UNIVERSITÀ DEGLI STUDI DI MILANO

DIPARTIMENTO DI SCIENZE
FARMACOLOGICHE E BIOMOLECOLARI

**Doctoral programme in Pharmacological Biomolecular Sciences, Experimental and Clinical– XXXIV Cycle**

SD BIO/14

# USE OF MENDELIAN RANDOMIZATION STUDIES TO IDENTIFY POSSIBLE PHARMACOLOGICAL TARGETS IN THE CARDIOVASCULAR AREA

PhD programme Coordinator:

Professor **Giuseppe Danilo Norata**

Tutor:

Professor **Alberico L. Catapano**

Co-tutor:

Dr. **Elena Tragni**

Candidate:

**Elena Olmastroni**

R12226

Academic year 2020-2021

*''Your genetics load the gun.*
*Your lifestyle pulls the trigger''*

*- Mehmet Oz -*

# TABLE OF CONTENTS

## Abbreviations mostly used

ASCVD - atherosclerotic cardiovascular disease

BMI - body mass index

CAD - coronary artery disease

CHD - coronary heart disease

CVD - cardiovascular disease

DM - diabetes mellitus

GWAS - genome-wide association study

IV - instrumental variable

LD - linkage disequilibrium

LDL-C - low-density lipoprotein cholesterol

Lp(a) - lipoprotein(a)

MCE - major coronary events

MR - Mendelian randomization

PGS - polygenic score

RCT - randomized controlled trial

SNP - single nucleotide polymorphism

## ABSTRACT

Different statistical approaches have been implemented to overcome the limitations that typically and differently influence both randomized clinical trials and observational studies. Mendelian randomization studies, in which functional genetic variants serve as tools ("instrumental variables") to approximate modifiable environmental exposures, have been developed and implemented in the context of observational epidemiological studies to strengthen causal inferential estimates in non-experimental situations. Since genetic variants are randomly transferred from parents to offspring at the time of gamete formation, they can realistically mimic the random allocation process of treatment in a randomized clinical trial, offering a strategy to eliminate, or at least reduce, the residual confounding typically affecting observational studies, thus allowing to obtain generalizable results for the entire population. If correctly conducted and carefully interpreted, Mendelian randomization studies can provide useful scientific evidence to support or reject causal hypotheses that verify the association between modifiable exposures and diseases. This kind of evidence may provide useful information to identify new potential drug targets, with a higher probability of success than approaches based on animal models or *in vitro* studies.

This thesis summarizes the history and context of Mendelian randomization, the main features of the study design, the assumptions for its correct use, and a brief discussion on the advantages and disadvantages of this approach. In addition, an overview of what the Mendelian randomization technique has contributed to date in the cardiovascular field has also been presented.

The methods and techniques discussed have been also practically applied on several studies conducted thanks to a collaboration established with Professor Brian A. Ference from the Cardiovascular Epidemiology Unit of the Department of Public Health and Primary Care (University of Cambridge). This agreement has allowed to access the UK Biobank, a prospective cohort study with deep genetic, physical, and health data, collected on about 500,000 volunteer participants recruited throughout the UK. The access to this large-scale biomedical database has been fundamental to carried out the projects presented in this thesis, which have provided key evidence to improve our knowledge about cardiovascular disease. First, we found that the increase of measured body mass index is a much stronger risk factor for type 2 diabetes than polygenic predisposition that leads to reversible metabolic changes that do not accumulate over time. Therefore, most cases of diabetes potentially can be prevented or reversed, leading to a major reduction of the prevalence of one of the most impactful risk factors for the development of cardiovascular disease. Second, we found that parental family history of coronary heart disease provides independent, complementary and additive information to the individual polygenic predisposition in the definition of the inherited genetic variation as well as to LDL cholesterol exposure in the estimation of the lifetime cardiovascular risk. In order to develop a simple, but powerful, algorithm to contextualize the frame of who will need to be treated, it is essential to retrieve information about parental family history of heart disease and individual polygenic predisposition to coronary artery disease, in addition to the measurement of all the other well-known cardiovascular risk factors, especially LDL cholesterol levels. Finally, we discovered three important evidence regarding lipoprotein(a), an independent risk factor for the development of coronary and cerebral

