 687.1	dermatologic
Complications of cardiacorvascular device; implant; and graft	OR=1.04 (0.75 to 1.44)	0.821	0.955	146946	1636	145310	phecode 854	injuries and poisonings
Stricture and stenosis of esophagus	OR=1.05 (0.71 to 1.54)	0.821	0.955	138053	1168	136885	phecode 530.3	digestive
Pneumonia	OR=0.98 (0.83 to 1.16)	0.825	0.958	157057	6642	150415	phecode 480	respiratory
Other disorders of male genital organs	OR=0.97 (0.75 to 1.26)	0.832	0.958	141295	2537	138758	phecode 608	genitourinary
Abnormal findings on examination of urine	OR=1.03 (0.77 to 1.38)	0.834	0.958	157211	2083	155128	phecode 598	genitourinary
Congenital anomalies of great vessels	OR=1.04 (0.72 to 1.51)	0.837	0.958	156725	1267	155458	phecode 747.13	congenital anomalies
Other disorders of eye	OR=1.03 (0.77 to 1.39)	0.839	0.958	155754	1967	153787	phecode 379	sense organs
Orthostatic hypotension	OR=1.05 (0.68 to 1.61)	0.840	0.958	146303	943	145360	phecode 458.1	circulatory system
Occlusion of cerebral arteries	OR=0.97 (0.76 to 1.25)	0.840	0.958	153888	2773	151115	phecode 433.2	circulatory system
Respiratory failure; insufficiency; arrest	OR=0.97 (0.71 to 1.32)	0.841	0.958	151605	1827	149778	phecode 509	respiratory
Anal and rectal polyp	OR=1.02 (0.83 to 1.25)	0.844	0.958	146264	4298	141966	phecode 565.1	digestive
Benign neoplasm of unspecified sites	OR=0.96 (0.62 to 1.48)	0.847	0.958	157211	900	156311	phecode 229	neoplasms
Disorders of mineral metabolism	OR=0.97 (0.67 to 1.38)	0.850	0.958	157211	1343	155868	phecode 275	endocrine metabolic
Other disorders of testis	OR=0.97 (0.73 to 1.3)	0.851	0.958	140795	2037	138758	phecode 603	genitourinary
Carditis	OR=0.97 (0.68 to 1.38)	0.852	0.958	156416	1376	155040	phecode 420	circulatory system
Cystitis	OR=1.04 (0.69 to 1.56)	0.859	0.961	145195	1046	144149	phecode 592.1	genitourinary
Cholelithiasis and cholecystitis	OR=1.02 (0.84 to 1.24)	0.861	0.961	156526	4666	151860	phecode 574	digestive
Cellulitis and abscess of armorhand	OR=1.02 (0.81 to 1.28)	0.866	0.961	154462	3370	151092	phecode 681.3	dermatologic
Cellulitis and abscess of foot; toe	OR=1.02 (0.81 to 1.28)	0.866	0.961	154462	3370	151092	phecode 681.6	dermatologic
Cellulitis and abscess of leg; except foot	OR=1.02 (0.81 to 1.28)	0.866	0.961	154462	3370	151092	phecode 681.5	dermatologic
Duodenal ulcer	OR=0.98 (0.72 to 1.32)	0.872	0.965	155249	1880	153369	phecode 531.3	digestive
Respiratory insufficiency	OR=1.03 (0.71 to 1.49)	0.876	0.965	151070	1292	149778	phecode 509.2	respiratory
Aortic valve disease	OR=1.03 (0.67 to 1.59)	0.878	0.965	149938	938	149000	phecode 394.3	circulatory system
Cerebrovascular disease	OR=1.01 (0.84 to 1.22)	0.881	0.965	156525	5410	151115	phecode 433	circulatory system
Chronic sinusitis	OR=1.03 (0.7 to 1.53)	0.881	0.965	149183	1126	148057	phecode 475	respiratory
Inflammatory bowel disease and other gastroenteritis and colitis	OR=1.02 (0.76 to 1.37)	0.882	0.965	127644	2082	125562	phecode 555	digestive
Psoriasis and related disorders	OR=0.98 (0.67 to 1.41)	0.896	0.977	143599	1268	142331	phecode 696	dermatologic
Fever of unknown origin	OR=1.02 (0.76 to 1.37)	0.897	0.977	157211	2001	155210	phecode 783	symptoms
Bronchiectasis	OR=1.03 (0.66 to 1.59)	0.911	0.984	143844	905	142939	phecode 496.3	respiratory
Disorder of skin and subcutaneous tissue NOS	OR=0.99 (0.76 to 1.28)	0.911	0.984	157211	2609	154602	phecode 689	dermatologic
Circulatory disease NEC	OR=0.99 (0.86 to 1.14)	0.916	0.984	154635	9275	145360	phecode 459.9	circulatory system
Secondary malignant neoplasm of liver	OR=1.02 (0.72 to 1.43)	0.918	0.984	139299	1492	137807	phecode 198.4	neoplasms
Gastritis and duodenitis	OR=0.99 (0.88 to 1.12)	0.918	0.984	155649	13055	142594	phecode 535	digestive
Symptoms involving skin and other integumentary tissue	OR=1.02 (0.69 to 1.52)	0.921	0.984	157211	1099	156112	phecode 782	symptoms
Other disorders of circulatory system	OR=0.99 (0.86 to 1.14)	0.922	0.984	154746	9386	145360	phecode 459	circulatory system
First degree AV block	OR=1.02 (0.65 to 1.6)	0.924	0.984	140890	879	140011	phecode 426.21	circulatory system
Lymphadenitis	OR=1.02 (0.71 to 1.46)	0.927	0.984	154930	1342	153588	phecode 289.4	hematopoietic
Psoriasis	OR=0.98 (0.68 to 1.43)	0.928	0.984	143566	1235	142331	phecode 696.4	dermatologic
Hydrocele	OR=1.02 (0.7 to 1.47)	0.929	0.984	140023	1265	138758	phecode 603.1	genitourinary
Other disorders of liver	OR=0.99 (0.81 to 1.22)	0.931	0.984	156496	4175	152321	phecode 573	digestive
Abdominal aortic aneurysm	OR=1.02 (0.66 to 1.58)	0.936	0.987	152477	902	151575	phecode 442.11	circulatory system
Irritable Bowel Syndrome	OR=0.99 (0.69 to 1.4)	0.938	0.987	126963	1401	125562	phecode 564.1	digestive

Other specified gastritis	OR=1.01 (0.81 to 1.26)	0.941	0.987	146248	3654	142594	phecode 535.8	digestive
Alcoholism	OR=1.01 (0.84 to 1.2)	0.943	0.987	146499	5736	140763	phecode 317.1	mental disorders
Phlebitis and thrombophlebitis of lower extremities	OR=0.99 (0.74 to 1.33)	0.945	0.987	143312	2039	141273	phecode 451.2	circulatory system
Other headache syndromes	OR=1.01 (0.79 to 1.29)	0.948	0.987	156671	2932	153739	phecode 339	neurological
Symptoms involving nervous and musculoskeletal systems	OR=1.01 (0.8 to 1.27)	0.949	0.987	157211	3214	153997	phecode 781	symptoms
Syncope and collapse	OR=1.01 (0.83 to 1.22)	0.952	0.988	157211	4965	152246	phecode 788	symptoms
Other upper respiratory disease	OR=0.99 (0.73 to 1.34)	0.960	0.988	149974	1917	148057	phecode 479	respiratory
Cardiomegaly	OR=1.01 (0.75 to 1.35)	0.963	0.988	154883	2067	152816	phecode 416	circulatory system
Visual disturbances	OR=1.01 (0.72 to 1.41)	0.963	0.988	157211	1577	155634	phecode 368	sense organs
Abnormality of gait	OR=0.99 (0.69 to 1.43)	0.966	0.988	156620	1324	155296	phecode 350.2	neurological
Other disorders of biliary tract	OR=1.01 (0.66 to 1.55)	0.968	0.988	152791	931	151860	phecode 575.8	digestive
Rheumatoid arthritis and other inflammatory polyarthropathies	OR=1.01 (0.72 to 1.4)	0.968	0.988	157211	1585	155626	phecode 714	musculoskeletal
Other disorders of bone and cartilage	OR=0.99 (0.71 to 1.39)	0.969	0.988	153714	1535	152179	phecode 733	musculoskeletal
Purpura and other hemorrhagic conditions	OR=1.01 (0.67 to 1.52)	0.972	0.988	156790	1034	155756	phecode 287	hematopoietic
Rheumatoid arthritis	OR=1.01 (0.71 to 1.43)	0.974	0.988	157211	1398	155813	phecode 714.1	musculoskeletal
Cardiac congenital anomalies	OR=1.01 (0.73 to 1.39)	0.974	0.988	157103	1645	155458	phecode 747.1	congenital anomalies
Nonrheumatic aortic valve disorders	OR=1 (0.74 to 1.37)	0.980	0.992	150841	1841	149000	phecode 395.2	circulatory system
Pulmonary collapse; interstitial and compensatory emphysema	OR=1 (0.7 to 1.44)	0.987	0.996	151099	1321	149778	phecode 508	respiratory
Edema	OR=1 (0.66 to 1.51)	0.992	0.998	157129	1017	156112	phecode 782.3	symptoms
Other chronic nonalcoholic liver disease	OR=1 (0.7 to 1.43)	0.995	0.998	153654	1333	152321	phecode 571.5	digestive
Bladder neck obstruction	OR=1 (0.72 to 1.38)	0.996	0.998	150802	1663	149139	phecode 596.1	genitourinary
Cardiac and circulatory congenital anomalies	OR=1 (0.73 to 1.37)	0.999	0.999	157211	1753	155458	phecode 747	congenital anomalies

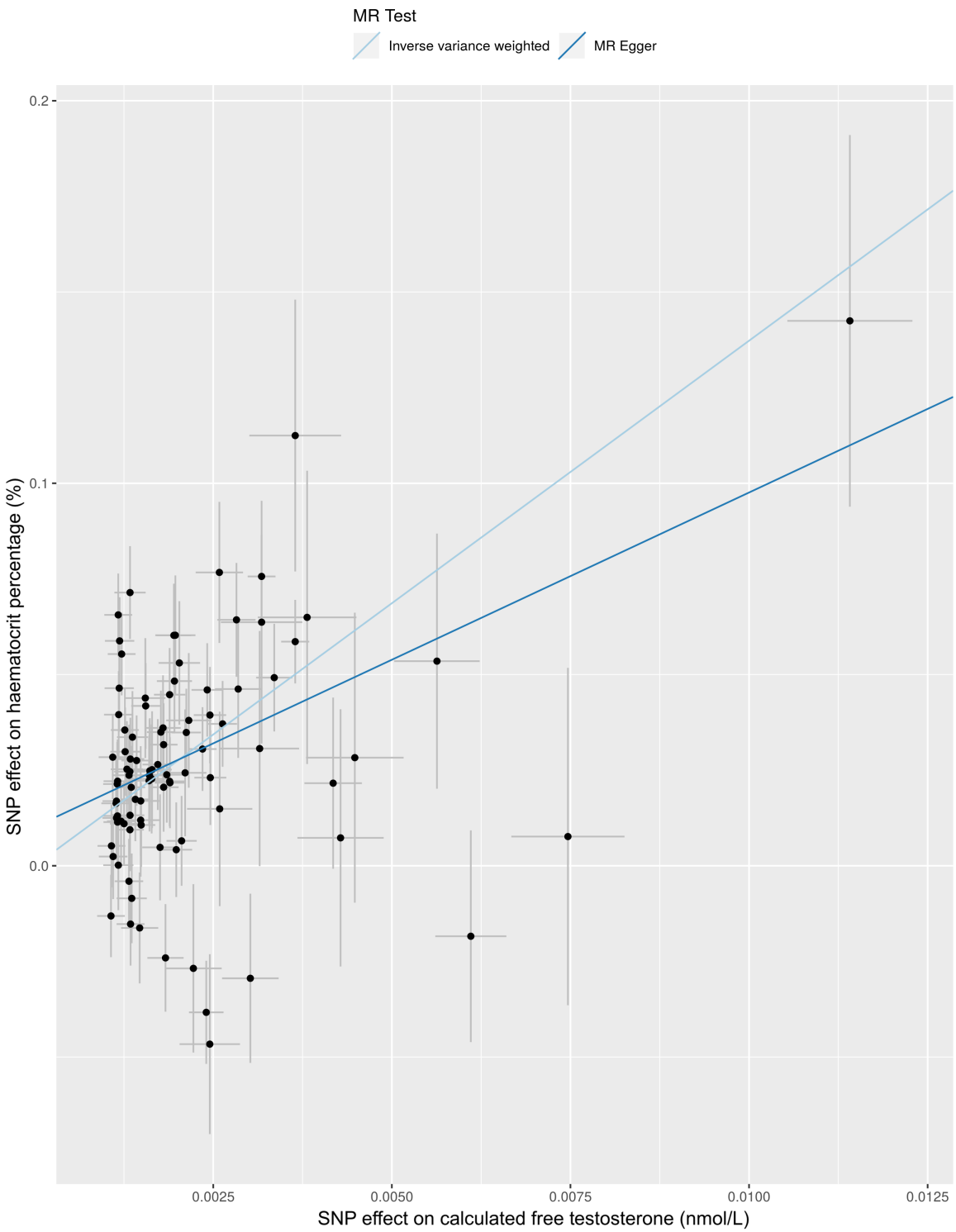


Figure 3 – Figure Supplement 1. Comparison of effect of calculated free testosterone on haematocrit percentage using Mendelian randomization with IVW and Egger regression methods.

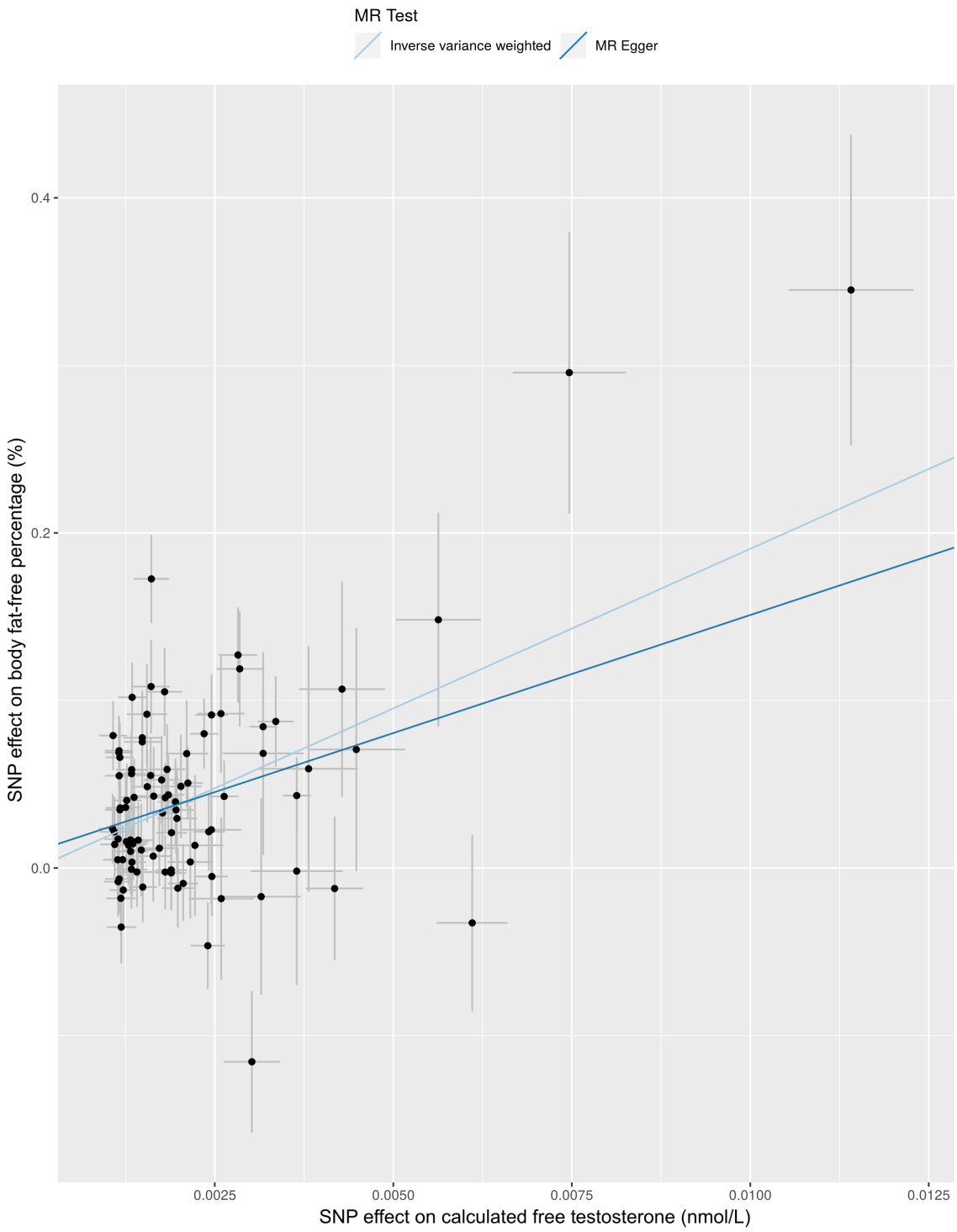


Figure 3 – Figure Supplement 2. Comparison of effect of calculated free testosterone on body fat-free percentage using Mendelian randomization with IVW and Egger regression methods.

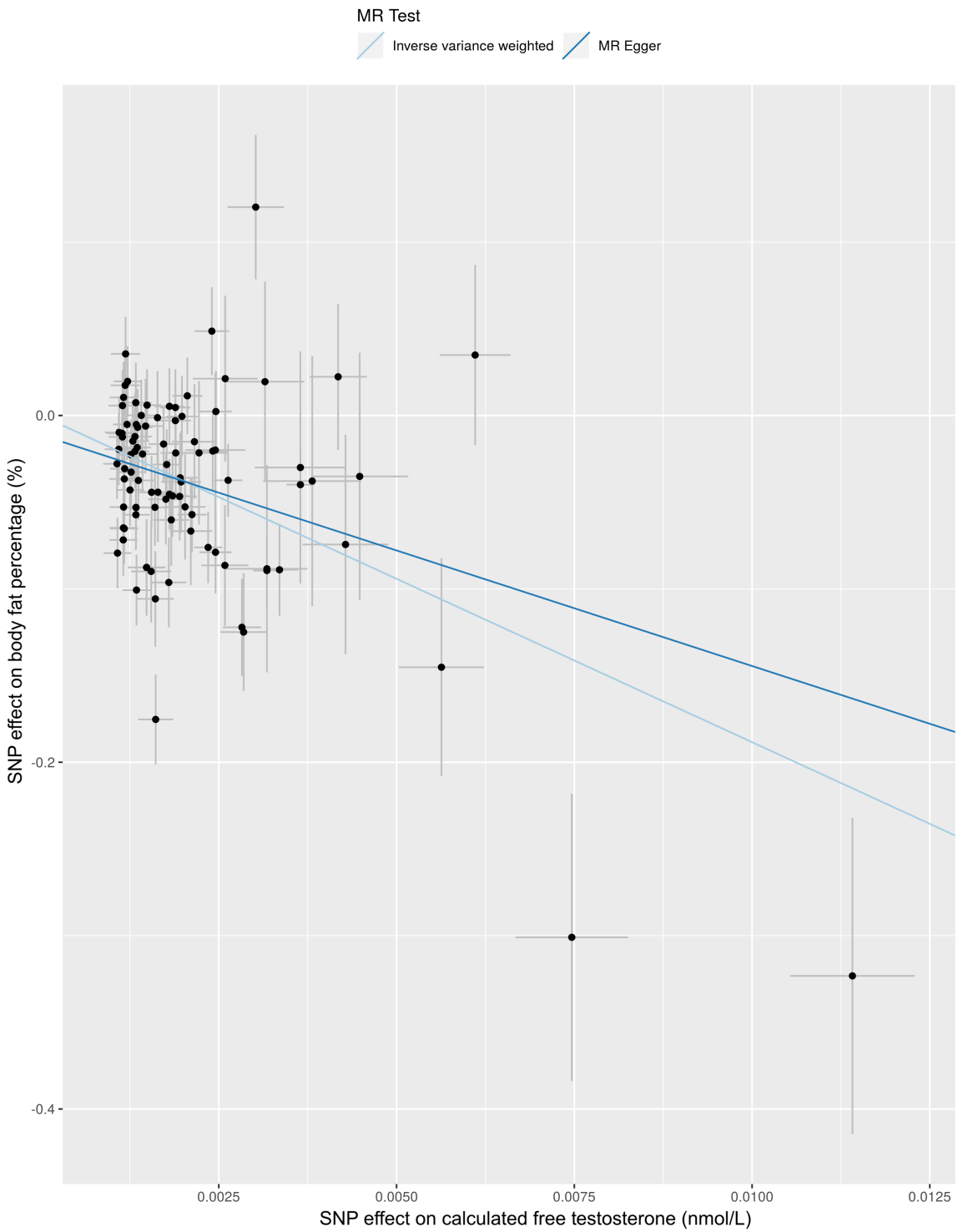


Figure 3 – Figure Supplement 3. Comparison of effect of calculated free testosterone on body fat percentage using Mendelian randomization with IVW and Egger regression methods.

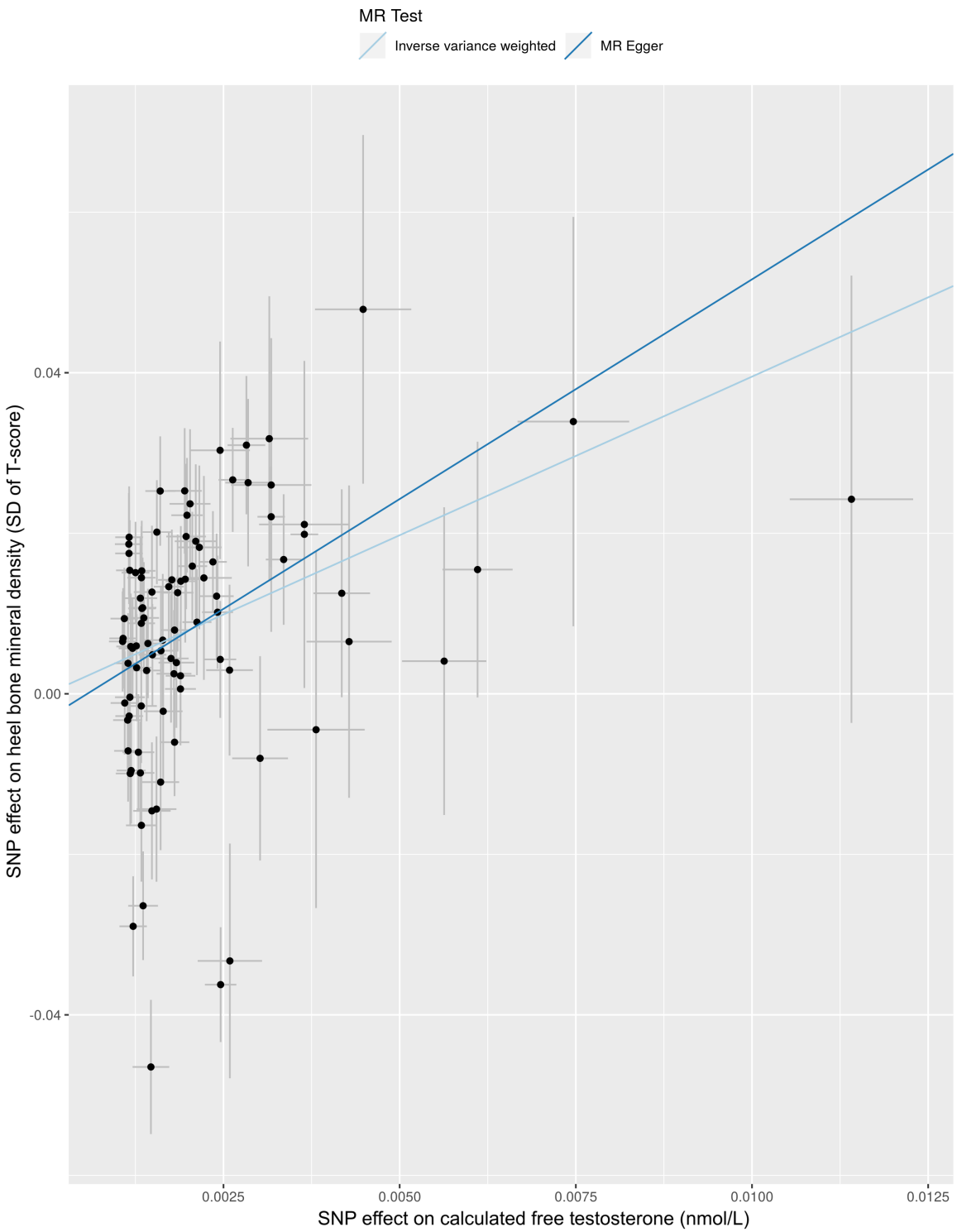


Figure 3 – Figure Supplement 4. Comparison of effect of calculated free testosterone on heel bone mineral density using Mendelian randomization with IVW and Egger regression methods.

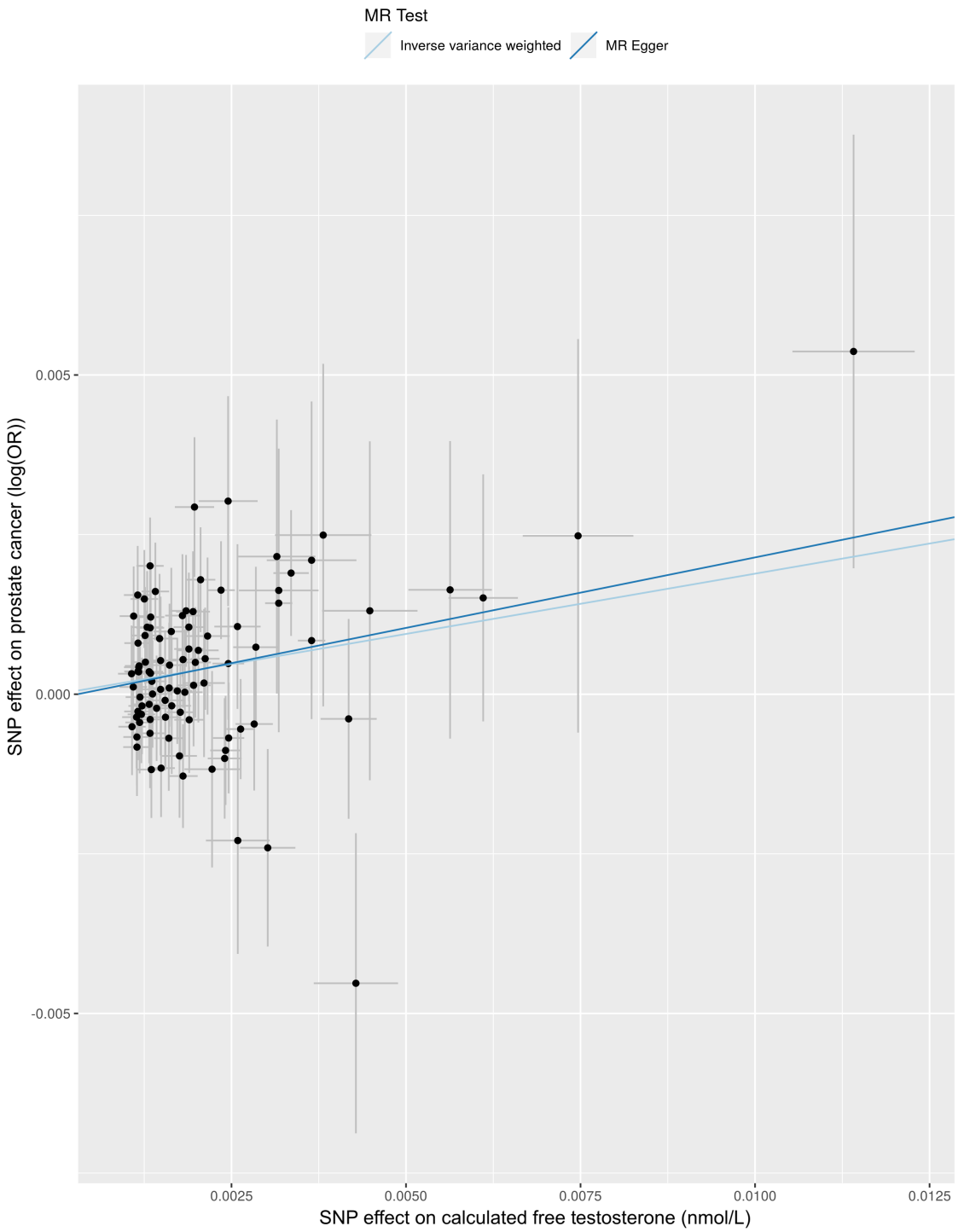


Figure 3 – Figure Supplement 5. Comparison of effect of calculated free testosterone on prostate cancer using Mendelian randomization with IVW and Egger regression methods.

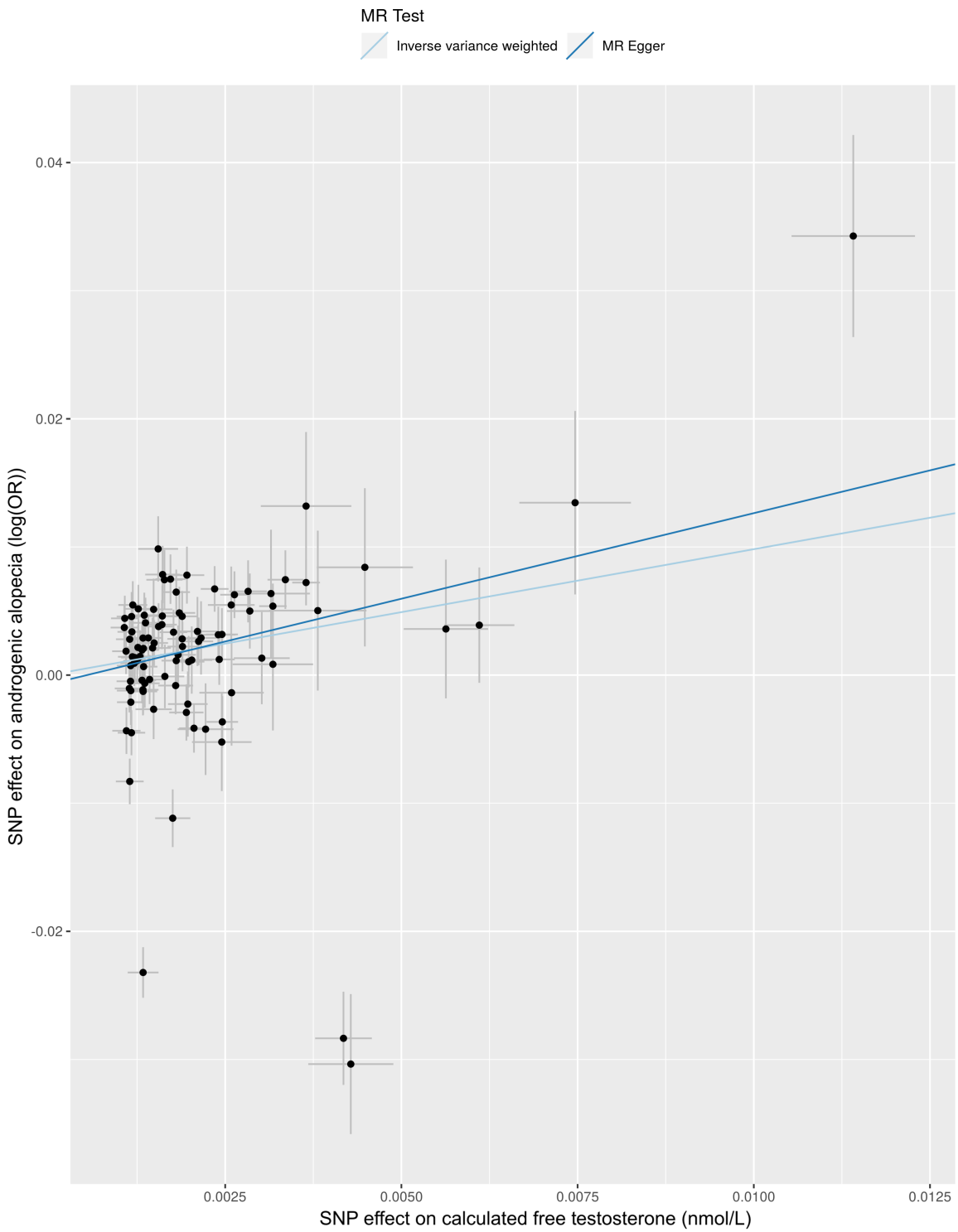


Figure 3 – Figure Supplement 6. Comparison of effect of calculated free testosterone on androgenic alopecia using Mendelian randomization with IVW and Egger regression methods.

Supplementary File 1 - Table 1. Characteristics at recruitment for study population of males from UK Biobank cohort study

Variable	UK Biobank Males (n=161,268)
Age at Recruitment (years)	57 (8.1)
Townsend Deprivation Index	-1.54 (2.98)
BMI (kg/m^2)	27.8 (4.2)
Biochemical Measurements	
Total Testosterone (nmol/L)	11.9 (3.7)
Sex Hormone-Binding Globulin (nmol/L)	39.9 (16.7)
Albumin (g/L)	45.5 (2.6)
Calculated Free Testosterone (nmol/L)	0.21 (0.06)
Assessment Centre	
Manchester	4435 (2.8%)
Oxford	4299 (2.7%)
Cardiff	6365 (3.9%)
Glasgow	5780 (3.6%)
Edinburgh	5702 (3.5%)
Stoke	7964 (4.9%)
Reading	9529 (5.9%)
Bury	9928 (6.2%)
Newcastle	13185 (8.2%)
Leeds	14794 (9.2%)
Bristol	14305 (8.9%)
Barts	2624 (1.6%)
Nottingham	11418 (7.1%)
Sheffield	10584 (6.6%)
Liverpool	11057 (6.9%)
Middlesborough	7329 (4.5%)
Hounslow	6635 (4.1%)
Croydon	6748 (4.2%)
Birmingham	7611 (4.7%)
Swansea	773 (0.5%)
Wrexham	203 (0.1%)

Values are expressed as mean (± SD) or count (%)
Abbreviations: BMI, body mass index

Supplementary File 1 - Table 2. Independent genetic variants associated with calculated free testosterone at genome-wide significance (p<5x10-8) and not associated with sex hormone-binding globulin in males

chr	rsid	position (hg19)	beta	se	effect allele	other allele	eaf	pval	gene	sample size
1	rs72664935	32274901	-0.00116	0.00020	T	C	0.424	4.80E-09	SPOCD1	161268
1	rs6656451	65990422	0.00126	0.00019	C	T	0.447	9.20E-11	LEPR	161268
1	rs41264945	155640115	-0.00245	0.00042	C	T	0.944	6.60E-09	YY1AP1	161268
1	rs12756999	163242095	-0.00129	0.00023	C	A	0.757	1.00E-08	RGS5	161268
1	rs2421985	172099136	0.00122	0.00019	T	C	0.519	3.60E-10	DNM3	161268
1	rs114654555	204117104	0.00318	0.00057	G	T	0.969	2.50E-08	ETNK2	161268
1	rs35737316	204161534	-0.00245	0.00023	C	T	0.754	2.10E-27	KISS1	161268
2	rs17529680	11727087	-0.00172	0.00021	G	T	0.706	6.50E-16	GREB1	161268
2	rs6729954	18286651	-0.00109	0.00020	A	T	0.566	2.90E-08	KCNS3;RDH14	161268
2	rs72862643	31533151	-0.00418	0.00040	T	C	0.937	2.60E-25	EHD3;XDH	161268
2	rs11124268	31951731	-0.00176	0.00025	C	A	0.811	1.90E-12	SRD5A2;LINC01946	161268
2	rs11901448	32096569	-0.00206	0.00021	C	T	0.443	1.90E-22	MEMO1	161268
2	rs7599125	32861555	-0.00115	0.00020	A	G	0.547	6.10E-09	TTC27	161268
2	rs112564689	33377229	-0.00428	0.00060	C	G	0.972	1.30E-12	LTBP1	161268
2	rs144956283	105867819	-0.00125	0.00020	A	T	0.505	2.20E-10	TGFBRAP1	161268
2	rs10192634	180500950	-0.00189	0.00022	C	T	0.730	6.10E-18	ZNF385B	161268
2	rs4233633	234619937	0.00211	0.00030	G	C	0.879	1.60E-12	UGT1A10;UGT1A6;UGT1A7;UGT1A8;UGT1A9	161268
3	rs1112195	24085166	-0.00108	0.00019	A	G	0.505	3.50E-08	NR1D2;LINC00691	161268
3	rs9824196	28807441	-0.00137	0.00022	G	T	0.280	2.40E-10	LINC00693;RBMS3-AS3	161268
3	rs56384446	53796386	0.00134	0.00020	A	G	0.614	2.20E-11	CACNA1D	161268
3	rs35018228	61296301	0.00119	0.00021	G	A	0.662	6.70E-09	FHIT;PTPRG	161268
3	rs7649178	61659697	0.00203	0.00029	G	A	0.869	3.00E-12	PTPRG	161268
3	rs11425772	88217876	0.00148	0.00026	C	CA	0.228	8.30E-09	C3orf38;EPHA3	161268
3	rs329926	107380821	0.00115	0.00019	A	G	0.477	3.90E-09	BBX	161268
3	rs7626226	138234576	0.00133	0.00019	C	T	0.532	7.10E-12	CEP70	161268
3	rs7610095	152035637	0.00135	0.00019	G	A	0.547	3.60E-12	MBNL1	161268
4	rs78627661	21951413	-0.00285	0.00032	G	C	0.900	1.50E-18	KCNIP4;LOC100505912	161268
4	rs6446913	73891050	0.00133	0.00022	G	A	0.740	1.70E-09	ADAMTS3;COX18	161268
4	rs188539658	103712252	0.00746	0.00079	C	T	0.983	4.30E-21	LOC102723704	161268
4	rs17289915	104491078	0.01141	0.00088	C	G	0.988	7.50E-39	LINC02428;TACR3	161268
4	rs565931739	104590005	0.00283	0.00027	C	CA	0.836	4.00E-26	TACR3	161268
5	rs72774885	95840231	-0.00148	0.00027	T	C	0.834	2.30E-08	LOC101929710	161268
5	rs950716	135680540	0.00197	0.00028	A	G	0.862	2.40E-12	TRPC7	161268
6	rs7748921	17415517	0.00117	0.00021	T	C	0.690	2.80E-08	CAP2	161268
6	rs1856502	100097384	-0.00177	0.00019	T	A	0.546	1.20E-19	PRDM13;MCHR2	161268
6	rs6902789	105358192	0.00263	0.00020	G	A	0.630	5.40E-39	HACE1;LIN28B-AS1	161268
6	rs3020421	152348122	0.00127	0.00021	G	A	0.674	1.00E-09	ESR1	161268
7	rs9986829	15019259	-0.00318	0.00019	G	A	0.494	6.30E-60	DGKB;AGMO	161268
7	rs10279715	40870935	0.00134	0.00020	A	G	0.539	5.90E-12	SUGCT	161268
7	rs145359938	78504431	0.00448	0.00068	G	C	0.976	5.30E-11	MAGI2	161268
8	rs7835492	21089517	-0.00164	0.00027	T	C	0.854	2.20E-09	LINC02153;LOC101929172	161268
8	rs7844586	61782304	-0.00185	0.00023	T	C	0.250	2.10E-16	CHD7;LOC100130298	161268
8	rs4507742	77899752	0.00155	0.00020	C	T	0.637	1.20E-14	PEX2	161268
8	rs11985714	113903152	0.00107	0.00019	C	T	0.472	3.50E-08	CSMD3	161268
9	rs72695843	4707078	-0.00222	0.00040	A	G	0.931	2.00E-08	CDC37L1	161268
9	rs12682723	24958749	0.00121	0.00020	G	A	0.570	8.00E-10	IZUMO3;TUSC1	161268
9	rs72729232	77104963	-0.00195	0.00024	T	G	0.795	9.20E-16	RORB-AS1	161268
9	rs2090409	108967088	0.00181	0.00021	C	A	0.684	6.60E-18	TMEM38B;MIR8081	161268
9	rs10982156	117088064	-0.00302	0.00040	T	A	0.930	2.40E-14	ORM1	161268
10	rs7912521	67262089	-0.00365	0.00020	T	C	0.585	5.00E-77	LOC101928887;LINC01515	161268
10	rs28613899	101723615	0.00116	0.00020	G	C	0.565	3.20E-09	DNMBP	161268
10	rs200372827	104890340	-0.00118	0.00020	T	TA	0.388	7.50E-09	NT5C2	161268
10	rs41313485	135055101	-0.00315	0.00055	G	A	0.968	1.10E-08	VENTX	161268
11	rs2957688	10364963	0.00117	0.00019	G	A	0.526	1.80E-09	CAND1.11	161268
11	rs11389722	28077708	0.00110	0.00020	G	GT	0.512	3.90E-08	KIF18A	161268
11	rs139807436	28728766	0.00365	0.00064	A	C	0.976	1.30E-08	MIR8068;LINC01616	161268
11	rs567203	28743701	0.00189	0.00021	T	C	0.322	2.70E-19	MIR8068;LINC01616	161268
11	rs142521478	29068121	0.00563	0.00060	G	A	0.972	4.50E-21	MIR8068;LINC01616	161268
11	rs35381476	29309284	0.00198	0.00022	C	CT	0.739	5.00E-19	MIR8068;LINC01616	161268
11	rs1148889	29595806	0.00381	0.00069	G	T	0.979	3.30E-08	MIR8068;LINC01616	161268
11	rs11604034	72364147	0.00161	0.00026	G	C	0.837	1.00E-09	PDE2A	161268
11	rs12796488	94131557	0.00335	0.00025	C	A	0.823	5.20E-40	GPR83	161268

11	rs4938576	118746769	-0.00116	0.00020	T	G	0.412	3.80E-09	DDX6;CXCR5	161268
11	rs10892924	122773715	-0.00235	0.00020	A	T	0.433	7.50E-33	C11orf63	161268
11	rs7110039	122834001	0.00116	0.00020	T	C	0.432	3.30E-09	C11orf63;BSX	161268
11	rs634554	125070392	0.00136	0.00021	C	A	0.308	1.10E-10	PKNOX2	161268
12	rs4765999	2945970	-0.00189	0.00021	C	T	0.668	7.70E-19	LOC100507424	161268
12	rs3819886	3001009	-0.00240	0.00024	C	G	0.798	3.30E-23	TULP3	161268
12	rs10082968	3099291	-0.00611	0.00050	G	A	0.960	1.20E-34	TEAD4	161268
12	rs141773786	57475601	0.00216	0.00032	A	G	0.894	7.60E-12	NEMP1;NAB2	161268
12	rs12810788	116196322	0.00180	0.00025	G	A	0.201	3.60E-13	TBX3;MED13L	161268
12	rs6486542	130952209	0.00149	0.00020	C	T	0.572	4.00E-14	RIMBP2	161268
13	rs1555589	100480664	0.00114	0.00021	G	A	0.340	3.30E-08	LOC101927437	161268
13	rs2038695	100559123	0.00141	0.00020	C	A	0.449	7.50E-13	LOC101927437	161268
13	rs3742223	112725196	0.00259	0.00033	T	C	0.903	6.00E-15	LINC00403	161268
14	rs1254269	60843017	-0.00160	0.00021	G	A	0.698	3.50E-14	LINC02322;C14orf39	161268
14	rs1812755	90007637	0.00196	0.00025	T	C	0.801	1.70E-15	FOXN3	161268
15	rs61733869	50769520	-0.00259	0.00046	G	A	0.953	1.40E-08	USP8	161268
15	rs6493485	51498587	0.00147	0.00026	G	T	0.171	1.60E-08	MIR4713HG	161268
15	rs6493487	51513729	0.00246	0.00022	G	A	0.255	3.10E-28	MIR4713HG	161268
15	rs12914625	57263440	-0.00134	0.00019	C	T	0.465	5.70E-12	TCF12	161268
15	rs35663835	60758053	-0.00155	0.00028	C	T	0.863	4.00E-08	ICE2	161268
16	rs2764772	20060653	-0.00212	0.00021	T	A	0.666	6.50E-25	GPR139	161268
16	rs74752114	28618068	0.00161	0.00025	G	A	0.727	7.10E-11	SULT1A1	161268
16	rs62041532	73922719	-0.00132	0.00024	T	G	0.787	2.70E-08	LINC01568;LOC101928035	161268
17	rs4542712	7469327	-0.00183	0.00025	T	C	0.821	4.90E-13	SENP3-EIF4A1	161268
17	rs2696641	43651550	-0.00134	0.00022	G	C	0.475	1.00E-09	LRRC37A4P;MAPK8IP1P2	161268
17	rs8076703	75612643	0.00143	0.00021	C	T	0.297	1.80E-11	LOC100507351;LINC01987	161268
18	rs2668776	44750365	0.00181	0.00019	C	T	0.470	1.90E-20	SKOR2	161268
19	rs55662444	48388635	-0.00132	0.00020	A	G	0.585	3.40E-11	SULT2A1	161268
20	rs2327121	8878250	-0.00118	0.00020	G	C	0.344	9.10E-09	PLCB1;PLCB4	161268
20	rs7265992	33525407	0.00164	0.00026	G	A	0.821	1.70E-10	GSS	161268
22	rs6009583	49677646	-0.00242	0.00022	C	T	0.733	3.80E-28	LINC01310;NONE	161268

Effect size (beta and standard error) in nmol/L of calculated free testosterone per copy of effect allele
Abbreviations: Chr, chromosome; rsID, rs identifier; SE, standard error; eaf, effect allele frequency; pval, p-value

Supplementary File 1 - Table 3. Results of Mendelian randomization analysis using Egger regression for 22 a priori outcomes relevant to testosterone supplementation

Outcome	P-value for Egger intercept	Effect using MR Egger (95% CI)	P-value
Haematocrit percentage	0.08	0.874 % (0.265 to 1.483)	0.006
Body fat-free percentage	0.33	1.409 % (0.323 to 2.495)	0.0127
Body fat percentage	0.28	-1.332656 % (-2.414 to -0.251)	0.0177
Heel bone mineral density	0.37	0.5473823 SD (0.183 to 0.912)	0.0041
Prostate cancer	0.79	OR = 1.62 (0.94 to 2.79)	0.0875
Androgenic alopecia	0.59	OR = 1.72 (0.98 to 3.01)	0.0632
Benign prostatic hyperplasia	0.94	OR = 1.38 (0.82 to 2.33)	0.2253
Depression	0.66	OR = 1.26 (0.64 to 2.5)	0.5106
Myocardial infarction	0.21	OR = 1.69 (0.99 to 2.89)	0.0561
Accelerometer-based physical activity	0.88	1.06 milligravity (-1.286 to 3.4)	0.3787
Glucose	0.96	-0.059 mmol/L (-0.26 to 0.142)	0.5661
Hemoglobin A1c	0.68	-0.101 mmol/mol (-1.316 to 1.11	0.8704
All stroke	0.37	OR = 0.88 (0.44 to 1.76)	0.7247
All fracture	0.41	OR = 1.1 (0.63 to 1.92)	0.7329
Handgrip strength	0.46	-0.224 kg (-1.725 to 1.278)	0.7709
Diastolic blood pressure	0.91	0.4 mmHg (-1.095 to 1.799)	0.6348
All dementia	0.60	OR = 0.86 (0.18 to 4.09)	0.8485
Ischemic stroke	0.35	OR = 0.59 (0.21 to 1.61)	0.3016
Systolic blood pressure	0.09	2.1 mmHg (-0.632 to 4.892)	0.134
Type 2 diabetes	0.95	OR = 1.04 (0.59 to 1.84)	0.898
Venous thromboembolism	0.06	OR = 0.51 (0.23 to 1.12)	0.0961
Heart failure	0.57	OR = 0.83 (0.41 to 1.7)	0.6177

Supplementary File 1 - Table 4. Results of Mendelian randomization analysis using MR-RAPS for effect of calculated free testosterone on 22 a priori outcomes relevant to testosterone supplementation

Outcome	Effect using MR RAPS (95% CI)	P-value
Haematocrit percentage	1.31 % (1.04 to 1.58)	2.11E-21
Body fat-free percentage	1.97 % (1.506 to 2.434)	7.76E-17
Body fat percentage	-1.95 % (-2.412 to -1.488)	1.28E-16
Heel bone mineral density	0.441 SD (0.308 to 0.574)	8.68E-11
Prostate cancer	OR = 1.53 (1.21 to 1.92)	2.86E-04
Androgenic alopecia	OR = 1.65 (1.32 to 2.06)	1.05E-05
Benign prostatic hyperplasia	OR = 1.42 (1.14 to 1.77)	0.00163
Depression	OR = 1.45 (1.09 to 1.93)	0.011
Myocardial infarction	OR = 1.30 (1.04 to 1.61)	0.020
Accelerometer-based physical activity	0.929 milligravity (0.004 to 1.854)	0.049
Glucose	-0.022 mmol/L (-0.091 to 0.046)	0.52
Hemoglobin A1c	-0.209 mmol/L (-0.648 to 0.23)	0.35
All stroke	OR = 1.22 (0.92 to 1.64)	0.17
All fracture	OR = 0.91 (0.71 to 1.15)	0.42
Handgrip strength	0.17 kg (-0.446 to 0.786)	0.59
Diastolic blood pressure	0.384 mmHg (-0.195 to 0.963)	0.19
All dementia	OR = 1.11 (0.59 to 2.1)	0.74
Ischemic stroke	OR = 0.92 (0.6 to 1.4)	0.69
Systolic blood pressure	0.267 mmHg (-0.892 to 1.426)	0.65
Type 2 diabetes	OR = 1.08 (0.85 to 1.38)	0.53
Venous thromboembolism	OR = 0.97 (0.69 to 1.36)	0.87
Heart failure	OR = 0.98 (0.73 to 1.32)	0.89

Abbreviations: CFT, calculated free testosterone; P, p-value; CI, confidence interval

Supplementary File 1 - Table 5. Results of Mendelian randomization analysis using MR-PRESSO for effect of calculated free testosterone on 22 a priori outcomes relevant to testosterone supplementation

Outcome	Global Test P-value	Effect using MR PRESSO (95% CI)	P-value	Distortion Test P-value
Haematocrit percentage	≤0.0001	1.394 % (1.182 to 1.606)	7.031E-22	0.8365
Body fat-free percentage	≤0.0001	1.917 % (1.538 to 2.296)	5.516E-16	0.9539
Body fat percentage	≤0.0001	-1.897 % (-2.272 to -1.522)	5.277E-16	0.9469
Heel bone mineral density	≤0.0001	0.4912 SD (0.38 to 0.602)	2.05E-13	0.1588
Androgenic alopecia	≤0.0001	OR = 1.79 (1.55 to 2.06)	2.45E-12	0.1359
Glucose	≤0.0001	-0.02062 mmol/L (-0.083 to 0.041)	0.516	0.0639
Hemoglobin A1c	≤0.0001	-0.2206 mmol/mol (-0.607 to 0.166)	0.2665	0.2268
Handgrip strength	≤0.0001	0.1428 kg (-0.423 to 0.709)	0.622	0.294
Systolic blood pressure	≤0.0001	0.1665 mmHg (-0.843 to 1.176)	0.7474	0.1588
Type 2 diabetes	≤0.0001	OR = 1.11 (0.9 to 1.39)	0.3342	0.6741
Diastolic blood pressure	0.0038	0.441 mmHg (-0.064 to 0.946)	0.09072	0.6254
Benign prostatic hyperplasia	0.026	OR = 1.4 (1.15 to 1.7)	0.001151	0.7902
All fracture	0.0723	NA	NA	NA
Venous thromboembolism	0.0805	NA	NA	NA
Myocardial infarction	0.0951	NA	NA	NA
All dementia	0.1866	NA	NA	NA
Accelerometer-based physical activity	0.2249	NA	NA	NA
Prostate cancer	0.4305	NA	NA	NA
Heart failure	0.4893	NA	NA	NA
Ischemic stroke	0.6653	NA	NA	NA
All stroke	0.8287	NA	NA	NA
Depression	0.8484	NA	NA	NA

[a] Distortion test evaluates significant differences in the causal estimates before and after correction for outliers

[b] Global test evaluates overall horizontal pleiotropy among all genetic variants

Abbreviations: CFT, calculated free testosterone; P, p-value; CI, confidence interval

Supplementary File 1 - Table 6. Associations of genetically predicted calculated free testosterone for 439 health outcomes across the human phenome excluding individuals on antihypertensive medication

Trait	Effect per 0.1 nmol/L increase CFT (95% CI)	P-value	FDR-adjusted p-value	Sample Size	Number of Cases	Number of Controls	Phecode	Category
Creatinine	0.118 (0.081 to 0.155)	3.19E-10	1.41E-07	112976	NA	NA	NA	biomarker
Apolipoprotein A	-0.021 (-0.031 to -0.012)	5.00E-06	1.10E-03	104233	NA	NA	NA	biomarker
HDL cholesterol	-0.083 (-0.123 to -0.044)	3.58E-05	4.82E-03	104382	NA	NA	NA	biomarker
C-reactive protein	-0.081 (-0.12 to -0.042)	4.37E-05	4.82E-03	112768	NA	NA	NA	biomarker
Other symptoms involving abdomen and pelvis	OR=2.15 (1.41 to 3.29)	3.96E-04	0.035	112195	963	111232	phecode 579	digestive
Umbilical hernia	OR=1.73 (1.26 to 2.39)	8.12E-04	0.043	101505	1690	99815	phecode 550.4	digestive
Spinal stenosis	OR=1.96 (1.32 to 2.91)	8.41E-04	0.043	115676	1112	114564	phecode 720	musculoskeletal
Fasciitis	OR=0.6 (0.44 to 0.81)	8.58E-04	0.043	110722	1947	108775	phecode 728.7	musculoskeletal
Albumin	-0.17 (-0.27 to -0.07)	8.82E-04	0.043	104429	NA	NA	NA	biomarker
Contracture of palmar fascia Dupuytrens disease	OR=0.6 (0.44 to 0.81)	1.12E-03	0.049	110604	1829	108775	phecode 728.71	musculoskeletal
Total protein	-0.26 (-0.42 to -0.1)	1.43E-03	0.053	104299	NA	NA	NA	biomarker
Phosphate	-0.011 (-0.017 to -0.004)	1.45E-03	0.053	104211	NA	NA	NA	biomarker
Abdominal hernia	OR=1.19 (1.07 to 1.32)	1.59E-03	0.054	118535	18720	99815	phecode 550	digestive
IGF1	0.322 (0.12 to 0.524)	1.76E-03	0.055	112446	NA	NA	NA	biomarker
Unspecified monoarthritis	OR=1.31 (1.1 to 1.55)	2.21E-03	0.065	112557	6314	106243	phecode 716.2	musculoskeletal
Inguinal hernia	OR=1.24 (1.08 to 1.42)	2.58E-03	0.071	109916	10101	99815	phecode 550.1	digestive
Calcium	-0.005 (-0.009 to -0.002)	3.19E-03	0.083	104386	NA	NA	NA	biomarker
Hypotension	OR=0.63 (0.46 to 0.86)	3.90E-03	0.093	113978	1823	112155	phecode 458	circulatory system
Disorders of muscle; ligament; and fascia	OR=0.66 (0.49 to 0.88)	4.11E-03	0.093	110944	2169	108775	phecode 728	musculoskeletal
Spondylosis and allied disorders	OR=1.54 (1.14 to 2.06)	4.24E-03	0.093	116588	2024	114564	phecode 721	musculoskeletal
Arthropathy NOS	OR=1.21 (1.06 to 1.38)	4.65E-03	0.098	117337	11094	106243	phecode 716.9	musculoskeletal
Back pain	OR=1.36 (1.1 to 1.69)	5.48E-03	0.106	118535	3745	114790	phecode 760	symptoms
Other arthropathies	OR=1.21 (1.06 to 1.38)	5.63E-03	0.106	117366	11123	106243	phecode 716	musculoskeletal
Other disorders of intestine	OR=1.29 (1.08 to 1.55)	5.78E-03	0.106	113894	5615	108279	phecode 569	digestive
Symptoms involving head and neck	OR=1.7 (1.16 to 2.5)	6.25E-03	0.110	118535	1185	117350	phecode 293	mental disorders
Hypotension NOS	OR=0.63 (0.45 to 0.88)	7.11E-03	0.119	113732	1577	112155	phecode 458.9	circulatory system
Inflammatory diseases of prostate	OR=1.55 (1.12 to 2.13)	7.29E-03	0.119	38517	1831	36686	phecode 601	genitourinary
Degenerative skin conditions and other dermatoses	OR=0.65 (0.48 to 0.9)	8.60E-03	0.135	117429	1778	115651	phecode 702	dermatologic
Hemiplegia	OR=0.44 (0.23 to 0.83)	0.011	0.170	112321	439	111882	phecode 342	neurological
Delirium dementia and amnestic and other cognitive disorders	OR=1.7 (1.11 to 2.62)	0.015	0.222	116902	948	115954	phecode 290	mental disorders
Osteoarthrosis	OR=1.22 (1.04 to 1.43)	0.016	0.226	118535	7314	111221	phecode 740	musculoskeletal
Alkaline phosphatase	-0.047 (-0.085 to -0.009)	0.016	0.226	113050	NA	NA	NA	biomarker
Internal derangement of knee	OR=1.23 (1.04 to 1.47)	0.018	0.235	117744	6027	111717	phecode 835	injuries and poisonings
Urethral stricture not specified as infectious	OR=1.58 (1.08 to 2.3)	0.018	0.235	114490	1217	113273	phecode 597.1	genitourinary
GERD	OR=1.21 (1.03 to 1.42)	0.019	0.239	112426	7371	105055	phecode 530.11	digestive
Mitral valve disease	OR=0.6 (0.39 to 0.92)	0.020	0.247	114991	969	114022	phecode 394.2	circulatory system
Prostatitis	OR=1.65 (1.08 to 2.54)	0.021	0.252	37682	996	36686	phecode 601.1	genitourinary
Chronic bronchitis	OR=1.34 (1.04 to 1.72)	0.024	0.254	112084	2819	109265	phecode 496.2	respiratory
Other disorders of peritoneum	OR=1.66 (1.07 to 2.57)	0.024	0.254	109183	904	108279	phecode 568	digestive
Obstructive chronic bronchitis	OR=1.34 (1.04 to 1.73)	0.024	0.254	112022	2757	109265	phecode 496.21	respiratory
Lipoma	OR=1.37 (1.04 to 1.8)	0.024	0.254	118243	2364	115879	phecode 214	neoplasms
Convulsions	OR=1.69 (1.07 to 2.68)	0.025	0.254	112698	816	111882	phecode 345.3	neurological
Personal history of diseases of digestive system	OR=1.24 (1.03 to 1.49)	0.025	0.254	102744	5389	97355	phecode 564.9	digestive
Sleep apnea	OR=1.44 (1.04 to 2)	0.027	0.264	118221	1647	116574	phecode 327.3	neurological
Total bilirubin	0.044 (0.005 to 0.083)	0.027	0.264	112560	NA	NA	NA	biomarker
Other disorders of urethra and urinary tract	OR=1.49 (1.04 to 2.13)	0.029	0.275	114644	1371	113273	phecode 597	genitourinary
Functional digestive disorders	OR=1.18 (1.02 to 1.36)	0.029	0.277	106299	8944	97355	phecode 564	digestive
Allergyoradverse effect of penicillin	OR=1.27 (1.02 to 1.58)	0.031	0.282	115509	3800	111709	phecode 960.2	injuries and poisonings
Duodenitis	OR=1.32 (1.02 to 1.7)	0.034	0.303	111801	2746	109055	phecode 535.6	digestive
Seborrheic keratosis	OR=0.64 (0.42 to 0.97)	0.035	0.306	117756	999	116757	phecode 702.2	dermatologic
Cancer; suspected or other	OR=1.17 (1.01 to 1.36)	0.035	0.306	114819	9023	105796	phecode 195	neoplasms
Malignant neoplasm; other	OR=1.17 (1.01 to 1.36)	0.037	0.312	114600	8804	105796	phecode 195.1	neoplasms
Fracture of lower limb	OR=0.73 (0.54 to 0.98)	0.038	0.312	114322	1927	112395	phecode 800	injuries and poisonings
Other peripheral nerve disorders	OR=1.33 (1.01 to 1.75)	0.039	0.322	116930	2354	114576	phecode 351	neurological
Noninfectious gastroenteritis	OR=1.26 (1.01 to 1.58)	0.041	0.332	100910	3555	97355	phecode 558	digestive

Diseases of the oral soft tissues; excluding lesions specific for gingiva and tongue	OR=1.47 (1.01 to 2.15)	0.044	0.343	118093	1226	116867	phecode 528	digestive
Osteoarthritis; localized	OR=1.22 (1.01 to 1.49)	0.044	0.343	115940	4719	111221	phecode 740.1	musculoskeletal
Peritoneal adhesions postoperative postinfection	OR=1.58 (1.01 to 2.49)	0.047	0.358	109128	849	108279	phecode 568.1	digestive
Nonrheumatic mitral valve disorders	OR=0.65 (0.42 to 1)	0.048	0.360	114977	955	114022	phecode 395.1	circulatory system
Other symptomsordisorders or the urinary system	OR=1.17 (1 to 1.36)	0.049	0.361	118535	7961	110574	phecode 599	genitourinary
Intestinal infection	OR=1.29 (1 to 1.65)	0.050	0.361	118535	2807	115728	phecode 008	infectious diseases
Hemorrhage of gastrointestinal tract	OR=1.34 (1 to 1.8)	0.053	0.362	113245	2013	111232	phecode 578.9	digestive
Essential hypertension	OR=1.12 (1 to 1.27)	0.053	0.362	118245	14230	104015	phecode 401.1	circulatory system
Visual disturbances	OR=0.66 (0.43 to 1.01)	0.055	0.362	118535	950	117585	phecode 368	sense organs
Gastrointestinal hemorrhage	OR=1.18 (1 to 1.39)	0.055	0.362	117762	6530	111232	phecode 578	digestive
Retinal detachments and defects	OR=0.7 (0.48 to 1.01)	0.055	0.362	115816	1276	114540	phecode 361	sense organs
Retinal detachment with retinal defect	OR=0.7 (0.48 to 1.01)	0.055	0.362	115816	1276	114540	phecode 361.1	sense organs
Abal findings examination of lungs	OR=1.54 (0.99 to 2.42)	0.057	0.372	118535	866	117669	phecode 514	respiratory
Osteoarthrosis NOS	OR=1.26 (0.99 to 1.6)	0.061	0.376	114295	3074	111221	phecode 740.9	musculoskeletal
Open wounds of head; neck; and trunk	OR=1.37 (0.99 to 1.91)	0.061	0.376	116186	1605	114581	phecode 870	injuries and poisonings
Osteoporosis	OR=0.61 (0.36 to 1.03)	0.063	0.376	118196	638	117558	phecode 743.1	musculoskeletal
Osteoporosis NOS	OR=0.61 (0.36 to 1.03)	0.063	0.376	118196	638	117558	phecode 743.11	musculoskeletal
Poisoning by antibiotics	OR=1.22 (0.99 to 1.5)	0.063	0.376	115843	4134	111709	phecode 960	injuries and poisonings
Hypertension	OR=1.12 (0.99 to 1.26)	0.063	0.376	118315	14300	104015	phecode 401	circulatory system
Benign neoplasm of colon	OR=1.16 (0.99 to 1.36)	0.065	0.376	98910	7577	91333	phecode 208	neoplasms
Other local infections of skin and subcutaneous tissue	OR=0.73 (0.52 to 1.02)	0.065	0.376	116217	1549	114668	phecode 686	dermatologic
Other diseases of the teeth and supporting structures	OR=1.54 (0.97 to 2.46)	0.066	0.376	116306	806	115500	phecode 525	digestive
Fracture of upper limb	OR=0.77 (0.59 to 1.02)	0.067	0.377	114769	2374	112395	phecode 803	injuries and poisonings
Symptoms involving digestive system	OR=1.22 (0.99 to 1.51)	0.068	0.381	101357	4002	97355	phecode 561	digestive
Intestinal obstruction without mention of hernia	OR=1.41 (0.97 to 2.05)	0.073	0.400	98600	1245	97355	phecode 560	digestive
Acute pulmonary heart disease	OR=1.36 (0.97 to 1.92)	0.077	0.412	117556	1491	116065	phecode 415.1	circulatory system
Pulmonary embolism and infarction; acute	OR=1.36 (0.97 to 1.92)	0.077	0.412	117556	1491	116065	phecode 415.11	circulatory system
Sciatica	OR=1.62 (0.95 to 2.78)	0.078	0.412	118066	596	117470	phecode 764	symptoms
Fracture of unspecified part of femur	OR=0.62 (0.37 to 1.06)	0.079	0.416	113013	618	112395	phecode 800.2	injuries and poisonings
Chronic airway obstruction	OR=1.23 (0.97 to 1.54)	0.083	0.433	112693	3428	109265	phecode 496	respiratory
Ability of gait	OR=0.66 (0.4 to 1.06)	0.088	0.445	118128	751	117377	phecode 350.2	neurological
Other disorders of prostate	OR=1.53 (0.94 to 2.5)	0.089	0.445	37448	762	36686	phecode 602	genitourinary
Malignant neoplasm of other and ill defined sites within the digestive organs and peritoneum	OR=1.28 (0.96 to 1.71)	0.089	0.445	92351	2158	90193	phecode 159	neoplasms
Poisoning by primarily systemic agents	OR=1.53 (0.94 to 2.51)	0.090	0.445	112422	713	111709	phecode 963	injuries and poisonings
Esophagitis; GERD and related diseases	OR=1.13 (0.98 to 1.29)	0.091	0.447	115274	10219	105055	phecode 530.1	digestive
Diseases of esophagus	OR=1.12 (0.98 to 1.27)	0.096	0.463	116340	11285	105055	phecode 530	digestive
Right bundle branch block	OR=0.65 (0.39 to 1.09)	0.102	0.488	109854	668	109186	phecode 426.31	circulatory system
Appendiceal conditions	OR=0.73 (0.49 to 1.07)	0.103	0.490	118535	1180	117355	phecode 540	digestive
Fracture of radius and ulna	OR=0.72 (0.49 to 1.07)	0.105	0.494	113535	1140	112395	phecode 803.2	injuries and poisonings
Abdominal pain	OR=1.12 (0.97 to 1.3)	0.108	0.498	118535	9247	109288	phecode 785	symptoms
Other diseases of blood and blood forming organs	OR=1.34 (0.94 to 1.91)	0.109	0.498	117523	1373	116150	phecode 289	hematopoietic
Sleep disorders	OR=1.28 (0.95 to 1.72)	0.110	0.498	118535	1961	116574	phecode 327	neurological
Other disorders of stomach and duodenum	OR=1.27 (0.95 to 1.69)	0.111	0.499	111171	2116	109055	phecode 537	digestive
Nausea and vomiting	OR=0.81 (0.63 to 1.05)	0.114	0.507	118535	2648	115887	phecode 789	symptoms
Lipoma of skin and subcutaneous tissue	OR=1.3 (0.94 to 1.81)	0.116	0.507	117475	1596	115879	phecode 214.1	neoplasms
Diaphragmatic hernia	OR=1.13 (0.97 to 1.32)	0.116	0.507	107545	7730	99815	phecode 550.2	digestive
Hypercholesterolemia	OR=1.12 (0.97 to 1.3)	0.125	0.533	117796	8894	108902	phecode 272.11	endocrine metabolic
Secondary malignancy of lymph nodes	OR=1.32 (0.93 to 1.88)	0.126	0.533	107187	1391	105796	phecode 198.1	neoplasms
Precordial pain	OR=0.73 (0.48 to 1.09)	0.127	0.533	111190	1049	110141	phecode 418.1	circulatory system
Anal and rectal conditions	OR=1.15 (0.96 to 1.39)	0.127	0.533	113627	5348	108279	phecode 565	digestive
Asthma	OR=1.13 (0.96 to 1.34)	0.133	0.555	116058	6793	109265	phecode 495	respiratory
Appendicitis	OR=0.74 (0.5 to 1.1)	0.137	0.557	118493	1138	117355	phecode 540.1	digestive
Pulmonary heart disease	OR=1.28 (0.92 to 1.78)	0.140	0.557	117694	1629	116065	phecode 415	circulatory system

Vitamin D	0.029 (-0.01 to 0.069)	0.142	0.557	109121	NA	NA	NA	biomarker
Other mental disorder	OR=1.1 (0.97 to 1.26)	0.143	0.557	116299	11056	105243	phecode 306	mental disorders
Hemorrhoids	OR=1.13 (0.96 to 1.32)	0.143	0.557	114575	7274	107301	phecode 455	circulatory system
Other biliary tract disease	OR=1.37 (0.9 to 2.09)	0.143	0.557	116054	980	115074	phecode 575	digestive
Cholelithiasis	OR=1.21 (0.94 to 1.57)	0.144	0.557	117744	2670	115074	phecode 574.1	digestive
Cystatin C	0.026 (-0.009 to 0.061)	0.144	0.557	113043	NA	NA	NA	biomarker
Open wounds of extremities	OR=1.23 (0.93 to 1.64)	0.146	0.558	116772	2191	114581	phecode 871	injuries and poisonings
Chronic liver disease and cirrhosis	OR=0.74 (0.49 to 1.11)	0.151	0.573	116576	1057	115519	phecode 571	digestive
Varicose veins	OR=1.2 (0.93 to 1.56)	0.160	0.597	109984	2683	107301	phecode 454	circulatory system
Osteoporosis; osteopenia and pathological fracture	OR=0.74 (0.48 to 1.13)	0.161	0.597	118535	977	117558	phecode 743	musculoskeletal
Hyperlipidemia	OR=1.11 (0.96 to 1.28)	0.161	0.597	118499	9597	108902	phecode 272.1	endocrine metabolic
Erythematous conditions	OR=1.48 (0.85 to 2.56)	0.165	0.603	117331	571	116760	phecode 695	dermatologic
Direct bilirubin	0.029 (-0.012 to 0.069)	0.165	0.603	104780	NA	NA	NA	biomarker
Anxiety disorder	OR=1.25 (0.91 to 1.71)	0.168	0.608	107024	1781	105243	phecode 300.1	mental disorders
Altered mental status	OR=1.37 (0.87 to 2.16)	0.171	0.611	116803	849	115954	phecode 292.4	mental disorders
Tachycardia NOS	OR=0.7 (0.43 to 1.16)	0.172	0.611	109882	696	109186	phecode 427.7	circulatory system
Disorders of mineral metabolism	OR=0.72 (0.45 to 1.15)	0.174	0.612	118535	795	117740	phecode 275	endocrine metabolic
Acute appendicitis	OR=0.75 (0.49 to 1.14)	0.178	0.621	118342	987	117355	phecode 540.11	digestive
Transient cerebral ischemia	OR=0.69 (0.4 to 1.19)	0.179	0.621	116037	585	115452	phecode 433.31	circulatory system
Frequency of urination and polyuria	OR=1.28 (0.89 to 1.85)	0.183	0.622	111873	1299	110574	phecode 599.5	genitourinary
Gout	OR=1.27 (0.89 to 1.8)	0.184	0.622	118327	1412	116915	phecode 274.1	endocrine metabolic
Disorders of lipoid metabolism	OR=1.1 (0.96 to 1.27)	0.185	0.622	118535	9633	108902	phecode 272	endocrine metabolic
Alanine aminotransferase	0.026 (-0.012 to 0.064)	0.186	0.622	112973	NA	NA	NA	biomarker
Osteoarthrosis; localized; primary	OR=1.2 (0.91 to 1.58)	0.187	0.622	113566	2345	111221	phecode 740.11	musculoskeletal
Anxiety disorders	OR=1.22 (0.91 to 1.64)	0.188	0.622	107239	1996	105243	phecode 300	mental disorders
Epistaxis or throat hemorrhage	OR=0.72 (0.43 to 1.19)	0.198	0.641	112820	687	112133	phecode 477	respiratory
Urinary obstruction	OR=1.31 (0.87 to 1.99)	0.198	0.641	111579	1005	110574	phecode 599.1	genitourinary
Septal DeviationsorTurbinate Hypertrophy	OR=1.22 (0.9 to 1.64)	0.199	0.641	114089	1956	112133	phecode 470	respiratory
Otitis media and Eustachian tube disorders	OR=1.41 (0.83 to 2.38)	0.199	0.641	118010	632	117378	phecode 381	sense organs
Cough	OR=1.34 (0.85 to 2.1)	0.203	0.648	115742	855	114887	phecode 512.8	respiratory
Spondylosis without myelopathy	OR=1.27 (0.88 to 1.83)	0.206	0.654	115849	1285	114564	phecode 721.1	musculoskeletal
Cholesterol	-0.027 (-0.068 to 0.015)	0.209	0.659	113057	NA	NA	NA	biomarker
Nasal polyps	OR=1.25 (0.88 to 1.77)	0.212	0.659	113558	1425	112133	phecode 471	respiratory
Renal colic	OR=0.74 (0.45 to 1.19)	0.213	0.659	116134	750	115384	phecode 594.8	genitourinary
Other abal blood chemistry	OR=1.16 (0.92 to 1.48)	0.214	0.659	118472	3111	115361	phecode 790.6	symptoms
Other chronic nonalcoholic liver disease	OR=0.74 (0.45 to 1.19)	0.215	0.659	116263	744	115519	phecode 571.5	digestive
Gout and other crystal arthropathies	OR=1.23 (0.89 to 1.71)	0.217	0.660	118535	1620	116915	phecode 274	endocrine metabolic
Urate	1.651 (-0.991 to 4.293)	0.221	0.663	112907	NA	NA	NA	biomarker
Hypothyroidism NOS	OR=1.23 (0.88 to 1.72)	0.222	0.663	118123	1556	116567	phecode 244.4	endocrine metabolic
Unspecified diffuse connective tissue disease	OR=0.73 (0.44 to 1.21)	0.223	0.663	110331	693	109638	phecode 709.7	dermatologic
Dermatitis due to solar radiation	OR=0.77 (0.5 to 1.18)	0.230	0.670	117504	948	116556	phecode 938	dermatologic
Type 1 diabetes	OR=0.71 (0.4 to 1.25)	0.230	0.670	114372	543	113829	phecode 250.1	endocrine metabolic
Blood in stool	OR=1.31 (0.84 to 2.06)	0.232	0.670	112099	867	111232	phecode 578.2	digestive
Secondary malignant neoplasm	OR=1.16 (0.91 to 1.49)	0.234	0.670	108697	2901	105796	phecode 198	neoplasms
Esophageal bleeding varicesorhemorrhage	OR=0.72 (0.42 to 1.23)	0.234	0.670	105667	612	105055	phecode 530.2	digestive
Colon cancer	OR=1.25 (0.86 to 1.8)	0.236	0.670	92581	1310	91271	phecode 153.2	neoplasms
Other disorders of bladder	OR=1.14 (0.92 to 1.42)	0.237	0.670	117067	3794	113273	phecode 596	genitourinary
Nonspecific findings on examination of blood	OR=1.15 (0.91 to 1.46)	0.237	0.670	118535	3174	115361	phecode 790	symptoms
Other disorders of soft tissues	OR=0.84 (0.62 to 1.13)	0.238	0.670	110770	1995	108775	phecode 729	musculoskeletal
Tobacco use disorder	OR=1.1 (0.94 to 1.3)	0.246	0.681	114482	6822	107660	phecode 318	mental disorders
Candidiasis	OR=1.35 (0.81 to 2.24)	0.248	0.681	118324	672	117652	phecode 112	infectious diseases
Chronic dermatitis due to solar radiation	OR=0.77 (0.5 to 1.2)	0.249	0.681	117472	916	116556	phecode 938.2	dermatologic
Chronic ulcer of skin	OR=1.34 (0.81 to 2.22)	0.249	0.681	118535	695	117840	phecode 707	dermatologic
Colorectal cancer	OR=1.19 (0.88 to 1.61)	0.254	0.690	93261	1990	91271	phecode 153	neoplasms
Actinic keratosis	OR=0.77 (0.49 to 1.21)	0.260	0.700	117622	865	116757	phecode 702.1	dermatologic
Malignant neoplasm of rectum; rectosigmoid junction; and anus	OR=1.27 (0.84 to 1.93)	0.261	0.700	92272	1001	91271	phecode 153.3	neoplasms
Cholelithiasis with other cholecystitis	OR=1.27 (0.84 to 1.92)	0.264	0.702	116076	1002	115074	phecode 574.12	digestive
Hemorrhage of rectum and anus	OR=1.13 (0.91 to 1.41)	0.275	0.719	114886	3654	111232	phecode 578.8	digestive
Fracture of ankle and foot	OR=0.75 (0.44 to 1.26)	0.275	0.719	113027	632	112395	phecode 801	injuries and poisonings

Other acute and subacute forms of ischemic heart disease	OR=1.36 (0.78 to 2.37)	0.276	0.719	109619	564	109055	phecode 411.9	circulatory system
Other open wound of head and face	OR=1.22 (0.85 to 1.73)	0.280	0.720	115967	1386	114581	phecode 870.3	injuries and poisonings
Bundle branch block	OR=0.81 (0.55 to 1.19)	0.282	0.720	110368	1182	109186	phecode 426.3	circulatory system
Epilepsy; recurrent seizures; convulsions	OR=1.19 (0.87 to 1.64)	0.282	0.720	113604	1722	111882	phecode 345	neurological
Hypothyroidism	OR=1.2 (0.86 to 1.66)	0.282	0.720	118205	1638	116567	phecode 244	endocrine metabolic
Staphylococcus infections	OR=0.8 (0.54 to 1.2)	0.288	0.724	114982	1076	113906	phecode 041.1	infectious diseases
Diffuse diseases of connective tissue	OR=0.77 (0.48 to 1.25)	0.290	0.724	110396	758	109638	phecode 709	dermatologic
Benign neoplasm of other parts of digestive system	OR=1.21 (0.85 to 1.72)	0.291	0.724	118535	1406	117129	phecode 211	neoplasms
Electrolyte imbalance	OR=1.22 (0.84 to 1.77)	0.292	0.724	117411	1266	116145	phecode 276.1	endocrine metabolic
Respiratory abalities	OR=1.14 (0.89 to 1.46)	0.292	0.724	118535	2888	115647	phecode 513	respiratory
Cardiac conduction disorders	OR=0.87 (0.66 to 1.13)	0.295	0.727	111671	2485	109186	phecode 426	circulatory system
Edema	OR=0.74 (0.41 to 1.33)	0.314	0.770	118486	510	117976	phecode 782.3	symptoms
Cancer of bladder	OR=0.82 (0.55 to 1.21)	0.317	0.772	117852	1129	116723	phecode 189.2	neoplasms
Cholecystitis without cholelithiasis	OR=0.77 (0.46 to 1.29)	0.319	0.772	115718	644	115074	phecode 574.3	digestive
Malaise and fatigue	OR=1.25 (0.8 to 1.96)	0.321	0.774	118535	869	117666	phecode 798	symptoms
Other non epithelial cancer of skin	OR=1.09 (0.92 to 1.28)	0.325	0.778	117363	6884	110479	phecode 172.2	neoplasms
Diseases of white blood cells	OR=1.21 (0.83 to 1.78)	0.326	0.778	117326	1176	116150	phecode 288	hematopoietic
Effects radiation NOS	OR=1.21 (0.83 to 1.77)	0.329	0.779	117516	1219	116297	phecode 990	injuries and poisonings
Redundant prepuce and phimosisorBXO	OR=1.17 (0.85 to 1.61)	0.330	0.779	108070	1735	106335	phecode 604.1	genitourinary
Dizziness and giddiness Light headedness and vertigo	OR=1.2 (0.83 to 1.73)	0.337	0.791	118129	1303	116826	phecode 386.9	sense organs
Postoperative infection	OR=1.2 (0.82 to 1.76)	0.340	0.791	118226	1207	117019	phecode 080	infectious diseases
Rheumatoid arthritis and other inflammatory polyarthropathies	OR=1.24 (0.8 to 1.91)	0.341	0.791	118535	921	117614	phecode 714	musculoskeletal
Other hypertrophic and atrophic conditions of skin	OR=1.2 (0.82 to 1.74)	0.347	0.799	118500	1247	117253	phecode 701	dermatologic
Peripheral enthesopathies and allied syndromes	OR=1.1 (0.9 to 1.35)	0.350	0.799	113228	4453	108775	phecode 726	musculoskeletal
Symptoms involving skin and other integumentary tissue	OR=0.77 (0.44 to 1.34)	0.352	0.799	118535	559	117976	phecode 782	symptoms
Abal heart sounds	OR=0.89 (0.68 to 1.15)	0.356	0.799	116683	2661	114022	phecode 396	circulatory system
Circulatory disease NEC	OR=0.91 (0.75 to 1.11)	0.357	0.799	117066	4911	112155	phecode 459.9	circulatory system
Bacterial enteritis	OR=1.26 (0.77 to 2.06)	0.357	0.799	116448	720	115728	phecode 008.5	infectious diseases
Disorders of fluid; electrolyte; and acid base balance	OR=1.14 (0.87 to 1.49)	0.358	0.799	118535	2390	116145	phecode 276	endocrine metabolic
Degeneration of intervertebral disc	OR=0.8 (0.49 to 1.29)	0.360	0.799	115316	752	114564	phecode 722.6	musculoskeletal
Musculoskeletal symptoms referable to limbs	OR=0.85 (0.6 to 1.2)	0.362	0.799	118535	1472	117063	phecode 771	symptoms
Diseases of pancreas	OR=1.23 (0.79 to 1.91)	0.364	0.799	118535	888	117647	phecode 577	digestive
Superficial injury without mention of infection	OR=1.15 (0.85 to 1.56)	0.365	0.799	118344	1870	116474	phecode 915	injuries and poisonings
Alcohol related disorders	OR=1.09 (0.91 to 1.31)	0.368	0.799	113106	5446	107660	phecode 317	mental disorders
Swelling of limb	OR=0.84 (0.58 to 1.22)	0.373	0.799	118335	1272	117063	phecode 771.1	symptoms
Other derangement of joint	OR=1.23 (0.78 to 1.95)	0.373	0.799	117323	825	116498	phecode 742.9	musculoskeletal
Complication due to other implant and internal device	OR=1.2 (0.8 to 1.81)	0.375	0.799	112474	1046	111428	phecode 859	injuries and poisonings
Ventral hernia	OR=1.19 (0.8 to 1.77)	0.380	0.799	100943	1128	99815	phecode 550.5	digestive
Skin cancer	OR=1.07 (0.92 to 1.25)	0.380	0.799	118484	8005	110479	phecode 172	neoplasms
Acquired foot deformities	OR=0.83 (0.54 to 1.27)	0.381	0.799	117328	961	116367	phecode 735	musculoskeletal
Nonspecific chest pain	OR=1.07 (0.92 to 1.24)	0.381	0.799	118535	8394	110141	phecode 418	circulatory system
Varicose veins of lower extremity	OR=1.13 (0.86 to 1.48)	0.383	0.799	109731	2430	107301	phecode 454.1	circulatory system
Reflux esophagitis	OR=1.1 (0.88 to 1.38)	0.383	0.799	108635	3580	105055	phecode 530.14	digestive
Rheumatic disease of the heart valves	OR=0.87 (0.63 to 1.2)	0.386	0.799	115707	1685	114022	phecode 394	circulatory system
Carbuncle and furuncle	OR=0.82 (0.51 to 1.3)	0.388	0.799	115480	812	114668	phecode 686.1	dermatologic
Occlusion of cerebral arteries	OR=1.17 (0.82 to 1.68)	0.389	0.799	116803	1351	115452	phecode 433.2	circulatory system
Localized superficial swelling; mass; or lump	OR=1.27 (0.74 to 2.2)	0.390	0.799	116996	575	116421	phecode 687.2	dermatologic
Alcoholism	OR=1.1 (0.88 to 1.37)	0.393	0.802	111466	3806	107660	phecode 317.1	mental disorders
First degree AV block	OR=0.76 (0.4 to 1.44)	0.396	0.804	109614	428	109186	phecode 426.21	circulatory system
Gamma glutamyltransferase	0.017 (-0.022 to 0.056)	0.399	0.806	112976	NA	NA	NA	biomarker
Atrial fibrillation and flutter	OR=1.08 (0.9 to 1.31)	0.400	0.806	114508	5322	109186	phecode 427.2	circulatory system
Cholelithiasis and cholecystitis	OR=1.11 (0.87 to 1.41)	0.407	0.813	118095	3021	115074	phecode 574	digestive
Other disorders of circulatory system	OR=0.92 (0.76 to 1.12)	0.408	0.813	117126	4971	112155	phecode 459	circulatory system