atherosclerosis: (i) the cumulative lifetime risk of major coronary events is comparable considering genetically and clinically determined Lp(a) concentrations, meaning that, in terms of cardiovascular risk prediction, it is reasonable to rely on measured levels, regardless the genotype; (ii) there is no significant association between high Lp(a) concentrations and the occurrence of venous thromboembolism events; (iii) an extra reduction of LDL cholesterol can overcome the extra cardiovascular risk due to high Lp(a) levels, and we quantitatively defined the additional LDL cholesterol reduction needed to abolish this risk.

At the end of this dissertation, the potential use of Mendelian randomization to inform the design of randomized controlled trials is also presented, as well as the possibility to use this approach to anticipate trials results in terms of predicting treatment efficacy and adverse effects, and to inform on potential repurposing of drugs.

## RIASSUNTO

Per superare i limiti metodologici che, tipicamente e in modo diverso, influenzano sia gli studi clinici randomizzati che gli studi osservazionali sono stati implementati diversi approcci statistici. Gli studi di randomizzazione Mendeliana, in cui varianti genetiche funzionali servono da strumenti ("variabili strumentali") per approssimare esposizioni ambientali modificabili, sono stati sviluppati e implementati nell'ambito di studi epidemiologici osservazionali per rafforzare le stime inferenziali causali in situazioni non sperimentali. Poiché le varianti genetiche vengono trasferite casualmente dai genitori alla prole al momento della formazione del gamete, queste possono realisticamente riprodurre il processo di allocazione casuale del trattamento in uno studio clinico randomizzato, offrendo una strategia per eliminare, o comunque ridurre, il confondimento residuo tipico degli studi osservazionali, rendendo così possibile ottenere risultati statistici che siano generalizzabili dal campione all'intera popolazione di indagine. Se condotti correttamente e interpretati attentamente, gli studi di randomizzazione Mendeliana possono fornire evidenze scientifiche utili per supportare o rifiutare ipotesi di causalità che verificano l'associazione tra l'esposizione a fattori di rischio modificabili e *outcome* correlati al rischio di malattia, in assenza di confondimento. Queste evidenze possono rappresentare informazioni utili per identificare nuovi potenziali bersagli farmacologici, con una probabilità di successo superiore rispetto ad approcci basati su studi sugli animali o *in vitro*.

Questa tesi riassume la storia e il contesto della randomizzazione Mendeliana, le caratteristiche principali del disegno di studio, gli assunti per un utilizzo corretto, e una breve discussione sui vantaggi e svantaggi di questo approccio. Viene, inoltre, presentata anche una

panoramica di quali sono state fino ad oggi le applicazioni della randomizzazione Mendeliana nell'ambito della malattia cardiovascolare.

I metodi e le tecniche discussi sono stati applicati in diversi progetti condotti grazie ad una collaborazione instaurata con il professor Brian A. Ference della *Cardiovascular Epidemiology Unit* del *Department of Public Health and Primary Care* (Università di Cambridge). Questo accordo ha consentito l'accesso alla UK Biobank, uno studio prospettico di coorte con dati genetici, fisici e clinici approfonditi, raccolti su circa 500.000 partecipanti volontari reclutati in tutto il Regno Unito. L'accesso a questa banca dati è stato essenziale per la conduzione dei progetti oggetto di questa tesi, che hanno fornito evidenze importanti per migliorare le nostre conoscenze sulla malattia cardiovascolare. In primo luogo, abbiamo osservato che l'aumento dell'indice di massa corporea misurato è un fattore di rischio molto più forte per il diabete di tipo 2 rispetto alla predisposizione poligenica, aumento che porta a cambiamenti metabolici reversibili che non si accumulano nel tempo. Pertanto, la maggior parte dei casi di diabete può potenzialmente essere prevenuta o invertita, determinando una grande riduzione della prevalenza di uno dei fattori di rischio più impattanti per lo sviluppo della malattia cardiovascolare. In secondo luogo, abbiamo evidenziato che la storia familiare di malattia coronarica fornisce informazioni indipendenti, complementari e additive alla predisposizione poligenica individuale nella definizione della variabilità genetica ereditata, nonché all'esposizione al colesterolo LDL nella stima del rischio cardiovascolare nel corso della vita. Quindi, al fine di sviluppare un semplice, ma potente, algoritmo per contestualizzare il quadro di chi dovrà essere trattato, è essenziale recuperare le informazioni sulla storia familiare di malattia cardiaca