Enthesopathy	OR=1.11 (0.86 to 1.43)	0.418	0.828	111530	2755	108775	phecode 726.1	musculoskeletal
Other specified cardiac dysrhythmias	OR=0.86 (0.59 to 1.25)	0.420	0.828	110432	1246	109186	phecode 427.3	circulatory system
Hypertensive heart andoror renal disease	OR=0.68 (0.26 to 1.77)	0.426	0.828	104204	189	104015	phecode 401.2	circulatory system
Other disorders of eyelids	OR=1.15 (0.82 to 1.62)	0.429	0.828	117280	1502	115778	phecode 374	sense organs
Abal movement	OR=0.85 (0.58 to 1.26)	0.429	0.828	118535	1158	117377	phecode 350	neurological
Rheumatoid arthritis	OR=1.2 (0.76 to 1.91)	0.432	0.828	118535	819	117716	phecode 714.1	musculoskeletal
Gastritis and duodenitis	OR=1.06 (0.91 to 1.23)	0.432	0.828	117526	8471	109055	phecode 535	digestive
Renal failure	OR=1.11 (0.86 to 1.43)	0.434	0.828	117464	2729	114735	phecode 585	genitourinary
Bladder neck obstruction	OR=1.17 (0.78 to 1.76)	0.435	0.828	114351	1078	113273	phecode 596.1	genitourinary
Nonrheumatic aortic valve disorders	OR=1.2 (0.76 to 1.88)	0.435	0.828	114888	866	114022	phecode 395.2	circulatory system
Hydronephrosis	OR=1.21 (0.75 to 1.94)	0.436	0.828	116163	779	115384	phecode 595	genitourinary
Ileostomy status	OR=1.23 (0.73 to 2.08)	0.438	0.828	97985	630	97355	phecode 559	digestive
Thrombocytopenia	OR=1.24 (0.71 to 2.17)	0.448	0.843	118204	561	117643	phecode 287.3	hematopoietic
Melanomas of skin	OR=1.15 (0.8 to 1.67)	0.451	0.843	111739	1260	110479	phecode 172.11	neoplasms
Melanomas of skin; dx or hx	OR=1.15 (0.8 to 1.67)	0.451	0.843	111739	1260	110479	phecode 172.1	neoplasms
Retention of urine	OR=1.09 (0.87 to 1.37)	0.457	0.850	114014	3440	110574	phecode 599.2	genitourinary
Congenital anomalies of great vessels	OR=1.23 (0.71 to 2.13)	0.459	0.851	118201	577	117624	phecode 747.13	congenital anomalies
Bacterial infection NOS	OR=1.09 (0.87 to 1.36)	0.470	0.862	117439	3533	113906	phecode 041	infectious diseases
Constipation	OR=1.1 (0.85 to 1.42)	0.470	0.862	100132	2777	97355	phecode 563	digestive
Disorders of penis	OR=1.11 (0.83 to 1.48)	0.471	0.862	108440	2105	106335	phecode 604	genitourinary
Aortic valve disease	OR=1.25 (0.68 to 2.3)	0.474	0.863	114487	465	114022	phecode 394.3	circulatory system
Diseases of the larynx and vocal cords	OR=1.19 (0.74 to 1.91)	0.478	0.864	112898	765	112133	phecode 473	respiratory
Ulcer of esophagus	OR=1.12 (0.82 to 1.51)	0.480	0.864	106980	1925	105055	phecode 530.12	digestive
Respiratory failure	OR=1.18 (0.74 to 1.89)	0.481	0.864	114970	799	114171	phecode 509.1	respiratory
Poisoning by analgesics; antipyretics; and antirheumatics	OR=1.13 (0.8 to 1.6)	0.482	0.864	113167	1458	111709	phecode 965	injuries and poisonings
Other and unspecified disc disorder	OR=1.13 (0.8 to 1.59)	0.491	0.864	116033	1469	114564	phecode 722.9	musculoskeletal
Neurological disorders	OR=1.12 (0.81 to 1.54)	0.493	0.864	117694	1740	115954	phecode 292	mental disorders
Pericarditis	OR=0.82 (0.45 to 1.47)	0.497	0.864	117883	500	117383	phecode 420.2	circulatory system
Complications of transplants and reattached limbs	OR=1.07 (0.89 to 1.28)	0.500	0.864	116695	5267	111428	phecode 851	injuries and poisonings
Other specified peripheral vascular diseases	OR=1.16 (0.76 to 1.77)	0.501	0.864	116771	959	115812	phecode 443.8	circulatory system
Other headache syndromes	OR=0.9 (0.67 to 1.22)	0.501	0.864	118153	1970	116183	phecode 339	neurological
Hemoptysis	OR=0.88 (0.62 to 1.27)	0.502	0.864	118471	1348	117123	phecode 516.1	respiratory
Orthostatic hypotension	OR=0.81 (0.45 to 1.49)	0.504	0.864	112639	484	112155	phecode 458.1	circulatory system
E coli	OR=1.19 (0.72 to 1.97)	0.506	0.864	114588	682	113906	phecode 041.4	infectious diseases
Shortness of breath	OR=0.89 (0.64 to 1.25)	0.509	0.864	116495	1608	114887	phecode 512.7	respiratory
Degeneration of macula and posterior pole of retina	OR=1.19 (0.7 to 2.03)	0.510	0.864	78616	632	77984	phecode 362.2	sense organs
Macular degeneration senile of retina NOS	OR=1.19 (0.7 to 2.03)	0.510	0.864	78616	632	77984	phecode 362.29	sense organs
Fracture of vertebral column without mention of spinal cord injury	OR=0.84 (0.51 to 1.4)	0.512	0.864	113076	681	112395	phecode 805	injuries and poisonings
Neoplasm of uncertain behavior	OR=0.87 (0.57 to 1.33)	0.514	0.864	106777	981	105796	phecode 199	neoplasms
Respiratory insufficiency	OR=1.18 (0.72 to 1.93)	0.514	0.864	114887	716	114171	phecode 509.2	respiratory
Superficial cellulitis and abscess	OR=1.09 (0.84 to 1.42)	0.517	0.864	117208	2540	114668	phecode 681	dermatologic
Heart valve disorders	OR=0.9 (0.66 to 1.23)	0.518	0.864	115898	1876	114022	phecode 395	circulatory system
Lymphadenitis	OR=1.16 (0.74 to 1.81)	0.520	0.864	117015	865	116150	phecode 289.4	hematopoietic
Irritable Bowel Syndrome	OR=0.87 (0.57 to 1.33)	0.523	0.864	98350	995	97355	phecode 564.1	digestive
Other inflammatory spondylopathies	OR=1.19 (0.7 to 2.01)	0.527	0.864	118535	622	117913	phecode 715	musculoskeletal
Other intestinal obstruction	OR=1.14 (0.75 to 1.74)	0.530	0.864	98343	988	97355	phecode 560.4	digestive
Other disorders of bone and cartilage	OR=0.88 (0.59 to 1.31)	0.530	0.864	116154	1083	115071	phecode 733	musculoskeletal
Hypovolemia	OR=1.14 (0.75 to 1.75)	0.533	0.864	117120	975	116145	phecode 276.5	endocrine metabolic
Hemorrhage or hematoma complicating a procedure	OR=1.12 (0.78 to 1.61)	0.533	0.864	112760	1332	111428	phecode 850	injuries and poisonings
Diabetic retinopathy	OR=1.29 (0.57 to 2.93)	0.537	0.864	78245	261	77984	phecode 250.7	endocrine metabolic
Malignant neoplasm of bladder	OR=0.87 (0.57 to 1.34)	0.538	0.864	117689	966	116723	phecode 189.21	neoplasms
Fracture of clavicle or scapula	OR=0.87 (0.55 to 1.36)	0.540	0.864	113261	866	112395	phecode 803.3	injuries and poisonings
Calculus of kidney	OR=1.12 (0.77 to 1.63)	0.541	0.864	116649	1265	115384	phecode 594.1	genitourinary
Cystitis and urethritis	OR=1.17 (0.7 to 1.95)	0.543	0.864	111086	667	110419	phecode 592	genitourinary
Paroxysmal supraventricular tachycardia	OR=0.85 (0.51 to 1.43)	0.544	0.864	109843	657	109186	phecode 427.11	circulatory system
Fracture of tibia and fibula	OR=0.86 (0.53 to 1.4)	0.549	0.864	113121	726	112395	phecode 800.3	injuries and poisonings
Diverticulosis	OR=1.05 (0.9 to 1.22)	0.551	0.864	105829	8474	97355	phecode 562.1	digestive

Diverticulosis and diverticulitis	OR=1.05 (0.9 to 1.22)	0.551	0.864	105829	8474	97355	phecode 562	digestive
Triglycerides	0.012 (-0.027 to 0.051)	0.553	0.864	112940	NA	NA	NA	biomarker
Other diseases of respiratory system; not elsewhere classified	OR=1.08 (0.84 to 1.37)	0.554	0.864	118535	3039	115496	phecode 519	respiratory
Other disorders of arteries and arterioles	OR=0.82 (0.43 to 1.57)	0.557	0.864	116228	416	115812	phecode 447	circulatory system
Disorders of refraction and accommodation; blindness and low vision	OR=1.14 (0.74 to 1.76)	0.560	0.864	118535	929	117606	phecode 367	sense organs
Pain in joint	OR=1.09 (0.82 to 1.45)	0.561	0.864	118535	2153	116382	phecode 745	musculoskeletal
Obesity	OR=0.93 (0.72 to 1.2)	0.564	0.864	118514	2780	115734	phecode 278.1	endocrine metabolic
Atrioventricular AV block	OR=0.87 (0.55 to 1.38)	0.567	0.864	110025	839	109186	phecode 426.2	circulatory system
Occlusion and stenosis of precerebral arteries	OR=1.21 (0.62 to 2.37)	0.569	0.864	115841	389	115452	phecode 433.1	circulatory system
Aortic aneurysm	OR=1.17 (0.68 to 2)	0.570	0.864	116417	605	115812	phecode 442.1	circulatory system
Nerve root and plexus disorders	OR=1.11 (0.77 to 1.59)	0.572	0.864	115922	1346	114576	phecode 353	neurological
Cardiac pacemaker in situ	OR=1.15 (0.71 to 1.84)	0.573	0.864	109969	783	109186	phecode 426.91	circulatory system
Derangement of joint; non traumatic	OR=1.13 (0.74 to 1.72)	0.573	0.864	117471	973	116498	phecode 742	musculoskeletal
Secondary malignancy of bone	OR=1.14 (0.73 to 1.78)	0.574	0.864	106680	884	105796	phecode 198.6	neoplasms
Overweight; obesity and other hyperalimentation	OR=0.93 (0.72 to 1.2)	0.578	0.867	118535	2801	115734	phecode 278	endocrine metabolic
Symptoms involving nervous and musculoskeletal systems	OR=1.08 (0.81 to 1.45)	0.588	0.878	118535	2068	116467	phecode 781	symptoms
Gastric ulcer	OR=0.9 (0.61 to 1.32)	0.593	0.878	117419	1176	116243	phecode 531.2	digestive
Other specified gastritis	OR=0.93 (0.71 to 1.22)	0.593	0.878	111426	2371	109055	phecode 535.8	digestive
Senile cataract	OR=0.93 (0.71 to 1.22)	0.595	0.878	115154	2515	112639	phecode 366.2	sense organs
Chemotherapy	OR=1.05 (0.87 to 1.28)	0.595	0.878	110590	4794	105796	phecode 197	neoplasms
Chronic renal failure CKD	OR=1.11 (0.75 to 1.66)	0.600	0.880	115832	1097	114735	phecode 585.3	genitourinary
Acute pancreatitis	OR=1.15 (0.68 to 1.95)	0.602	0.880	118273	626	117647	phecode 577.1	digestive
Dislocation	OR=1.13 (0.72 to 1.76)	0.605	0.880	112578	861	111717	phecode 830	injuries and poisonings
Urinary tract infection	OR=1.07 (0.84 to 1.35)	0.607	0.880	113538	3119	110419	phecode 591	genitourinary
Inflammation of eyelids	OR=1.13 (0.71 to 1.79)	0.607	0.880	116595	817	115778	phecode 371.3	sense organs
Apolipoprotein B	0.002 (-0.007 to 0.011)	0.613	0.884	112313	NA	NA	NA	biomarker
Pain in limb	OR=1.09 (0.78 to 1.53)	0.613	0.884	118535	1521	117014	phecode 773	symptoms
Peripheral vascular disease; unspecified	OR=1.13 (0.69 to 1.85)	0.624	0.895	116539	727	115812	phecode 443.9	circulatory system
Chronic sinusitis	OR=1.12 (0.71 to 1.78)	0.626	0.895	112939	806	112133	phecode 475	respiratory
Viral infection	OR=1.13 (0.69 to 1.83)	0.629	0.895	117630	733	116897	phecode 079	infectious diseases
Pneumonia	OR=1.05 (0.85 to 1.3)	0.633	0.895	118432	3989	114443	phecode 480	respiratory
Iron deficiency anemias	OR=0.93 (0.69 to 1.25)	0.634	0.895	115914	2000	113914	phecode 280	hematopoietic
Iron deficiency anemias; unspecified or not due to blood loss	OR=0.93 (0.69 to 1.25)	0.634	0.895	115914	2000	113914	phecode 280.1	hematopoietic
Ganglion and cyst of synovium; tendon; and bursa	OR=0.89 (0.54 to 1.46)	0.638	0.895	109467	692	108775	phecode 727.4	musculoskeletal
Unstable angina intermediate coronary syndrome	OR=1.1 (0.73 to 1.66)	0.638	0.895	110113	1058	109055	phecode 411.1	circulatory system
Aspartate aminotransferase	0.009 (-0.029 to 0.048)	0.642	0.896	112601	NA	NA	NA	biomarker
LDL direct	-0.008 (-0.039 to 0.024)	0.642	0.896	112819	NA	NA	NA	biomarker
Other disorders of testis	OR=0.92 (0.65 to 1.31)	0.645	0.897	107732	1397	106335	phecode 603	genitourinary
Complication of internal orthopedic device	OR=1.08 (0.76 to 1.54)	0.661	0.916	112841	1413	111428	phecode 858	injuries and poisonings
Other disorders of the kidney and ureters	OR=1.09 (0.73 to 1.64)	0.664	0.916	115814	1079	114735	phecode 586	genitourinary
Disturbance of skin sensation	OR=1.11 (0.69 to 1.77)	0.665	0.916	117209	788	116421	phecode 687.4	dermatologic
Atopicorcontact dermatitis due to other or unspecified	OR=1.12 (0.67 to 1.86)	0.667	0.916	117229	673	116556	phecode 939	dermatologic
Cardiac pacemakerordevice in situ	OR=1.1 (0.7 to 1.74)	0.674	0.921	110035	849	109186	phecode 426.9	circulatory system
Inflammatory bowel disease and other gastroenteritis and colitis	OR=1.08 (0.76 to 1.51)	0.676	0.921	98852	1497	97355	phecode 555	digestive
Other disorders of eye	OR=1.08 (0.74 to 1.59)	0.678	0.921	117591	1188	116403	phecode 379	sense organs
Psoriasis vulgaris	OR=1.12 (0.65 to 1.91)	0.680	0.921	109225	603	108622	phecode 696.41	dermatologic
Paroxysmal tachycardia; unspecified	OR=0.91 (0.59 to 1.4)	0.681	0.921	110132	946	109186	phecode 427.1	circulatory system
Other retinal disorders	OR=1.08 (0.75 to 1.56)	0.684	0.922	79312	1328	77984	phecode 362	sense organs
Abal sputum	OR=0.93 (0.65 to 1.32)	0.685	0.922	118535	1412	117123	phecode 516	respiratory
Purpura and other hemorrhagic conditions	OR=1.11 (0.66 to 1.88)	0.694	0.925	118276	633	117643	phecode 287	hematopoietic
Hematuria	OR=1.04 (0.86 to 1.25)	0.694	0.925	115639	5220	110419	phecode 593	genitourinary
Fever of unknown origin	OR=1.08 (0.74 to 1.56)	0.696	0.925	118535	1277	117258	phecode 783	symptoms
Peripheral vascular disease	OR=1.09 (0.72 to 1.65)	0.696	0.925	116826	1014	115812	phecode 443	circulatory system

Cardiac congenital anomalies	OR=1.09 (0.69 to 1.73)	0.705	0.933	118458	834	117624	phecode 747.1	congenital anomalies
Other anemias	OR=0.95 (0.74 to 1.22)	0.709	0.935	116809	2895	113914	phecode 285	hematopoietic
Other disorders of male genital organs	OR=0.94 (0.69 to 1.29)	0.710	0.935	108144	1809	106335	phecode 608	genitourinary
Synovitis and tenosynovitis	OR=1.07 (0.75 to 1.51)	0.712	0.935	110231	1456	108775	phecode 727.1	musculoskeletal
Cancer of urinary organs incl kidney and bladder	OR=0.94 (0.69 to 1.29)	0.720	0.940	118535	1812	116723	phecode 189	neoplasms
Other diseases of respiratory system; NEC	OR=1.05 (0.82 to 1.34)	0.721	0.940	118346	2850	115496	phecode 519.8	respiratory
Streptococcus infection	OR=0.9 (0.51 to 1.61)	0.724	0.941	114425	519	113906	phecode 041.2	infectious diseases
Inflammation of the eye	OR=1.08 (0.71 to 1.64)	0.725	0.941	116764	986	115778	phecode 371	sense organs
Sepsis	OR=0.93 (0.63 to 1.38)	0.732	0.944	118535	1144	117391	phecode 994.2	injuries and poisonings
Sepsis and SIRS	OR=0.93 (0.63 to 1.38)	0.732	0.944	118535	1144	117391	phecode 994	injuries and poisonings
Decreased white blood cell count	OR=1.07 (0.7 to 1.64)	0.739	0.947	117121	971	116150	phecode 288.1	hematopoietic
Neutropenia	OR=1.07 (0.7 to 1.64)	0.739	0.947	117121	971	116150	phecode 288.11	hematopoietic
Carditis	OR=0.92 (0.57 to 1.49)	0.743	0.947	118145	762	117383	phecode 420	circulatory system
Diseases of hair and hair follicles	OR=0.95 (0.69 to 1.3)	0.746	0.947	118220	1726	116494	phecode 704	dermatologic
Angina pectoris	OR=0.97 (0.79 to 1.19)	0.746	0.947	113465	4410	109055	phecode 411.3	circulatory system
Heart valve replaced	OR=1.11 (0.59 to 2.08)	0.747	0.947	114462	440	114022	phecode 395.6	circulatory system
Vertiginous syndromes and other disorders of vestibular system	OR=1.05 (0.76 to 1.45)	0.751	0.949	118535	1709	116826	phecode 386	sense organs
Sebaceous cyst	OR=1.04 (0.8 to 1.35)	0.754	0.949	118283	2608	115675	phecode 706.2	dermatologic
Emphysema	OR=1.08 (0.65 to 1.79)	0.757	0.951	109953	688	109265	phecode 496.1	respiratory
Diabetes mellitus	OR=0.97 (0.8 to 1.18)	0.759	0.951	118535	4706	113829	phecode 250	endocrine metabolic
Urinary calculus	OR=0.96 (0.74 to 1.24)	0.764	0.955	118049	2665	115384	phecode 594	genitourinary
Atherosclerosis	OR=1.1 (0.57 to 2.13)	0.772	0.958	116215	403	115812	phecode 440	circulatory system
Other ability of urination	OR=1.07 (0.65 to 1.77)	0.777	0.958	111278	704	110574	phecode 599.9	genitourinary
Other chronic ischemic heart disease; unspecified	OR=0.98 (0.84 to 1.14)	0.779	0.958	116939	7884	109055	phecode 411.8	circulatory system
Diseases of sebaceous glands	OR=1.04 (0.8 to 1.34)	0.779	0.958	118300	2625	115675	phecode 706	dermatologic
Fracture of hand or wrist	OR=0.95 (0.66 to 1.37)	0.780	0.958	113709	1314	112395	phecode 804	injuries and poisonings
Cancer of other lymphoid; histiocytic tissue	OR=0.94 (0.62 to 1.43)	0.782	0.958	117624	1006	116618	phecode 202	neoplasms
Bronchiectasis	OR=1.08 (0.62 to 1.89)	0.782	0.958	109829	564	109265	phecode 496.3	respiratory
Phlebitis and thrombophlebitis of lower extremities	OR=1.05 (0.73 to 1.51)	0.788	0.962	108653	1352	107301	phecode 451.2	circulatory system
Fracture of ribs	OR=0.94 (0.58 to 1.52)	0.791	0.962	113146	751	112395	phecode 807	injuries and poisonings
Leukemia	OR=1.06 (0.67 to 1.69)	0.793	0.962	117429	811	116618	phecode 204	neoplasms
Cardiac dysrhythmias	OR=1.02 (0.88 to 1.19)	0.798	0.962	117352	8166	109186	phecode 427	circulatory system
Symptoms affecting skin	OR=1.04 (0.78 to 1.39)	0.798	0.962	118535	2114	116421	phecode 687	dermatologic
Hereditary retinal dystrophies	OR=1.11 (0.51 to 2.43)	0.799	0.962	78268	284	77984	phecode 362.7	sense organs
Glaucoma	OR=1.04 (0.75 to 1.45)	0.803	0.962	116158	1618	114540	phecode 365	sense organs
Abal findings on examination of urine	OR=1.04 (0.73 to 1.48)	0.809	0.962	118535	1419	117116	phecode 598	genitourinary
Symptoms and disorders of the joints	OR=1.05 (0.71 to 1.55)	0.809	0.962	117639	1141	116498	phecode 741	musculoskeletal
Non Hodgkins lymphoma	OR=0.95 (0.6 to 1.5)	0.810	0.962	117444	826	116618	phecode 202.2	neoplasms
Benign neoplasm of unspecified sites	OR=0.94 (0.56 to 1.56)	0.811	0.962	118535	668	117867	phecode 229	neoplasms
Pleurisy; pleural effusion	OR=1.03 (0.79 to 1.36)	0.814	0.962	116570	2399	114171	phecode 507	respiratory
Ischemic Heart Disease	OR=0.98 (0.85 to 1.14)	0.814	0.962	118315	9260	109055	phecode 411	circulatory system
Skull and face fracture and other intercranial injury	OR=1.04 (0.74 to 1.47)	0.821	0.967	118535	1465	117070	phecode 819	injuries and poisonings
Secondary malignant neoplasm of liver	OR=0.95 (0.62 to 1.46)	0.824	0.967	106763	967	105796	phecode 198.4	neoplasms
Cardiomegaly	OR=0.95 (0.62 to 1.47)	0.824	0.967	116993	928	116065	phecode 416	circulatory system
Other aneurysm	OR=1.05 (0.65 to 1.71)	0.828	0.967	116560	748	115812	phecode 442	circulatory system
Urinary incontinence	OR=0.95 (0.59 to 1.52)	0.830	0.967	111356	782	110574	phecode 599.4	genitourinary
Intervertebral disc disorders	OR=1.03 (0.77 to 1.39)	0.834	0.967	116539	1975	114564	phecode 722	musculoskeletal
Complications of cardiacorvascular device; implant; and graft	OR=1.05 (0.64 to 1.74)	0.837	0.967	112127	699	111428	phecode 854	injuries and poisonings
Peptic ulcer excl esophageal	OR=1.03 (0.78 to 1.36)	0.839	0.967	118535	2292	116243	phecode 531	digestive
Degenerative disease of the spinal cord	OR=1.05 (0.64 to 1.73)	0.840	0.967	112579	697	111882	phecode 334	neurological
Abal results of function study of liver	OR=1.04 (0.7 to 1.54)	0.842	0.967	116659	1140	115519	phecode 573.7	digestive
Secondary malignancy of respiratory organs	OR=1.05 (0.66 to 1.67)	0.846	0.967	106607	811	105796	phecode 198.2	neoplasms
Cancer of bronchus; lung	OR=0.96 (0.62 to 1.48)	0.848	0.967	118327	922	117405	phecode 165.1	neoplasms
Cancer within the respiratory system	OR=1.04 (0.7 to 1.54)	0.849	0.967	118523	1118	117405	phecode 165	neoplasms
Other acquired deformities of limbs	OR=1.05 (0.62 to 1.8)	0.850	0.967	116967	600	116367	phecode 736	musculoskeletal
Syncope and collapse	OR=0.98 (0.77 to 1.25)	0.854	0.967	118535	2972	115563	phecode 788	symptoms

Phlebitis and thrombophlebitis	OR=1.03 (0.73 to 1.47)	0.855	0.967	108740	1439	107301	phecode 451	circulatory system
Cerebrovascular disease	OR=1.02 (0.79 to 1.33)	0.855	0.967	118107	2655	115452	phecode 433	circulatory system
Other disorders of biliary tract	OR=1.05 (0.61 to 1.8)	0.863	0.967	115665	591	115074	phecode 575.8	digestive
Benign neoplasm of skin	OR=0.98 (0.73 to 1.3)	0.863	0.967	118364	2131	116233	phecode 216	neoplasms
Left bundle branch block	OR=1.05 (0.58 to 1.9)	0.863	0.967	109688	502	109186	phecode 426.32	circulatory system
Aphakia and other disorders of lens	OR=1.05 (0.6 to 1.84)	0.864	0.967	116962	559	116403	phecode 379.3	sense organs
Cerebral artery occlusion; with cerebral infarction	OR=1.05 (0.61 to 1.78)	0.869	0.968	116065	613	115452	phecode 433.21	circulatory system
Other symptoms of respiratory system	OR=0.98 (0.79 to 1.23)	0.870	0.968	118535	3648	114887	phecode 512	respiratory
Calculus of bile duct	OR=0.96 (0.54 to 1.68)	0.877	0.971	115620	546	115074	phecode 574.2	digestive
Acute upper respiratory infections of multiple or unspecified sites	OR=0.96 (0.56 to 1.64)	0.881	0.971	118493	602	117891	phecode 465	respiratory
Other disorders of synovium; tendon; and bursa	OR=1.02 (0.79 to 1.32)	0.882	0.971	111503	2728	108775	phecode 727	musculoskeletal
Cataract	OR=0.99 (0.82 to 1.18)	0.883	0.971	118535	5896	112639	phecode 366	sense organs
Stricture and stenosis of esophagus	OR=1.04 (0.65 to 1.66)	0.883	0.971	105849	794	105055	phecode 530.3	digestive
Abdominal aortic aneurysm	OR=0.95 (0.49 to 1.86)	0.886	0.971	116201	389	115812	phecode 442.11	circulatory system
Anal and rectal polyp	OR=1.02 (0.79 to 1.31)	0.892	0.971	111137	2858	108279	phecode 565.1	digestive
Cellulitis and abscess of armorhand	OR=0.98 (0.73 to 1.32)	0.898	0.971	116649	1981	114668	phecode 681.3	dermatologic
Cellulitis and abscess of foot; toe	OR=0.98 (0.73 to 1.32)	0.898	0.971	116649	1981	114668	phecode 681.6	dermatologic
Cellulitis and abscess of leg; except foot	OR=0.98 (0.73 to 1.32)	0.898	0.971	116649	1981	114668	phecode 681.5	dermatologic
Cerebral ischemia	OR=0.98 (0.67 to 1.42)	0.898	0.971	116678	1226	115452	phecode 433.3	circulatory system
Other disorders of liver	OR=0.98 (0.76 to 1.28)	0.898	0.971	118088	2569	115519	phecode 573	digestive
Ulcerative colitis	OR=1.02 (0.69 to 1.52)	0.905	0.975	98472	1117	97355	phecode 555.2	digestive
Hydrocele	OR=1.03 (0.65 to 1.61)	0.914	0.979	107190	855	106335	phecode 603.1	genitourinary
Lipoprotein A	0.002 (-0.041 to 0.046)	0.915	0.979	90248	NA	NA	NA	biomarker
Pulmonary collapse; interstitial and compensatory emphysema	OR=0.98 (0.61 to 1.56)	0.920	0.979	114965	794	114171	phecode 508	respiratory
Palpitations	OR=0.98 (0.61 to 1.56)	0.921	0.979	109968	782	109186	phecode 427.9	circulatory system
Hearing loss	OR=0.98 (0.71 to 1.36)	0.924	0.979	118518	1665	116853	phecode 389	sense organs
Substance addiction and disorders	OR=0.99 (0.76 to 1.29)	0.926	0.979	110152	2492	107660	phecode 316	mental disorders
Coronary atherosclerosis	OR=0.99 (0.84 to 1.17)	0.929	0.979	115660	6605	109055	phecode 411.4	circulatory system
Other forms of chronic heart disease	OR=1.02 (0.63 to 1.65)	0.931	0.979	109814	759	109055	phecode 414	circulatory system
Septicemia	OR=0.99 (0.7 to 1.38)	0.933	0.979	115435	1529	113906	phecode 038	infectious diseases
Diseases of hard tissues of teeth	OR=0.98 (0.65 to 1.48)	0.934	0.979	116538	1038	115500	phecode 521	digestive
Cardiac and circulatory congenital anomalies	OR=1.02 (0.66 to 1.58)	0.935	0.979	118535	911	117624	phecode 747	congenital anomalies
Respiratory failure; insufficiency; arrest	OR=1.02 (0.67 to 1.55)	0.935	0.979	115147	976	114171	phecode 509	respiratory
Hyposmolality andoror hyponatremia	OR=0.98 (0.57 to 1.68)	0.937	0.979	116739	594	116145	phecode 276.12	endocrine metabolic
Acquired toe deformities	OR=1.02 (0.6 to 1.72)	0.945	0.982	117001	634	116367	phecode 735.2	musculoskeletal
Disorder of skin and subcutaneous tissue NOS	OR=1.01 (0.74 to 1.39)	0.946	0.982	118535	1760	116775	phecode 689	dermatologic
Rash and other nonspecific skin eruption	OR=0.98 (0.59 to 1.65)	0.947	0.982	117067	646	116421	phecode 687.1	dermatologic
Bacterial pneumonia	OR=0.99 (0.76 to 1.29)	0.951	0.982	116978	2535	114443	phecode 480.1	respiratory
Inflammatory and toxic neuropathy	OR=0.98 (0.56 to 1.72)	0.951	0.982	118278	552	117726	phecode 357	neurological
Dental caries	OR=0.99 (0.65 to 1.49)	0.953	0.982	116527	1027	115500	phecode 521.1	digestive
Cardiac arrest and ventricular fibrillation	OR=1.02 (0.54 to 1.91)	0.958	0.984	109621	435	109186	phecode 427.4	circulatory system
Psoriasis and related disorders	OR=1.01 (0.63 to 1.62)	0.961	0.984	109402	780	108622	phecode 696	dermatologic
Hematemesis	OR=1.01 (0.61 to 1.69)	0.962	0.984	111896	664	111232	phecode 578.1	digestive
Other upper respiratory disease	OR=1.01 (0.7 to 1.44)	0.967	0.988	113495	1362	112133	phecode 479	respiratory
Dysphagia	OR=0.99 (0.73 to 1.35)	0.970	0.988	106951	1896	105055	phecode 532	digestive
Urea	0.001 (-0.047 to 0.049)	0.972	0.988	112971	NA	NA	NA	biomarker
Renal failure NOS	OR=1.01 (0.52 to 1.96)	0.979	0.993	115131	396	114735	phecode 585.2	genitourinary
Acute renal failure	OR=1 (0.73 to 1.37)	0.988	0.995	116538	1803	114735	phecode 585.1	genitourinary
Duodenal ulcer	OR=1 (0.68 to 1.47)	0.989	0.995	117411	1168	116243	phecode 531.3	digestive
Cystitis	OR=1 (0.58 to 1.7)	0.990	0.995	111030	611	110419	phecode 592.1	genitourinary
Calculus of ureter	OR=1 (0.66 to 1.5)	0.993	0.995	116430	1046	115384	phecode 594.3	genitourinary
Psoriasis	OR=1 (0.62 to 1.62)	0.993	0.995	109381	759	108622	phecode 696.4	dermatologic
Pneumococcal pneumonia	OR=1 (0.76 to 1.32)	0.999	0.999	116697	2254	114443	phecode 480.11	respiratory

Supplementary File 1 - Table 7. Associations of genetically predicted calculated free testosterone for 439 health outcomes across the human phenome excluding individuals on cholesterol-lowering medication

Trait	Effect per 0.1 nmol/L increase CFT (95% CI)	P-value	FDR-adjusted p-value	Sample Size	Number of Cases	Number of Controls	Phecode
Creatinine	0.119 (0.082 to 0.156)	3.23E-10	1.42E-07	115799	NA	NA	NA
C-reactive protein	-0.08 (-0.119 to -0.042)	3.71E-05	8.17E-03	115569	NA	NA	NA
Apolipoprotein A	-0.019 (-0.028 to -0.01)	5.97E-05	8.78E-03	106807	NA	NA	NA
HDL cholesterol	-0.077 (-0.116 to -0.038)	1.19E-04	0.012	106972	NA	NA	NA
Umbilical hernia	OR=1.83 (1.34 to 2.49)	1.39E-04	0.012	103930	1814	102116	phecode 550.4
Hypotension	OR=0.56 (0.41 to 0.76)	2.20E-04	0.016	116879	1909	114970	phecode 458
Degenerative skin conditions and other dermatoses	OR=0.56 (0.41 to 0.77)	3.46E-04	0.022	120338	1785	118553	phecode 702
IGF1	0.364 (0.163 to 0.565)	3.95E-04	0.022	115263	NA	NA	NA
Essential hypertension	OR=1.19 (1.07 to 1.32)	8.25E-04	0.038	121158	21233	99925	phecode 401.1
Hypotension NOS	OR=0.57 (0.41 to 0.79)	8.73E-04	0.038	116614	1644	114970	phecode 458.9
Hypertension	OR=1.19 (1.07 to 1.31)	9.54E-04	0.038	121244	21319	99925	phecode 401
Spinal stenosis	OR=1.88 (1.28 to 2.77)	1.25E-03	0.046	118554	1175	117379	phecode 720
Hypercholesterolemia	OR=1.31 (1.11 to 1.55)	1.62E-03	0.055	120894	6512	114382	phecode 272.11
Inguinal hernia	OR=1.24 (1.08 to 1.42)	2.01E-03	0.063	112581	10465	102116	phecode 550.1
Hemiplegia	OR=0.38 (0.2 to 0.71)	2.43E-03	0.070	115099	454	114645	phecode 342
Abdominal hernia	OR=1.17 (1.06 to 1.3)	2.56E-03	0.070	121474	19358	102116	phecode 550
Seborrheic keratosis	OR=0.53 (0.35 to 0.81)	3.17E-03	0.081	120693	1004	119689	phecode 702.2
Back pain	OR=1.38 (1.11 to 1.71)	3.34E-03	0.081	121474	3861	117613	phecode 760
Disorders of lipoid metabolism	OR=1.27 (1.08 to 1.5)	3.66E-03	0.081	121474	7092	114382	phecode 272
Internal derangement of knee	OR=1.29 (1.09 to 1.53)	3.76E-03	0.081	120667	6223	114444	phecode 835
Hyperlipidemia	OR=1.27 (1.08 to 1.49)	3.88E-03	0.081	121447	7065	114382	phecode 272.1
Fasciitis	OR=0.66 (0.49 to 0.89)	6.35E-03	0.122	113534	1970	111564	phecode 728.7
Inflammatory diseases of prostate	OR=1.55 (1.13 to 2.13)	6.35E-03	0.122	39044	1872	37172	phecode 601
Other disorders of urethra and urinary tract	OR=1.61 (1.14 to 2.27)	6.93E-03	0.123	117396	1476	115920	phecode 597
Albumin	-0.136 (-0.235 to -0.037)	7.00E-03	0.123	107018	NA	NA	NA
GERD	OR=1.24 (1.06 to 1.45)	7.65E-03	0.130	115190	7426	107764	phecode 530.11
Phosphate	-0.009 (-0.015 to -0.002)	9.09E-03	0.148	106811	NA	NA	NA
Contracture of palmar fascia Dupuytrens disease	OR=0.67 (0.49 to 0.91)	0.012	0.180	113412	1848	111564	phecode 728.71
Fracture of upper limb	OR=0.71 (0.54 to 0.93)	0.012	0.180	117584	2437	115147	phecode 803
Lipoma	OR=1.4 (1.07 to 1.84)	0.013	0.180	121170	2429	118741	phecode 214
Other symptoms involving abdomen and pelvis	OR=1.69 (1.11 to 2.56)	0.014	0.180	114923	999	113924	phecode 579
Urethral stricture not specified as infectious	OR=1.58 (1.1 to 2.28)	0.014	0.180	117220	1300	115920	phecode 597.1
Convulsions	OR=1.75 (1.11 to 2.74)	0.015	0.180	115498	853	114645	phecode 345.3
Osteoporosis	OR=0.53 (0.32 to 0.88)	0.015	0.180	121122	673	120449	phecode 743.1
Osteoporosis NOS	OR=0.53 (0.32 to 0.88)	0.015	0.180	121122	673	120449	phecode 743.11
Gout	OR=1.49 (1.08 to 2.05)	0.015	0.180	121261	1712	119549	phecode 274.1
Prostatitis	OR=1.69 (1.11 to 2.57)	0.015	0.180	38199	1027	37172	phecode 601.1
Chronic dermatitis due to solar radiation	OR=0.58 (0.37 to 0.9)	0.016	0.180	120352	904	119448	phecode 938.2
Unspecified monoarthritis	OR=1.22 (1.04 to 1.44)	0.016	0.184	115270	6797	108473	phecode 716.2
Disorders of muscle; ligament; and fascia	OR=0.71 (0.53 to 0.94)	0.018	0.198	113758	2194	111564	phecode 728
Mitral valve disease	OR=0.61 (0.41 to 0.93)	0.020	0.209	117822	1046	116776	phecode 394.2
Actinic keratosis	OR=0.59 (0.37 to 0.92)	0.020	0.209	120555	866	119689	phecode 702.1
Chronic bronchitis	OR=1.34 (1.04 to 1.73)	0.022	0.211	114823	2819	112004	phecode 496.2
Obstructive chronic bronchitis	OR=1.35 (1.04 to 1.74)	0.022	0.211	114769	2765	112004	phecode 496.21
Gout and other crystal arthropathies	OR=1.42 (1.05 to 1.92)	0.022	0.211	121474	1925	119549	phecode 274
Dermatitis due to solar radiation	OR=0.6 (0.39 to 0.93)	0.023	0.211	120381	933	119448	phecode 938
Total protein	-0.185 (-0.344 to -0.026)	0.023	0.211	106886	NA	NA	NA
Spondylosis and allied disorders	OR=1.41 (1.05 to 1.88)	0.023	0.211	119424	2045	117379	phecode 721
Cancer; suspected or other	OR=1.18 (1.02 to 1.36)	0.024	0.219	117604	9438	108166	phecode 195
Cerebral artery occlusion; with cerebral infarction	OR=0.54 (0.32 to 0.93)	0.025	0.220	118973	611	118362	phecode 433.21
Total bilirubin	0.044 (0.005 to 0.082)	0.026	0.221	115368	NA	NA	NA
Other local infections of skin and subcutaneous tissue	OR=0.69 (0.49 to 0.96)	0.026	0.221	118948	1611	117337	phecode 686
Nonrheumatic mitral valve disorders	OR=0.63 (0.41 to 0.95)	0.027	0.223	117803	1027	116776	phecode 395.1
Other disorders of prostate	OR=1.71 (1.05 to 2.79)	0.030	0.244	37956	784	37172	phecode 602
Lipoma of skin and subcutaneous tissue	OR=1.43 (1.03 to 1.98)	0.030	0.244	120398	1657	118741	phecode 214.1
Tobacco use disorder	OR=1.2 (1.02 to 1.42)	0.032	0.249	117178	6575	110603	phecode 318
Malignant neoplasm; other	OR=1.17 (1.01 to 1.35)	0.033	0.251	117382	9216	108166	phecode 195.1
Fracture of unspecified part of femur	OR=0.57 (0.34 to 0.96)	0.033	0.251	115801	654	115147	phecode 800.2
Calcium	-0.004 (-0.007 to 0)	0.034	0.251	106982	NA	NA	NA
Other disorders of peritoneum	OR=1.59 (1.03 to 2.44)	0.035	0.254	111884	943	110941	phecode 568

Fracture of lower limb	OR=0.73 (0.55 to 0.98)	0.038	0.276	117176	2029	115147	phecode	800
Sciatica	OR=1.73 (1.03 to 2.92)	0.039	0.277	120988	632	120356	phecode	764
Intestinal obstruction without mention of hernia	OR=1.47 (1.02 to 2.12)	0.040	0.277	100944	1313	99631	phecode	560
Fracture of radius and ulna	OR=0.67 (0.45 to 0.98)	0.040	0.277	116318	1171	115147	phecode	803.2
Other symptomsordisorders or the urinary system	OR=1.17 (1.01 to 1.36)	0.042	0.286	121474	8452	113022	phecode	599
Secondary malignant neoplasm	OR=1.28 (1.01 to 1.64)	0.043	0.286	111236	3070	108166	phecode	198
Peripheral vascular disease; unspecified	OR=1.75 (1.02 to 3.01)	0.043	0.286	119389	590	118799	phecode	443.9
Right bundle branch block	OR=0.6 (0.36 to 0.99)	0.044	0.288	112256	693	111563	phecode	426.31
Urinary tract infection	OR=1.26 (1 to 1.59)	0.052	0.331	116204	3340	112864	phecode	591
Urinary obstruction	OR=1.47 (0.99 to 2.19)	0.054	0.343	114139	1117	113022	phecode	599.1
Peritoneal adhesions postoperative postinfection	OR=1.54 (0.99 to 2.39)	0.057	0.353	111827	886	110941	phecode	568.1
Other disorders of intestine	OR=1.19 (0.99 to 1.42)	0.058	0.353	116735	5794	110941	phecode	569
Other and unspecified disc disorder	OR=1.38 (0.99 to 1.94)	0.060	0.360	118897	1518	117379	phecode	722.9
Esophagitis; GERD and related diseases	OR=1.14 (0.99 to 1.3)	0.061	0.360	118118	10354	107764	phecode	530.1
Duodenitis	OR=1.27 (0.99 to 1.64)	0.061	0.360	114579	2784	111795	phecode	535.6
Allergyoradverse effect of penicillin	OR=1.22 (0.99 to 1.51)	0.066	0.386	118367	3882	114485	phecode	960.2
Hydronephrosis	OR=1.51 (0.96 to 2.37)	0.073	0.419	118881	853	118028	phecode	595
Symptoms involving head and neck	OR=1.41 (0.97 to 2.05)	0.074	0.419	121474	1231	120243	phecode	293
Alkaline phosphatase	-0.034 (-0.072 to 0.004)	0.076	0.423	115860	NA	NA	NA	
Esophageal bleeding varicesorhemorrhage	OR=0.63 (0.38 to 1.05)	0.078	0.427	108450	686	107764	phecode	530.2
Delirium dementia and amnestic and other cognitive disorders	OR=1.45 (0.96 to 2.21)	0.080	0.434	119811	1007	118804	phecode	290
Other peripheral nerve disorders	OR=1.27 (0.97 to 1.66)	0.081	0.434	119802	2468	117334	phecode	351
Erythematous conditions	OR=1.61 (0.94 to 2.75)	0.083	0.442	120153	600	119553	phecode	695
Arthropathy NOS	OR=1.12 (0.98 to 1.28)	0.085	0.443	120189	11716	108473	phecode	716.9
Chronic airway obstruction	OR=1.22 (0.97 to 1.54)	0.085	0.443	115434	3430	112004	phecode	496
Gamma glutamyltransferase	0.033 (-0.005 to 0.072)	0.090	0.459	115794	NA	NA	NA	
Alanine aminotransferase	0.033 (-0.005 to 0.071)	0.093	0.472	115780	NA	NA	NA	
Sleep apnea	OR=1.3 (0.95 to 1.77)	0.096	0.482	121139	1818	119321	phecode	327.3
Visual disturbances	OR=0.7 (0.46 to 1.07)	0.100	0.484	121474	1006	120468	phecode	368
Other arthropathies	OR=1.11 (0.98 to 1.27)	0.100	0.484	120221	11748	108473	phecode	716
Secondary malignancy of lymph nodes	OR=1.33 (0.95 to 1.88)	0.100	0.484	109648	1482	108166	phecode	198.1
Intervertebral disc disorders	OR=1.28 (0.95 to 1.71)	0.102	0.484	119426	2047	117379	phecode	722
Osteoarthrosis NOS	OR=1.22 (0.96 to 1.54)	0.102	0.484	116919	3244	113675	phecode	740.9
Abal heart sounds	OR=0.81 (0.63 to 1.04)	0.104	0.484	119519	2743	116776	phecode	396
Functional digestive disorders	OR=1.13 (0.98 to 1.3)	0.105	0.484	108831	9200	99631	phecode	564
Altered mental status	OR=1.44 (0.92 to 2.26)	0.107	0.484	119678	874	118804	phecode	292.4
Direct bilirubin	0.033 (-0.007 to 0.072)	0.108	0.484	107392	NA	NA	NA	
Acute pulmonary heart disease	OR=1.32 (0.94 to 1.84)	0.109	0.484	120369	1558	118811	phecode	415.1
Pulmonary embolism and infarction; acute	OR=1.32 (0.94 to 1.84)	0.109	0.484	120369	1558	118811	phecode	415.11
Diseases of the oral soft tissues; excluding lesions specific for gingiva and tongue	OR=1.35 (0.93 to 1.97)	0.110	0.484	120982	1254	119728	phecode	528
Tachycardia NOS	OR=0.67 (0.41 to 1.1)	0.112	0.489	112295	732	111563	phecode	427.7
Peptic ulcer excl esophageal	OR=0.8 (0.61 to 1.05)	0.114	0.489	121474	2348	119126	phecode	531
Diseases of pancreas	OR=1.42 (0.92 to 2.21)	0.114	0.489	121474	901	120573	phecode	577
Osteoporosis; osteopenia and pathological fracture	OR=0.72 (0.47 to 1.09)	0.116	0.494	121474	1025	120449	phecode	743
Benign neoplasm of colon	OR=1.13 (0.97 to 1.32)	0.118	0.495	101110	7856	93254	phecode	208
Other specified peripheral vascular diseases	OR=1.43 (0.91 to 2.27)	0.123	0.510	119630	831	118799	phecode	443.8
Diseases of esophagus	OR=1.11 (0.97 to 1.26)	0.126	0.517	119230	11466	107764	phecode	530
Osteoarthrosis	OR=1.13 (0.97 to 1.32)	0.127	0.517	121474	7799	113675	phecode	740
Other acute and subacute forms of ischemic heart disease	OR=1.55 (0.88 to 2.73)	0.128	0.517	113645	542	113103	phecode	411.9
Other biliary tract disease	OR=1.37 (0.91 to 2.05)	0.130	0.522	118799	1063	117736	phecode	575
Other anemias	OR=0.83 (0.65 to 1.06)	0.134	0.532	119722	3085	116637	phecode	285
Degeneration of macula and posterior pole of retina	OR=1.49 (0.88 to 2.52)	0.138	0.537	81342	642	80700	phecode	362.2
Macular degeneration senile of retina NOS	OR=1.49 (0.88 to 2.52)	0.138	0.537	81342	642	80700	phecode	362.29
Duodenal ulcer	OR=0.75 (0.51 to 1.1)	0.142	0.546	120344	1218	119126	phecode	531.3
Anal and rectal conditions	OR=1.15 (0.96 to 1.38)	0.143	0.546	116383	5442	110941	phecode	565
Gastrointestinal hemorrhage	OR=1.13 (0.96 to 1.33)	0.144	0.548	120663	6739	113924	phecode	578
Ability of gait	OR=0.71 (0.44 to 1.14)	0.153	0.575	121091	787	120304	phecode	350.2
Cardiomegaly	OR=0.74 (0.5 to 1.12)	0.155	0.580	119874	1063	118811	phecode	416
Poisoning by antibiotics	OR=1.16 (0.94 to 1.43)	0.157	0.581	118722	4237	114485	phecode	960
Abdominal pain	OR=1.11 (0.96 to 1.27)	0.160	0.585	121474	9393	112081	phecode	785
Fracture of clavicle or scapula	OR=0.73 (0.47 to 1.14)	0.161	0.585	116023	876	115147	phecode	803.3

Orthostatic hypotension	OR=0.66 (0.37 to 1.18)	0.162	0.585	115481	511	114970	phecode 458.1
Abal findings examination of lungs	OR=1.38 (0.88 to 2.15)	0.164	0.585	121474	866	120608	phecode 514
Personal history of diseases of digestive system	OR=1.14 (0.95 to 1.37)	0.165	0.585	105175	5544	99631	phecode 564.9
First degree AV block	OR=0.64 (0.34 to 1.2)	0.167	0.590	112010	447	111563	phecode 426.21
Other disorders of soft tissues	OR=0.82 (0.62 to 1.09)	0.178	0.610	113708	2144	111564	phecode 729
Symptoms involving digestive system	OR=1.16 (0.94 to 1.42)	0.178	0.610	103712	4081	99631	phecode 561
Fracture of tibia and fibula	OR=0.72 (0.44 to 1.16)	0.179	0.610	115899	752	115147	phecode 800.3
Nasal polyps	OR=1.26 (0.9 to 1.78)	0.183	0.610	116293	1480	114813	phecode 471
Pulmonary heart disease	OR=1.24 (0.9 to 1.71)	0.184	0.610	120510	1699	118811	phecode 415
Occlusion and stenosis of precerebral arteries	OR=1.58 (0.8 to 3.12)	0.186	0.610	118737	375	118362	phecode 433.1
Other specified cardiac dysrhythmias	OR=0.78 (0.53 to 1.13)	0.188	0.610	112808	1245	111563	phecode 427.3
Poisoning by primarily systemic agents	OR=1.39 (0.85 to 2.28)	0.189	0.610	115201	716	114485	phecode 963
Cholelithiasis with other cholecystitis	OR=1.3 (0.88 to 1.93)	0.190	0.610	118863	1127	117736	phecode 574.12
Other disorders of bone and cartilage	OR=0.77 (0.52 to 1.14)	0.190	0.610	119014	1121	117893	phecode 733
Otitis media and Eustachian tube disorders	OR=1.4 (0.84 to 2.32)	0.193	0.610	120959	675	120284	phecode 381
Vitamin D	0.026 (-0.013 to 0.064)	0.193	0.610	111844	NA	NA	NA
Frequency of urination and polyuria	OR=1.27 (0.89 to 1.81)	0.194	0.610	114389	1367	113022	phecode 599.5
Secondary malignancy of bone	OR=1.33 (0.86 to 2.06)	0.194	0.610	109093	927	108166	phecode 198.6
Postoperative infection	OR=1.27 (0.88 to 1.84)	0.195	0.610	121145	1304	119841	phecode 080
Retention of urine	OR=1.16 (0.93 to 1.45)	0.196	0.610	116708	3686	113022	phecode 599.2
Heart valve disorders	OR=0.82 (0.61 to 1.11)	0.197	0.610	118770	1994	116776	phecode 395
Hemorrhage of gastrointestinal tract	OR=1.21 (0.9 to 1.61)	0.202	0.622	116027	2103	113924	phecode 578.9
Peripheral vascular disease	OR=1.33 (0.85 to 2.08)	0.206	0.631	119676	877	118799	phecode 443
Type 1 diabetes	OR=0.64 (0.32 to 1.28)	0.208	0.632	117632	363	117269	phecode 250.1
Intestinal infection	OR=1.17 (0.92 to 1.5)	0.211	0.636	121474	2938	118536	phecode 008
Circulatory disease NEC	OR=0.89 (0.73 to 1.07)	0.213	0.638	119934	4964	114970	phecode 459.9
Open wounds of head; neck; and trunk	OR=1.23 (0.89 to 1.69)	0.215	0.638	119102	1682	117420	phecode 870
Epilepsy; recurrent seizures; convulsions	OR=1.22 (0.89 to 1.68)	0.216	0.638	116382	1737	114645	phecode 345
Bundle branch block	OR=0.79 (0.54 to 1.15)	0.221	0.650	112799	1236	111563	phecode 426.3
Sleep disorders	OR=1.19 (0.9 to 1.59)	0.224	0.650	121474	2153	119321	phecode 327
Other disorders of bladder	OR=1.14 (0.92 to 1.41)	0.226	0.650	119881	3961	115920	phecode 596
Bacterial pneumonia	OR=0.85 (0.66 to 1.1)	0.228	0.650	119825	2655	117170	phecode 480.1
Gastric ulcer	OR=0.79 (0.54 to 1.16)	0.230	0.650	120324	1198	119126	phecode 531.2
Other disorders of circulatory system	OR=0.89 (0.74 to 1.08)	0.230	0.650	119994	5024	114970	phecode 459
Ventral hernia	OR=1.26 (0.86 to 1.83)	0.231	0.650	103355	1239	102116	phecode 550.5
Hemorrhoids	OR=1.1 (0.94 to 1.29)	0.233	0.650	117280	7463	109817	phecode 455
Rheumatic disease of the heart valves	OR=0.83 (0.61 to 1.13)	0.234	0.650	118594	1818	116776	phecode 394
Precordial pain	OR=0.78 (0.52 to 1.18)	0.237	0.650	114327	1022	113305	phecode 418.1
Urate	1.611 (-1.064 to 4.287)	0.238	0.650	115723	NA	NA	NA
Retinal detachments and defects	OR=0.81 (0.56 to 1.15)	0.239	0.650	118711	1371	117340	phecode 361
Retinal detachment with retinal defect	OR=0.81 (0.56 to 1.15)	0.239	0.650	118711	1371	117340	phecode 361.1
Other inflammatory spondylopathies	OR=1.35 (0.82 to 2.25)	0.241	0.650	121474	674	120800	phecode 715
Candidiasis	OR=1.35 (0.82 to 2.23)	0.242	0.650	121251	688	120563	phecode 112
Acute renal failure	OR=0.84 (0.63 to 1.13)	0.245	0.651	119127	2117	117010	phecode 585.1
Other hypertrophic and atrophic conditions of skin	OR=1.24 (0.86 to 1.79)	0.248	0.655	121436	1290	120146	phecode 701
Other disorders of synovium; tendon; and bursa	OR=1.16 (0.9 to 1.49)	0.256	0.667	114313	2749	111564	phecode 727
Redundant prepuce and phimosisorBXO	OR=1.2 (0.88 to 1.64)	0.258	0.667	110583	1756	108827	phecode 604.1
Septal DeviationsorTurbinate Hypertrophy	OR=1.19 (0.88 to 1.6)	0.259	0.667	116780	1967	114813	phecode 470
Cholecystitis without cholelithiasis	OR=0.75 (0.45 to 1.24)	0.263	0.667	118406	670	117736	phecode 574.3
Other disorders of the kidney and ureters	OR=1.24 (0.85 to 1.79)	0.263	0.667	118283	1273	117010	phecode 586
Fracture of vertebral column without mention of spinal cord injury	OR=0.75 (0.46 to 1.24)	0.265	0.667	115840	693	115147	phecode 805
Alcohol related disorders	OR=1.11 (0.93 to 1.32)	0.268	0.667	116319	5716	110603	phecode 317
Synovitis and tenosynovitis	OR=1.22 (0.86 to 1.73)	0.268	0.667	112986	1422	111564	phecode 727.1
Renal colic	OR=0.77 (0.49 to 1.22)	0.268	0.667	118845	817	118028	phecode 594.8
Effects radiation NOS	OR=1.23 (0.85 to 1.8)	0.270	0.667	120372	1252	119120	phecode 990
Acute pancreatitis	OR=1.34 (0.8 to 2.26)	0.272	0.667	121208	635	120573	phecode 577.1
Other diseases of blood and blood forming organs	OR=1.21 (0.86 to 1.7)	0.273	0.667	120442	1512	118930	phecode 289
Cholesterol	-0.022 (-0.061 to 0.017)	0.273	0.667	115873	NA	NA	NA
Benign neoplasm of other parts of digestive system	OR=1.21 (0.86 to 1.7)	0.274	0.667	121474	1503	119971	phecode 211
Diaphragmatic hernia	OR=1.09 (0.93 to 1.27)	0.285	0.688	109966	7850	102116	phecode 550.2
Cystatin C	0.019 (-0.016 to 0.054)	0.285	0.688	115851	NA	NA	NA
Appendiceal conditions	OR=0.82 (0.56 to 1.19)	0.295	0.701	121474	1202	120272	phecode 540
Dysphagia	OR=0.85 (0.63 to 1.15)	0.296	0.701	109708	1944	107764	phecode 532
Pericarditis	OR=0.74 (0.42 to 1.31)	0.297	0.701	120761	533	120228	phecode 420.2

Chronic liver disease and cirrhosis	OR=0.81 (0.55 to 1.2)	0.298	0.701	119428	1169	118259	phecode 571
Carditis	OR=0.78 (0.49 to 1.24)	0.299	0.701	121038	810	120228	phecode 420
Diseases of white blood cells	OR=1.22 (0.83 to 1.79)	0.301	0.703	120126	1196	118930	phecode 288
Nerve root and plexus disorders	OR=1.2 (0.85 to 1.71)	0.303	0.704	118754	1420	117334	phecode 353
Carbuncle and furuncle	OR=0.79 (0.5 to 1.24)	0.307	0.706	118186	849	117337	phecode 686.1
Gastritis and duodenitis	OR=1.08 (0.93 to 1.25)	0.308	0.706	120408	8613	111795	phecode 535
Anxiety disorder	OR=1.17 (0.86 to 1.6)	0.310	0.707	109873	1840	108033	phecode 300.1
Asthma	OR=1.09 (0.92 to 1.28)	0.311	0.707	118994	6990	112004	phecode 495
Cardiac conduction disorders	OR=0.87 (0.67 to 1.14)	0.314	0.709	114084	2521	111563	phecode 426
Atrioventricular AV block	OR=0.79 (0.51 to 1.24)	0.315	0.709	112442	879	111563	phecode 426.2
Staphylococcus infections	OR=0.82 (0.56 to 1.21)	0.322	0.721	117706	1157	116549	phecode 041.1
Diseases of the larynx and vocal cords	OR=1.26 (0.79 to 2.02)	0.324	0.721	115613	800	114813	phecode 473
Apolipoprotein B	0.004 (-0.004 to 0.013)	0.330	0.731	115123	NA	NA	NA
Anxiety disorders	OR=1.15 (0.86 to 1.55)	0.335	0.734	110081	2048	108033	phecode 300
Abal sputum	OR=1.18 (0.84 to 1.66)	0.337	0.734	121474	1530	119944	phecode 516
Bacterial enteritis	OR=1.26 (0.78 to 2.04)	0.337	0.734	119294	758	118536	phecode 008.5
Other chronic nonalcoholic liver disease	OR=0.8 (0.5 to 1.27)	0.338	0.734	119077	818	118259	phecode 571.5
Paroxysmal supraventricular tachycardia	OR=0.78 (0.47 to 1.3)	0.340	0.734	112226	663	111563	phecode 427.11
Diabetes mellitus	OR=0.91 (0.74 to 1.11)	0.352	0.758	121474	4205	117269	phecode 250
Other headache syndromes	OR=0.87 (0.65 to 1.17)	0.359	0.762	121094	2025	119069	phecode 339
Hemorrhage of rectum and anus	OR=1.11 (0.89 to 1.38)	0.359	0.762	117702	3778	113924	phecode 578.8
Appendicitis	OR=0.83 (0.57 to 1.23)	0.360	0.762	121430	1158	120272	phecode 540.1
Noninfectious gastroenteritis	OR=1.11 (0.89 to 1.38)	0.369	0.775	103265	3634	99631	phecode 558
Acquired foot deformities	OR=0.82 (0.54 to 1.26)	0.370	0.775	120200	988	119212	phecode 735
Acute appendicitis	OR=0.83 (0.54 to 1.25)	0.372	0.775	121271	999	120272	phecode 540.11
Other intestinal obstruction	OR=1.2 (0.8 to 1.81)	0.373	0.775	100682	1051	99631	phecode 560.4
Hypothyroidism NOS	OR=1.17 (0.83 to 1.65)	0.376	0.775	121042	1490	119552	phecode 244.4
Complication of internal orthopedic device	OR=0.86 (0.61 to 1.21)	0.376	0.775	115577	1491	114086	phecode 858
Spondylosis without myelopathy	OR=1.18 (0.82 to 1.7)	0.380	0.780	118685	1306	117379	phecode 721.1
Cerebral ischemia	OR=1.18 (0.81 to 1.73)	0.382	0.780	119599	1237	118362	phecode 433.3
Unspecified diffuse connective tissue disease	OR=0.8 (0.47 to 1.34)	0.389	0.791	112998	650	112348	phecode 709.7
Inflammatory and toxic neuropathy	OR=0.79 (0.46 to 1.36)	0.392	0.791	121201	587	120614	phecode 357
Melanomas of skin	OR=1.17 (0.81 to 1.68)	0.397	0.791	114460	1323	113137	phecode 172.11
Melanomas of skin; dx or hx	OR=1.17 (0.81 to 1.68)	0.397	0.791	114460	1323	113137	phecode 172.1
Hematuria	OR=1.08 (0.9 to 1.3)	0.399	0.791	118370	5506	112864	phecode 593
Other forms of chronic heart disease	OR=0.81 (0.51 to 1.31)	0.400	0.791	113871	768	113103	phecode 414
Osteoarthritis; localized	OR=1.09 (0.9 to 1.31)	0.400	0.791	118729	5054	113675	phecode 740.1
Aspartate aminotransferase	0.016 (-0.022 to 0.054)	0.407	0.796	115412	NA	NA	NA
Rash and other nonspecific skin eruption	OR=0.81 (0.49 to 1.34)	0.408	0.796	119935	680	119255	phecode 687.1
Swelling of limb	OR=0.86 (0.6 to 1.23)	0.409	0.796	121251	1388	119863	phecode 771.1
Hypovolemia	OR=0.84 (0.56 to 1.27)	0.410	0.796	119831	1054	118777	phecode 276.5
Pneumococcal pneumonia	OR=0.89 (0.68 to 1.17)	0.412	0.797	119534	2364	117170	phecode 480.11
Irritable Bowel Syndrome	OR=0.84 (0.56 to 1.27)	0.416	0.797	100652	1021	99631	phecode 564.1
Hemoptysis	OR=1.15 (0.82 to 1.63)	0.418	0.797	121404	1460	119944	phecode 516.1
Other non epithelial cancer of skin	OR=1.07 (0.91 to 1.26)	0.419	0.797	120234	7097	113137	phecode 172.2
Diffuse diseases of connective tissue	OR=0.82 (0.5 to 1.33)	0.421	0.797	113076	728	112348	phecode 709
Cancer of urinary organs incl kidney and bladder	OR=1.13 (0.84 to 1.53)	0.421	0.797	121474	1952	119522	phecode 189
Other diseases of respiratory system; not elsewhere classified	OR=1.1 (0.87 to 1.4)	0.427	0.805	121474	3148	118326	phecode 519
Reflux esophagitis	OR=1.09 (0.88 to 1.37)	0.429	0.805	111385	3621	107764	phecode 530.14
Cholelithiasis	OR=1.1 (0.86 to 1.41)	0.435	0.813	120642	2906	117736	phecode 574.1
Psoriasis and related disorders	OR=0.84 (0.54 to 1.31)	0.437	0.814	112103	872	111231	phecode 696
Electrolyte imbalance	OR=1.14 (0.81 to 1.61)	0.443	0.821	120263	1486	118777	phecode 276.1
Other diseases of the teeth and supporting structures	OR=1.2 (0.75 to 1.9)	0.447	0.824	119154	805	118349	phecode 525
Superficial cellulitis and abscess	OR=1.1 (0.85 to 1.42)	0.458	0.832	120102	2765	117337	phecode 681
Psoriasis	OR=0.84 (0.53 to 1.33)	0.459	0.832	112077	846	111231	phecode 696.4
Varicose veins	OR=1.1 (0.86 to 1.41)	0.460	0.832	112640	2823	109817	phecode 454
Obesity	OR=0.91 (0.72 to 1.16)	0.461	0.832	121452	3147	118305	phecode 278.1
Complications of transplants and reattached limbs	OR=1.07 (0.89 to 1.28)	0.462	0.832	119610	5524	114086	phecode 851
Dislocation	OR=1.18 (0.76 to 1.85)	0.463	0.832	115316	872	114444	phecode 830
Overweight; obesity and other hyperalimentation	OR=0.92 (0.72 to 1.16)	0.464	0.832	121474	3169	118305	phecode 278
Cystitis and urethritis	OR=1.2 (0.73 to 1.96)	0.469	0.833	113587	723	112864	phecode 592
Other symptoms of respiratory system	OR=0.92 (0.74 to 1.15)	0.469	0.833	121474	3728	117746	phecode 512
Lymphadenitis	OR=1.17 (0.76 to 1.79)	0.472	0.836	119884	954	118930	phecode 289.4

Purpura and other hemorrhagic conditions	OR=0.83 (0.51 to 1.37)	0.475	0.837	121188	700	120488	phecode 287
Fracture of ankle and foot	OR=0.83 (0.5 to 1.38)	0.477	0.837	115815	668	115147	phecode 801
Cardiac pacemakerordevice in situ	OR=1.19 (0.74 to 1.9)	0.478	0.837	112347	784	111563	phecode 426.9
Other upper respiratory disease	OR=1.13 (0.8 to 1.61)	0.487	0.847	116209	1396	114813	phecode 479
Diseases of hair and hair follicles	OR=0.89 (0.65 to 1.23)	0.488	0.847	121144	1768	119376	phecode 704
Abal movement	OR=0.87 (0.59 to 1.29)	0.497	0.852	121474	1170	120304	phecode 350
Colon cancer	OR=1.13 (0.79 to 1.62)	0.497	0.852	94548	1360	93188	phecode 153.2
Other abal blood chemistry	OR=1.08 (0.86 to 1.37)	0.499	0.852	121400	3222	118178	phecode 790.6
Other open wound of head and face	OR=1.13 (0.8 to 1.59)	0.499	0.852	118870	1450	117420	phecode 870.3
Triglycerides	0.013 (-0.025 to 0.052)	0.502	0.852	115746	NA	NA	NA
Respiratory failure; insufficiency; arrest	OR=0.87 (0.57 to 1.31)	0.503	0.852	117969	1031	116938	phecode 509
Psoriasis vulgaris	OR=0.84 (0.5 to 1.41)	0.507	0.857	111889	658	111231	phecode 696.41
Stricture and stenosis of esophagus	OR=0.86 (0.54 to 1.36)	0.512	0.862	108592	828	107764	phecode 530.3
Nonspecific findings on examination of blood	OR=1.08 (0.86 to 1.37)	0.516	0.864	121474	3296	118178	phecode 790
Cardiac pacemaker in situ	OR=1.17 (0.72 to 1.91)	0.518	0.864	112301	738	111563	phecode 426.91
Constipation	OR=1.09 (0.85 to 1.39)	0.520	0.864	102498	2867	99631	phecode 563
Lipoprotein A	-0.014 (-0.056 to 0.029)	0.523	0.864	92883	NA	NA	NA
Acquired toe deformities	OR=0.85 (0.51 to 1.42)	0.526	0.864	119868	656	119212	phecode 735.2
Paroxysmal tachycardia; unspecified	OR=0.87 (0.57 to 1.34)	0.530	0.864	112513	950	111563	phecode 427.1
Peripheral enthesopathies and allied syndromes	OR=1.07 (0.87 to 1.3)	0.530	0.864	115999	4435	111564	phecode 726
Hypertensive heart andoror renal disease	OR=1.23 (0.64 to 2.37)	0.532	0.864	100332	407	99925	phecode 401.2
Other diseases of respiratory system; NEC	OR=1.08 (0.85 to 1.38)	0.536	0.864	121283	2957	118326	phecode 519.8
Substance addiction and disorders	OR=1.08 (0.84 to 1.4)	0.537	0.864	113303	2700	110603	phecode 316
Other disorders of arteries and arterioles	OR=1.24 (0.63 to 2.44)	0.538	0.864	119177	378	118799	phecode 447
Disorders of penis	OR=1.09 (0.82 to 1.46)	0.540	0.864	110967	2140	108827	phecode 604
Other retinal disorders	OR=1.12 (0.78 to 1.61)	0.540	0.864	82043	1343	80700	phecode 362
Urinary incontinence	OR=1.15 (0.73 to 1.83)	0.542	0.864	113849	827	113022	phecode 599.4
Cataract	OR=1.06 (0.89 to 1.26)	0.544	0.864	121474	6090	115384	phecode 366
Vertiginous syndromes and other disorders of vestibular system	OR=0.91 (0.66 to 1.25)	0.545	0.864	121474	1721	119753	phecode 386
Respiratory abalities	OR=1.08 (0.84 to 1.38)	0.548	0.867	121474	2926	118548	phecode 513
Inflammation of eyelids	OR=1.15 (0.73 to 1.8)	0.555	0.869	119446	850	118596	phecode 371.3
Skin cancer	OR=1.05 (0.9 to 1.22)	0.557	0.869	121420	8283	113137	phecode 172
Unstable angina intermediate coronary syndrome	OR=1.15 (0.72 to 1.83)	0.560	0.869	113903	800	113103	phecode 411.1
Other disorders of stomach and duodenum	OR=1.09 (0.82 to 1.44)	0.564	0.869	114044	2249	111795	phecode 537
Poisoning by analgesics; antipyretics; and antirheumatics	OR=1.11 (0.78 to 1.56)	0.564	0.869	115962	1477	114485	phecode 965
Fracture of ribs	OR=0.87 (0.54 to 1.4)	0.566	0.869	115910	763	115147	phecode 807
Occlusion of cerebral arteries	OR=0.9 (0.63 to 1.29)	0.568	0.869	119720	1358	118362	phecode 433.2
Disorders of mineral metabolism	OR=0.88 (0.57 to 1.37)	0.568	0.869	121474	906	120568	phecode 275
Other abality of urination	OR=1.15 (0.71 to 1.86)	0.569	0.869	113770	748	113022	phecode 599.9
Shortness of breath	OR=0.91 (0.65 to 1.26)	0.573	0.871	119367	1621	117746	phecode 512.7
Ulcerative colitis	OR=0.89 (0.61 to 1.32)	0.576	0.871	100790	1159	99631	phecode 555.2
Open wounds of extremities	OR=1.08 (0.82 to 1.43)	0.577	0.871	119651	2231	117420	phecode 871
Nausea and vomiting	OR=0.93 (0.72 to 1.2)	0.581	0.874	121474	2760	118714	phecode 789
Non Hodgkins lymphoma	OR=1.13 (0.72 to 1.78)	0.584	0.876	120282	852	119430	phecode 202.2
Nonspecific chest pain	OR=1.04 (0.9 to 1.21)	0.588	0.879	121474	8169	113305	phecode 418
Thrombocytopenia	OR=0.86 (0.51 to 1.47)	0.591	0.880	121110	622	120488	phecode 287.3
Bronchiectasis	OR=1.16 (0.67 to 2)	0.593	0.880	112591	587	112004	phecode 496.3
Other chronic ischemic heart disease; unspecified	OR=1.05 (0.89 to 1.24)	0.596	0.881	119836	6733	113103	phecode 411.8
Bladder neck obstruction	OR=1.11 (0.75 to 1.65)	0.600	0.885	117044	1124	115920	phecode 596.1
Renal failure	OR=0.94 (0.74 to 1.19)	0.602	0.885	120249	3239	117010	phecode 585
Musculoskeletal symptoms referable to limbs	OR=0.92 (0.66 to 1.28)	0.613	0.889	121474	1611	119863	phecode 771
Disturbance of skin sensation	OR=1.13 (0.71 to 1.8)	0.619	0.889	120046	791	119255	phecode 687.4
Chronic renal failure CKD	OR=0.91 (0.64 to 1.31)	0.620	0.889	118379	1369	117010	phecode 585.3
Other disorders of eye	OR=0.91 (0.63 to 1.32)	0.622	0.889	120505	1257	119248	phecode 379
Edema	OR=0.87 (0.5 to 1.51)	0.622	0.889	121426	574	120852	phecode 782.3
Other specified gastritis	OR=0.93 (0.71 to 1.22)	0.622	0.889	114222	2427	111795	phecode 535.8
Respiratory insufficiency	OR=0.89 (0.54 to 1.44)	0.623	0.889	117678	740	116938	phecode 509.2
Pneumonia	OR=0.95 (0.77 to 1.17)	0.624	0.889	121366	4196	117170	phecode 480
Leukemia	OR=0.89 (0.57 to 1.4)	0.627	0.889	120293	863	119430	phecode 204
Diverticulosis	OR=0.96 (0.83 to 1.12)	0.627	0.889	108491	8860	99631	phecode 562.1
Diverticulosis and diverticulitis	OR=0.96 (0.83 to 1.12)	0.627	0.889	108491	8860	99631	phecode 562
Fracture of hand or wrist	OR=0.91 (0.64 to 1.31)	0.629	0.889	116468	1321	115147	phecode 804
Phlebitis and thrombophlebitis	OR=0.92 (0.66 to 1.29)	0.634	0.893	111359	1542	109817	phecode 451

Cancer of other lymphoid; histiocytic tissue	OR=1.1 (0.73 to 1.66)	0.640	0.897	120468	1038	119430	phecode 202
Glaucoma	OR=1.08 (0.78 to 1.49)	0.648	0.897	119013	1673	117340	phecode 365
Other disorders of eyelids	OR=1.08 (0.77 to 1.51)	0.649	0.897	120146	1550	118596	phecode 374
Symptoms affecting skin	OR=0.94 (0.71 to 1.24)	0.649	0.897	121474	2219	119255	phecode 687
Cough	OR=1.11 (0.71 to 1.74)	0.650	0.897	118598	852	117746	phecode 512.8
Hypothyroidism	OR=1.08 (0.77 to 1.51)	0.652	0.897	121125	1573	119552	phecode 244
Blood in stool	OR=1.11 (0.72 to 1.71)	0.653	0.897	114842	918	113924	phecode 578.2
Abal findings on examination of urine	OR=0.92 (0.65 to 1.31)	0.654	0.897	121474	1456	120018	phecode 598
Complications of cardiacorvascular device; implant; and graft	OR=1.12 (0.67 to 1.89)	0.657	0.897	114733	647	114086	phecode 854
Decreased white blood cell count	OR=1.1 (0.72 to 1.67)	0.661	0.897	119927	997	118930	phecode 288.1
Neutropenia	OR=1.1 (0.72 to 1.67)	0.661	0.897	119927	997	118930	phecode 288.11
Alcoholism	OR=1.05 (0.85 to 1.3)	0.663	0.897	114546	3943	110603	phecode 317.1
Hearing loss	OR=0.93 (0.67 to 1.29)	0.663	0.897	121457	1674	119783	phecode 389
Neoplasm of uncertain behavior	OR=0.91 (0.6 to 1.38)	0.668	0.898	109190	1024	108166	phecode 199
Secondary malignancy of respiratory organs	OR=0.91 (0.58 to 1.42)	0.668	0.898	109023	857	108166	phecode 198.2
Diabetic retinopathy	OR=1.22 (0.48 to 3.12)	0.671	0.899	80900	200	80700	phecode 250.7
Atherosclerosis	OR=1.16 (0.58 to 2.34)	0.676	0.903	119154	355	118799	phecode 440
Atopicorcontact dermatitis due to other or unspecified	OR=1.11 (0.68 to 1.81)	0.680	0.906	120170	722	119448	phecode 939
Secondary malignant neoplasm of liver	OR=1.09 (0.72 to 1.64)	0.686	0.907	109206	1040	108166	phecode 198.4
Malignant neoplasm of other and ill defined sites within the digestive organs and peritoneum	OR=1.06 (0.8 to 1.4)	0.687	0.907	94346	2278	92068	phecode 159
Aphakia and other disorders of lens	OR=1.12 (0.65 to 1.93)	0.687	0.907	119838	590	119248	phecode 379.3
Pain in joint	OR=0.94 (0.71 to 1.25)	0.690	0.908	121474	2192	119282	phecode 745
Hemorrhage or hematoma complicating a procedure	OR=0.93 (0.65 to 1.33)	0.694	0.908	115468	1382	114086	phecode 850
Bacterial infection NOS	OR=1.04 (0.84 to 1.3)	0.695	0.908	120302	3753	116549	phecode 041
Left bundle branch block	OR=1.12 (0.63 to 2)	0.696	0.908	112085	522	111563	phecode 426.32
Palpitations	OR=0.91 (0.57 to 1.45)	0.701	0.912	112363	800	111563	phecode 427.9
Chronic sinusitis	OR=1.09 (0.69 to 1.72)	0.703	0.912	115655	842	114813	phecode 475
Heart valve replaced	OR=1.12 (0.62 to 2.04)	0.708	0.915	117261	485	116776	phecode 395.6
Diseases of hard tissues of teeth	OR=1.08 (0.72 to 1.61)	0.717	0.923	119420	1071	118349	phecode 521
E coli	OR=1.09 (0.68 to 1.76)	0.720	0.923	117308	759	116549	phecode 041.4
Hematemesis	OR=1.1 (0.66 to 1.82)	0.721	0.923	114601	677	113924	phecode 578.1
Atrial fibrillation and flutter	OR=1.03 (0.86 to 1.24)	0.724	0.923	117384	5821	111563	phecode 427.2
Benign neoplasm of skin	OR=0.95 (0.72 to 1.26)	0.726	0.923	121288	2205	119083	phecode 216
Ischemic Heart Disease	OR=1.03 (0.88 to 1.2)	0.728	0.923	121203	8100	113103	phecode 411
Dental caries	OR=1.07 (0.72 to 1.61)	0.729	0.923	119406	1057	118349	phecode 521.1
Enthesopathy	OR=1.05 (0.81 to 1.35)	0.732	0.924	114326	2762	111564	phecode 726.1
Symptoms involving skin and other integumentary tissue	OR=0.91 (0.54 to 1.55)	0.733	0.924	121474	622	120852	phecode 782
Malaise and fatigue	OR=0.93 (0.6 to 1.43)	0.740	0.928	121474	929	120545	phecode 798
Phlebitis and thrombophlebitis of lower extremities	OR=0.94 (0.67 to 1.34)	0.743	0.928	111267	1450	109817	phecode 451.2
Cardiac and circulatory congenital anomalies	OR=0.93 (0.6 to 1.44)	0.744	0.928	121474	920	120554	phecode 747
Calculus of bile duct	OR=0.91 (0.53 to 1.57)	0.745	0.928	118333	597	117736	phecode 574.2
Chemotherapy	OR=1.03 (0.85 to 1.25)	0.748	0.929	113140	4974	108166	phecode 197
Cardiac arrest and ventricular fibrillation	OR=1.11 (0.6 to 2.05)	0.750	0.929	112021	458	111563	phecode 427.4
Ulcer of esophagus	OR=1.05 (0.78 to 1.41)	0.754	0.930	109763	1999	107764	phecode 530.12
Symptoms involving nervous and musculoskeletal systems	OR=0.96 (0.72 to 1.27)	0.759	0.930	121474	2159	119315	phecode 781
Cellulitis and abscess of armorhand	OR=0.96 (0.72 to 1.27)	0.762	0.930	119519	2182	117337	phecode 681.3
Cellulitis and abscess of foot; toe	OR=0.96 (0.72 to 1.27)	0.762	0.930	119519	2182	117337	phecode 681.6
Cellulitis and abscess of leg; except foot	OR=0.96 (0.72 to 1.27)	0.762	0.930	119519	2182	117337	phecode 681.5
Hydrocele	OR=1.07 (0.69 to 1.65)	0.772	0.933	109742	915	108827	phecode 603.1
Disorders of refraction and accommodation; blindness and low vision	OR=1.06 (0.7 to 1.63)	0.774	0.933	121474	969	120505	phecode 367
Renal failure NOS	OR=1.09 (0.6 to 2)	0.774	0.933	117487	477	117010	phecode 585.2
Angina pectoris	OR=1.03 (0.82 to 1.3)	0.775	0.933	116549	3446	113103	phecode 411.3
Superficial injury without mention of infection	OR=1.04 (0.77 to 1.41)	0.779	0.933	121281	1939	119342	phecode 915
Calculus of kidney	OR=1.05 (0.74 to 1.5)	0.781	0.933	119422	1394	118028	phecode 594.1
Cardiac congenital anomalies	OR=0.94 (0.59 to 1.48)	0.782	0.933	121394	840	120554	phecode 747.1
Colorectal cancer	OR=1.04 (0.78 to 1.4)	0.782	0.933	95259	2071	93188	phecode 153
Osteoarthrosis; localized; primary	OR=1.04 (0.8 to 1.35)	0.783	0.933	116206	2531	113675	phecode 740.11

Other disorders of biliary tract	OR=1.07 (0.63 to 1.83)	0.791	0.936	118352	616	117736	phecode	575.8
Iron deficiency anemias	OR=0.96 (0.72 to 1.29)	0.791	0.936	118719	2082	116637	phecode	280
Iron deficiency anemias; unspecified or not due to blood loss	OR=0.96 (0.72 to 1.29)	0.791	0.936	118719	2082	116637	phecode	280.1
Rheumatoid arthritis	OR=0.94 (0.61 to 1.47)	0.794	0.936	121474	889	120585	phecode	714.1
Cerebrovascular disease	OR=1.03 (0.8 to 1.34)	0.799	0.936	120987	2625	118362	phecode	433
LDL direct	-0.004 (-0.034 to 0.026)	0.801	0.936	115619	NA	NA	NA	
Malignant neoplasm of rectum; rectosigmoid junction; and anus	OR=1.05 (0.7 to 1.59)	0.802	0.936	94237	1049	93188	phecode	153.3
Aortic valve disease	OR=1.08 (0.6 to 1.94)	0.802	0.936	117282	506	116776	phecode	394.3
Disorder of skin and subcutaneous tissue NOS	OR=0.96 (0.71 to 1.31)	0.804	0.936	121474	1860	119614	phecode	689
Fever of unknown origin	OR=1.05 (0.73 to 1.5)	0.808	0.936	121474	1328	120146	phecode	783
Ganglion and cyst of synovium; tendon; and bursa	OR=1.06 (0.65 to 1.74)	0.810	0.936	112269	705	111564	phecode	727.4
Senile cataract	OR=1.03 (0.79 to 1.34)	0.811	0.936	117981	2597	115384	phecode	366.2
Complication due to other implant and internal device	OR=0.95 (0.64 to 1.42)	0.815	0.938	115194	1108	114086	phecode	859
Anal and rectal polyp	OR=1.03 (0.8 to 1.31)	0.824	0.946	113903	2962	110941	phecode	565.1
Pulmonary collapse; interstitial and compensatory emphysema	OR=0.95 (0.6 to 1.5)	0.828	0.947	117785	847	116938	phecode	508
Cancer within the respiratory system	OR=1.04 (0.7 to 1.55)	0.829	0.947	121462	1136	120326	phecode	165
Skull and face fracture and other intercranial injury	OR=1.04 (0.74 to 1.45)	0.837	0.950	121474	1519	119955	phecode	819
Viral infection	OR=1.05 (0.65 to 1.71)	0.841	0.950	120536	739	119797	phecode	079
Aortic aneurysm	OR=1.05 (0.62 to 1.79)	0.844	0.950	119425	626	118799	phecode	442.1
Congenital anomalies of great vessels	OR=0.95 (0.55 to 1.64)	0.844	0.950	121134	580	120554	phecode	747.13
Rheumatoid arthritis and other inflammatory polyarthropathies	OR=0.96 (0.63 to 1.46)	0.845	0.950	121474	1004	120470	phecode	714
Epistaxis or throat hemorrhage	OR=1.05 (0.66 to 1.67)	0.847	0.950	115611	798	114813	phecode	477
Abal results of function study of liver	OR=1.04 (0.71 to 1.52)	0.847	0.950	119455	1196	118259	phecode	573.7
Respiratory failure	OR=0.96 (0.61 to 1.51)	0.849	0.950	117787	849	116938	phecode	509.1
Varicose veins of lower extremity	OR=1.03 (0.79 to 1.34)	0.852	0.950	112359	2542	109817	phecode	454.1
Urea	0.004 (-0.044 to 0.053)	0.857	0.950	115789	NA	NA	NA	
Emphysema	OR=0.95 (0.58 to 1.58)	0.858	0.950	112693	689	112004	phecode	496.1
Cancer of bladder	OR=0.97 (0.66 to 1.42)	0.858	0.950	120718	1196	119522	phecode	189.2
Other mental disorder	OR=1.01 (0.89 to 1.16)	0.861	0.950	119162	11129	108033	phecode	306
Septicemia	OR=0.97 (0.7 to 1.34)	0.862	0.950	118221	1672	116549	phecode	038
Pain in limb	OR=1.03 (0.73 to 1.44)	0.868	0.951	121474	1542	119932	phecode	773
Streptococcus infection	OR=1.05 (0.61 to 1.81)	0.868	0.951	117127	578	116549	phecode	041.2
Neurological disorders	OR=1.03 (0.75 to 1.41)	0.869	0.951	120592	1788	118804	phecode	292
Transient cerebral ischemia	OR=0.96 (0.54 to 1.68)	0.877	0.957	118911	549	118362	phecode	433.31
Malignant neoplasm of bladder	OR=1.03 (0.68 to 1.56)	0.883	0.960	120554	1032	119522	phecode	189.21
Degeneration of intervertebral disc	OR=1.04 (0.65 to 1.66)	0.885	0.960	118166	787	117379	phecode	722.6
Hyposmolality andoror hyponatremia	OR=0.96 (0.58 to 1.6)	0.886	0.960	119460	683	118777	phecode	276.12
Ileostomy status	OR=1.03 (0.62 to 1.73)	0.897	0.963	100290	659	99631	phecode	559
Degenerative disease of the spinal cord	OR=1.03 (0.64 to 1.68)	0.897	0.963	115385	740	114645	phecode	334
Cardiac dysrhythmias	OR=0.99 (0.85 to 1.15)	0.897	0.963	120253	8690	111563	phecode	427
Chronic ulcer of skin	OR=0.97 (0.59 to 1.58)	0.897	0.963	121474	733	120741	phecode	707
Localized superficial swelling; mass; or lump	OR=1.03 (0.61 to 1.75)	0.905	0.963	119881	626	119255	phecode	687.2
Inflammation of the eye	OR=1.03 (0.68 to 1.55)	0.905	0.963	119626	1030	118596	phecode	371
Diseases of sebaceous glands	OR=0.98 (0.76 to 1.27)	0.906	0.963	121235	2681	118554	phecode	706
Cholelithiasis and cholecystitis	OR=1.01 (0.8 to 1.28)	0.908	0.963	120996	3260	117736	phecode	574
Dizziness and giddiness Light headedness and vertigo	OR=0.98 (0.68 to 1.41)	0.908	0.963	121043	1290	119753	phecode	386.9
Nonrheumatic aortic valve disorders	OR=0.98 (0.63 to 1.51)	0.916	0.968	117688	912	116776	phecode	395.2
Other derangement of joint	OR=1.02 (0.65 to 1.61)	0.919	0.968	120175	850	119325	phecode	742.9
Other acquired deformities of limbs	OR=1.03 (0.61 to 1.73)	0.922	0.968	119843	631	119212	phecode	736
Pleurisy; pleural effusion	OR=0.99 (0.76 to 1.29)	0.922	0.968	119435	2497	116938	phecode	507
Hereditary retinal dystrophies	OR=1.04 (0.43 to 2.5)	0.925	0.968	80929	229	80700	phecode	362.7
Disorders of fluid; electrolyte; and acid base balance	OR=1.01 (0.78 to 1.31)	0.926	0.968	121474	2697	118777	phecode	276
Sebaceous cyst	OR=0.99 (0.76 to 1.28)	0.934	0.974	121216	2662	118554	phecode	706.2
Coronary atherosclerosis	OR=1.01 (0.84 to 1.21)	0.942	0.980	118697	5594	113103	phecode	411.4
Cystitis	OR=1.01 (0.61 to 1.69)	0.955	0.986	113538	674	112864	phecode	592.1
Sepsis	OR=0.99 (0.68 to 1.44)	0.959	0.986	121474	1230	120244	phecode	994.2
Sepsis and SIRS	OR=0.99 (0.68 to 1.44)	0.959	0.986	121474	1230	120244	phecode	994