dei genitori e la predisposizione poligenica individuale alla malattia coronarica, oltre che alla misurazione di tutti gli altri ben noti fattori di rischio cardiovascolare, soprattutto i livelli di colesterolo LDL. Infine, abbiamo ottenuto tre evidenze importanti riguardanti la lipoproteina(a), un fattore di rischio indipendente per lo sviluppo dell'aterosclerosi coronarica e cerebrale: (i) il rischio cumulativo di eventi coronarici maggiori nel corso della vita è comparabile considerando le concentrazioni di Lp(a) determinate geneticamente e clinicamente, il che significa che, in termini di previsione del rischio cardiovascolare, è ragionevole fare affidamento sui livelli misurati, indipendentemente dal genotipo; (ii) non esiste un'associazione significativa tra alte concentrazioni di Lp(a) e il verificarsi di eventi di tromboembolismo venoso; (iii) una diminuzione aggiuntiva del colesterolo LDL può determinare la riduzione del rischio cardiovascolare causato dagli elevati livelli di Lp(a), e abbiamo definito quantitativamente la riduzione supplementare del colesterolo LDL necessaria per abolire questo rischio.

Alla fine di questo elaborato, viene anche presentato l'uso potenziale della randomizzazione Mendeliana per informare il disegno di trial clinici controllati e randomizzati, così come la possibilità di usare questo approccio per anticipare i risultati di tali studi, in termini di previsione dell'efficacia del trattamento e degli effetti avversi, e per proporre il possibile riposizionamento di farmaci sul mercato.

# CHAPTER 1

## Medical Research

## 1.1 Shortcomings of classical epidemiology

Epidemiology is the study of patterns of health and disease at the population level. A fundamental problem in epidemiological research is the distinction between correlation and causation (1). If we want to address basic medical questions, such as to determine disease aetiology (what is the cause of a disease?), to assess the impact of a public health intervention (what would be the result of a change in treatment?), to inform public policy, to prioritize healthcare resources, to advise treatment practice, or to counsel on the impact of lifestyle choices, then we have to answer questions of cause and effect.

The optimal way to address these questions is by appropriate study design, such as the use of randomized trials and prospective data (2). However, such designs are not always possible, and often causal questions must be answered using only observational data. Unfortunately, interpreting the association between a risk factor and a disease outcome in observational data as a causal association relies on untestable and often implausible assumptions, such as the absence of unmeasured confounding and of reverse causation. This has led to several high-profile cases where a risk factor has been widely advocated as an important factor in disease prevention from observational data, only to be later discredited when the evidence from randomized trials did not support a causal interpretation to the findings (3).

Therefore, in the following paragraphs the main characteristics of randomized controlled trials (RCTs) and observational studies will be described, in order to explain why more robust approaches are needed for assessing causal relationship using observational data. This will lead to introduce a relatively new genetic approach to

epidemiology that offers opportunities to deal with some of the difficulties of conventional epidemiology and to estimate causal associations of modifiable (non-genetic) risk factors using observational data.

## 1.2 Randomized controlled trials

The "gold standard" for the empirical testing of a scientific hypothesis in clinical research is a RCT. This design involves the random allocation of different treatment regimens to experimental units (usually individuals) in a population (4). The researchers decide randomly as to which participants in the trial receive the new treatment and which receive placebo or a reference treatment (standard or best care). In its simplest form, one "active treatment" (for example, intervention on a risk factor) is compared against a "control treatment" (no intervention), and the average outcomes in each of the arms of the trial are contrasted (**Figure 1**).

Here the risk factor (which we will often refer to as the "exposure" variable) is a putative causal risk factor. We seek to assess whether the risk factor is a cause of the outcome, and estimate (if appropriate) the magnitude of the causal effect. This is because the act of randomization balances participant characteristics (both observed and unobserved) between the groups. It makes groups comparable through the data collected during the research process, allowing attribution of any differences in outcome to the study intervention (5).

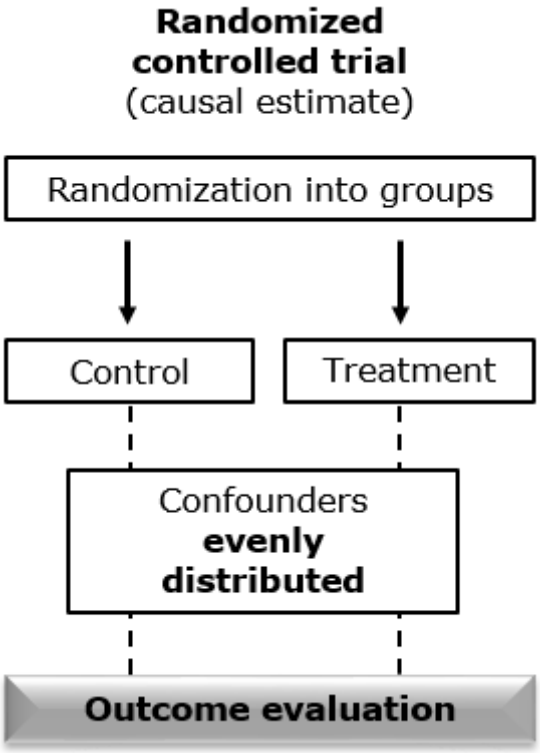



**Figure 1.** Randomized controlled trial design.

It is important to ensure that, at the time of recruitment, there is no knowledge of which group the participant will be allocated to; this is known as concealment. This is often ensured by using automated randomization systems (e.g. computer generated). RCTs are often blinded, so that participants as well as doctors, nurses or researchers do not know what treatment each participant is receiving, further minimizing bias. This avoids confounding factors linked to subjective perception of therapeutic effects because patients and health-care providers managing the therapy administration are unaware of the treatment assigned.

Trial results can support at least two of the causality criteria: the reversibility (if the cause is deleted, then the effect should disappear) and biological gradient (greater exposure generally leads to greater incidence of the effect).

Because of these conditions, it is then possible to confidently attribute cause and effect. That's why a properly designed study is well-regarded as being a true measure of efficacy. RCTs are therefore described as having high internal validity (the experimental design ensures that strong cause and effect conclusions can be drawn from the results).

While randomized trials are in principle the best way of determining the causal status of a particular risk factor, they are also characterized by limitations and are not always feasible, appropriate, or ethical (6). Indeed, although RCTs are precious tools, their applicability is restricted to ideal conditions, and this limits their ability to portray what happens in the real world (external validity).

One of the more relevant limitations of RCTs is a direct consequence of rigorous patient selection based on strict inclusion and exclusion criteria. The reason behind such selective enrolment is to minimize the presence of confounding factors (7). Thus, a "trial" population differs from the unselected general population (8). As an example, RCTs do not consider a number of factors that potentially influence outcomes in real-life, such as tobacco smoking, concomitant diseases, adherence to treatment, and many others, thus results may not always mimic real-life treatment situation.

Moreover, some specific categories, such as older individuals, are often underrepresented in RCTs. They are frequently excluded as a result of direct and indirect exclusion criteria based on the presence of comorbidities and polypharmacy. The consequence is that participants enrolled in clinical trials often do not represent the older patients in general practice setting (9).

RCTs are expensive and time-consuming, especially when the outcome is rare or requires a long follow-up period to be observed. Even if the randomized controlled trial only uses one baseline group

and a single treatment demographic, the length of the research requires a significant investment. It is one of the most expensive methods of collecting data in terms of time and money (10).

Timing is another constraint with RCTs, which frequently have duration of weeks or few months or rarely few years. Such a timeframe minimizes problems with management, costs, and patient withdraws (11), but is significantly shorter than the usual treatment period for chronic diseases. This limits the ability to provide reliable information on long-term treatment, especially regarding the safety of the therapy (12).