Other disorders of testis	OR=0.99 (0.7 to 1.4)	0.959	0.986	110301	1474	108827	phecode 603
Symptoms and disorders of the joints	OR=1.01 (0.69 to 1.47)	0.961	0.986	120544	1219	119325	phecode 741
Benign neoplasm of unspecified sites	OR=0.99 (0.6 to 1.62)	0.961	0.986	121474	703	120771	phecode 229
Abdominal aortic aneurysm	OR=0.98 (0.49 to 1.99)	0.965	0.988	119155	356	118799	phecode 442.11
Inflammatory bowel disease and other gastroenteritis and colitis	OR=0.99 (0.71 to 1.39)	0.971	0.991	101189	1558	99631	phecode 555
Acute upper respiratory infections of multiple or unspecified sites	OR=1.01 (0.6 to 1.71)	0.976	0.993	121428	625	120803	phecode 465
Calculus of ureter	OR=1.01 (0.68 to 1.49)	0.978	0.993	119165	1137	118028	phecode 594.3
Other aneurysm	OR=1.01 (0.62 to 1.62)	0.979	0.993	119566	767	118799	phecode 442
Other disorders of male genital organs	OR=1 (0.73 to 1.35)	0.984	0.995	110693	1866	108827	phecode 608
Other disorders of liver	OR=1 (0.77 to 1.29)	0.990	0.995	121011	2752	118259	phecode 573
Syncope and collapse	OR=1 (0.79 to 1.27)	0.991	0.995	121474	3064	118410	phecode 788
Derangement of joint; non traumatic	OR=1 (0.66 to 1.52)	0.992	0.995	120336	1011	119325	phecode 742
Urinary calculus	OR=1 (0.78 to 1.28)	0.992	0.995	120952	2924	118028	phecode 594
Cancer of bronchus; lung	OR=1 (0.65 to 1.54)	0.995	0.995	121260	934	120326	phecode 165.1
Other disorders of male genital organs	OR=1 (0.73 to 1.35)	0.984	0.995	110693	1866	108827	phecode 608
Other disorders of liver	OR=1 (0.77 to 1.29)	0.990	0.995	121011	2752	118259	phecode 573
Syncope and collapse	OR=1 (0.79 to 1.27)	0.991	0.995	121474	3064	118410	phecode 788
Derangement of joint; non traumatic	OR=1 (0.66 to 1.52)	0.992	0.995	120336	1011	119325	phecode 742
Urinary calculus	OR=1 (0.78 to 1.28)	0.992	0.995	120952	2924	118028	phecode 594
Cancer of bronchus; lung	OR=1 (0.65 to 1.54)	0.995	0.995	121260	934	120326	phecode 165.1

Supplementary File 1 - Table 8. Independent genetic variants associated with total testosterone at genome-wide significance (p<5x10-8) and not associated with sex hormone-binding globulin in 175,421 males from UK Biobank

chr	rsid	position (hg19)	beta	se	effect allele	other allele	eaf	pval	gene	sample size
1	rs12035604	66056052	-0.077	0.014	T	C	0.743	4.00E-08	LEPR	175421
1	rs537115525	155572575	-0.148	0.027	T	C	0.945	3.40E-08	ASH1L-AS1;MSTO1	175421
1	rs34702488	163256609	-0.089	0.016	T	A	0.825	3.20E-08	RGS5	175421
1	rs537444	172134469	0.071	0.012	C	G	0.579	8.70E-09	DNM3	175421
1	rs35737316	204161534	-0.137	0.014	C	T	0.754	5.30E-22	KISS1	175421
2	rs72862643	31533151	-0.250	0.025	T	C	0.937	2.70E-23	EHD3;XDH	175421
2	rs11124268	31951731	-0.100	0.016	C	A	0.811	1.30E-10	SRD5A2;LINC01946	175421
2	rs11901448	32096569	-0.106	0.013	C	T	0.444	8.10E-16	MEMO1	175421
2	rs112564689	33377229	-0.309	0.038	C	G	0.972	3.20E-16	LTBP1	175421
2	rs10192634	180500950	-0.080	0.014	C	T	0.730	6.80E-09	ZNF385B	175421
2	rs12468274	234627914	0.193	0.028	T	C	0.950	6.60E-12	UGT1A4	175421
3	rs6785560	88269441	0.103	0.017	G	A	0.155	8.00E-10	C3orf38;EPHA3	175421
3	rs7626226	138234576	0.085	0.012	C	T	0.532	3.10E-12	CEP70	175421
3	rs79223973	151992162	-0.121	0.020	A	C	0.900	3.00E-09	MBNL1	175421
4	rs7679843	22028079	-0.155	0.021	C	G	0.905	3.90E-13	KCNIP4;LOC100505912	175421
4	rs115872353	103983087	0.277	0.044	T	G	0.980	2.00E-10	SLC9B2	175421
4	rs17289915	104491078	0.485	0.055	C	G	0.987	7.00E-19	LINC02428;TACR3	175421
4	rs565931739	104590005	0.155	0.017	C	CA	0.836	2.40E-20	TACR3	175421
5	rs1979835	135689839	0.110	0.018	G	A	0.862	5.10E-10	TRPC7	175421
6	rs7758796	100105028	-0.088	0.012	G	A	0.547	9.50E-13	PRDM13;MCHR2	175421
6	rs6902789	105358192	0.116	0.013	G	A	0.630	3.60E-20	HACE1;LIN28B-AS1	175421
6	rs7755185	152339615	0.072	0.013	A	G	0.689	3.70E-08	ESR1	175421
7	rs9986829	15019259	-0.146	0.012	G	A	0.494	7.00E-33	DGKB;AGMO	175421
7	rs10279715	40870935	0.073	0.012	A	G	0.539	2.00E-09	SUGCT	175421
7	rs10229569	99183704	-0.093	0.017	G	A	0.840	2.10E-08	ZNF655;TMEM225B	175421
8	rs7835492	21089517	-0.096	0.017	T	C	0.853	2.30E-08	LINC02153;LOC101929172	175421
8	rs35006173	77911193	0.084	0.013	A	AT	0.611	3.60E-11	PEX2	175421
9	rs2090409	108967088	0.095	0.013	C	A	0.684	3.90E-13	TMEM38B;MIR8081	175421
10	rs571568325	67293894	-0.174	0.012	G	GA	0.595	8.40E-45	LOC101928887;LINC01515	175421
11	rs76160029	28884695	0.248	0.037	G	A	0.971	1.20E-11	MIR8068;LINC01616	175421
11	rs7478970	29118542	-0.108	0.015	G	A	0.782	2.30E-13	MIR8068;LINC01616	175421
11	rs2933170	29135383	0.079	0.014	A	G	0.729	7.40E-09	MIR8068;LINC01616	175421
11	rs35934227	29458062	0.280	0.045	C	T	0.981	3.70E-10	MIR8068;LINC01616	175421
11	rs12796488	94131557	0.160	0.016	C	A	0.823	5.90E-24	GPR83	175421
11	rs34669210	122772285	-0.109	0.012	T	TA	0.433	1.40E-18	C11orf63	175421
11	rs7107502	122830214	0.071	0.012	G	A	0.431	7.70E-09	C11orf63	175421
11	rs634554	125070392	0.076	0.013	C	A	0.308	7.90E-09	PKNOX2	175421
12	rs7959682	2944356	-0.085	0.015	T	A	0.529	1.40E-08	NRIP2	175421
12	rs11835185	3057794	-0.135	0.015	C	T	0.801	1.00E-18	TULP3;TEAD4	175421
12	rs12318430	3115206	-0.347	0.031	C	A	0.961	1.80E-28	TEAD4	175421
12	rs12810788	116196322	0.100	0.015	G	A	0.201	1.10E-10	TBX3;MED13L	175421

12	rs6486542	130952209	0.074	0.012	C	T	0.571	2.30E-09	RIMBP2	175421
13	rs12429920	112720354	-0.070	0.012	G	A	0.524	7.10E-09	LINC00403	175421
14	rs1812755	90007637	0.094	0.015	T	C	0.800	9.00E-10	FOXN3	175421
15	rs28892005	51519945	0.101	0.013	A	AAAG	0.350	1.90E-15	MIR4713HG	175421
15	rs17238845	56936160	-0.068	0.012	A	G	0.406	4.50E-08	ZNF280D	175421
16	rs2764772	20060653	-0.122	0.013	T	A	0.666	3.50E-21	GPR139	175421
16	rs2077412	28621311	0.096	0.014	T	C	0.301	1.10E-11	SULT1A1	175421
17	rs2696641	43651550	-0.076	0.014	G	C	0.475	2.60E-08	LRRC37A4P;MAPK8IP1P2	175421
18	rs2684837	44749884	0.085	0.012	A	C	0.559	3.80E-12	SKOR2	175421
19	rs142936065	48449301	-0.073	0.013	G	GTT	0.483	2.40E-08	SNAR-A12;SNAR-C3	175421
22	rs11703376	49678713	-0.122	0.014	C	T	0.731	1.00E-18	LINC01310;NONE	175421

Effect size (beta and standard error) in nmol/L of calculated free testosterone per copy of effect allele
Abbreviations: Chr, chromosome; rsID, rs identifier; SE, standard error; eaf, effect allele frequency; pval, p-value

Supplementary File 1 - Table 9. All Mendelian randomization analyses of total testosterone on 22 a priori outcomes

	IVW		Egger			MR-RAPS		MR-PRESSO				
Outcome	Effect per 1 nmol/L increase in total testosterone (95% CI)	P-value	P-value for Egger intercept	Effect per 1 nmol/L increase in total testosterone (95% CI)	P-value	Effect per 1 nmol/L increase in total testosterone (95% CI)	P-value	Global Test P-value	Effect per 1 nmol/L increase in total testosterone (95% CI)	P-value	Distortion Test P-value	Sample Size (Cases/Controls)
Haematocrit percentage	**0.256 % (0.193 to 0.319)**	**1.87E-15**	0.11	0.127 % (-0.04 to 0.294)	0.14	**0.232 % (0.158 to 0.306)**	**9.76E-10**	≤0.0001	**0.2612 % (0.211 to 0.312)**	**3.17E-13**	0.81	152872
Body fat-free percentage	**0.408 % (0.3 to 0.516)**	**1.66E-13**	0.56	0.327 % (0.034 to 0.62)	0.03	**0.447 % (0.322 to 0.572)**	**2.19E-12**	≤0.0001	**0.4155 % (0.313 to 0.518)**	**2.58E-10**	0.8856	154254
Body fat percentage	**-0.395 % (-0.503 to -0.287)**	**7.59E-13**	0.54	-0.31 % (-0.602 to -0.018)	0.04	**-0.426 % (-0.551 to -0.301)**	**2.36E-11**	≤0.0001	**-0.3764 % (-0.473 to -0.28)**	**8.32E-10**	0.7025	153772
Benign prostatic hyperplasia	**OR = 1.09 (1.04 to 1.14)**	**3.57E-04**	0.49	OR = 1.04 (0.92 to 1.18)	0.51	**OR = 1.10 (1.05 to 1.15)**	**1.26E-04**	0.1086	NA	NA	NA	10894/146316
Heel bone mineral density T-score	**0.0666 SD (0.028 to 0.105)**	**7.11E-04**	0.23	0.126 SD (0.023 to 0.229)	0.02	**0.08 SD (0.051 to 0.109)**	**6.68E-08**	≤0.0001	**0.0943 SD (0.071 to 0.118)**	**3.76E-10**	0.1483	90676
Hemoglobin A1c	**-0.17 mmol/L (-0.274 to -0.066)**	**1.27E-03**	0.39	-0.0555 mmol/L (-0.334 to 0.223)	0.70	-0.124 mmol/L (-0.215 to -0.033)	7.85E-03	≤0.0001	-0.1063 mmol/L (-0.19 to -0.023)	0.01612	0.036	149828
Glucose	-0.0256 mmol/L (-0.045 to -0.006)	8.99E-03	0.85	-0.0208 mmol/L (-0.073 to 0.031)	0.44	-0.015 mmol/L (-0.03 to 0)	0.05	≤0.0001	-0.01373 mmol/L (-0.027 to 0)	0.04776	0.0244	138307
Depression	OR = 1.08 (1.02 to 1.15)	0.014	0.82	OR = 1.06 (0.9 to 1.26)	0.48	OR = 1.08 (1.01 to 1.16)	0.02	0.8715	NA	NA	NA	4725/152485
Accelerometer-based physical activity	0.266 milligravity (0.042 to 0.49)	0.020	0.50	0.0736 milligravity (-0.532 to 0.679)	0.81	0.276 milligravity (0.053 to 0.499)	0.02	0.1331	NA	NA	NA	30439
Prostate cancer	OR = 1.07 (1.01 to 1.13)	0.029	0.91	OR = 1.06 (0.9 to 1.24)	0.49	OR = 1.06 (1 to 1.13)	0.05	0.0213	NA	NA	NA	7586/149624
Androgenic alopecia	OR = 1.06 (1 to 1.13)	0.051	0.68	OR = 1.03 (0.87 to 1.21)	0.74	OR = 1.08 (0.99 to 1.17)	0.07	≤0.0001	**OR = 1.12 (1.08 to 1.15)**	**1.46E-08**	0.2131	70283/85756
Type 2 diabetes	OR = 0.97 (0.91 to 1.02)	0.200	0.41	OR = 1.02 (0.88 to 1.18)	0.77	OR = 0.98 (0.92 to 1.03)	0.39	0.0011	OR = 0.98 (0.94 to 1.03)	0.5362	0.2032	11079/146131
Myocardial infarction	OR = 1.02 (0.98 to 1.08)	0.320	0.97	OR = 1.02 (0.9 to 1.16)	0.74	OR = 1.03 (0.98 to 1.09)	0.23	0.2035	NA	NA	NA	9398/147812
Venous thromboembolism	OR = 1.02 (0.95 to 1.09)	0.604	0.08	OR = 0.87 (0.73 to 1.05)	0.15	OR = 1.02 (0.95 to 1.09)	0.65	0.6821	NA	NA	NA	4127/153083
All fracture	OR = 0.99 (0.94 to 1.04)	0.611	0.54	OR = 0.95 (0.82 to 1.09)	0.45	OR = 0.99 (0.93 to 1.05)	0.75	0.0427	NA	NA	NA	9133/148077
Handgrip strength	0.0306 kg (-0.105 to 0.166)	0.657	0.61	-0.0581 kg (-0.424 to 0.307)	0.76	0.029 kg (-0.102 to 0.16)	0.67	≤0.0001	-0.009878 kg (-0.132 to 0.112)	0.8744	0.0892	156400
Ischemic stroke	OR = 1.02 (0.93 to 1.12)	0.663	0.37	OR = 0.92 (0.72 to 1.18)	0.50	OR = 1.01 (0.92 to 1.12)	0.77	0.9149	NA	NA	NA	2122/155088
Diastolic blood pressure	0.0187 mmHg (-0.114 to 0.151)	0.782	0.87	0.0475 mmHg (-0.312 to 0.407)	0.80	0.023 mmHg (-0.114 to 0.159)	0.75	0.0164	0.03997 mmHg (-0.084 to 0.164)	0.5304	0.8204	148384
All stroke	OR = 1.01 (0.95 to 1.07)	0.812	0.69	OR = 0.98 (0.82 to 1.16)	0.78	OR = 1.01 (0.95 to 1.08)	0.68	0.7977	NA	NA	NA	4569/152641
Heart failure	OR = 1 (0.93 to 1.07)	0.894	0.49	OR = 0.94 (0.78 to 1.13)	0.49	OR = 0.99 (0.91 to 1.06)	0.71	2.29E-01	NA	NA	NA	4288/152922
All dementia	OR = 1.01 (0.86 to 1.17)	0.937	0.43	OR = 0.86 (0.57 to 1.3)	0.48	OR = 0.97 (0.84 to 1.13)	0.69	0.085	NA	NA	NA	1003/156207
Systolic blood pressure	-0.0105 mmHg (-0.271 to 0.25)	0.937	0.10	0.534 mmHG (-0.152 to 1.22)	0.13	0.087 mmHg (-0.173 to 0.347)	0.51	≤0.0001	0.06339 mmHg (-0.164 to 0.29)	0.5866	0.2225	148383

Bolded rows are significant adjusting for multiple hypothesis testing using Bonferroni correction

Supplementary File 1 - Table 10. Associations of genetically predicted total testosterone for 439 health outcomes across the human phenome.

Trait	Effect per 1 nmol/L increase total testosterone (95% CI)	P-value	FDR-adjusted p-value	Sample Size	Number of Cases	Number of Controls	Phecode	Category
Total bilirubin	**0.034 (0.024 to 0.045)**	**1.26E-10**	**5.55E-08**	**149300**	**NA**	**NA**	**NA**	**biomarker**
C-reactive protein	**-0.033 (-0.043 to -0.023)**	**5.42E-10**	**1.19E-07**	**149547**	**NA**	**NA**	**NA**	**biomarker**
Creatinine	**0.033 (0.022 to 0.043)**	**8.81E-10**	**1.29E-07**	**149849**	**NA**	**NA**	**NA**	**biomarker**
Phosphate	**-0.005 (-0.006 to -0.003)**	**2.11E-07**	**2.32E-05**	**138207**	**NA**	**NA**	**NA**	**biomarker**
Inguinal hernia	**OR=1.1 (1.06 to 1.14)**	**4.56E-07**	**4.01E-05**	**144145**	**13906**	**130239**	**phecode 550.1**	**digestive**
Direct bilirubin	**0.026 (0.015 to 0.037)**	**2.59E-06**	**1.90E-04**	**139800**	**NA**	**NA**	**NA**	**biomarker**
Alkaline phosphatase	**-0.021 (-0.032 to -0.011)**	**6.57E-05**	**3.52E-03**	**149937**	**NA**	**NA**	**NA**	**biomarker**
Spinal stenosis	**OR=1.21 (1.1 to 1.33)**	**7.07E-05**	**3.52E-03**	**152836**	**1917**	**150919**	**phecode 720**	**musculoskeletal**
Abdominal hernia	**OR=1.06 (1.03 to 1.09)**	**7.20E-05**	**3.52E-03**	**157211**	**26972**	**130239**	**phecode 550**	**digestive**
Umbilical hernia	**OR=1.17 (1.08 to 1.27)**	**8.41E-05**	**3.57E-03**	**132863**	**2624**	**130239**	**phecode 550.4**	**digestive**
Degenerative disease of the spinal cord	**OR=1.28 (1.13 to 1.46)**	**8.91E-05**	**3.57E-03**	**148165**	**1062**	**147103**	**phecode 334**	**neurological**
Other symptoms or disorders or the urinary system	OR=1.07 (1.03 to 1.12)	2.93E-04	0.010	157211	12410	144801	phecode 599	genitourinary
Essential hypertension	OR=1.05 (1.02 to 1.07)	3.00E-04	0.010	156766	40809	115957	phecode 401.1	circulatory system
Hypertension	OR=1.05 (1.02 to 1.07)	3.33E-04	0.010	156917	40960	115957	phecode 401	circulatory system
IGF1	0.1 (0.045 to 0.156)	4.24E-04	0.012	149151	NA	NA	NA	biomarker
Hypercholesterolemia	OR=1.06 (1.03 to 1.09)	4.28E-04	0.012	155658	19758	135900	phecode 272.11	endocrine metabolic
Other disorders of peritoneum	OR=1.22 (1.09 to 1.36)	6.72E-04	0.017	143289	1323	141966	phecode 568	digestive
Symptoms involving digestive system	OR=1.1 (1.04 to 1.16)	7.78E-04	0.019	131291	5729	125562	phecode 561	digestive
Hyperlipidemia	OR=1.05 (1.02 to 1.09)	9.42E-04	0.022	157148	21248	135900	phecode 272.1	endocrine metabolic
Disorders of lipoid metabolism	OR=1.05 (1.02 to 1.09)	1.03E-03	0.022	157211	21311	135900	phecode 272	endocrine metabolic
Spondylosis and allied disorders	OR=1.13 (1.05 to 1.21)	1.04E-03	0.022	154215	3296	150919	phecode 721	musculoskeletal
Prostatitis	OR=1.2 (1.07 to 1.34)	1.33E-03	0.027	43064	1466	41598	phecode 601.1	genitourinary
Candidiasis	OR=1.23 (1.08 to 1.39)	1.49E-03	0.028	156902	1064	155838	phecode 112	infectious diseases
Internal derangement of knee	OR=1.08 (1.03 to 1.13)	1.88E-03	0.034	156183	7941	148242	phecode 835	injuries and poisonings
Spondylosis without myelopathy	OR=1.15 (1.05 to 1.26)	2.47E-03	0.043	153008	2089	150919	phecode 721.1	musculoskeletal
Peritoneal adhesions postoperative postinfection	OR=1.19 (1.06 to 1.34)	2.97E-03	0.050	143200	1234	141966	phecode 568.1	digestive
Anal and rectal conditions	OR=1.07 (1.02 to 1.13)	3.38E-03	0.054	149752	7786	141966	phecode 565	digestive
Other symptoms involving abdomen and pelvis	OR=1.17 (1.05 to 1.3)	3.49E-03	0.054	147838	1496	146342	phecode 579	digestive
Urinary obstruction	OR=1.16 (1.05 to 1.29)	3.59E-03	0.054	146437	1636	144801	phecode 599.1	genitourinary
Erythematous conditions	OR=1.22 (1.07 to 1.41)	3.95E-03	0.057	155296	881	154415	phecode 695	dermatologic
Unspecified monoarthritis	OR=1.06 (1.02 to 1.11)	3.98E-03	0.057	147634	9832	137802	phecode 716.2	musculoskeletal
Inflammatory diseases of prostate	OR=1.13 (1.04 to 1.22)	4.53E-03	0.062	44293	2695	41598	phecode 601	genitourinary
Urinary tract infection	OR=1.09 (1.03 to 1.15)	4.92E-03	0.066	149462	5313	144149	phecode 591	genitourinary
Septal DeviationsorTurbinate Hypertrophy	OR=1.12 (1.03 to 1.22)	5.87E-03	0.076	150593	2536	148057	phecode 470	respiratory
Nonspecific findings on examination of blood	OR=1.08 (1.02 to 1.15)	6.36E-03	0.080	157211	5239	151972	phecode 790	symptoms
Angina pectoris	OR=1.06 (1.02 to 1.1)	7.12E-03	0.085	147250	11008	136242	phecode 411.3	circulatory system
Other disorders of intestine	OR=1.06 (1.02 to 1.11)	7.19E-03	0.085	150634	8668	141966	phecode 569	digestive
Intestinal infection	OR=1.09 (1.02 to 1.16)	7.47E-03	0.086	157211	4483	152728	phecode 008	infectious diseases
Other abnormal blood chemistry	OR=1.08 (1.02 to 1.15)	7.59E-03	0.086	157098	5126	151972	phecode 790.6	symptoms
E coli	OR=1.17 (1.04 to 1.32)	7.85E-03	0.086	150693	1204	149489	phecode 041.4	infectious diseases
Respiratory abnormalities	OR=1.08 (1.02 to 1.15)	0.011	0.113	157211	4784	152427	phecode 513	respiratory
SHBG	0.014 (0.003 to 0.024)	0.011	0.113	137408	NA	NA	NA	biomarker
Sleep apnea	OR=1.1 (1.02 to 1.18)	0.012	0.124	156734	3124	153610	phecode 327.3	neurological
Retention of urine	OR=1.07 (1.02 to 1.14)	0.013	0.126	150344	5543	144801	phecode 599.2	genitourinary
Peripheral vascular disease; unspecified	OR=1.13 (1.03 to 1.24)	0.013	0.126	153464	1889	151575	phecode 443.9	circulatory system
Other specified peripheral vascular diseases	OR=1.12 (1.02 to 1.22)	0.013	0.126	153785	2210	151575	phecode 443.8	circulatory system
Hematuria	OR=1.06 (1.01 to 1.11)	0.014	0.128	152438	8289	144149	phecode 593	genitourinary
Chronic airway obstruction	OR=1.07 (1.01 to 1.13)	0.014	0.129	148957	6018	142939	phecode 496	respiratory
Hydronephrosis	OR=1.15 (1.03 to 1.3)	0.015	0.137	153509	1254	152255	phecode 595	genitourinary
GERD	OR=1.05 (1.01 to 1.09)	0.018	0.154	147762	10877	136885	phecode 530.11	digestive
Other disorders of male genital organs	OR=1.1 (1.02 to 1.2)	0.018	0.156	141295	2537	138758	phecode 608	genitourinary

Atopicorcontact dermatitis due to other or unspecified	OR=1.17 (1.03 to 1.32)	0.019	0.156	155217	1029	154188	phecode 939	dermatologic
Cancer; suspected or other	OR=1.05 (1.01 to 1.08)	0.019	0.156	151649	13842	137807	phecode 195	neoplasms
Bacterial infection NOS	OR=1.07 (1.01 to 1.13)	0.020	0.158	155365	5876	149489	phecode 041	infectious diseases
Other peripheral nerve disorders	OR=1.08 (1.01 to 1.16)	0.020	0.158	154804	3798	151006	phecode 351	neurological
Personal history of diseases of digestive system	OR=1.06 (1.01 to 1.11)	0.020	0.158	133795	8233	125562	phecode 564.9	digestive
Other diseases of blood and blood forming organs	OR=1.11 (1.02 to 1.21)	0.020	0.158	155768	2180	153588	phecode 289	hematopoietic
Chronic bronchitis	OR=1.07 (1.01 to 1.14)	0.022	0.170	148033	5094	142939	phecode 496.2	respiratory
Frequency of urination and polyuria	OR=1.11 (1.01 to 1.22)	0.024	0.176	146763	1962	144801	phecode 599.5	genitourinary
Other chronic ischemic heart disease; unspecified	OR=1.04 (1 to 1.07)	0.026	0.191	154252	18010	136242	phecode 411.8	circulatory system
Diseases of esophagus	OR=1.04 (1 to 1.07)	0.027	0.194	153704	16819	136885	phecode 530	digestive
Congenital anomalies of great vessels	OR=1.14 (1.01 to 1.28)	0.028	0.196	156725	1267	155458	phecode 747.13	congenital anomalies
Other symptoms of respiratory system	OR=1.06 (1.01 to 1.12)	0.029	0.200	157211	6102	151109	phecode 512	respiratory
Hemiplegia	OR=0.86 (0.75 to 0.99)	0.030	0.204	148019	916	147103	phecode 342	neurological
Other abnormality of urination	OR=1.14 (1.01 to 1.29)	0.030	0.204	145926	1125	144801	phecode 599.9	genitourinary
Complications of transplants and reattached limbs	OR=1.05 (1 to 1.1)	0.031	0.207	154168	8858	145310	phecode 851	injuries and poisonings
Other retinal disorders	OR=1.09 (1.01 to 1.18)	0.034	0.221	96487	2665	93822	phecode 362	sense organs
Gamma glutamyltransferase	0.011 (0.001 to 0.022)	0.034	0.221	149840	NA	NA	NA	biomarker
Malignant neoplasm; other	OR=1.04 (1 to 1.08)	0.035	0.221	151312	13505	137807	phecode 195.1	neoplasms
Coronary atherosclerosis	OR=1.04 (1 to 1.08)	0.036	0.223	151051	14809	136242	phecode 411.4	circulatory system
Benign neoplasm of colon	OR=1.04 (1 to 1.09)	0.037	0.223	128626	11711	116915	phecode 208	neoplasms
Obstructive chronic bronchitis	OR=1.07 (1 to 1.13)	0.037	0.223	147942	5003	142939	phecode 496.21	respiratory
Right bundle branch block	OR=0.88 (0.79 to 0.99)	0.037	0.223	141247	1236	140011	phecode 426.31	circulatory system
Other disorders of prostate	OR=1.14 (1.01 to 1.28)	0.038	0.223	42816	1218	41598	phecode 602	genitourinary
Ischemic Heart Disease	OR=1.03 (1 to 1.07)	0.038	0.223	156847	20605	136242	phecode 411	circulatory system
Peripheral vascular disease	OR=1.09 (1 to 1.19)	0.039	0.225	153963	2388	151575	phecode 443	circulatory system
Lipoma	OR=1.08 (1 to 1.16)	0.039	0.225	156819	3221	153598	phecode 214	neoplasms
Other inflammatory spondylopathies	OR=1.14 (1.01 to 1.3)	0.040	0.228	157211	1022	156189	phecode 715	musculoskeletal
Shortness of breath	OR=1.08 (1 to 1.17)	0.041	0.230	153980	2871	151109	phecode 512.7	respiratory
Hypothyroidism	OR=1.08 (1 to 1.17)	0.044	0.244	156694	2715	153979	phecode 244	endocrine metabolic
Acute pancreatitis	OR=1.14 (1 to 1.31)	0.046	0.252	156801	950	155851	phecode 577.1	digestive
Inflammatory bowel disease and other gastroenteritis and colitis	OR=1.1 (1 to 1.2)	0.047	0.253	127644	2082	125562	phecode 555	digestive
Gout	OR=1.08 (1 to 1.16)	0.048	0.253	156915	3131	153784	phecode 274.1	endocrine metabolic
Hemorrhage or hematoma complicating a procedure	OR=1.09 (1 to 1.19)	0.049	0.256	147619	2309	145310	phecode 850	injuries and poisonings
Actinic keratosis	OR=0.89 (0.8 to 1)	0.050	0.256	155924	1297	154627	phecode 702.1	dermatologic
Functional digestive disorders	OR=1.04 (1 to 1.08)	0.051	0.260	139238	13676	125562	phecode 564	digestive
Other disorders of bladder	OR=1.06 (1 to 1.12)	0.051	0.260	154968	5829	149139	phecode 596	genitourinary
Gout and other crystal arthropathies	OR=1.07 (1 to 1.15)	0.058	0.292	157211	3427	153784	phecode 274	endocrine metabolic
Sepsis	OR=1.09 (1 to 1.2)	0.060	0.293	157211	1954	155257	phecode 994.2	injuries and poisonings
Sepsis and SIRS	OR=1.09 (1 to 1.2)	0.060	0.293	157211	1954	155257	phecode 994	injuries and poisonings
Postoperative infection	OR=1.09 (0.99 to 1.19)	0.065	0.311	156723	2028	154695	phecode 080	infectious diseases
Esophagitis; GERD and related diseases	OR=1.03 (1 to 1.07)	0.065	0.311	152090	15205	136885	phecode 530.1	digestive
Enthesopathy	OR=1.06 (1 to 1.13)	0.067	0.314	147072	4046	143026	phecode 726.1	musculoskeletal
Hypothyroidism NOS	OR=1.08 (0.99 to 1.17)	0.067	0.314	156567	2588	153979	phecode 244.4	endocrine metabolic
Sciatica	OR=1.13 (0.99 to 1.29)	0.073	0.327	156502	934	155568	phecode 764	symptoms
Osteoporosis; osteopenia and pathological fracture	OR=0.91 (0.82 to 1.01)	0.074	0.327	157211	1482	155729	phecode 743	musculoskeletal
Osteoporosis	OR=0.89 (0.78 to 1.01)	0.074	0.327	156733	1004	155729	phecode 743.1	musculoskeletal
Osteoporosis NOS	OR=0.89 (0.78 to 1.01)	0.074	0.327	156733	1004	155729	phecode 743.11	musculoskeletal
Degenerative skin conditions and other dermatoses	OR=0.93 (0.86 to 1.01)	0.074	0.327	155651	2584	153067	phecode 702	dermatologic
Other non epithelial cancer of skin	OR=1.04 (1 to 1.08)	0.075	0.327	155540	10010	145530	phecode 172.2	neoplasms

Derangement of joint; non traumatic	OR=1.11 (0.99 to 1.24)	0.075	0.327	155611	1273	154338	phecode 742	musculoskeletal
Peripheral enthesopathies and allied syndromes	OR=1.05 (0.99 to 1.1)	0.080	0.344	149433	6407	143026	phecode 726	musculoskeletal
Other disorders of bone and cartilage	OR=1.1 (0.99 to 1.22)	0.081	0.346	153714	1535	152179	phecode 733	musculoskeletal
Diseases of pancreas	OR=1.1 (0.99 to 1.23)	0.082	0.346	157211	1360	155851	phecode 577	digestive
Poisoning by analgesics; antipyretics; and antirheumatics	OR=1.08 (0.99 to 1.18)	0.083	0.347	148899	2287	146612	phecode 965	injuries and poisonings
Arthropathy NOS	OR=1.03 (1 to 1.06)	0.087	0.358	155282	17480	137802	phecode 716.9	musculoskeletal
Nerve root and plexus disorders	OR=1.08 (0.99 to 1.19)	0.089	0.358	153006	2000	151006	phecode 353	neurological
Total protein	-0.038 (-0.082 to 0.006)	0.089	0.358	138299	NA	NA	NA	biomarker
Other hypertrophic and atrophic conditions of skin	OR=1.09 (0.99 to 1.2)	0.090	0.358	157153	1777	155376	phecode 701	dermatologic
Other arthropathies	OR=1.03 (1 to 1.06)	0.090	0.358	155325	17523	137802	phecode 716	musculoskeletal
Rash and other nonspecific skin eruption	OR=1.12 (0.98 to 1.27)	0.091	0.358	154935	1003	153932	phecode 687.1	dermatologic
Secondary malignancy of lymph nodes	OR=1.08 (0.99 to 1.18)	0.091	0.358	139898	2091	137807	phecode 198.1	neoplasms
Sleep disorders	OR=1.06 (0.99 to 1.14)	0.093	0.363	157211	3601	153610	phecode 327	neurological
Otitis media and Eustachian tube disorders	OR=1.12 (0.98 to 1.28)	0.095	0.363	156494	954	155540	phecode 381	sense organs
Diseases of white blood cells	OR=1.09 (0.99 to 1.2)	0.095	0.363	155271	1683	153588	phecode 288	hematopoietic
Bladder neck obstruction	OR=1.09 (0.98 to 1.2)	0.096	0.363	150802	1663	149139	phecode 596.1	genitourinary
Pleurisy; pleural effusion	OR=1.06 (0.99 to 1.13)	0.097	0.363	153858	4080	149778	phecode 507	respiratory
Lipoprotein A	0.01 (-0.002 to 0.022)	0.102	0.381	118783	NA	NA	NA	biomarker
Chronic dermatitis due to solar radiation	OR=0.91 (0.82 to 1.02)	0.107	0.395	155541	1353	154188	phecode 938.2	dermatologic
Constipation	OR=1.05 (0.99 to 1.12)	0.108	0.395	130028	4466	125562	phecode 563	digestive
Other and unspecified disc disorder	OR=1.07 (0.98 to 1.17)	0.112	0.407	153055	2136	150919	phecode 722.9	musculoskeletal
Cancer of other lymphoid; histiocytic tissue	OR=1.09 (0.98 to 1.22)	0.114	0.409	155784	1437	154347	phecode 202	neoplasms
Dermatitis due to solar radiation	OR=0.92 (0.82 to 1.02)	0.114	0.409	155587	1399	154188	phecode 938	dermatologic
Intervertebral disc disorders	OR=1.06 (0.98 to 1.15)	0.118	0.418	153832	2913	150919	phecode 722	musculoskeletal
Hydrocele	OR=1.1 (0.98 to 1.23)	0.119	0.419	140023	1265	138758	phecode 603.1	genitourinary
Symptoms involving nervous and musculoskeletal systems	OR=0.94 (0.88 to 1.02)	0.120	0.419	157211	3214	153997	phecode 781	symptoms
Renal colic	OR=0.91 (0.81 to 1.03)	0.124	0.427	153392	1137	152255	phecode 594.8	genitourinary
Disorders of fluid; electrolyte; and acid base balance	OR=1.05 (0.99 to 1.12)	0.124	0.427	157211	4637	152574	phecode 276	endocrine metabolic
Nonrheumatic aortic valve disorders	OR=1.08 (0.98 to 1.19)	0.126	0.429	150841	1841	149000	phecode 395.2	circulatory system
Acute appendicitis	OR=1.09 (0.97 to 1.23)	0.131	0.442	156962	1267	155695	phecode 540.11	digestive
Skin cancer	OR=1.03 (0.99 to 1.07)	0.132	0.442	157141	11611	145530	phecode 172	neoplasms
Cystitis and urethritis	OR=1.1 (0.97 to 1.24)	0.133	0.442	145262	1113	144149	phecode 592	genitourinary
Melanomas of skin	OR=1.08 (0.98 to 1.19)	0.135	0.442	147340	1810	145530	phecode 172.11	neoplasms
Melanomas of skin; dx or hx	OR=1.08 (0.98 to 1.19)	0.135	0.442	147340	1810	145530	phecode 172.1	neoplasms
Other biliary tract disease	OR=1.08 (0.97 to 1.2)	0.138	0.444	153390	1530	151860	phecode 575	digestive
Degeneration of macula and posterior pole of retina	OR=1.1 (0.97 to 1.25)	0.139	0.444	94875	1053	93822	phecode 362.2	sense organs
Macular degeneration senile of retina NOS	OR=1.1 (0.97 to 1.25)	0.139	0.444	94875	1053	93822	phecode 362.29	sense organs
Calcium	-0.001 (-0.002 to 0)	0.139	0.444	138426	NA	NA	NA	biomarker
Diseases of the oral soft tissues; excluding lesions specific for gingiva and tongue	OR=1.08 (0.98 to 1.19)	0.141	0.446	156503	1744	154759	phecode 528	digestive
Other diseases of respiratory system; NEC	OR=1.05 (0.98 to 1.11)	0.144	0.451	156921	4822	152099	phecode 519.8	respiratory
Secondary malignancy of bone	OR=1.09 (0.97 to 1.22)	0.146	0.453	139152	1345	137807	phecode 198.6	neoplasms
Secondary malignancy of respiratory organs	OR=0.92 (0.82 to 1.03)	0.147	0.453	139057	1250	137807	phecode 198.2	neoplasms
Cataract	OR=1.03 (0.99 to 1.08)	0.148	0.453	157211	9843	147368	phecode 366	sense organs
Benign neoplasm of other parts of digestive system	OR=1.07 (0.98 to 1.16)	0.150	0.453	157211	2238	154973	phecode 211	neoplasms
Pain in limb	OR=1.06 (0.98 to 1.16)	0.150	0.453	157211	2384	154827	phecode 773	symptoms

Neoplasm of uncertain behavior	OR=0.93 (0.83 to 1.03)	0.152	0.453	139321	1514	137807	phecode 199	neoplasms
Diabetes mellitus	OR=0.97 (0.93 to 1.01)	0.153	0.453	157211	12038	145173	phecode 250	endocrine metabolic
Urea	-0.01 (-0.024 to 0.004)	0.153	0.453	149832	NA	NA	NA	biomarker
Hypertensive heart andoror renal disease	OR=1.1 (0.97 to 1.24)	0.154	0.453	117021	1064	115957	phecode 401.2	circulatory system
Abdominal aortic aneurysm	OR=0.91 (0.79 to 1.04)	0.154	0.453	152477	902	151575	phecode 442.11	circulatory system
Cystitis	OR=1.1 (0.97 to 1.24)	0.158	0.456	145195	1046	144149	phecode 592.1	genitourinary
Other diseases of respiratory system; not elsewhere classified	OR=1.04 (0.98 to 1.11)	0.158	0.456	157211	5112	152099	phecode 519	respiratory
Poisoning by primarily systemic agents	OR=1.09 (0.97 to 1.24)	0.159	0.456	147670	1058	146612	phecode 963	injuries and poisonings
Tobacco use disorder	OR=1.03 (0.99 to 1.07)	0.160	0.456	151119	10356	140763	phecode 318	mental disorders
Cardiac congenital anomalies	OR=1.07 (0.97 to 1.19)	0.162	0.457	157103	1645	155458	phecode 747.1	congenital anomalies
Anxiety disorder	OR=1.06 (0.98 to 1.14)	0.162	0.457	138279	2725	135554	phecode 300.1	mental disorders
Senile cataract	OR=1.05 (0.98 to 1.12)	0.164	0.460	151653	4285	147368	phecode 366.2	sense organs
Atherosclerosis	OR=1.09 (0.96 to 1.24)	0.166	0.463	152626	1051	151575	phecode 440	circulatory system
Dislocation	OR=1.09 (0.96 to 1.23)	0.172	0.474	149324	1082	148242	phecode 830	injuries and poisonings
Malignant neoplasm of other and ill defined sites within the digestive organs and peritoneum	OR=1.05 (0.98 to 1.13)	0.173	0.474	118550	3285	115265	phecode 159	neoplasms
Duodenitis	OR=1.05 (0.98 to 1.11)	0.174	0.474	146811	4217	142594	phecode 535.6	digestive
Pain in joint	OR=1.05 (0.98 to 1.13)	0.179	0.486	157211	3215	153996	phecode 745	musculoskeletal
Back pain	OR=1.04 (0.98 to 1.1)	0.188	0.507	157211	5797	151414	phecode 760	symptoms
Other disorders of arteries and arterioles	OR=1.09 (0.96 to 1.24)	0.193	0.517	152579	1004	151575	phecode 447	circulatory system
Dental caries	OR=1.07 (0.96 to 1.19)	0.198	0.528	154465	1476	152989	phecode 521.1	digestive
Altered mental status	OR=1.07 (0.96 to 1.19)	0.200	0.528	154362	1491	152871	phecode 292.4	mental disorders
Poisoning by antibiotics	OR=1.04 (0.98 to 1.09)	0.200	0.528	152812	6200	146612	phecode 960	injuries and poisonings
Fracture of upper limb	OR=0.95 (0.89 to 1.03)	0.207	0.541	152093	3153	148940	phecode 803	injuries and poisonings
Other local infections of skin and subcutaneous tissue	OR=0.95 (0.87 to 1.03)	0.209	0.543	153426	2334	151092	phecode 686	dermatologic
Delirium dementia and amnestic and other cognitive disorders	OR=1.07 (0.96 to 1.18)	0.210	0.544	154523	1652	152871	phecode 290	mental disorders
Transient cerebral ischemia	OR=0.93 (0.83 to 1.04)	0.212	0.545	152337	1222	151115	phecode 433.31	circulatory system
Cardiac and circulatory congenital anomalies	OR=1.06 (0.96 to 1.17)	0.219	0.556	157211	1753	155458	phecode 747	congenital anomalies
Fracture of radius and ulna	OR=0.94 (0.84 to 1.04)	0.220	0.556	150452	1512	148940	phecode 803.2	injuries and poisonings
Other derangement of joint	OR=1.08 (0.95 to 1.23)	0.221	0.556	155391	1053	154338	phecode 742.9	musculoskeletal
Diseases of hard tissues of teeth	OR=1.07 (0.96 to 1.19)	0.221	0.556	154484	1495	152989	phecode 521	digestive
Other mental disorder	OR=1.02 (0.99 to 1.05)	0.224	0.559	154074	18520	135554	phecode 306	mental disorders
Intestinal obstruction without mention of hernia	OR=1.06 (0.96 to 1.17)	0.227	0.560	127446	1884	125562	phecode 560	digestive
Calculus of ureter	OR=0.94 (0.85 to 1.04)	0.227	0.560	153872	1617	152255	phecode 594.3	genitourinary
Symptoms involving head and neck	OR=1.06 (0.96 to 1.17)	0.229	0.560	157211	1709	155502	phecode 293	mental disorders
Tachycardia NOS	OR=0.93 (0.82 to 1.05)	0.230	0.560	141157	1146	140011	phecode 427.7	circulatory system
Other disorders of eyelids	OR=1.05 (0.97 to 1.15)	0.230	0.560	155466	2335	153131	phecode 374	sense organs
Hemorrhage of gastrointestinal tract	OR=1.05 (0.97 to 1.13)	0.233	0.564	149364	3022	146342	phecode 578.9	digestive
Cancer of urinary organs incl kidney and bladder	OR=0.96 (0.89 to 1.03)	0.239	0.572	157211	3057	154154	phecode 189	neoplasms
Diseases of the larynx and vocal cords	OR=1.07 (0.95 to 1.21)	0.239	0.572	149221	1164	148057	phecode 473	respiratory
Anal and rectal polyp	OR=1.04 (0.97 to 1.11)	0.241	0.572	146264	4298	141966	phecode 565.1	digestive
Fracture of unspecified part of femur	OR=0.92 (0.8 to 1.06)	0.242	0.572	149838	898	148940	phecode 800.2	injuries and poisonings
Cerebral artery occlusion; with cerebral infarction	OR=0.93 (0.83 to 1.05)	0.243	0.572	152417	1302	151115	phecode 433.21	circulatory system
Other open wound of head and face	OR=0.95 (0.86 to 1.04)	0.245	0.573	153838	1980	151858	phecode 870.3	injuries and poisonings
Psoriasis	OR=1.07 (0.95 to 1.2)	0.247	0.574	143566	1235	142331	phecode 696.4	dermatologic
Open wounds of extremities	OR=1.05 (0.97 to 1.13)	0.256	0.593	154725	2867	151858	phecode 871	injuries and poisonings
Abnormal results of function study of liver	OR=1.06 (0.96 to 1.16)	0.265	0.609	154132	1811	152321	phecode 573.7	digestive
Paroxysmal tachycardia; unspecified	OR=0.95 (0.86 to 1.04)	0.277	0.628	141775	1764	140011	phecode 427.1	circulatory system
Hemorrhoids	OR=1.02 (0.98 to 1.07)	0.278	0.628	151566	10293	141273	phecode 455	circulatory system

Paroxysmal supraventricular tachycardia	OR=0.93 (0.83 to 1.06)	0.278	0.628	141140	1129	140011	phecode 427.11	circulatory system
Other upper respiratory disease	OR=1.05 (0.96 to 1.16)	0.279	0.628	149974	1917	148057	phecode 479	respiratory
Electrolyte imbalance	OR=1.05 (0.96 to 1.13)	0.280	0.628	155178	2604	152574	phecode 276.1	endocrine metabolic
Other disorders of urethra and urinary tract	OR=1.05 (0.96 to 1.15)	0.283	0.631	151230	2091	149139	phecode 597	genitourinary
Bacterial enteritis	OR=1.07 (0.95 to 1.2)	0.286	0.635	153897	1169	152728	phecode 008.5	infectious diseases
Ulcerative colitis	OR=1.06 (0.95 to 1.17)	0.293	0.649	127122	1560	125562	phecode 555.2	digestive
Inflammatory and toxic neuropathy	OR=0.94 (0.83 to 1.06)	0.297	0.653	156804	1127	155677	phecode 357	neurological
Cholecystitis without cholelithiasis	OR=1.07 (0.94 to 1.22)	0.300	0.657	152857	997	151860	phecode 574.3	digestive
Anxiety disorders	OR=1.04 (0.97 to 1.12)	0.304	0.657	138570	3016	135554	phecode 300	mental disorders
Varicose veins	OR=1.04 (0.97 to 1.11)	0.306	0.657	144964	3691	141273	phecode 454	circulatory system
Hypovolemia	OR=1.05 (0.95 to 1.16)	0.307	0.657	154337	1763	152574	phecode 276.5	endocrine metabolic
Other aneurysm	OR=0.95 (0.85 to 1.05)	0.309	0.657	153158	1583	151575	phecode 442	circulatory system
Non Hodgkins lymphoma	OR=1.06 (0.94 to 1.2)	0.309	0.657	155517	1170	154347	phecode 202.2	neoplasms
Other disorders of liver	OR=1.03 (0.97 to 1.1)	0.309	0.657	156496	4175	152321	phecode 573	digestive
First degree AV block	OR=0.93 (0.81 to 1.07)	0.313	0.663	140890	879	140011	phecode 426.21	circulatory system
Psoriasis and related disorders	OR=1.06 (0.95 to 1.19)	0.318	0.669	143599	1268	142331	phecode 696	dermatologic
Seborrheic keratosis	OR=0.95 (0.85 to 1.06)	0.322	0.669	156046	1419	154627	phecode 702.2	dermatologic
Unstable angina intermediate coronary syndrome	OR=1.04 (0.96 to 1.12)	0.322	0.669	139239	2997	136242	phecode 411.1	circulatory system
Aspartate aminotransferase	0.005 (-0.005 to 0.016)	0.323	0.669	149354	NA	NA	NA	biomarker
Chronic ulcer of skin	OR=1.06 (0.95 to 1.18)	0.334	0.682	157211	1420	155791	phecode 707	dermatologic
Cellulitis and abscess of armorhand	OR=0.97 (0.9 to 1.04)	0.335	0.682	154462	3370	151092	phecode 681.3	dermatologic
Cellulitis and abscess of foot; toe	OR=0.97 (0.9 to 1.04)	0.335	0.682	154462	3370	151092	phecode 681.6	dermatologic
Cellulitis and abscess of leg; except foot	OR=0.97 (0.9 to 1.04)	0.335	0.682	154462	3370	151092	phecode 681.5	dermatologic
Chronic renal failure CKD	OR=1.04 (0.96 to 1.11)	0.338	0.685	152156	3242	148914	phecode 585.3	genitourinary
Aortic aneurysm	OR=0.95 (0.85 to 1.06)	0.341	0.688	152901	1326	151575	phecode 442.1	circulatory system
Abnormal sputum	OR=1.04 (0.96 to 1.13)	0.343	0.688	157211	2581	154630	phecode 516	respiratory
Gastric ulcer	OR=1.04 (0.95 to 1.14)	0.346	0.691	155429	2060	153369	phecode 531.2	digestive
Other anemias	OR=0.97 (0.92 to 1.03)	0.349	0.695	154546	5218	149328	phecode 285	hematopoietic
Septicemia	OR=1.04 (0.96 to 1.13)	0.354	0.699	152151	2662	149489	phecode 038	infectious diseases
Allergyoradverse effect of penicillin	OR=1.03 (0.97 to 1.08)	0.355	0.699	152298	5686	146612	phecode 960.2	injuries and poisonings
Cardiac conduction disorders	OR=0.97 (0.92 to 1.03)	0.356	0.699	145047	5036	140011	phecode 426	circulatory system
Asthma	OR=1.02 (0.98 to 1.06)	0.361	0.699	152957	10018	142939	phecode 495	respiratory
Other disorders of testis	OR=1.04 (0.95 to 1.14)	0.363	0.699	140795	2037	138758	phecode 603	genitourinary
Fracture of tibia and fibula	OR=1.06 (0.93 to 1.21)	0.364	0.699	149884	944	148940	phecode 800.3	injuries and poisonings
Other intestinal obstruction	OR=1.05 (0.94 to 1.17)	0.365	0.699	127050	1488	125562	phecode 560.4	digestive
Gastritis and duodenitis	OR=1.02 (0.98 to 1.06)	0.365	0.699	155649	13055	142594	phecode 535	digestive
Carditis	OR=1.05 (0.94 to 1.18)	0.366	0.699	156416	1376	155040	phecode 420	circulatory system
Benign neoplasm of skin	OR=0.97 (0.9 to 1.04)	0.370	0.705	156940	2940	154000	phecode 216	neoplasms
Triglycerides	-0.005 (-0.015 to 0.006)	0.372	0.706	149776	NA	NA	NA	biomarker
Nonspecific chest pain	OR=1.02 (0.98 to 1.05)	0.377	0.712	157211	14178	143033	phecode 418	circulatory system
Lipoma of skin and subcutaneous tissue	OR=1.04 (0.95 to 1.14)	0.385	0.725	155732	2134	153598	phecode 214.1	neoplasms
Streptococcus infection	OR=1.06 (0.93 to 1.22)	0.387	0.725	150400	911	149489	phecode 041.2	infectious diseases
Appendicitis	OR=1.05 (0.94 to 1.17)	0.389	0.726	157158	1463	155695	phecode 540.1	digestive
Reflux esophagitis	OR=1.03 (0.97 to 1.09)	0.396	0.736	142036	5151	136885	phecode 530.14	digestive
Other diseases of the teeth and supporting structures	OR=1.05 (0.93 to 1.19)	0.400	0.737	154150	1161	152989	phecode 525	digestive
Psoriasis vulgaris	OR=1.06 (0.93 to 1.2)	0.403	0.737	143312	981	142331	phecode 696.41	dermatologic
Renal failure	OR=1.02 (0.97 to 1.08)	0.406	0.737	155360	6446	148914	phecode 585	genitourinary
Cystatin C	-0.004 (-0.014 to 0.006)	0.408	0.737	149927	NA	NA	NA	biomarker
Fasciitis	OR=0.97 (0.89 to 1.05)	0.409	0.737	145804	2778	143026	phecode 728.7	musculoskeletal
Respiratory insufficiency	OR=1.05 (0.94 to 1.18)	0.410	0.737	151070	1292	149778	phecode 509.2	respiratory
Staphylococcus infections	OR=1.04 (0.95 to 1.15)	0.410	0.737	151291	1802	149489	phecode 041.1	infectious diseases
Mitral valve disease	OR=0.96 (0.87 to 1.06)	0.411	0.737	150902	1902	149000	phecode 394.2	circulatory system
Emphysema	OR=1.05 (0.93 to 1.19)	0.412	0.737	144084	1145	142939	phecode 496.1	respiratory
Bundle branch block	OR=0.97 (0.89 to 1.05)	0.418	0.745	142464	2453	140011	phecode 426.3	circulatory system
Diaphragmatic hernia	OR=1.02 (0.98 to 1.06)	0.421	0.745	141891	11652	130239	phecode 550.2	digestive

Nonrheumatic mitral valve disorders	OR=0.96 (0.87 to 1.06)	0.422	0.745	150859	1859	149000	phecode 395.1	circulatory system
Symptoms involving skin and other integumentary tissue	OR=1.05 (0.93 to 1.19)	0.431	0.752	157211	1099	156112	phecode 782	symptoms
Epistaxis or throat hemorrhage	OR=1.05 (0.94 to 1.17)	0.431	0.752	149441	1384	148057	phecode 477	respiratory
Fever of unknown origin	OR=0.96 (0.88 to 1.06)	0.432	0.752	157211	2001	155210	phecode 783	symptoms
Ventral hernia	OR=1.04 (0.94 to 1.15)	0.434	0.752	131993	1754	130239	phecode 550.5	digestive
Neurological disorders	OR=1.03 (0.96 to 1.11)	0.436	0.752	155814	2943	152871	phecode 292	mental disorders
Contracture of palmar fascia Dupuytrens disease	OR=0.97 (0.89 to 1.05)	0.439	0.752	145632	2606	143026	phecode 728.71	musculoskeletal
Decreased white blood cell count	OR=1.04 (0.94 to 1.17)	0.440	0.752	154969	1381	153588	phecode 288.1	hematopoietic
Neutropenia	OR=1.04 (0.94 to 1.17)	0.440	0.752	154969	1381	153588	phecode 288.11	hematopoietic
Lymphadenitis	OR=1.04 (0.93 to 1.17)	0.442	0.752	154930	1342	153588	phecode 289.4	hematopoietic
Acquired foot deformities	OR=0.96 (0.86 to 1.07)	0.447	0.752	155349	1384	153965	phecode 735	musculoskeletal
Degeneration of intervertebral disc	OR=1.05 (0.93 to 1.18)	0.448	0.752	152067	1148	150919	phecode 722.6	musculoskeletal
Cough	OR=1.04 (0.93 to 1.17)	0.448	0.752	152453	1344	151109	phecode 512.8	respiratory
Convulsions	OR=1.05 (0.93 to 1.18)	0.451	0.752	148322	1219	147103	phecode 345.3	neurological
Acute upper respiratory infections of multiple or unspecified sites	OR=1.05 (0.92 to 1.2)	0.453	0.752	157150	942	156208	phecode 465	respiratory
Edema	OR=1.05 (0.92 to 1.19)	0.453	0.752	157129	1017	156112	phecode 782.3	symptoms
Hemoptysis	OR=1.03 (0.95 to 1.12)	0.454	0.752	157105	2475	154630	phecode 516.1	respiratory
Apolipoprotein A	-0.001 (-0.003 to 0.002)	0.455	0.752	138185	NA	NA	NA	biomarker
Abdominal pain	OR=1.01 (0.98 to 1.05)	0.457	0.752	157211	13297	143914	phecode 785	symptoms
Diseases of hair and hair follicles	OR=1.03 (0.95 to 1.12)	0.460	0.754	156720	2347	154373	phecode 704	dermatologic
Fracture of ribs	OR=1.05 (0.92 to 1.19)	0.466	0.762	149986	1046	148940	phecode 807	injuries and poisonings
Cholelithiasis with other cholecystitis	OR=1.04 (0.94 to 1.15)	0.471	0.767	153385	1525	151860	phecode 574.12	digestive
Stricture and stenosis of esophagus	OR=1.04 (0.93 to 1.18)	0.475	0.771	138053	1168	136885	phecode 530.3	digestive
Malaise and fatigue	OR=1.04 (0.93 to 1.16)	0.478	0.772	157211	1440	155771	phecode 798	symptoms
Vitamin D	0.004 (-0.007 to 0.014)	0.479	0.772	145566	NA	NA	NA	biomarker
Fracture of ankle and foot	OR=0.95 (0.83 to 1.09)	0.482	0.774	149820	880	148940	phecode 801	injuries and poisonings
Other specified gastritis	OR=1.02 (0.96 to 1.1)	0.493	0.789	146248	3654	142594	phecode 535.8	digestive
Substance addiction and disorders	OR=1.02 (0.96 to 1.09)	0.497	0.789	144599	3836	140763	phecode 316	mental disorders
Unspecified diffuse connective tissue disease	OR=0.96 (0.84 to 1.09)	0.498	0.789	145036	1019	144017	phecode 709.7	dermatologic
Appendiceal conditions	OR=1.04 (0.93 to 1.15)	0.500	0.789	157211	1516	155695	phecode 540	digestive
Effects radiation NOS	OR=1.03 (0.94 to 1.14)	0.501	0.789	155439	1852	153587	phecode 990	injuries and poisonings
Other disorders of synovium; tendon; and bursa	OR=1.02 (0.96 to 1.09)	0.502	0.789	146899	3873	143026	phecode 727	musculoskeletal
Acquired toe deformities	OR=0.95 (0.83 to 1.09)	0.505	0.790	154873	908	153965	phecode 735.2	musculoskeletal
Aphakia and other disorders of lens	OR=1.05 (0.92 to 1.19)	0.506	0.790	154756	969	153787	phecode 379.3	sense organs
Heart valve disorders	OR=1.02 (0.96 to 1.09)	0.511	0.792	152826	3826	149000	phecode 395	circulatory system
Diffuse diseases of connective tissue	OR=0.96 (0.85 to 1.08)	0.513	0.792	145150	1133	144017	phecode 709	dermatologic
Respiratory failure; insufficiency; arrest	OR=1.03 (0.94 to 1.14)	0.513	0.792	151605	1827	149778	phecode 509	respiratory
Cancer of bladder	OR=0.97 (0.88 to 1.07)	0.516	0.794	156022	1868	154154	phecode 189.2	neoplasms
Ileostomy status	OR=1.04 (0.91 to 1.2)	0.528	0.805	126476	914	125562	phecode 559	digestive
Other disorders of stomach and duodenum	OR=1.02 (0.95 to 1.1)	0.530	0.805	145936	3342	142594	phecode 537	digestive
Hyposmolality andoror hyponatremia	OR=1.04 (0.92 to 1.18)	0.531	0.805	153659	1085	152574	phecode 276.12	endocrine metabolic
Disorder of skin and subcutaneous tissue NOS	OR=1.03 (0.95 to 1.11)	0.532	0.805	157211	2609	154602	phecode 689	dermatologic
Irritable Bowel Syndrome	OR=0.97 (0.87 to 1.08)	0.532	0.805	126963	1401	125562	phecode 564.1	digestive
Thrombocytopenia	OR=1.04 (0.91 to 1.19)	0.535	0.806	156674	918	155756	phecode 287.3	hematopoietic
Orthostatic hypotension	OR=1.04 (0.91 to 1.19)	0.536	0.806	146303	943	145360	phecode 458.1	circulatory system
Inflammation of the eye	OR=0.97 (0.86 to 1.08)	0.541	0.808	154480	1349	153131	phecode 371	sense organs
Localized superficial swelling; mass; or lump	OR=0.96 (0.83 to 1.1)	0.542	0.808	154808	876	153932	phecode 687.2	dermatologic
Glaucoma	OR=1.03 (0.94 to 1.11)	0.546	0.811	153095	2469	150626	phecode 365	sense organs
HDL cholesterol	-0.003 (-0.014 to 0.008)	0.549	0.812	138394	NA	NA	NA	biomarker
Blood in stool	OR=1.03 (0.93 to 1.15)	0.550	0.812	147799	1457	146342	phecode 578.2	digestive
Fracture of clavicle or scapula	OR=1.04 (0.92 to 1.17)	0.552	0.812	150048	1108	148940	phecode 803.3	injuries and poisonings

Atrial fibrillation and flutter	OR=1.01 (0.97 to 1.06)	0.556	0.816	150383	10372	140011	phecode 427.2	circulatory system
Symptoms affecting skin	OR=1.02 (0.95 to 1.1)	0.563	0.823	157211	3279	153932	phecode 687	dermatologic
Ganglion and cyst of synovium; tendon; and bursa	OR=0.96 (0.84 to 1.1)	0.565	0.823	143973	947	143026	phecode 727.4	musculoskeletal
Leukemia	OR=1.03 (0.92 to 1.16)	0.570	0.827	155553	1206	154347	phecode 204	neoplasms
Malignant neoplasm of rectum; rectosigmoid junction; and anus	OR=1.03 (0.93 to 1.15)	0.572	0.828	118320	1494	116826	phecode 153.3	neoplasms
Pulmonary collapse; interstitial and compensatory emphysema	OR=1.03 (0.92 to 1.16)	0.580	0.837	151099	1321	149778	phecode 508	respiratory
Syncope and collapse	OR=0.98 (0.93 to 1.04)	0.586	0.840	157211	4965	152246	phecode 788	symptoms
Superficial injury without mention of infection	OR=1.02 (0.95 to 1.1)	0.589	0.840	156951	2791	154160	phecode 915	injuries and poisonings
Alcoholism	OR=0.98 (0.93 to 1.04)	0.590	0.840	146499	5736	140763	phecode 317.1	mental disorders
Renal failure NOS	OR=1.04 (0.91 to 1.18)	0.591	0.840	149863	949	148914	phecode 585.2	genitourinary
Cancer of bronchus; lung	OR=1.03 (0.92 to 1.15)	0.592	0.840	156871	1458	155413	phecode 165.1	neoplasms
Other disorders of the kidney and ureters	OR=1.02 (0.94 to 1.12)	0.594	0.840	150972	2058	148914	phecode 586	genitourinary
Aortic valve disease	OR=1.04 (0.91 to 1.19)	0.596	0.840	149938	938	149000	phecode 394.3	circulatory system
Abnormality of gait	OR=1.03 (0.92 to 1.15)	0.600	0.841	156620	1324	155296	phecode 350.2	neurological
Cardiac pacemakerordevice in situ	OR=0.98 (0.89 to 1.07)	0.602	0.841	141903	1892	140011	phecode 426.9	circulatory system
Abnormal findings examination of lungs	OR=1.03 (0.92 to 1.15)	0.603	0.841	157211	1391	155820	phecode 514	respiratory
Symptoms and disorders of the joints	OR=1.03 (0.93 to 1.13)	0.604	0.841	156044	1706	154338	phecode 741	musculoskeletal
Cardiomegaly	OR=1.02 (0.94 to 1.12)	0.607	0.843	154883	2067	152816	phecode 416	circulatory system
Complication due to other implant and internal device	OR=0.97 (0.88 to 1.08)	0.612	0.845	146977	1667	145310	phecode 859	injuries and poisonings
Gastrointestinal hemorrhage	OR=1.01 (0.97 to 1.06)	0.612	0.845	156012	9670	146342	phecode 578	digestive
Swelling of limb	OR=1.02 (0.94 to 1.12)	0.619	0.848	156888	2127	154761	phecode 771.1	symptoms
Other acute and subacute forms of ischemic heart disease	OR=1.03 (0.92 to 1.16)	0.620	0.848	137463	1221	136242	phecode 411.9	circulatory system
Open wounds of head; neck; and trunk	OR=0.98 (0.9 to 1.07)	0.620	0.848	154137	2279	151858	phecode 870	injuries and poisonings
Pericarditis	OR=1.04 (0.9 to 1.19)	0.622	0.848	155919	879	155040	phecode 420.2	circulatory system
Occlusion and stenosis of precerebral arteries	OR=1.03 (0.9 to 1.18)	0.626	0.848	152030	915	151115	phecode 433.1	circulatory system
Malignant neoplasm of bladder	OR=0.97 (0.88 to 1.08)	0.627	0.848	155780	1626	154154	phecode 189.21	neoplasms
Peptic ulcer excl esophageal	OR=1.02 (0.95 to 1.09)	0.632	0.849	157211	3842	153369	phecode 531	digestive
Other disorders of eye	OR=0.98 (0.89 to 1.07)	0.632	0.849	155754	1967	153787	phecode 379	sense organs
Vertiginous syndromes and other disorders of vestibular system	OR=0.98 (0.91 to 1.06)	0.633	0.849	157211	2838	154373	phecode 386	sense organs
Heart valve replaced	OR=1.03 (0.91 to 1.17)	0.639	0.855	150046	1046	149000	phecode 395.6	circulatory system
Type 1 diabetes	OR=0.98 (0.88 to 1.08)	0.645	0.859	146702	1529	145173	phecode 250.1	endocrine metabolic
Fracture of vertebral column without mention of spinal cord injury	OR=1.03 (0.9 to 1.18)	0.649	0.859	149858	918	148940	phecode 805	injuries and poisonings
Carbuncle and furuncle	OR=0.97 (0.86 to 1.1)	0.652	0.859	152286	1194	151092	phecode 686.1	dermatologic
Diverticulosis	OR=0.99 (0.95 to 1.03)	0.653	0.859	139014	13452	125562	phecode 562.1	digestive
Diverticulosis and diverticulitis	OR=0.99 (0.95 to 1.03)	0.653	0.859	139014	13452	125562	phecode 562	digestive
Fracture of hand or wrist	OR=0.98 (0.88 to 1.08)	0.655	0.859	150596	1656	148940	phecode 804	injuries and poisonings
Cardiac arrest and ventricular fibrillation	OR=1.03 (0.9 to 1.18)	0.656	0.859	140944	933	140011	phecode 427.4	circulatory system
Varicose veins of lower extremity	OR=1.02 (0.95 to 1.09)	0.660	0.862	144622	3349	141273	phecode 454.1	circulatory system
Other forms of chronic heart disease	OR=1.02 (0.92 to 1.13)	0.666	0.865	137837	1595	136242	phecode 414	circulatory system
Synovitis and tenosynovitis	OR=1.02 (0.93 to 1.12)	0.666	0.865	145109	2083	143026	phecode 727.1	musculoskeletal
Cerebrovascular disease	OR=0.99 (0.93 to 1.05)	0.671	0.869	156525	5410	151115	phecode 433	circulatory system
Cholelithiasis and cholecystitis	OR=1.01 (0.95 to 1.08)	0.677	0.874	156526	4666	151860	phecode 574	digestive
Urethral stricture not specified as infectious	OR=1.02 (0.93 to 1.12)	0.690	0.886	150977	1838	149139	phecode 597.1	genitourinary
Chronic sinusitis	OR=1.03 (0.91 to 1.16)	0.691	0.886	149183	1126	148057	phecode 475	respiratory
Cholesterol	-0.002 (-0.014 to 0.009)	0.696	0.888	149940	NA	NA	NA	biomarker
Hematemesis	OR=1.02 (0.9 to 1.16)	0.700	0.888	147411	1069	146342	phecode 578.1	digestive
Acute renal failure	OR=1.01 (0.95 to 1.08)	0.700	0.888	152907	3993	148914	phecode 585.1	genitourinary
Colorectal cancer	OR=1.02 (0.94 to 1.1)	0.701	0.888	119810	2984	116826	phecode 153	neoplasms
Colon cancer	OR=1.02 (0.93 to 1.12)	0.705	0.889	118823	1997	116826	phecode 153.2	neoplasms
Diabetic retinopathy	OR=1.03 (0.9 to 1.18)	0.711	0.889	94735	913	93822	phecode 250.7	endocrine metabolic

Fracture of lower limb	OR=1.02 (0.94 to 1.1)	0.711	0.889	151628	2688	148940	phecode 800	injuries and poisonings
Left bundle branch block	OR=1.02 (0.91 to 1.15)	0.715	0.889	141209	1198	140011	phecode 426.32	circulatory system
Pneumonia	OR=0.99 (0.94 to 1.04)	0.716	0.889	157057	6642	150415	phecode 480	respiratory
Iron deficiency anemias	OR=1.01 (0.94 to 1.09)	0.717	0.889	152701	3373	149328	phecode 280	hematopoietic
Iron deficiency anemias; unspecified or not due to blood loss	OR=1.01 (0.94 to 1.09)	0.717	0.889	152701	3373	149328	phecode 280.1	hematopoietic
Osteoarthrosis	OR=1.01 (0.97 to 1.05)	0.717	0.889	157211	11721	145490	phecode 740	musculoskeletal
Bacterial pneumonia	OR=0.99 (0.93 to 1.05)	0.726	0.893	154608	4193	150415	phecode 480.1	respiratory
Palpitations	OR=0.98 (0.87 to 1.1)	0.727	0.893	141308	1297	140011	phecode 427.9	circulatory system
Disorders of mineral metabolism	OR=0.98 (0.88 to 1.1)	0.730	0.893	157211	1343	155868	phecode 275	endocrine metabolic
Abnormal findings on examination of urine	OR=1.02 (0.93 to 1.11)	0.732	0.893	157211	2083	155128	phecode 598	genitourinary
Hypotension	OR=0.99 (0.92 to 1.06)	0.733	0.893	148804	3444	145360	phecode 458	circulatory system
Retinal detachments and defects	OR=0.98 (0.89 to 1.08)	0.734	0.893	152471	1845	150626	phecode 361	sense organs
Retinal detachment with retinal defect	OR=0.98 (0.89 to 1.08)	0.734	0.893	152471	1845	150626	phecode 361.1	sense organs
Redundant prepuce and phimosisorBXO	OR=1.01 (0.93 to 1.1)	0.740	0.897	141251	2493	138758	phecode 604.1	genitourinary
Nasal polyps	OR=1.02 (0.93 to 1.11)	0.744	0.899	150012	1955	148057	phecode 471	respiratory
Hearing loss	OR=1.01 (0.93 to 1.1)	0.746	0.899	157185	2573	154612	phecode 389	sense organs
Dizziness and giddiness Light headedness and vertigo	OR=1.01 (0.93 to 1.11)	0.747	0.899	156589	2216	154373	phecode 386.9	sense organs
Purpura and other hemorrhagic conditions	OR=1.02 (0.9 to 1.16)	0.751	0.901	156790	1034	155756	phecode 287	hematopoietic
Rheumatoid arthritis and other inflammatory polyarthropathies	OR=0.98 (0.89 to 1.09)	0.755	0.902	157211	1585	155626	phecode 714	musculoskeletal
Respiratory failure	OR=1.02 (0.91 to 1.13)	0.759	0.905	151257	1479	149778	phecode 509.1	respiratory
Overweight; obesity and other hyperalimentation	OR=0.99 (0.94 to 1.05)	0.765	0.910	157211	5734	151477	phecode 278	endocrine metabolic
Obesity	OR=0.99 (0.94 to 1.05)	0.773	0.916	157184	5707	151477	phecode 278.1	endocrine metabolic
Urinary calculus	OR=0.99 (0.93 to 1.06)	0.775	0.917	156430	4175	152255	phecode 594	genitourinary
Osteoarthritis; localized	OR=1.01 (0.96 to 1.06)	0.779	0.919	152917	7427	145490	phecode 740.1	musculoskeletal
Pulmonary heart disease	OR=0.99 (0.91 to 1.07)	0.783	0.919	155334	2518	152816	phecode 415	circulatory system
Benign neoplasm of unspecified sites	OR=1.02 (0.89 to 1.17)	0.784	0.919	157211	900	156311	phecode 229	neoplasms
Disorders of penis	OR=1.01 (0.94 to 1.09)	0.791	0.925	141790	3032	138758	phecode 604	genitourinary
Urinary incontinence	OR=0.98 (0.88 to 1.11)	0.795	0.927	146048	1247	144801	phecode 599.4	genitourinary
Bronchiectasis	OR=1.02 (0.89 to 1.17)	0.800	0.927	143844	905	142939	phecode 496.3	respiratory
Cardiac pacemaker in situ	OR=0.99 (0.89 to 1.09)	0.802	0.927	141740	1729	140011	phecode 426.91	circulatory system
Other headache syndromes	OR=1.01 (0.94 to 1.09)	0.803	0.927	156671	2932	153739	phecode 339	neurological
Cancer within the respiratory system	OR=1.01 (0.92 to 1.12)	0.807	0.927	157193	1780	155413	phecode 165	neoplasms
Sebaceous cyst	OR=1.01 (0.94 to 1.08)	0.810	0.927	156823	3563	153260	phecode 706.2	dermatologic
Other disorders of circulatory system	OR=1.01 (0.96 to 1.05)	0.813	0.927	154746	9386	145360	phecode 459	circulatory system
Secondary malignant neoplasm	OR=1.01 (0.95 to 1.07)	0.814	0.927	142238	4431	137807	phecode 198	neoplasms
Acute pulmonary heart disease	OR=1.01 (0.93 to 1.1)	0.815	0.927	155042	2226	152816	phecode 415.1	circulatory system
Pulmonary embolism and infarction; acute	OR=1.01 (0.93 to 1.1)	0.815	0.927	155042	2226	152816	phecode 415.11	circulatory system
Disturbance of skin sensation	OR=0.99 (0.88 to 1.11)	0.817	0.927	155152	1220	153932	phecode 687.4	dermatologic
Circulatory disease NEC	OR=1.01 (0.96 to 1.05)	0.818	0.927	154635	9275	145360	phecode 459.9	circulatory system
Other disorders of biliary tract	OR=0.99 (0.86 to 1.13)	0.832	0.941	152791	931	151860	phecode 575.8	digestive
Phlebitis and thrombophlebitis	OR=0.99 (0.91 to 1.08)	0.836	0.943	143446	2173	141273	phecode 451	circulatory system
Calculus of bile duct	OR=1.01 (0.88 to 1.16)	0.845	0.947	152759	899	151860	phecode 574.2	digestive
Atrioventricular AV block	OR=0.99 (0.9 to 1.09)	0.846	0.947	141731	1720	140011	phecode 426.2	circulatory system
Cholelithiasis	OR=1.01 (0.94 to 1.07)	0.850	0.947	156001	4141	151860	phecode 574.1	digestive
Other disorders of soft tissues	OR=0.99 (0.92 to 1.07)	0.854	0.947	146197	3171	143026	phecode 729	musculoskeletal
Other chronic nonalcoholic liver disease	OR=0.99 (0.88 to 1.11)	0.854	0.947	153654	1333	152321	phecode 571.5	digestive
Disorders of refraction and accommodation; blindness and low vision	OR=1.01 (0.91 to 1.13)	0.855	0.947	157211	1411	155800	phecode 367	sense organs
Inflammation of eyelids	OR=0.99 (0.87 to 1.12)	0.856	0.947	154216	1085	153131	phecode 371.3	sense organs
Diseases of sebaceous glands	OR=1.01 (0.94 to 1.08)	0.857	0.947	156848	3588	153260	phecode 706	dermatologic
Secondary malignant neoplasm of liver	OR=0.99 (0.89 to 1.1)	0.863	0.948	139299	1492	137807	phecode 198.4	neoplasms
Rheumatic disease of the heart valves	OR=0.99 (0.92 to 1.07)	0.864	0.948	152296	3296	149000	phecode 394	circulatory system

Complication of internal orthopedic device	OR=0.99 (0.91 to 1.08)	0.865	0.948	147475	2165	145310	phecode 858	injuries and poisonings
Calculus of kidney	OR=0.99 (0.91 to 1.09)	0.868	0.948	154296	2041	152255	phecode 594.1	genitourinary
Dysphagia	OR=0.99 (0.92 to 1.07)	0.869	0.948	139768	2883	136885	phecode 532	digestive
Epilepsy; recurrent seizures; convulsions	OR=1.01 (0.93 to 1.09)	0.878	0.949	149628	2525	147103	phecode 345	neurological
Viral infection	OR=0.99 (0.87 to 1.12)	0.880	0.949	155858	1051	154807	phecode 079	infectious diseases
Other acquired deformities of limbs	OR=0.99 (0.86 to 1.13)	0.880	0.949	154886	921	153965	phecode 736	musculoskeletal
Skull and face fracture and other intercranial injury	OR=1.01 (0.92 to 1.11)	0.881	0.949	157211	1949	155262	phecode 819	injuries and poisonings
Abnormal movement	OR=1.01 (0.92 to 1.11)	0.881	0.949	157211	1915	155296	phecode 350	neurological
Visual disturbances	OR=0.99 (0.89 to 1.1)	0.882	0.949	157211	1577	155634	phecode 368	sense organs
Abnormal heart sounds	OR=1 (0.94 to 1.06)	0.886	0.951	153532	4532	149000	phecode 396	circulatory system
Superficial cellulitis and abscess	OR=1 (0.93 to 1.06)	0.888	0.951	155296	4204	151092	phecode 681	dermatologic
LDL direct	-0.001 (-0.01 to 0.008)	0.895	0.955	149626	NA	NA	NA	biomarker
Cardiac dysrhythmias	OR=1 (0.97 to 1.04)	0.897	0.955	154977	14966	140011	phecode 427	circulatory system
Hemorrhage of rectum and anus	OR=1 (0.95 to 1.06)	0.900	0.956	151592	5250	146342	phecode 578.8	digestive
Alanine aminotransferase	-0.001 (-0.011 to 0.01)	0.903	0.956	149830	NA	NA	NA	biomarker
Esophageal bleeding varicesorhemorrhage	OR=1.01 (0.88 to 1.15)	0.903	0.956	137865	980	136885	phecode 530.2	digestive
Osteoarthrosis; localized; primary	OR=1 (0.94 to 1.07)	0.909	0.959	149212	3722	145490	phecode 740.11	musculoskeletal
Occlusion of cerebral arteries	OR=1 (0.92 to 1.08)	0.918	0.965	153888	2773	151115	phecode 433.2	circulatory system
Hereditary retinal dystrophies	OR=1.01 (0.88 to 1.15)	0.919	0.965	94759	937	93822	phecode 362.7	sense organs
Precordial pain	OR=1 (0.9 to 1.1)	0.923	0.965	144876	1843	143033	phecode 418.1	circulatory system
Disorders of muscle; ligament; and fascia	OR=1 (0.93 to 1.08)	0.925	0.965	146118	3092	143026	phecode 728	musculoskeletal
Osteoarthrosis NOS	OR=1 (0.94 to 1.06)	0.925	0.965	150594	5104	145490	phecode 740.9	musculoskeletal
Hypotension NOS	OR=1 (0.92 to 1.07)	0.928	0.965	148355	2995	145360	phecode 458.9	circulatory system
Nausea and vomiting	OR=1 (0.94 to 1.06)	0.940	0.975	157211	4168	153043	phecode 789	symptoms
Complications of cardiacorvascular device; implant; and graft	OR=1 (0.9 to 1.1)	0.943	0.977	146946	1636	145310	phecode 854	injuries and poisonings
Phlebitis and thrombophlebitis of lower extremities	OR=1 (0.92 to 1.1)	0.953	0.983	143312	2039	141273	phecode 451.2	circulatory system
Chronic liver disease and cirrhosis	OR=1 (0.91 to 1.1)	0.954	0.983	154165	1844	152321	phecode 571	digestive
Urate	0.02 (-0.731 to 0.771)	0.958	0.985	149755	NA	NA	NA	biomarker
Other specified cardiac dysrhythmias	OR=1 (0.91 to 1.09)	0.966	0.991	142197	2186	140011	phecode 427.3	circulatory system
Noninfectious gastroenteritis	OR=1 (0.95 to 1.06)	0.972	0.993	130942	5380	125562	phecode 558	digestive
Duodenal ulcer	OR=1 (0.91 to 1.1)	0.976	0.993	155249	1880	153369	phecode 531.3	digestive
Musculoskeletal symptoms referable to limbs	OR=1 (0.92 to 1.09)	0.977	0.993	157211	2450	154761	phecode 771	symptoms
Chemotherapy	OR=1 (0.95 to 1.05)	0.979	0.993	144944	7137	137807	phecode 197	neoplasms
Rheumatoid arthritis	OR=1 (0.9 to 1.11)	0.980	0.993	157211	1398	155813	phecode 714.1	musculoskeletal
Alcohol related disorders	OR=1 (0.95 to 1.05)	0.982	0.993	149004	8241	140763	phecode 317	mental disorders
Pneumococcal pneumonia	OR=1 (0.93 to 1.07)	0.987	0.994	154129	3714	150415	phecode 480.11	respiratory
Apolipoprotein B	0 (-0.003 to 0.002)	0.989	0.994	148956	NA	NA	NA	biomarker
Cerebral ischemia	OR=1 (0.92 to 1.09)	0.990	0.994	153702	2587	151115	phecode 433.3	circulatory system
Ulcer of esophagus	OR=1 (0.93 to 1.08)	0.994	0.994	139816	2931	136885	phecode 530.12	digestive
Albumin	0.0001 (-0.0132 to 0.0133)	0.994	0.994	138474	NA	NA	NA	biomarker

Bolded rows are significant adjusting for multiple hypothesis testing using Bonferroni correction

Supplementary File 1 - Table 11. Definitions for 22 health outcomes with suspected relevance with testosterone supplementation.