Additionally, in some cases, a targeted treatment which has an effect only on the risk factor of interest may not be available. Moreover, many "treatments" cannot be randomly allocated for practical or ethical reasons (classically, RCTs of the effects of parachutes on the survival of sky-divers).

Although RCTs have immensely contributed to development of health services in the last 50 years, other methods with greater external validity (or greater potential for generalizability) should also be considered for determining impact of interventions.

This is not intending to undermine the value of RCTs, but rather to point out some of its limitations and recognize the benefits of other alternative methodologies for establishing intervention impact (13).

As there is an urgent need for clinical information beyond that obtainable from classical RCTs, it is not surprising that there has been an impressive increase in the number of real-life studies.

## 1.3 Observational epidemiology

Scientific hypotheses are often assessed using observational data (14). Real-life studies have been defined by the European Working

Group on Relative Effectiveness as a way to analyse medical data collected under real-life conditions (15). In essence, they are conducted in everyday settings, and for this reason, they provide insights into the effectiveness of a medical condition/intervention. They can be observational or descriptive, i.e. non-interventional, or they can evaluate therapeutic interventions in usual care settings (15). Numerous strategies have been described for collecting data to inform real world clinical decisions. Examples include databases, patient and population surveys, chart reviews, registries, and observational studies (**Table 1**).

**Table 1.** Main types of real-life studies, their features, including the source of data, and their applications.

| **Type** | **Characteristics** | **Application** |
|---|---|---|
| Databases | Cross-sectional or longitudinal analysis of previously collected data. | Retrospective data analysis on various topics. |
| Population surveys | Questionnaires, patient health status and opinion assessment. | Epidemiological studies. |
| Patient chart reviews | In depth evaluation of previously collected data, particularly focusing on diagnosis and treatment. | Assessment of disease management for planning guidelines. |
| Registries | A medical institute record of all patients treated for a specific disease. | Analysis of a medical centre experience/management/changes in the treatment of a disease. |
| Observational data | Prospective or retrospective data collection, usually on population cohorts, over a long follow-up period. | Examination of medical intervention effectiveness, including safety and tolerability |

Observational studies are the most widely used. In particular, observational studies are designed to monitor and describe real-life management/treatment of clinical conditions, without interference from the strict rules that limit the generalizability of RCTs. They can be both prospective and retrospective and consist of cohort, case-control, and cross-sectional studies (12). By nature, they can assess large populations over long follow-up periods. In prospective cohort studies a group of people without the outcome of interest is selected. The investigator then measures a variety of variables that might be relevant to the development of the condition. Over a period of time, the people in the sample are observed to see whether they develop the outcome of interest. In single cohort studies, those people who do not develop the outcome of interest are used as internal controls. Where two cohorts are used, one group has been exposed to or treated with the agent of interest and the other has not, thereby acting as an external control (16). Conversely, retrospective cohort studies are based on data regarding exposures and events occurred in the past. The methodology is the same, but the study is performed *post hoc*. The cohort is "followed up" retrospectively (16).

Real-life studies reflect how treatments/interventions are administered in everyday clinical practice. They do not use inclusion and exclusion criteria to allocate a treatment: being nonselective, they can include all the variables that can influence outcomes under real-life conditions. This qualifying characteristic may lead to discrepancies with the results obtained by RCTs for a given outcome. Real-life studies are less affected by logistical and ethical constraints that limit the feasibility of RCTs. The absence of these limitations endows observational studies with the power to assess outcomes such as hospitalization and mortality (17), and enables pragmatic trials to estimate cost-effectiveness under real-life conditions (18).

The characteristics of real-life studies make them useful tools for evaluating complex therapies, and policies for prescribing and management (19). In addition, at variance with RCTs, which usually compare treatments and placebo, real-life studies are able to integrate alternative health care options, providing information on the management of complex interventions, which are very common in the routine health care setting and influence the clinical decision process (20). In addition, also the nature and characteristics of real-life studies make them the appropriate setting to assess safety. Their large scale and long duration in non-selected populations favour the identification of rare adverse events or interactions with other treatments. An additional advantage of real-life studies is their natural practice setting, e.g. physicians' offices or clinics, which ensures a high external validity (20). Involvement of patients from different settings increases the variability of the results, but also reproduces the complexity of the health care system more reliably than the controlled conditions in RCTs (20). Finally, real-life studies have the benefit of longer durations than RCTs. This timing permits a more appropriate assessment of the long-term effects of medical interventions, and the identification of late side effects, by following the natural course of a disease (20). Thus, real-life studies have many positive features and play a key role in the investigation of clinical conditions and interventional opportunities. These intrinsic values explain the increasing importance of such studies in clinical research.

Despite the fact that observational studies offer the opportunity to study a "real-life situation", they can only create new hypotheses rather than infer a causal relationship between an exposure and a disease outcome. The limitations of real-life studies are often intrinsically associated with their characteristic design. The proximity

to the real world, with all its complexity, including different disease manifestations, comorbidities, variable treatment adherence, and multiple therapies, often dilutes the magnitude of a treatment effect (7). The lack of patient selection, one of the most distinctive characteristics of real-life studies, makes it impossible to avoid unmeasured confounding factors (21), while the absence of blinding and randomization does not always allow factors potentially influencing the outcomes to be properly balanced (**Figure 2**) (12).

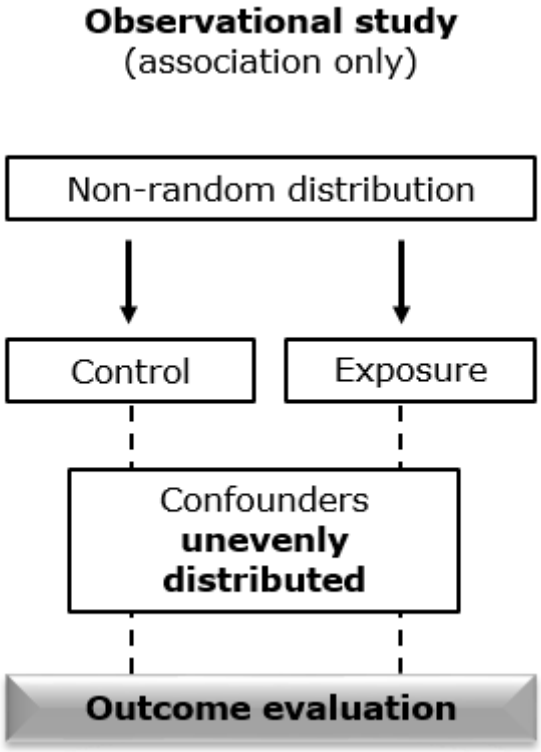



**Figure 2.** Observational study design.

Under real-life conditions, in the absence of randomization, severity of the underlying disease influences treatment decisions (22). This results in "confounding by indication", meaning that the perception of a different prognosis leads the physician to preferentially prescribe one of the available treatments. As a result, prognostic factors are

not equally distributed among the patients treated with different therapies and the comparison is affected by disease severity, when patients treated with one medication are systematically different from the other groups in terms of illness stage (23). In addition, as it is common clinical practice to follow severe patients more closely, this difference in monitoring influences the results of real-life studies. Finally, in observational studies, the choice to treat is not strictly regulated and can be made at any time during the disease history/evolution.

Although differences between groups can be minimized with statistical effective methods, the absence of randomization clearly limits the reliability of conclusions from real-life studies anyway.

In many cases, differences between the average outcomes in the two groups have been interpreted as evidence for the causal role of the risk factor. However, such a conclusion could confuse correlation with causation. Interpreting an association between an exposure and a disease outcome in observational data as a causal relationship relies on untestable and usually implausible assumptions, such as the absence of unmeasured confounding and of reverse causation (24), as mentioned before. Hence, more robust approaches are needed for assessing the causal relationship using observational data.