Trait	Field ID	ICD10 codes (if applicable)	Total Sample (Cases/Controls)	Notes
Outcomes with Expected Clinical Benefits				
All Fracture	41270; 40001; 40002	S02-; S12-; S22-; S32-; S42-; S52-; S62-; S72-; S82-; S92-; T02-	9,133/148,098	Derived from ICD-10 codes in hospital inpatient episode, and death registry records
Body Fat Percentage	23100; 23098	NA	154,095	Body fat mass divided by weight; measured by BIA
Body Fat-free Percentage	23101; 23098	NA	154,262	Body fat-free mass divided by weight; measured by BIA
Dementia	42018	NA	1,003/156,228	Derived from algorithmically-defined outcomes as recommended by the UK Biobank adjudication committee
Depression	2090; 2100; 41270	F32-; F33-; F34-; F38-; F39-	4,725/152,506	
Handgrip Strength	46; 47	NA	156,403	Average of handgrip strength for both hands
Heel Bone Mineral Density T-score	3148	NA	90,597	
Accelerometer-based Physical Activity	90012	NA	30,439	
Outcomes with Potential Adverse Effects				
All-cause Stroke	42006	NA	4,569/152,662	Derived from algorithmically-defined outcomes as recommended by the UK Biobank adjudication committee
Ischaemic Stroke	42008	NA	2,122/155,109	Derived from algorithmically-defined outcomes as recommended by the UK Biobank adjudication committee
Androgenic Alopecia	2395	NA	70,283/85,757	
Benign Prostatic Hyperplasia	41270; 40001; 40002	N40-	10,894/146,337	Derived from ICD-10 codes in hospital inpatient episode, and death registry records
Diastolic Blood Pressure	4079	NA	145,156	Average of two consecutive measures for blood pressure
Glucose	30740	NA	138,308	
Hematocrit Percentage	30030	NA	152,893	
Hemoglobin A1c	30750	NA	149,829	
Heart Failure	41270; 40001; 40002	I50-	4,288/152,943	Derived from ICD-10 codes in hospital inpatient episode, and death registry records
Prostate Cancer	70; 40001; 40002; 40	C61-	7,586/149,645	Derived from ICD-10 codes in hospital inpatient episode, death registry, and cancer registry records
Myocardial Infarction	42000	NA	9,398/147,833	Derived from algorithmically-defined outcomes as recommended by the UK Biobank adjudication committee
Systolic Blood Pressure	4080	NA	145,155	Average of two consecutive measures for blood pressure
Type 2 Diabetes	41270; 40001; 40002	E11-	11,079/146,152	Derived from ICD-10 codes in hospital inpatient episode, and death registry records
Venous Thromboembolism	41270; 40001; 40002	I26-; I80-; I81-; I82-	4,127/153,104	Derived from ICD-10 codes in hospital inpatient episode, and death registry records

Supplementary File 1 - Table 12. Definitions for 439 phenome-wide health outcomes.

Trait	Field ID	Phecode	Units	Transformation
Alanine aminotransferase	30620	NA	SD	Quantile normalized
Albumin	30600	NA	g/L	NA
Alkaline phosphatase	30610	NA	SD	Quantile normalized
Apolipoprotein A	30630	NA	g/L	NA
Apolipoprotein B	30640	NA	g/L	NA
Aspartate aminotransferase	30650	NA	SD	Quantile normalized
C-reactive protein	30710	NA	SD	Quantile normalized
Calcium	30680	NA	mmol/L	NA
Cholesterol	30690	NA	mmol/L	NA
Creatinine	30700	NA	SD	Quantile normalized
Cystatin C	30720	NA	SD	Quantile normalized
Direct bilirubin	30660	NA	SD	Quantile normalized
Gamma glutamyltransferase	30730	NA	SD	Quantile normalized
HDL cholesterol	30760	NA	SD	Quantile normalized
IGF1	30770	NA	nmol/L	NA
LDL direct	30780	NA	mmol/L	NA
Lipoprotein A	30790	NA	SD	Quantile normalized
Phosphate	30810	NA	mmol/L	NA
Total bilirubin	30840	NA	SD	Quantile normalized
Total protein	30860	NA	g/L	NA
Triglycerides	30870	NA	SD	Quantile normalized
Urate	30880	NA	umol/L	NA
Urea	30670	NA	mmol/L	NA
Vitamin D	30890	NA	SD	Quantile normalized
Intestinal infection	41270; 40006; 40001; 40002	008	NA	NA
Bacterial enteritis	41270; 40006; 40001; 40002	008.5	NA	NA
Septicemia	41270; 40006; 40001; 40002	038	NA	NA
Bacterial infection NOS	41270; 40006; 40001; 40002	041	NA	NA
Staphylococcus infections	41270; 40006; 40001; 40002	041.1	NA	NA
Streptococcus infection	41270; 40006; 40001; 40002	041.2	NA	NA
E coli	41270; 40006; 40001; 40002	041.4	NA	NA
Viral infection	41270; 40006; 40001; 40002	079	NA	NA
Postoperative infection	41270; 40006; 40001; 40002	080	NA	NA
Candidiasis	41270; 40006; 40001; 40002	112	NA	NA
Colorectal cancer	41270; 40006; 40001; 40002	153	NA	NA
Colon cancer	41270; 40006; 40001; 40002	153.2	NA	NA
Malignant neoplasm of rectum; rectosigmoid junction; and anus	41270; 40006; 40001; 40002	153.3	NA	NA
Malignant neoplasm of other and ill defined sites within the digestive organs and peritoneum	41270; 40006; 40001; 40002	159	NA	NA
Cancer within the respiratory system	41270; 40006; 40001; 40002	165	NA	NA
Cancer of bronchus; lung	41270; 40006; 40001; 40002	165.1	NA	NA
Skin cancer	41270; 40006; 40001; 40002	172	NA	NA
Melanomas of skin; dx or hx	41270; 40006; 40001; 40002	172.1	NA	NA
Melanomas of skin	41270; 40006; 40001; 40002	172.11	NA	NA
Other non epithelial cancer of skin	41270; 40006; 40001; 40002	172.2	NA	NA
Cancer of urinary organs incl kidney and bladder	41270; 40006; 40001; 40002	189	NA	NA
Cancer of bladder	41270; 40006; 40001; 40002	189.2	NA	NA
Malignant neoplasm of bladder	41270; 40006; 40001; 40002	189.21	NA	NA
Cancer; suspected or other	41270; 40006; 40001; 40002	195	NA	NA
Malignant neoplasm; other	41270; 40006; 40001; 40002	195.1	NA	NA
Chemotherapy	41270; 40006; 40001; 40002	197	NA	NA
Secondary malignant neoplasm	41270; 40006; 40001; 40002	198	NA	NA
Secondary malignancy of lymph nodes	41270; 40006; 40001; 40002	198.1	NA	NA
Secondary malignancy of respiratory organs	41270; 40006; 40001; 40002	198.2	NA	NA
Secondary malignant neoplasm of liver	41270; 40006; 40001; 40002	198.4	NA	NA

Secondary malignancy of bone	41270; 40006; 40001; 40002	198.6	NA	NA
Neoplasm of uncertain behavior	41270; 40006; 40001; 40002	199	NA	NA
Cancer of other lymphoid; histiocytic tissue	41270; 40006; 40001; 40002	202	NA	NA
Non Hodgkins lymphoma	41270; 40006; 40001; 40002	202.2	NA	NA
Leukemia	41270; 40006; 40001; 40002	204	NA	NA
Benign neoplasm of colon	41270; 40006; 40001; 40002	208	NA	NA
Benign neoplasm of other parts of digestive system	41270; 40006; 40001; 40002	211	NA	NA
Lipoma	41270; 40006; 40001; 40002	214	NA	NA
Lipoma of skin and subcutaneous tissue	41270; 40006; 40001; 40002	214.1	NA	NA
Benign neoplasm of skin	41270; 40006; 40001; 40002	216	NA	NA
Benign neoplasm of unspecified sites	41270; 40006; 40001; 40002	229	NA	NA
Hypothyroidism	41270; 40006; 40001; 40002	244	NA	NA
Hypothyroidism NOS	41270; 40006; 40001; 40002	244.4	NA	NA
Diabetes mellitus	41270; 40006; 40001; 40002	250	NA	NA
Type 1 diabetes	41270; 40006; 40001; 40002	250.1	NA	NA
Diabetic retinopathy	41270; 40006; 40001; 40002	250.7	NA	NA
Disorders of lipoid metabolism	41270; 40006; 40001; 40002	272	NA	NA
Hyperlipidemia	41270; 40006; 40001; 40002	272.1	NA	NA
Hypercholesterolemia	41270; 40006; 40001; 40002	272.11	NA	NA
Gout and other crystal arthropathies	41270; 40006; 40001; 40002	274	NA	NA
Gout	41270; 40006; 40001; 40002	274.1	NA	NA
Disorders of mineral metabolism	41270; 40006; 40001; 40002	275	NA	NA
Disorders of fluid; electrolyte; and acid base balance	41270; 40006; 40001; 40002	276	NA	NA
Electrolyte imbalance	41270; 40006; 40001; 40002	276.1	NA	NA
Hyposmolality andoror hyponatremia	41270; 40006; 40001; 40002	276.12	NA	NA
Hypovolemia	41270; 40006; 40001; 40002	276.5	NA	NA
Overweight; obesity and other hyperalimentation	41270; 40006; 40001; 40002	278	NA	NA
Obesity	41270; 40006; 40001; 40002	278.1	NA	NA
Iron deficiency anemias	41270; 40006; 40001; 40002	280	NA	NA
Iron deficiency anemias; unspecified or not due to blood loss	41270; 40006; 40001; 40002	280.1	NA	NA
Other anemias	41270; 40006; 40001; 40002	285	NA	NA
Purpura and other hemorrhagic conditions	41270; 40006; 40001; 40002	287	NA	NA
Thrombocytopenia	41270; 40006; 40001; 40002	287.3	NA	NA
Diseases of white blood cells	41270; 40006; 40001; 40002	288	NA	NA
Decreased white blood cell count	41270; 40006; 40001; 40002	288.1	NA	NA
Neutropenia	41270; 40006; 40001; 40002	288.11	NA	NA
Other diseases of blood and blood forming organs	41270; 40006; 40001; 40002	289	NA	NA
Lymphadenitis	41270; 40006; 40001; 40002	289.4	NA	NA
Delirium dementia and amnestic and other cognitive disorders	41270; 40006; 40001; 40002	290	NA	NA
Neurological disorders	41270; 40006; 40001; 40002	292	NA	NA
Altered mental status	41270; 40006; 40001; 40002	292.4	NA	NA
Symptoms involving head and neck	41270; 40006; 40001; 40002	293	NA	NA
Anxiety disorders	41270; 40006; 40001; 40002	300	NA	NA
Anxiety disorder	41270; 40006; 40001; 40002	300.1	NA	NA
Other mental disorder	41270; 40006; 40001; 40002	306	NA	NA
Substance addiction and disorders	41270; 40006; 40001; 40002	316	NA	NA
Alcohol related disorders	41270; 40006; 40001; 40002	317	NA	NA
Alcoholism	41270; 40006; 40001; 40002	317.1	NA	NA
Tobacco use disorder	41270; 40006; 40001; 40002	318	NA	NA
Sleep disorders	41270; 40006; 40001; 40002	327	NA	NA
Sleep apnea	41270; 40006; 40001; 40002	327.3	NA	NA
Degenerative disease of the spinal cord	41270; 40006; 40001; 40002	334	NA	NA
Other headache syndromes	41270; 40006; 40001; 40002	339	NA	NA
Hemiplegia	41270; 40006; 40001; 40002	342	NA	NA
Epilepsy; recurrent seizures; convulsions	41270; 40006; 40001; 40002	345	NA	NA

Convulsions	41270; 40006; 40001; 40002	345.3	NA	NA
Abnormal movement	41270; 40006; 40001; 40002	350	NA	NA
Abnormality of gait	41270; 40006; 40001; 40002	350.2	NA	NA
Other peripheral nerve disorders	41270; 40006; 40001; 40002	351	NA	NA
Nerve root and plexus disorders	41270; 40006; 40001; 40002	353	NA	NA
Inflammatory and toxic neuropathy	41270; 40006; 40001; 40002	357	NA	NA
Retinal detachments and defects	41270; 40006; 40001; 40002	361	NA	NA
Retinal detachment with retinal defect	41270; 40006; 40001; 40002	361.1	NA	NA
Other retinal disorders	41270; 40006; 40001; 40002	362	NA	NA
Degeneration of macula and posterior pole of retina	41270; 40006; 40001; 40002	362.2	NA	NA
Macular degeneration senile of retina NOS	41270; 40006; 40001; 40002	362.29	NA	NA
Hereditary retinal dystrophies	41270; 40006; 40001; 40002	362.7	NA	NA
Glaucoma	41270; 40006; 40001; 40002	365	NA	NA
Cataract	41270; 40006; 40001; 40002	366	NA	NA
Senile cataract	41270; 40006; 40001; 40002	366.2	NA	NA
Disorders of refraction and accommodation; blindness and low vision	41270; 40006; 40001; 40002	367	NA	NA
Visual disturbances	41270; 40006; 40001; 40002	368	NA	NA
Inflammation of the eye	41270; 40006; 40001; 40002	371	NA	NA
Inflammation of eyelids	41270; 40006; 40001; 40002	371.3	NA	NA
Other disorders of eyelids	41270; 40006; 40001; 40002	374	NA	NA
Other disorders of eye	41270; 40006; 40001; 40002	379	NA	NA
Aphakia and other disorders of lens	41270; 40006; 40001; 40002	379.3	NA	NA
Otitis media and Eustachian tube disorders	41270; 40006; 40001; 40002	381	NA	NA
Vertiginous syndromes and other disorders of vestibular system	41270; 40006; 40001; 40002	386	NA	NA
Dizziness and giddiness Light headedness and vertigo	41270; 40006; 40001; 40002	386.9	NA	NA
Hearing loss	41270; 40006; 40001; 40002	389	NA	NA
Rheumatic disease of the heart valves	41270; 40006; 40001; 40002	394	NA	NA
Mitral valve disease	41270; 40006; 40001; 40002	394.2	NA	NA
Aortic valve disease	41270; 40006; 40001; 40002	394.3	NA	NA
Heart valve disorders	41270; 40006; 40001; 40002	395	NA	NA
Nonrheumatic mitral valve disorders	41270; 40006; 40001; 40002	395.1	NA	NA
Nonrheumatic aortic valve disorders	41270; 40006; 40001; 40002	395.2	NA	NA
Heart valve replaced	41270; 40006; 40001; 40002	395.6	NA	NA
Abnormal heart sounds	41270; 40006; 40001; 40002	396	NA	NA
Hypertension	41270; 40006; 40001; 40002	401	NA	NA
Essential hypertension	41270; 40006; 40001; 40002	401.1	NA	NA
Hypertensive heart andoror renal disease	41270; 40006; 40001; 40002	401.2	NA	NA
Ischemic Heart Disease	41270; 40006; 40001; 40002	411	NA	NA
Unstable angina intermediate coronary syndrome	41270; 40006; 40001; 40002	411.1	NA	NA
Angina pectoris	41270; 40006; 40001; 40002	411.3	NA	NA
Coronary atherosclerosis	41270; 40006; 40001; 40002	411.4	NA	NA
Other chronic ischemic heart disease; unspecified	41270; 40006; 40001; 40002	411.8	NA	NA
Other acute and subacute forms of ischemic heart disease	41270; 40006; 40001; 40002	411.9	NA	NA
Other forms of chronic heart disease	41270; 40006; 40001; 40002	414	NA	NA
Pulmonary heart disease	41270; 40006; 40001; 40002	415	NA	NA
Acute pulmonary heart disease	41270; 40006; 40001; 40002	415.1	NA	NA
Pulmonary embolism and infarction; acute	41270; 40006; 40001; 40002	415.11	NA	NA
Cardiomegaly	41270; 40006; 40001; 40002	416	NA	NA
Nonspecific chest pain	41270; 40006; 40001; 40002	418	NA	NA
Precordial pain	41270; 40006; 40001; 40002	418.1	NA	NA
Carditis	41270; 40006; 40001; 40002	420	NA	NA
Pericarditis	41270; 40006; 40001; 40002	420.2	NA	NA
Cardiac conduction disorders	41270; 40006; 40001; 40002	426	NA	NA
Atrioventricular AV block	41270; 40006; 40001; 40002	426.2	NA	NA
First degree AV block	41270; 40006; 40001; 40002	426.21	NA	NA

Bundle branch block	41270; 40006; 40001; 40002	426.3	NA	NA
Right bundle branch block	41270; 40006; 40001; 40002	426.31	NA	NA
Left bundle branch block	41270; 40006; 40001; 40002	426.32	NA	NA
Cardiac pacemakerordevice in situ	41270; 40006; 40001; 40002	426.9	NA	NA
Cardiac pacemaker in situ	41270; 40006; 40001; 40002	426.91	NA	NA
Cardiac dysrhythmias	41270; 40006; 40001; 40002	427	NA	NA
Paroxysmal tachycardia; unspecified	41270; 40006; 40001; 40002	427.1	NA	NA
Paroxysmal supraventricular tachycardia	41270; 40006; 40001; 40002	427.11	NA	NA
Atrial fibrillation and flutter	41270; 40006; 40001; 40002	427.2	NA	NA
Other specified cardiac dysrhythmias	41270; 40006; 40001; 40002	427.3	NA	NA
Cardiac arrest and ventricular fibrillation	41270; 40006; 40001; 40002	427.4	NA	NA
Tachycardia NOS	41270; 40006; 40001; 40002	427.7	NA	NA
Palpitations	41270; 40006; 40001; 40002	427.9	NA	NA
Cerebrovascular disease	41270; 40006; 40001; 40002	433	NA	NA
Occlusion and stenosis of precerebral arteries	41270; 40006; 40001; 40002	433.1	NA	NA
Occlusion of cerebral arteries	41270; 40006; 40001; 40002	433.2	NA	NA
Cerebral artery occlusion; with cerebral infarction	41270; 40006; 40001; 40002	433.21	NA	NA
Cerebral ischemia	41270; 40006; 40001; 40002	433.3	NA	NA
Transient cerebral ischemia	41270; 40006; 40001; 40002	433.31	NA	NA
Atherosclerosis	41270; 40006; 40001; 40002	440	NA	NA
Other aneurysm	41270; 40006; 40001; 40002	442	NA	NA
Aortic aneurysm	41270; 40006; 40001; 40002	442.1	NA	NA
Abdominal aortic aneurysm	41270; 40006; 40001; 40002	442.11	NA	NA
Peripheral vascular disease	41270; 40006; 40001; 40002	443	NA	NA
Other specified peripheral vascular diseases	41270; 40006; 40001; 40002	443.8	NA	NA
Peripheral vascular disease; unspecified	41270; 40006; 40001; 40002	443.9	NA	NA
Other disorders of arteries and arterioles	41270; 40006; 40001; 40002	447	NA	NA
Phlebitis and thrombophlebitis	41270; 40006; 40001; 40002	451	NA	NA
Phlebitis and thrombophlebitis of lower extremities	41270; 40006; 40001; 40002	451.2	NA	NA
Varicose veins	41270; 40006; 40001; 40002	454	NA	NA
Varicose veins of lower extremity	41270; 40006; 40001; 40002	454.1	NA	NA
Hemorrhoids	41270; 40006; 40001; 40002	455	NA	NA
Hypotension	41270; 40006; 40001; 40002	458	NA	NA
Orthostatic hypotension	41270; 40006; 40001; 40002	458.1	NA	NA
Hypotension NOS	41270; 40006; 40001; 40002	458.9	NA	NA
Other disorders of circulatory system	41270; 40006; 40001; 40002	459	NA	NA
Circulatory disease NEC	41270; 40006; 40001; 40002	459.9	NA	NA
Acute upper respiratory infections of multiple or unspecified sites	41270; 40006; 40001; 40002	465	NA	NA
Septal DeviationsorTurbinate Hypertrophy	41270; 40006; 40001; 40002	470	NA	NA
Nasal polyps	41270; 40006; 40001; 40002	471	NA	NA
Diseases of the larynx and vocal cords	41270; 40006; 40001; 40002	473	NA	NA
Chronic sinusitis	41270; 40006; 40001; 40002	475	NA	NA
Epistaxis or throat hemorrhage	41270; 40006; 40001; 40002	477	NA	NA
Other upper respiratory disease	41270; 40006; 40001; 40002	479	NA	NA
Pneumonia	41270; 40006; 40001; 40002	480	NA	NA
Bacterial pneumonia	41270; 40006; 40001; 40002	480.1	NA	NA
Pneumococcal pneumonia	41270; 40006; 40001; 40002	480.11	NA	NA
Asthma	41270; 40006; 40001; 40002	495	NA	NA
Chronic airway obstruction	41270; 40006; 40001; 40002	496	NA	NA
Emphysema	41270; 40006; 40001; 40002	496.1	NA	NA
Chronic bronchitis	41270; 40006; 40001; 40002	496.2	NA	NA
Obstructive chronic bronchitis	41270; 40006; 40001; 40002	496.21	NA	NA
Bronchiectasis	41270; 40006; 40001; 40002	496.3	NA	NA
Pleurisy; pleural effusion	41270; 40006; 40001; 40002	507	NA	NA
Pulmonary collapse; interstitial and compensatory emphysema	41270; 40006; 40001; 40002	508	NA	NA

Respiratory failure; insufficiency; arrest	41270; 40006; 40001; 40002	509	NA	NA
Respiratory failure	41270; 40006; 40001; 40002	509.1	NA	NA
Respiratory insufficiency	41270; 40006; 40001; 40002	509.2	NA	NA
Other symptoms of respiratory system	41270; 40006; 40001; 40002	512	NA	NA
Shortness of breath	41270; 40006; 40001; 40002	512.7	NA	NA
Cough	41270; 40006; 40001; 40002	512.8	NA	NA
Respiratory abnormalities	41270; 40006; 40001; 40002	513	NA	NA
Abnormal findings examination of lungs	41270; 40006; 40001; 40002	514	NA	NA
Abnormal sputum	41270; 40006; 40001; 40002	516	NA	NA
Hemoptysis	41270; 40006; 40001; 40002	516.1	NA	NA
Other diseases of respiratory system; not elsewhere classified	41270; 40006; 40001; 40002	519	NA	NA
Other diseases of respiratory system; NEC	41270; 40006; 40001; 40002	519.8	NA	NA
Diseases of hard tissues of teeth	41270; 40006; 40001; 40002	521	NA	NA
Dental caries	41270; 40006; 40001; 40002	521.1	NA	NA
Other diseases of the teeth and supporting structures	41270; 40006; 40001; 40002	525	NA	NA
Diseases of the oral soft tissues; excluding lesions specific for gingiva and tongue	41270; 40006; 40001; 40002	528	NA	NA
Diseases of esophagus	41270; 40006; 40001; 40002	530	NA	NA
Esophagitis; GERD and related diseases	41270; 40006; 40001; 40002	530.1	NA	NA
GERD	41270; 40006; 40001; 40002	530.11	NA	NA
Ulcer of esophagus	41270; 40006; 40001; 40002	530.12	NA	NA
Reflux esophagitis	41270; 40006; 40001; 40002	530.14	NA	NA
Esophageal bleeding varicesorhemorrhage	41270; 40006; 40001; 40002	530.2	NA	NA
Stricture and stenosis of esophagus	41270; 40006; 40001; 40002	530.3	NA	NA
Peptic ulcer excl esophageal	41270; 40006; 40001; 40002	531	NA	NA
Gastric ulcer	41270; 40006; 40001; 40002	531.2	NA	NA
Duodenal ulcer	41270; 40006; 40001; 40002	531.3	NA	NA
Dysphagia	41270; 40006; 40001; 40002	532	NA	NA
Gastritis and duodenitis	41270; 40006; 40001; 40002	535	NA	NA
Duodenitis	41270; 40006; 40001; 40002	535.6	NA	NA
Other specified gastritis	41270; 40006; 40001; 40002	535.8	NA	NA
Other disorders of stomach and duodenum	41270; 40006; 40001; 40002	537	NA	NA
Appendiceal conditions	41270; 40006; 40001; 40002	540	NA	NA
Appendicitis	41270; 40006; 40001; 40002	540.1	NA	NA
Acute appendicitis	41270; 40006; 40001; 40002	540.11	NA	NA
Abdominal hernia	41270; 40006; 40001; 40002	550	NA	NA
Inguinal hernia	41270; 40006; 40001; 40002	550.1	NA	NA
Diaphragmatic hernia	41270; 40006; 40001; 40002	550.2	NA	NA
Umbilical hernia	41270; 40006; 40001; 40002	550.4	NA	NA
Ventral hernia	41270; 40006; 40001; 40002	550.5	NA	NA
Inflammatory bowel disease and other gastroenteritis and colitis	41270; 40006; 40001; 40002	555	NA	NA
Ulcerative colitis	41270; 40006; 40001; 40002	555.2	NA	NA
Noninfectious gastroenteritis	41270; 40006; 40001; 40002	558	NA	NA
Ileostomy status	41270; 40006; 40001; 40002	559	NA	NA
Intestinal obstruction without mention of hernia	41270; 40006; 40001; 40002	560	NA	NA
Other intestinal obstruction	41270; 40006; 40001; 40002	560.4	NA	NA
Symptoms involving digestive system	41270; 40006; 40001; 40002	561	NA	NA
Diverticulosis and diverticulitis	41270; 40006; 40001; 40002	562	NA	NA
Diverticulosis	41270; 40006; 40001; 40002	562.1	NA	NA
Constipation	41270; 40006; 40001; 40002	563	NA	NA
Functional digestive disorders	41270; 40006; 40001; 40002	564	NA	NA
Irritable Bowel Syndrome	41270; 40006; 40001; 40002	564.1	NA	NA
Personal history of diseases of digestive system	41270; 40006; 40001; 40002	564.9	NA	NA
Anal and rectal conditions	41270; 40006; 40001; 40002	565	NA	NA
Anal and rectal polyp	41270; 40006; 40001; 40002	565.1	NA	NA

Other disorders of peritoneum	41270; 40006; 40001; 40002	568	NA	NA
Peritoneal adhesions postoperative postinfection	41270; 40006; 40001; 40002	568.1	NA	NA
Other disorders of intestine	41270; 40006; 40001; 40002	569	NA	NA
Chronic liver disease and cirrhosis	41270; 40006; 40001; 40002	571	NA	NA
Other chronic nonalcoholic liver disease	41270; 40006; 40001; 40002	571.5	NA	NA
Other disorders of liver	41270; 40006; 40001; 40002	573	NA	NA
Abnormal results of function study of liver	41270; 40006; 40001; 40002	573.7	NA	NA
Cholelithiasis and cholecystitis	41270; 40006; 40001; 40002	574	NA	NA
Cholelithiasis	41270; 40006; 40001; 40002	574.1	NA	NA
Cholelithiasis with other cholecystitis	41270; 40006; 40001; 40002	574.12	NA	NA
Calculus of bile duct	41270; 40006; 40001; 40002	574.2	NA	NA
Cholecystitis without cholelithiasis	41270; 40006; 40001; 40002	574.3	NA	NA
Other biliary tract disease	41270; 40006; 40001; 40002	575	NA	NA
Other disorders of biliary tract	41270; 40006; 40001; 40002	575.8	NA	NA
Diseases of pancreas	41270; 40006; 40001; 40002	577	NA	NA
Acute pancreatitis	41270; 40006; 40001; 40002	577.1	NA	NA
Gastrointestinal hemorrhage	41270; 40006; 40001; 40002	578	NA	NA
Hematemesis	41270; 40006; 40001; 40002	578.1	NA	NA
Blood in stool	41270; 40006; 40001; 40002	578.2	NA	NA
Hemorrhage of rectum and anus	41270; 40006; 40001; 40002	578.8	NA	NA
Hemorrhage of gastrointestinal tract	41270; 40006; 40001; 40002	578.9	NA	NA
Other symptoms involving abdomen and pelvis	41270; 40006; 40001; 40002	579	NA	NA
Renal failure	41270; 40006; 40001; 40002	585	NA	NA
Acute renal failure	41270; 40006; 40001; 40002	585.1	NA	NA
Renal failure NOS	41270; 40006; 40001; 40002	585.2	NA	NA
Chronic renal failure CKD	41270; 40006; 40001; 40002	585.3	NA	NA
Other disorders of the kidney and ureters	41270; 40006; 40001; 40002	586	NA	NA
Urinary tract infection	41270; 40006; 40001; 40002	591	NA	NA
Cystitis and urethritis	41270; 40006; 40001; 40002	592	NA	NA
Cystitis	41270; 40006; 40001; 40002	592.1	NA	NA
Hematuria	41270; 40006; 40001; 40002	593	NA	NA
Urinary calculus	41270; 40006; 40001; 40002	594	NA	NA
Calculus of kidney	41270; 40006; 40001; 40002	594.1	NA	NA
Calculus of ureter	41270; 40006; 40001; 40002	594.3	NA	NA
Renal colic	41270; 40006; 40001; 40002	594.8	NA	NA
Hydronephrosis	41270; 40006; 40001; 40002	595	NA	NA
Other disorders of bladder	41270; 40006; 40001; 40002	596	NA	NA
Bladder neck obstruction	41270; 40006; 40001; 40002	596.1	NA	NA
Other disorders of urethra and urinary tract	41270; 40006; 40001; 40002	597	NA	NA
Urethral stricture not specified as infectious	41270; 40006; 40001; 40002	597.1	NA	NA
Abnormal findings on examination of urine	41270; 40006; 40001; 40002	598	NA	NA
Other symptomsordisorders or the urinary system	41270; 40006; 40001; 40002	599	NA	NA
Urinary obstruction	41270; 40006; 40001; 40002	599.1	NA	NA
Retention of urine	41270; 40006; 40001; 40002	599.2	NA	NA
Urinary incontinence	41270; 40006; 40001; 40002	599.4	NA	NA
Frequency of urination and polyuria	41270; 40006; 40001; 40002	599.5	NA	NA
Other abnormality of urination	41270; 40006; 40001; 40002	599.9	NA	NA
Inflammatory diseases of prostate	41270; 40006; 40001; 40002	601	NA	NA
Prostatitis	41270; 40006; 40001; 40002	601.1	NA	NA
Other disorders of prostate	41270; 40006; 40001; 40002	602	NA	NA
Other disorders of testis	41270; 40006; 40001; 40002	603	NA	NA
Hydrocele	41270; 40006; 40001; 40002	603.1	NA	NA
Disorders of penis	41270; 40006; 40001; 40002	604	NA	NA
Redundant prepuce and phimosisorBXO	41270; 40006; 40001; 40002	604.1	NA	NA
Other disorders of male genital organs	41270; 40006; 40001; 40002	608	NA	NA
Superficial cellulitis and abscess	41270; 40006; 40001; 40002	681	NA	NA
Cellulitis and abscess of armorhand	41270; 40006; 40001; 40002	681.3	NA	NA

Cellulitis and abscess of leg; except foot	41270; 40006; 40001; 40002	681.5	NA	NA
Cellulitis and abscess of foot; toe	41270; 40006; 40001; 40002	681.6	NA	NA
Other local infections of skin and subcutaneous tissue	41270; 40006; 40001; 40002	686	NA	NA
Carbuncle and furuncle	41270; 40006; 40001; 40002	686.1	NA	NA
Symptoms affecting skin	41270; 40006; 40001; 40002	687	NA	NA
Rash and other nonspecific skin eruption	41270; 40006; 40001; 40002	687.1	NA	NA
Localized superficial swelling; mass; or lump	41270; 40006; 40001; 40002	687.2	NA	NA
Disturbance of skin sensation	41270; 40006; 40001; 40002	687.4	NA	NA
Disorder of skin and subcutaneous tissue NOS	41270; 40006; 40001; 40002	689	NA	NA
Erythematous conditions	41270; 40006; 40001; 40002	695	NA	NA
Psoriasis and related disorders	41270; 40006; 40001; 40002	696	NA	NA
Psoriasis	41270; 40006; 40001; 40002	696.4	NA	NA
Psoriasis vulgaris	41270; 40006; 40001; 40002	696.41	NA	NA
Other hypertrophic and atrophic conditions of skin	41270; 40006; 40001; 40002	701	NA	NA
Degenerative skin conditions and other dermatoses	41270; 40006; 40001; 40002	702	NA	NA
Actinic keratosis	41270; 40006; 40001; 40002	702.1	NA	NA
Seborrheic keratosis	41270; 40006; 40001; 40002	702.2	NA	NA
Diseases of hair and hair follicles	41270; 40006; 40001; 40002	704	NA	NA
Diseases of sebaceous glands	41270; 40006; 40001; 40002	706	NA	NA
Sebaceous cyst	41270; 40006; 40001; 40002	706.2	NA	NA
Chronic ulcer of skin	41270; 40006; 40001; 40002	707	NA	NA
Diffuse diseases of connective tissue	41270; 40006; 40001; 40002	709	NA	NA
Unspecified diffuse connective tissue disease	41270; 40006; 40001; 40002	709.7	NA	NA
Rheumatoid arthritis and other inflammatory polyarthropathies	41270; 40006; 40001; 40002	714	NA	NA
Rheumatoid arthritis	41270; 40006; 40001; 40002	714.1	NA	NA
Other inflammatory spondylopathies	41270; 40006; 40001; 40002	715	NA	NA
Other arthropathies	41270; 40006; 40001; 40002	716	NA	NA
Unspecified monoarthritis	41270; 40006; 40001; 40002	716.2	NA	NA
Arthropathy NOS	41270; 40006; 40001; 40002	716.9	NA	NA
Spinal stenosis	41270; 40006; 40001; 40002	720	NA	NA
Spondylosis and allied disorders	41270; 40006; 40001; 40002	721	NA	NA
Spondylosis without myelopathy	41270; 40006; 40001; 40002	721.1	NA	NA
Intervertebral disc disorders	41270; 40006; 40001; 40002	722	NA	NA
Degeneration of intervertebral disc	41270; 40006; 40001; 40002	722.6	NA	NA
Other and unspecified disc disorder	41270; 40006; 40001; 40002	722.9	NA	NA
Peripheral enthesopathies and allied syndromes	41270; 40006; 40001; 40002	726	NA	NA
Enthesopathy	41270; 40006; 40001; 40002	726.1	NA	NA
Other disorders of synovium; tendon; and bursa	41270; 40006; 40001; 40002	727	NA	NA
Synovitis and tenosynovitis	41270; 40006; 40001; 40002	727.1	NA	NA
Ganglion and cyst of synovium; tendon; and bursa	41270; 40006; 40001; 40002	727.4	NA	NA
Disorders of muscle; ligament; and fascia	41270; 40006; 40001; 40002	728	NA	NA
Fasciitis	41270; 40006; 40001; 40002	728.7	NA	NA
Contracture of palmar fascia Dupuytrens disease	41270; 40006; 40001; 40002	728.71	NA	NA
Other disorders of soft tissues	41270; 40006; 40001; 40002	729	NA	NA
Other disorders of bone and cartilage	41270; 40006; 40001; 40002	733	NA	NA
Acquired foot deformities	41270; 40006; 40001; 40002	735	NA	NA
Acquired toe deformities	41270; 40006; 40001; 40002	735.2	NA	NA
Other acquired deformities of limbs	41270; 40006; 40001; 40002	736	NA	NA
Osteoarthrosis	41270; 40006; 40001; 40002	740	NA	NA
Osteoarthritis; localized	41270; 40006; 40001; 40002	740.1	NA	NA
Osteoarthrosis; localized; primary	41270; 40006; 40001; 40002	740.11	NA	NA
Osteoarthrosis NOS	41270; 40006; 40001; 40002	740.9	NA	NA
Symptoms and disorders of the joints	41270; 40006; 40001; 40002	741	NA	NA
Derangement of joint; non traumatic	41270; 40006; 40001; 40002	742	NA	NA
Other derangement of joint	41270; 40006; 40001; 40002	742.9	NA	NA
Osteoporosis; osteopenia and pathological fracture	41270; 40006; 40001; 40002	743	NA	NA

Osteoporosis	41270; 40006; 40001; 40002	743.1	NA	NA
Osteoporosis NOS	41270; 40006; 40001; 40002	743.11	NA	NA
Pain in joint	41270; 40006; 40001; 40002	745	NA	NA
Cardiac and circulatory congenital anomalies	41270; 40006; 40001; 40002	747	NA	NA
Cardiac congenital anomalies	41270; 40006; 40001; 40002	747.1	NA	NA
Congenital anomalies of great vessels	41270; 40006; 40001; 40002	747.13	NA	NA
Back pain	41270; 40006; 40001; 40002	760	NA	NA
Sciatica	41270; 40006; 40001; 40002	764	NA	NA
Musculoskeletal symptoms referable to limbs	41270; 40006; 40001; 40002	771	NA	NA
Swelling of limb	41270; 40006; 40001; 40002	771.1	NA	NA
Pain in limb	41270; 40006; 40001; 40002	773	NA	NA
Symptoms involving nervous and musculoskeletal systems	41270; 40006; 40001; 40002	781	NA	NA
Symptoms involving skin and other integumentary tissue	41270; 40006; 40001; 40002	782	NA	NA
Edema	41270; 40006; 40001; 40002	782.3	NA	NA
Fever of unknown origin	41270; 40006; 40001; 40002	783	NA	NA
Abdominal pain	41270; 40006; 40001; 40002	785	NA	NA
Syncope and collapse	41270; 40006; 40001; 40002	788	NA	NA
Nausea and vomiting	41270; 40006; 40001; 40002	789	NA	NA
Nonspecific findings on examination of blood	41270; 40006; 40001; 40002	790	NA	NA
Other abnormal blood chemistry	41270; 40006; 40001; 40002	790.6	NA	NA
Malaise and fatigue	41270; 40006; 40001; 40002	798	NA	NA
Fracture of lower limb	41270; 40006; 40001; 40002	800	NA	NA
Fracture of unspecified part of femur	41270; 40006; 40001; 40002	800.2	NA	NA
Fracture of tibia and fibula	41270; 40006; 40001; 40002	800.3	NA	NA
Fracture of ankle and foot	41270; 40006; 40001; 40002	801	NA	NA
Fracture of upper limb	41270; 40006; 40001; 40002	803	NA	NA
Fracture of radius and ulna	41270; 40006; 40001; 40002	803.2	NA	NA
Fracture of clavicle or scapula	41270; 40006; 40001; 40002	803.3	NA	NA
Fracture of hand or wrist	41270; 40006; 40001; 40002	804	NA	NA
Fracture of vertebral column without mention of spinal cord injury	41270; 40006; 40001; 40002	805	NA	NA
Fracture of ribs	41270; 40006; 40001; 40002	807	NA	NA
Skull and face fracture and other intercranial injury	41270; 40006; 40001; 40002	819	NA	NA
Dislocation	41270; 40006; 40001; 40002	830	NA	NA
Internal derangement of knee	41270; 40006; 40001; 40002	835	NA	NA
Hemorrhage or hematoma complicating a procedure	41270; 40006; 40001; 40002	850	NA	NA
Complications of transplants and reattached limbs	41270; 40006; 40001; 40002	851	NA	NA
Complications of cardiacorvascular device; implant; and graft	41270; 40006; 40001; 40002	854	NA	NA
Complication of internal orthopedic device	41270; 40006; 40001; 40002	858	NA	NA
Complication due to other implant and internal device	41270; 40006; 40001; 40002	859	NA	NA
Open wounds of head; neck; and trunk	41270; 40006; 40001; 40002	870	NA	NA
Other open wound of head and face	41270; 40006; 40001; 40002	870.3	NA	NA
Open wounds of extremities	41270; 40006; 40001; 40002	871	NA	NA
Superficial injury without mention of infection	41270; 40006; 40001; 40002	915	NA	NA
Dermatitis due to solar radiation	41270; 40006; 40001; 40002	938	NA	NA
Chronic dermatitis due to solar radiation	41270; 40006; 40001; 40002	938.2	NA	NA
Atopicorcontact dermatitis due to other or unspecified	41270; 40006; 40001; 40002	939	NA	NA
Poisoning by antibiotics	41270; 40006; 40001; 40002	960	NA	NA
Allergyoradverse effect of penicillin	41270; 40006; 40001; 40002	960.2	NA	NA
Poisoning by primarily systemic agents	41270; 40006; 40001; 40002	963	NA	NA
Poisoning by analgesics; antipyretics; and antirheumatics	41270; 40006; 40001; 40002	965	NA	NA
Effects radiation NOS	41270; 40006; 40001; 40002	990	NA	NA
Sepsis and SIRS	41270; 40006; 40001; 40002	994	NA	NA
Sepsis	41270; 40006; 40001; 40002	994.2	NA	NA

APPENDIX C:

SUPPLEMENTARY DATA FOR CHAPTER 5

Supplemental Methods

Genotyping Quality Control and Imputation in ORIGIN

Poor quality SNPs were excluded using standard thresholds for low call rate (<99%), deviation from Hardy-Weinberg equilibrium ($P<1x10^{-6}$), and low minor allele frequency (<0.01). Poor quality samples were excluded using standard measures for low call rate (<99%), sex or ethnicity mismatches, and cryptic relatedness. Samples belonging to ethnicities with small sample sizes were excluded (n<500). Quality control was done using PLINK (1) or GCTA (2) software. Imputation was performed using the 1000 Genomes Project (3) as a reference panel and IMPUTE2 for software (4) resulting in complete coverage of the genome with over 30 million SNPs. INFO score was calculated by IMPUTE2, and a cut-off (<0.7) was used to remove SNPs with low confidence (4).

Study outcomes – ORIGIN

GFR was estimated based on serum creatinine measurements using the CKD-Epidemiology collaboration (CKD-EPI) equation (5). Albuminuria categories were defined to be three mutually exclusive classifications for urinary ACR corresponding to normoalbuminuria (less than 30 mg/g [3.4 mg/mmol]), elevated albuminuria (between 30 mg/g [3.4 mg/mmol] and 300 mg/g [33.9 mg/mmol]), or highly elevated (greater than 300 mg/g [33.9 mg/mmol]). Patients with an $eGFR_{crea}$ < 60 mL/min per $1.73m^2$ were classified as having stage 3 CKD as per the 2013 Kidney Disease Improving Global Outcomes (6). ESRD was defined as an $eGFR_{crea}$ <15 mL/min per $1.73m^2$, chronic dialysis or having had a renal transplant on follow-up (6). Prior renal disease was defined as either $eGFR_{crea}$ < 60 mL/min per 1.73 m^2, or elevated albuminuria, or highly elevated albuminuria at baseline.

Sensitivity Analyses for Reverse MR

Validation was performed using a “leave-one-out” approach in which the MR analysis was repeated after excluding each of the SNPs. Non-significant biomarker associations resulting from such an approach suggest that the excluded SNP disproportionally contributed to the originally detected association; these biomarkers were flagged as

potential false positives. To identify whether any of the significant biomarkers were affected by alternative measures of kidney function, the reverse MR analysis was also performed using genetic variants associated with estimated GFR based on cystatin-C ($eGFR_{cys}$) and CKD itself. In the CKDGen Consortium, 5 variants were associated with CKD (7), and 5 variants were associated with $eGFR_{cys}$, of which 1 variant was located in the *CST3* gene encoding cystatin-C and excluded leaving 4 variants for subsequent analyses. This variant (rs3827143) likely doesn't reflect kidney function since it affects $eGFR_{cys}$ by altering cystatin-C rather than kidney filtration.

To determine whether identified biomarkers for kidney function were independent of established renal risk factors, we repeated the reverse MR technique using genetic variants associated at GWS level ($P<5x10^{-8}$) with established CKD risk factors instead of with $eGFR_{crea}$. These genetic associations were obtained from summary statistics of publicly available consortium datasets of risk factors including: type 2 diabetes (8); systolic blood pressure, diastolic blood pressure, hypertension (9); coronary artery disease, myocardial infarction (10); body mass index (11), waist-to-hip ratio adjusted for BMI (12); fasting glucose, fasting insulin (13), glycated hemoglobin (14); HDL cholesterol, LDL cholesterol, total cholesterol, triglycerides (15), as well as nominally significant variants ($P<1x10^{-6}$) for urinary ACR (16). ORIGIN was not part of these consortia. The reverse MR analysis was performed separately for each of these risk factors, representing the effect of that risk factor on the biomarker.

Identification of SNPs for Traditional MR

Observed regional associations located hundreds of kilobases from known loci required selection of a 300 Kb window (17). Therefore, we selected SNPs located within 300 Kb of the gene encoding any significant biomarkers from the reverse MR, based on the Reference Sequence gene list on the University of California, Santa Cruz Genome Table Browser (18). Gene names were identified using the GeneCards Encyclopedia (19). SNPs with minor allele frequency below 0.05 or absent from the CKDGen database were removed (7).

To arrive at an independent set of SNPs, they were pruned for linkage disequilibrium at a threshold of $r^2<0.1$ using 1000 Genomes data in Europeans. Genetic association of these independent SNPs with $eGFR_{crea}$ was determined from the publicly-available summary statistics of the CKDGen consortium (7). Similar to the reverse MR, a two-sample MR analysis was then performed using these independent SNPs for each biomarker. However, in this analysis, the independent variable was the effect of SNPs on the biomarker, while the dependent variable was the effect of SNPs on $eGFR_{crea}$.

Association of TFF3 in the full ORIGIN biomarker sub-study

In the full ORIGIN biomarker sub-study, there were 8,197 participants with available ethnicity information, of whom 34% had prior renal disease (defined as either $eGFR_{crea}$ < 60 mL/min per 1.73 m^2, or microalbuminuria, or macroalbuminuria) and a mean $eGFR_{crea}$ at baseline of 75.0 mL/min per 1.73 m^2. To confirm our reverse MR and epidemiological associations for incident CKD held true with baseline renal characteristics, we investigated the association of the significant biomarkers with $eGFR_{crea}$ and prior renal disease at baseline. For these ancillary analyses we used all participants in the ORIGIN biomarker sub-study with ethnicity information available. Linear regression models were used to evaluate the relationship between $eGFR_{crea}$ as well as prior renal disease at baseline (independent variable) on each significant biomarker (dependent variable) identified in the reverse MR analysis. Both models were adjusted for age, sex, and ethnicity.

Supplemental References

1. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: A tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007;81:559–75.
2. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: A tool for genome-wide complex trait analysis. Am J Hum Genet 2011;88:76–82.
3. Gibbs RA, Boerwinkle E, Doddapaneni H, Han Y, Korchina V, Kovar C, et al. A global reference for human genetic variation. Nature 2015;526:68–74.
4. Howie BN, Donnelly P, Marchini J. A Flexible and Accurate Genotype Imputation Method for the Next Generation of Genome-Wide Association Studies. Schork NJ, editor. PLoS Genet 2009;5:e1000529.
5. Levey AS, Stevens LA, Schmid CH, Zhang YL, Castro AF, Feldman HI, et al. A

new equation to estimate glomerular filtration rate. Ann Intern Med 2009;150:604–12.
6. Group KDIGO (KDIGO) CW. KDIGO Clinical Practice Guideline for the Evaluation and Management of Chronic Kidney Disease. Kidney Int Suppl 2013;3:1–150.
7. Pattaro C, Teumer A, Gorski M, Chu AY, Li M, Mijatovic V, et al. Genetic associations at 53 loci highlight cell types and biological pathways relevant for kidney function. Nat Commun 2016;7:10023.
8. Mahajan A, Go MJ, Zhang W, Below JE, Gaulton KJ, Ferreira T, et al. Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. Nat Genet 2014;46:234–44.
9. Ehret GB, Munroe PB, Rice KM, Bochud M, Johnson AD, Chasman DI, et al. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. Nature 2011;478:103–9.
10. Nikpay M, Goel A, Won H-H, Hall LM, Willenborg C, Kanoni S, et al. A comprehensive 1000 Genomes–based genome-wide association meta-analysis of coronary artery disease. Nat Genet 2015;47:1121–30.
11. Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR, et al. Genetic studies of body mass index yield new insights for obesity biology. Nature 2015;518:197–206.
12. Shungin D, Winkler TW, Croteau-Chonka DC, Ferreira T, Locke AE, Mägi R, et al. New genetic loci link adipose and insulin biology to body fat distribution. Nature 2015;518:187–96.
13. Manning AK, Hivert M-F, Scott RA, Grimsby JL, Bouatia-Naji N, Chen H, et al. A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. Nat Genet 2012;44:659–69.
14. Soranzo N, Sanna S, Wheeler E, Gieger C, Radke D, Dupuis J, et al. Common variants at 10 genomic loci influence hemoglobin A1C levels via glycemic and nonglycemic pathways. Diabetes 2010;59:3229–39.
15. Willer CJ, Schmidt EM, Sengupta S, Peloso GM, Gustafsson S, Kanoni S, et al. Discovery and refinement of loci associated with lipid levels. Nat Genet 2013;45:1274–83.
16. Teumer A, Tin A, Sorice R, Gorski M, Yeo NC, Chu AY, et al. Genome-wide association studies identify genetic loci associated with Albuminuria in diabetes. Diabetes 2016;65:803–17.
17. Paré G, Asma S, Deng WQ. Contribution of large region joint associations to complex traits genetics. Shriner D, editor. PLoS Genet 2015;11:e1005103.
18. Hsu F, Kent JW, Clawson H, Kuhn RM, Diekhans M, Haussler D. The UCSC known genes. Bioinformatics 2006;22:1036–46.
19. Rebhan M, Chalifa-Caspi V, Prilusky J, Lancet D. GeneCards: A novel functional genomics compendium with automated data mining and query reformulation support. Bioinformatics 1998;14:656–64.

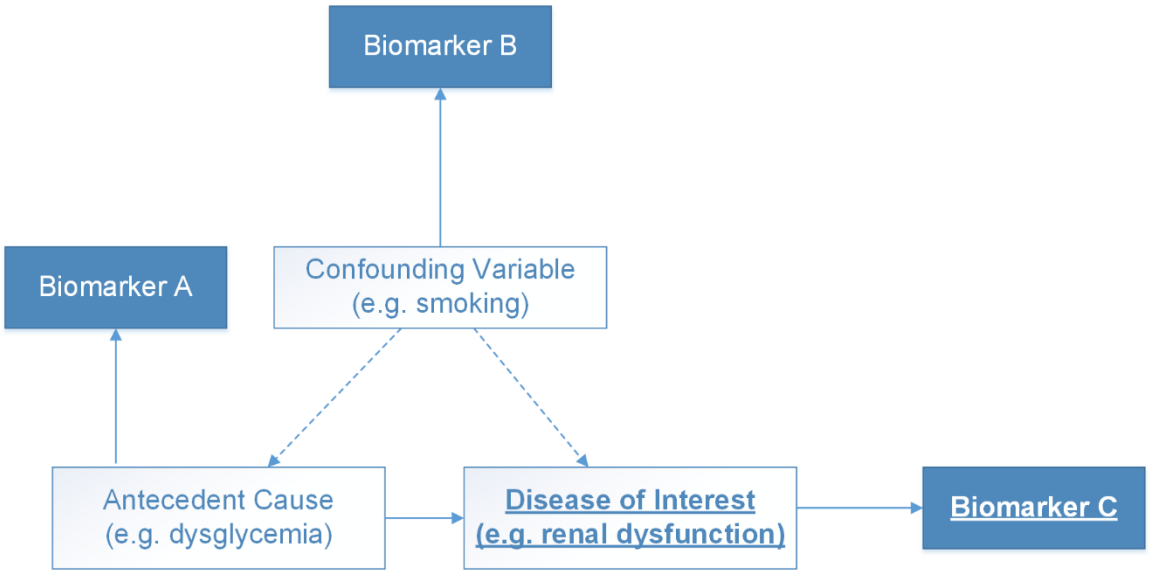
Supplemental Results

Associations of TFF3 in the full ORIGIN biomarker sub-study

The relationship between baseline $eGFR_{crea}$ and TFF3 levels was assessed in the full ORIGIN biomarker sub-study with a sample of N=8,197. After adjusting for age, sex, and ethnicity, higher levels of TFF3 were independently linked to lower levels of $eGFR_{crea}$ (β = -1.64 SD of TFF3 per 1 unit increase in log-transformed $eGFR_{crea}$; 95% CI = -1.72 to -1.56; $p = 3.75x10^{-321}$). Consistently, after adjusting for the same variables, renal disease at baseline was a determinant for higher levels of TFF3 (β=0.58 SD higher in individuals with renal disease at baseline; 95% CI= 0.53 to 0.62; $p=2.45x10^{-147}$).

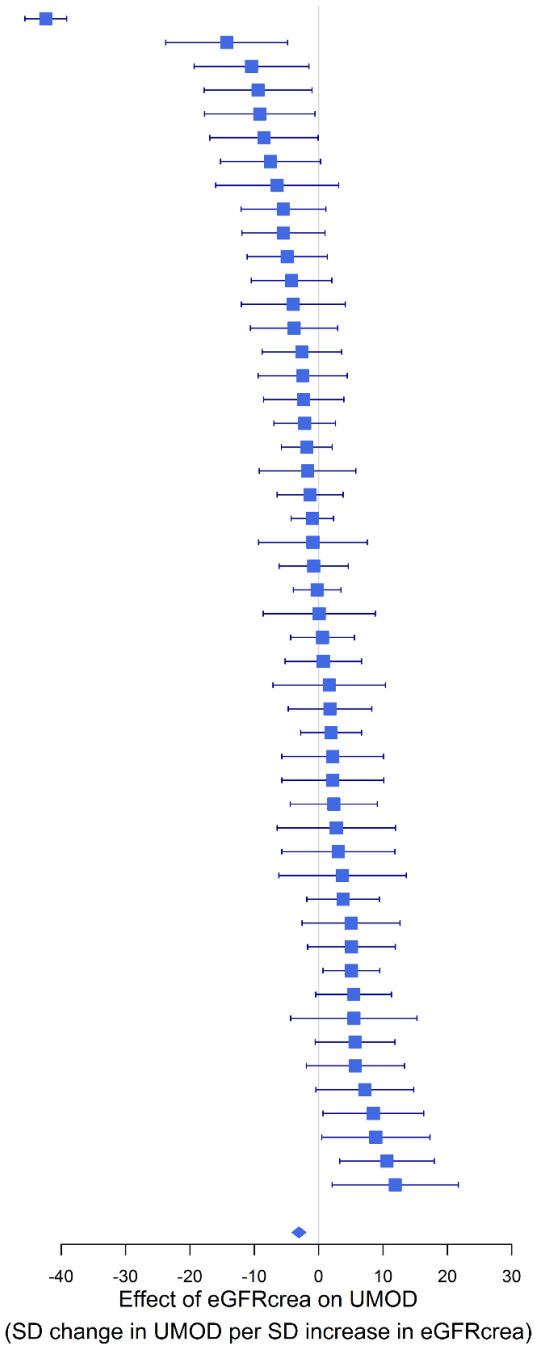
Supplemental Figures

Supplemental Figure 1. Schematic representation of the different types of biomarkers that can be found in the serum of individuals with diabetic renal disease. Individuals with diabetic renal disease host several different types of serum biomarkers, which are difficult to distinguish from observation alone. Biomarker A represents biomarkers associated with the antecedent cause of disease (e.g., dysglycemia). Biomarker B represents biomarkers associated with a confounding variable associated with both dysglycemia and renal dysfunction (e.g., obesity). Biomarker C represents the biomarker of interest that is affected by renal dysfunction itself.



Supplemental Figure 2. Effect of $eGFR_{crea}$ on uromodulin using Mendelian randomization

SNP	Beta [95% CI]	P-value
rs13329952	-42.41 [-45.66, -39.15]	5.06e-144
rs3750082	-14.29 [-23.77, -4.81]	3.13e-03
rs164748	-10.43 [-19.35, -1.52]	2.18e-02
rs476633	-9.42 [-17.80, -1.04]	2.76e-02
rs11666497	-9.14 [-17.72, -0.56]	3.68e-02
rs3850625	-8.49 [-16.91, -0.07]	4.81e-02
rs228611	-7.48 [-15.26, 0.30]	5.95e-02
rs4667594	-6.45 [-16.02, 3.11]	1.86e-01
rs848490	-5.47 [-12.07, 1.12]	1.04e-01
rs12460876	-5.47 [-11.93, 1.00]	9.73e-02
rs1260326	-4.88 [-11.10, 1.34]	1.24e-01
rs4744712	-4.20 [-10.46, 2.05]	1.88e-01
rs1800615	-3.94 [-12.02, 4.14]	3.39e-01
rs10774021	-3.81 [-10.59, 2.97]	2.70e-01
rs3758086	-2.60 [-8.79, 3.58]	4.10e-01
rs807601	-2.48 [-9.41, 4.45]	4.83e-01
rs9472135	-2.32 [-8.56, 3.92]	4.66e-01
rs10277115	-2.17 [-6.95, 2.60]	3.72e-01
rs17319721	-1.85 [-5.79, 2.09]	3.57e-01
rs1044261	-1.73 [-9.24, 5.79]	6.52e-01
rs11959928	-1.33 [-6.46, 3.81]	6.13e-01
rs2467853	-0.99 [-4.29, 2.32]	5.59e-01
rs163160	-0.89 [-9.33, 7.54]	8.36e-01
rs963837	-0.77 [-6.15, 4.60]	7.78e-01
rs7805747	-0.23 [-3.93, 3.47]	9.04e-01
rs716877	0.09 [-8.62, 8.80]	9.84e-01
rs6546838	0.60 [-4.36, 5.55]	8.13e-01
rs316009	0.72 [-5.22, 6.67]	8.12e-01
rs2712184	1.64 [-7.10, 10.38]	7.13e-01
rs10491967	1.76 [-4.73, 8.25]	5.95e-01
rs6420094	1.93 [-2.80, 6.66]	4.24e-01
rs8091180	2.18 [-5.73, 10.08]	5.89e-01
rs10994860	2.20 [-5.73, 10.12]	5.87e-01
rs2861422	2.35 [-4.42, 9.11]	4.96e-01
rs9682041	2.74 [-6.46, 11.94]	5.59e-01
rs2802729	3.06 [-5.73, 11.85]	4.95e-01
rs12136063	3.71 [-6.19, 13.61]	4.63e-01
rs491567	3.79 [-1.84, 9.43]	1.87e-01
rs267734	5.03 [-2.58, 12.64]	1.95e-01
rs17216707	5.09 [-1.72, 11.91]	1.43e-01
rs11657044	5.10 [0.68, 9.51]	2.36e-02
rs9916302	5.43 [-0.45, 11.31]	7.02e-02
rs7759001	5.47 [-4.35, 15.28]	2.75e-01
rs1394125	5.65 [-0.54, 11.84]	7.38e-02
rs6088580	5.70 [-1.92, 13.33]	1.43e-01
rs6459680	7.17 [-0.43, 14.77]	6.43e-02
rs1106766	8.50 [0.67, 16.33]	3.34e-02
rs6795744	8.88 [0.47, 17.29]	3.85e-02
rs4014195	10.61 [3.28, 17.95]	4.59e-03
rs10513801	11.90 [2.10, 21.70]	1.73e-02
Overall	**-3.05 [-4.10, -2.00]**	**<0.0001**



Supplemental Tables

Supplemental Table 1. List of $eGFR_{crea}$-associated SNPs from CKDGen consortium GWAS meta-analysis that matched in ORIGIN genetic panel.

Chr	SNP	Position (hg19)	Effect Allele	Other Allele	Effect Allele Frequency	Beta	Standard Error	*P*-value
1	rs1800615	15832281	t	c	0.231	-0.0058	0.00092	1.90E-09
1	rs12136063	110014170	a	g	0.714	0.0049	0.00092	2.30E-07
1	rs267734	150951477	t	c	0.792	-0.0079	0.0011	4.00E-13
1	rs3850625	201016296	a	g	0.097	0.008	0.0014	6.40E-09
1	rs2802729	243501763	a	c	0.406	-0.005	0.00092	7.40E-08
2	rs807601	15793014	t	g	0.319	0.0064	0.00092	6.60E-12
2	rs1260326	27730940	t	c	0.42	0.0068	0.00092	3.40E-14
2	rs6546838	73679280	a	g	0.748	-0.0093	0.001	7.70E-20
2	rs4667594	170008506	a	t	0.509	-0.0045	0.00092	2.40E-07
2	rs2712184	217682779	a	c	0.541	-0.0049	0.00092	2.70E-08
3	rs6795744	13906850	a	g	0.133	0.0071	0.0012	9.60E-09
3	rs2861422	141724644	t	c	0.248	0.0074	0.001	9.10E-14
3	rs9682041	170091902	t	c	0.898	-0.0067	0.0013	3.80E-07
3	rs10513801	185822353	t	g	0.898	0.007	0.0013	9.30E-08
4	rs17319721	77368847	a	g	0.42	-0.011	0.00092	1.30E-37
4	rs228611	103561709	a	g	0.447	-0.0055	0.00092	4.70E-10
5	rs11959928	39397132	a	t	0.431	-0.0083	0.00092	1.70E-20
5	rs6420094	176817636	a	g	0.633	0.0096	0.001	4.90E-22
6	rs7759001	27341409	a	g	0.783	-0.0053	0.001	2.60E-07
6	rs9472135	43809802	t	c	0.734	-0.008	0.001	3.30E-15
6	rs316009	160675764	t	c	0.097	0.013	0.0014	4.40E-19
7	rs10277115	1285195	a	t	0.215	0.0095	0.0014	1.10E-10
7	rs3750082	32919927	a	t	0.314	0.0049	0.00092	2.50E-07
7	rs848490	77555005	c	g	0.727	0.0073	0.001	7.80E-13
7	rs7805747	151407801	a	g	0.296	-0.013	0.0011	8.00E-29
7	rs6459680	156258568	t	g	0.8	-0.0065	0.001	2.00E-10
8	rs3758086	23714992	a	g	0.446	-0.0071	0.00092	1.70E-15
9	rs4744712	71434707	a	c	0.383	-0.0071	0.00092	4.30E-15
10	rs1044261	1065710	t	c	0.084	-0.011	0.0016	1.20E-11
10	rs10994860	52645424	t	c	0.186	0.0075	0.0011	1.20E-10
11	rs163160	2789955	a	g	0.857	0.0067	0.0011	9.70E-09
11	rs963837	30749090	t	c	0.544	-0.0078	0.00092	5.70E-18
11	rs4014195	65506822	c	g	0.673	0.0061	0.00092	2.20E-11

12	rs10774021	349298	t	c	0.695	-0.0063	0.00092	4.80E-12
12	rs10491967	3368093	a	g	0.097	-0.0092	0.0014	3.00E-10
12	rs1106766	57809456	t	c	0.246	0.0062	0.0011	4.70E-08
13	rs716877	72347448	c	g	0.376	0.0049	0.00092	6.20E-08
15	rs476633	41392134	c	g	0.669	0.0051	0.00092	8.90E-09
15	rs2467853	45698793	t	g	0.672	0.013	0.00092	1.00E-42
15	rs491567	53946593	a	c	0.792	-0.0084	0.001	2.90E-15
15	rs1394125	76158983	a	g	0.35	-0.0073	0.001	5.50E-14
16	rs13329952	20366507	t	c	0.783	-0.016	0.0011	9.50E-43
16	rs164748	89708292	c	g	0.554	0.0047	0.00092	9.30E-08
17	rs9916302	37499949	t	c	0.754	-0.008	0.001	4.80E-15
17	rs11657044	59450105	t	c	0.18	-0.011	0.0012	7.90E-22
18	rs8091180	77164243	a	g	0.598	-0.0054	0.001	3.50E-07
19	rs12460876	33356891	t	c	0.577	-0.0066	0.00092	1.90E-13
19	rs11666497	38464262	t	c	0.156	-0.0064	0.0012	8.60E-08
20	rs6088580	33285053	c	g	0.51	-0.0055	0.00092	7.20E-10
20	rs17216707	52732362	t	c	0.788	-0.0084	0.0011	6.00E-13

Abbreviations: Chr, chromosome; SNP, single nucleotide polymorphism; hg19, human genome version 19

Supplemental Table 2. Population characteristics for the genetic and biomarker sub-studies of the ORIGIN clinical trial.

Variable	Biomarker Sub-study with Normal Baseline Kidney Measures (n=5,300)	Genetic Sub-study (n=4,147)
Age (years)	62.42 (7.43)	63.45 (7.98)
Sex (% male)	68.13	64.14
Ethnicity (%)		
European	60.49	46.56
Latin	28.62	53.44
Black	4.85	0
South Asian	5.51	0
Other Asian	0.53	0
Smoking Status (%)		
Never	38.04	39.84
Former	49.17	48.76
Current	12.79	11.41
Prior CVD (% yes)	61.09	53.29
Hypertension (% yes)	76.08	82.90
Prior renal disease (% yes)	NA	35.62
Prior diabetes mellitus (% yes)	79.98	87.56
Incident CKD (% yes)	25.53	34.07
Worsening of Albuminuria Category (% of CKD)	72.43	57.68
Doubling of Serum Creatinine (% of CKD)	4.88	6.02
$eGFR_{crea}$ < 60 mL/min per 1.73 m^2 (% of CKD)	38.06	57.47
Body mass index (kg/m^2)	30.07 (5.23)	30.45 (5.33)
Serum creatinine (mg/dL)	0.92 (0.17)	1.02 (0.26)
(µmol/L)	81.42 (14.84)	90.09 (22.98)
$eGFR_{crea}$ (mL/min per 1.73 m^2)	81.96 (13.24)	75.91 (21.07)
Urinary ACR (mg/g)	5.93 (6.02)	64.51 (249.56)
(mg/mmol)	0.67 (0.68)	7.29 (28.20)
FPG (mg/dL)	130.08 (34.05)	136.56 (39.09)
(mmol/L)	7.22 (1.89)	7.58 (2.17)
HDL (mg/dL)	45.56 (11.97)	45.17 (12.36)
(mmol/L)	1.18 (0.31)	1.17 (0.32)

Data are presented as mean (SD) or median [IQR] unless otherwise stated

Abbreviations: ACR, albumin-to-creatinine ratio; CVD, cardiovascular disease; CKD, chronic kidney disease; $eGFR_{crea}$, creatinine-based estimated glomerular filtration rate (MDRD formula); FPG, fasting plasma glucose; HDL, high-density lipoprotein; IQR, interquartile range; SD, standard deviation

Supplemental Table 3. Summary of top Mendelian randomization biomarkers for $eGFR_{crea}$

Biomarker	**Effect of $eGFR_{crea}$ on biomarker [a]** ***(95% CI)***	***P*-value**
Uromodulin	-3.05 *(-2.00 – -4.10)*	<0.00001
Trefoil Factor 3	1.86 *(0.95 – 2.76)*	0.00008
Beta-2-Microglobulin	1.47 *(0.56 – 2.38)*	0.0016
Mesothelin	1.39 *(0.50 – 2.29)*	0.0020

a. SD change in biomarker per 1 unit decrease in log-transformed $eGFR_{crea}$

Supplemental Table 4. Reverse Mendelian randomization for other metabolic and kidney-related traits with serum trefoil factor 3.

Trait	Beta	95% CI	*P*-value	Number of GWS SNPs
Body mass index	0.02	-0.20 - 0.25	0.848	77
Chronic kidney disease	0.18	0.02 - 0.34	0.028	4
Coronary artery disease	-0.1	-0.23 - 0.02	0.092	52
Diastolic blood pressure	-0.02	-0.05 - 0.01	0.219	27
$eGFR_{cys}$	-0.69	-2.33 - 0.94	0.409	4
Fasting glucose	0.26	-0.10 - 0.63	0.157	34
Fasting glucose adjusted for BMI	0.24	-0.10 - 0.58	0.162	37
Fasting insulin adjusted for BMI	-0.59	-1.39 - 0.20	0.143	15
Haemoglobin A1c	0.12	-0.46 - 0.70	0.678	11
HDL	0.08	-0.05 - 0.20	0.227	76
Hypertension	0.19	0 - 0.38	0.053	11
LDL	0.03	-0.09 - 0.15	0.638	58
Myocardial infarction	-0.09	-0.22 - 0.04	0.162	52
Systolic blood pressure	-0.01	-0.03 - 0.01	0.168	26
Total cholesterol	0.05	-0.07 - 0.17	0.433	73
Triglycerides	-0.02	-0.17 - 0.13	0.77	42
Type 2 diabetes	0.04	-0.03 - 0.11	0.284	66
Urinary albumin-creatinine ratio ($p<1x10^{-6}$)	-0.04	-0.60 - 0.53	0.893	5
Waist-hip ratio adjusted for BMI	0.07	-0.23 - 0.36	0.649	39

Abbreviations: BMI, body mass index; CI, confidence interval; $eGFR_{cys}$, cystatin C-based estimated glomerular filtration rate; GWS, genome-wide significant; HDL, high density lipoprotein cholesterol; LDL, low density lipoprotein cholesterol; p, p-value; SNP, single nucleotide polymorphism

Supplemental Table 5. Adjusted odds ratios of incident CKD according to TFF3 quartiles

		Age, Sex, and Ethnicity-Adjusted		Laboratory Model [a]		Full Model [b]	
TFF3 Quartiles	**N**	**OR [c] *(95% CI)***	***P*-value**	**OR [c] *(95% CI)***	***P*-value**	**OR [c] *(95% CI)***	***P*-value**
Lower	1310	**1**	-	**1**	-	**1**	-
Lower Middle	1280	1.23 *(0.93 - 1.37)*	0.213	1.12 *(0.92 - 1.36)*	0.256	1.12 *(0.92 - 1.36)*	0.271
Upper Middle	1357	1.47 *(1.22 - 1.77)*	3.84E-05	1.40 *(1.16 - 1.69)*	4.70E-04	1.41 *(1.16 - 1.70)*	4.50E-04
Upper	1353	1.86 *(1.54 - 2.24)*	6.40E-11	1.68 *(1.39 - 2.04)*	1.14E-07	1.74 *(1.43 - 2.12)*	3.58E-08
***P* Trend**		0.476 *(0.344 - 0.608)*	1.69E-12	0.400 *(0.263 - 0.537)*	1.04E-08	0.422 *(0.283 - 0.562)*	3.04E-09

a. Adjusted for age, sex, ethnicity, baseline $eGFR_{crea}$ and natural log-transformed urinary ACR
b. Adjusted for age, sex, ethnicity, baseline $eGFR_{crea}$, natural log-transformed urinary ACR, fasting plasma glucose, systolic blood pressure, BMI, prior diabetes, prior cardiovascular disease, anti-hypertensive drug use, and smoking status
c. Odds of incident CKD relative to lower quartile per 1 SD increase in TFF3

Abbreviations: ACR, albumin-to-creatinine ratio; BMI, body mass index; CI, confidence interval; CKD, chronic kidney disease; $eGFR_{crea}$, creatinine-based estimated glomerular filtration rate; OR, odds ratio; SD, standard deviation; TFF3, trefoil factor 3

Supplemental Table 6. Kaplan-Meier table for freedom from incident CKD stratified by serum TFF3 quartiles.

	Lower			Lower Middle			Upper Middle			Upper		
Time Interval (year)	**At risk (n)**	**Events (n)**	**Freedom [95% CI]**	**At risk (n)**	**Events (n)**	**Freedom [95% CI]**	**At risk (n)**	**Events (n)**	**Freedom [95% CI]**	**At risk (n)**	**Events (n)**	**Freedom [95% CI]**
0 - 2	1310	105	92.0 [90.4 - 93.3]	1280	117	90.9 [89.1 - 92.3]	1357	170	87.5 [85.6 - 89.1]	1353	208	84.6 [82.6 - 86.4]
2 - EUF	1174	151	80.2 [77.9 – 82.2]	1120	164	77.6 [75.1 - 79.8]	1142	202	72.0 [69.5 - 74.3]	1057	236	65.7 [63.1 - 68.3]

Abbreviations: CI, confidence interval; EUF, end of usual follow-up

Supplemental Table 7. Association of TFF3 with changes in $eGFR_{crea}$ from baseline to end of usual follow-up in a subset of ORIGIN with normal kidney measures

	Age, Sex, and Ethnicity		Laboratory Model [a]		Full Model [b]	
	Effect of TFF3 on ΔeGFR [c] ***(95% CI)***	***P-value***	***Effect of TFF3 on ΔeGFR*** [c] ***(95% CI)***	***P-value***	***Effect of TFF3 on ΔeGFR*** [c] ***(95% CI)***	***P-value***
ΔeGFR	-2.12 *(-2.96 - -1.28)*	8.36E-07	-4.27 *(-5.07 - -3.48)*	1.40E-25	-4.33 *(-5.13 - -3.53)*	7.92E-26

a. Adjusted for age, sex, ethnicity, baseline $eGFR_{crea}$ and natural log-transformed urinary albumin-to-creatinine ratio.

b. Adjusted for age, sex, ethnicity, baseline $eGFR_{crea}$, natural log-transformed urinary albumin-to-creatinine ratio, fasting plasma glucose, systolic blood pressure, body mass index, prior diabetes, prior cardiovascular disease, anti-hypertensive drug use, and smoking status.

c. mL/min per 1.73 m^2 $eGFR_{crea}$ per 1 SD increase in serum trefoil factor 3.

Supplemental Table 8. Comparison of discrimination capacity for incident CKD risk models using integrated discrimination improvement (IDI) scores.

Model A	Model B	IDI[c] (95% CI)	*P*-value
$eGFR_{crea}$	TFF3	0.005 (0.002 - 0.007)	7.48E-05
	$eGFR_{crea}$ + TFF3	0.006 (0.004 - 0.008)	1.59E-08
$eGFR_{crea}$ + clinical risk factors[a]	$eGFR_{crea}$ + clinical risk factors[a] + TFF3	0.006 (0.004 – 0.008)	1.05E-08
$eGFR_{crea}$ + clinical risk factors[a] + ln(ACR)	$eGFR_{crea}$ + clinical risk factors[a] + ln(ACR) + TFF3	0.004 (0.002 – 0.006)	6.89E-06
Full model[b]	Full $model_2$ + TFF3	0.005 (0.003 - 0.007)	1.32E-06

a. Prior renal disease + prior cardiovascular disease + prior diabetes + SBP + smoking status + anti-hypertensive drug use
b. Baseline $eGFR_{crea}$ + natural log-transformed baseline urinary ACR + prior renal disease + prior cardiovascular disease + prior diabetes + systolic blood pressure + smoking status + anti-hypertensive drug use + BMI + fasting plasma glucose
c. IDI comparing model B to model A

Abbreviations: ACR, albumin-to-creatinine ratio; BMI, body mass index; CKD, chronic kidney disease; $eGFR_{crea}$, creatinine-based estimated glomerular filtration rate (CKD-EPI formula); IDI, integrated discrimination improvement; TFF3, trefoil factor 3

Supplemental Table 9. Summary-level associations of eGFRcrea SNPs with uromodulin and TFF3 in ORIGIN trial.

Chromosome	Biomarker	SNP	Position (hg19)	OA	EA	*P*-value	EAF	Effect[a]	SE
1	Uromodulin	rs1800615	15832281	C	T	0.341	0.30	0.023	0.024
1	Uromodulin	rs12136063	110014170	G	A	0.456	0.74	0.018	0.025
1	Uromodulin	rs267734	150951477	T	C	0.192	0.15	0.040	0.031
1	Uromodulin	rs3850625	201016296	G	A	0.048	0.11	-0.067	0.034
1	Uromodulin	rs2802729	243501763	C	A	0.487	0.44	-0.016	0.022
2	Uromodulin	rs807601	15793014	G	T	0.483	0.36	-0.016	0.023
2	Uromodulin	rs1260326	27730940	T	C	0.120	0.59	0.033	0.022
2	Uromodulin	rs6546838	73679280	A	G	0.821	0.29	0.005	0.024
2	Uromodulin	rs4667594	170008506	T	A	0.197	0.57	0.028	0.022
2	Uromodulin	rs2712184	217682779	C	A	0.703	0.60	-0.008	0.022
3	Uromodulin	rs6795744	13906850	G	A	0.037	0.16	0.063	0.030
3	Uromodulin	rs2861422	141724644	C	T	0.495	0.23	0.017	0.026
3	Uromodulin	rs9682041	170091902	C	T	0.568	0.85	-0.018	0.031
3	Uromodulin	rs10513801	185822353	T	G	0.017	0.11	-0.083	0.035
4	Uromodulin	rs17319721	77368847	G	A	0.358	0.39	0.020	0.022
4	Uromodulin	rs228611	103561709	G	A	0.060	0.43	0.041	0.022
5	Uromodulin	rs11959928	39397132	T	A	0.611	0.44	0.011	0.022
5	Uromodulin	rs6420094	176817636	A	G	0.426	0.31	-0.018	0.023
6	Uromodulin	rs7759001	27341409	G	A	0.283	0.78	-0.029	0.027
6	Uromodulin	rs9472135	43809802	T	C	0.478	0.26	-0.018	0.026
6	Uromodulin	rs316009	160675764	T	C	0.818	0.92	-0.009	0.039
7	Uromodulin	rs10277115	1285195	A	T	0.381	0.70	0.021	0.023
7	Uromodulin	rs3750082	32919927	T	A	0.003	0.30	-0.071	0.024
7	Uromodulin	rs848490	77555005	G	C	0.099	0.74	-0.040	0.025
7	Uromodulin	rs7805747	151407801	G	A	0.917	0.26	0.003	0.025
7	Uromodulin	rs6459680	156258568	G	T	0.068	0.75	-0.047	0.026
8	Uromodulin	rs3758086	23714992	G	A	0.400	0.37	0.019	0.022
9	Uromodulin	rs4744712	71434707	A	C	0.183	0.64	-0.030	0.023

10	Uromodulin	rs1044261	1065710	C	T	0.654	0.07	0.019	0.042
10	Uromodulin	rs10994860	52645424	C	T	0.592	0.16	0.016	0.030
11	Uromodulin	rs163160	2789955	A	G	0.825	0.17	0.006	0.029
11	Uromodulin	rs963837	30749090	T	C	0.779	0.45	-0.006	0.022
11	Uromodulin	rs4014195	65506822	C	G	0.005	0.33	-0.065	0.023
12	Uromodulin	rs10774021	349298	C	T	0.269	0.54	0.024	0.022
12	Uromodulin	rs10491967	3368093	G	A	0.591	0.15	-0.016	0.030
12	Uromodulin	rs1106766	57809456	C	T	0.034	0.24	0.053	0.025
13	Uromodulin	rs716877	72347448	C	G	0.994	0.53	0.000	0.022
15	Uromodulin	rs476633	41392134	C	G	0.027	0.43	0.048	0.022
15	Uromodulin	rs2467853	45698793	T	G	0.557	0.48	0.013	0.022
15	Uromodulin	rs491567	53946593	A	C	0.190	0.26	0.032	0.024
15	Uromodulin	rs1394125	76158983	G	A	0.070	0.32	-0.041	0.023
16	Uromodulin	rs13329952	20366507	T	C	0.000	0.21	-0.679	0.026
16	Uromodulin	rs164748	89708292	C	G	0.022	0.45	0.049	0.021
17	Uromodulin	rs9916302	37499949	T	C	0.070	0.28	0.043	0.024
17	Uromodulin	rs11657044	59450105	T	C	0.023	0.74	0.057	0.025
18	Uromodulin	rs8091180	77164243	G	A	0.598	0.59	-0.012	0.022
19	Uromodulin	rs12460876	33356891	T	C	0.098	0.41	-0.036	0.022
19	Uromodulin	rs11666497	38464262	C	T	0.038	0.18	0.058	0.028
20	Uromodulin	rs6088580	33285053	G	C	0.154	0.43	-0.031	0.022
20	Uromodulin	rs17216707	52732362	T	C	0.137	0.26	0.043	0.029
1	Trefoil Factor 3	rs1800615	15832281	C	T	0.809	0.30	0.006	0.023
1	Trefoil Factor 3	rs12136063	110014170	G	A	0.615	0.74	-0.012	0.024
1	Trefoil Factor 3	rs267734	150951477	T	C	0.551	0.15	-0.017	0.029
1	Trefoil Factor 3	rs3850625	201016296	G	A	0.904	0.11	-0.004	0.033
1	Trefoil Factor 3	rs2802729	243501763	C	A	0.536	0.44	0.013	0.021
2	Trefoil Factor 3	rs807601	15793014	G	T	0.196	0.36	0.028	0.022

2	Trefoil Factor 3	rs1260326	27730940	T	C	0.499	0.59	0.014	0.021
2	Trefoil Factor 3	rs6546838	73679280	A	G	0.122	0.29	-0.035	0.023
2	Trefoil Factor 3	rs4667594	170008506	T	A	0.032	0.57	0.045	0.021
2	Trefoil Factor 3	rs2712184	217682779	C	A	0.093	0.60	-0.035	0.021
3	Trefoil Factor 3	rs6795744	13906850	G	A	0.532	0.16	0.018	0.029
3	Trefoil Factor 3	rs2861422	141724644	C	T	0.092	0.23	-0.041	0.024
3	Trefoil Factor 3	rs9682041	170091902	C	T	0.356	0.85	-0.028	0.030
3	Trefoil Factor 3	rs10513801	185822353	T	G	0.301	0.11	0.035	0.034
4	Trefoil Factor 3	rs17319721	77368847	G	A	0.176	0.39	0.029	0.021
4	Trefoil Factor 3	rs228611	103561709	G	A	0.756	0.43	-0.007	0.021
5	Trefoil Factor 3	rs11959928	39397132	T	A	0.151	0.44	0.030	0.021
5	Trefoil Factor 3	rs6420094	176817636	A	G	0.688	0.31	0.009	0.022
6	Trefoil Factor 3	rs7759001	27341409	G	A	0.314	0.78	-0.026	0.026
6	Trefoil Factor 3	rs9472135	43809802	T	C	0.803	0.26	-0.006	0.024
6	Trefoil Factor 3	rs316009	160675764	T	C	0.247	0.92	0.044	0.038
7	Trefoil Factor 3	rs10277115	1285195	A	T	0.267	0.70	0.025	0.023
7	Trefoil Factor 3	rs3750082	32919927	T	A	0.123	0.30	0.035	0.023

7	Trefoil Factor 3	rs848490	77555005	G	C	0.318	0.74	-0.023	0.024
7	Trefoil Factor 3	rs7805747	151407801	G	A	0.113	0.26	0.038	0.024
7	Trefoil Factor 3	rs6459680	156258568	G	T	0.148	0.75	0.036	0.025
8	Trefoil Factor 3	rs3758086	23714992	G	A	0.674	0.37	0.009	0.021
9	Trefoil Factor 3	rs4744712	71434707	A	C	0.096	0.64	0.036	0.022
10	Trefoil Factor 3	rs1044261	1065710	C	T	0.705	0.07	0.015	0.040
10	Trefoil Factor 3	rs10994860	52645424	C	T	0.693	0.16	-0.012	0.029
11	Trefoil Factor 3	rs163160	2789955	A	G	0.266	0.17	0.031	0.027
11	Trefoil Factor 3	rs963837	30749090	T	C	0.214	0.45	-0.026	0.021
11	Trefoil Factor 3	rs4014195	65506822	C	G	0.987	0.33	0.000	0.022
12	Trefoil Factor 3	rs10774021	349298	C	T	0.892	0.54	-0.003	0.021
12	Trefoil Factor 3	rs10491967	3368093	G	A	0.107	0.15	-0.047	0.029
12	Trefoil Factor 3	rs1106766	57809456	C	T	0.076	0.24	-0.042	0.024
13	Trefoil Factor 3	rs716877	72347448	C	G	0.127	0.53	0.032	0.021
15	Trefoil Factor 3	rs476633	41392134	C	G	0.169	0.43	0.029	0.021
15	Trefoil Factor 3	rs2467853	45698793	T	G	0.798	0.48	0.005	0.021
15	Trefoil Factor 3	rs491567	53946593	A	C	0.105	0.26	-0.038	0.023

15	Trefoil Factor 3	rs1394125	76158983	G	A	0.031	0.32	0.047	0.022
16	Trefoil Factor 3	rs13329952	20366507	T	C	0.096	0.21	-0.042	0.025
16	Trefoil Factor 3	rs164748	89708292	C	G	0.501	0.45	-0.014	0.021
17	Trefoil Factor 3	rs9916302	37499949	T	C	0.324	0.28	-0.023	0.023
17	Trefoil Factor 3	rs11657044	59450105	T	C	0.316	0.74	-0.024	0.024
18	Trefoil Factor 3	rs8091180	77164243	G	A	0.435	0.59	0.017	0.021
19	Trefoil Factor 3	rs12460876	33356891	T	C	0.841	0.41	0.004	0.021
19	Trefoil Factor 3	rs11666497	38464262	C	T	0.409	0.18	0.022	0.027
20	Trefoil Factor 3	rs6088580	33285053	G	C	0.650	0.43	0.009	0.021
20	Trefoil Factor 3	rs17216707	52732362	T	C	0.038	0.26	-0.058	0.028

a. per copy of effect allele

Abbreviations: EA, effect allele; EAF, effect allele frequency; hg19, human genome version 19; OA, other allele; SE, standard error; SNP, single nucleotide polymorphism