## 1.4 The rise of genetic epidemiology

The concept of inherited characteristics goes back to the dawn of time, although the mechanism for inheritance was long unknown. When Charles Darwin proposed his theory of evolution in 1859, one of its major problems was the lack of an underlying mechanism for heredity. Gregor Mendel in 1866 proposed two laws of inheritance: the law of segregation, that when any individual produces gametes

(sex cells), the copies of a gene separate so that each gamete receives only one copy; and the law of independent assortment, that "unlinked or distantly linked segregating gene pairs assort independently at meiosis". These laws are summarized by the term "Mendelian inheritance" (which gives Mendelian randomization its name), specifically due to the second law, the law of independent assortment (25). The two areas of evolution and Mendelian inheritance were brought together through the 1910s-30s in the "modern evolutionary synthesis", by amongst others Ronald Fisher, who helped to develop population genetics. The link between genetics and disease was established by Linus Pauling in 1949, who linked a specific genetic mutation in patients with sickle-cell anaemia to a demonstrated change in the haemoglobin of the red-blood cells of affected individuals (26). The discovery of the structure of deoxyribonucleic acid (DNA) in 1953 gave rise to the birth of molecular biology, which led to greater understanding of the genetic code. The Human Genome Project was established in 1990, leading to the publication of the entirety of the human genetic code by 2003 (27, 28). Recently, technological advances have reduced the cost of DNA sequencing, so that it is now economically viable to measure genetic information for a large number of individuals.

As the knowledge of the human genome developed, the search for genetic determinants of disease expanded from monogenetic disorders (that is, disorders which are due to a single mutated gene), to polygenic and multifactorial disorders, where the burden of disease risk is not due to a single gene, but to multiple genes combined with lifestyle and environmental factors. These diseases, such as cancers, diabetes, and coronary heart disease (CHD), tend to cluster within families, but also depend on other factors, such as diet or blood pressure. Several genetic factors have been found which relate to

diseases, especially through the increased use of whole-genome scans known as genome wide association studies (GWAS). GWAS is an approach used in genetics research to associate specific genetic variations with particular diseases. The method involves scanning the genomes from many different people and looking for genetic markers that can be used to predict the presence of a disease. Once such genetic markers are identified, they can be used to understand how genes contribute to the disease and to develop better prevention and treatment strategies, increasing public and professional awareness of the genetic contribution to some diseases.

Genetic epidemiology (which is the study of the role of genetic factors in health and disease) is a relatively new science. It has the potential to improve the understanding of environmental as well as genetic determinants of disease, using biochemical markers —and in particular, genomic and epigenomic information— to increase the resolving power of traditional epidemiologic methods (29). Genetics has advanced by using epidemiological methods, and this new field of epidemiology has also advanced by drawing on genetic principles. Genetic epidemiology studies can be grouped into those that seek to identify new genetic variants that cause disease (gene discovery studies) or those that aim to understand the importance of these variants in terms of the frequency or size of effect (gene characterisation studies). The former studies are often conducted in special groups of people or populations with particularly high disease incidence or risk, whereas the latter are typically conducted in representative groups with careful sampling so that results are generalizable to the wider population.

The clinical outcomes of genetic epidemiology are (30):

- improved understanding of the aetiological gene–environment interactions for many of the major conditions (for example,

CHD, diabetes and cancers);

- increased genetic testing, from pre-conception onwards;
- describing new taxonomies of pathophysiology and disease, based on molecular classifications rather than signs and symptoms, with distinct information about prognosis and treatment;
- leading to targeted prevention and prognosis based on genetic or molecular factors;
- linking with pharmacogenetics, towards targeted therapeutic drug strategies;
- determining the heritability and familial aggregation of diseases or intermediate phenotypes, and thus directing further gene discovery studies.

All these aspects link to the ultimate goal of developing specific intervention strategies for genetic primary care conditions ("personalised medicine").

Genetic epidemiological studies require the enrolment of large population, mainly for three main reasons. First, association studies based on small, highly selected samples, which attempt to identify genes associated with specific traits, can often be underpowered, resulting in false negative results (31). This is partly because the linkage between gene markers and specific genes may be relatively weak, partly because the prevalence of a gene polymorphism may be relatively rare or variable, and partly because the genetic contribution to the trait may be complex or weak. Secondly, in order to examine the interaction between genetic and environmental factors, very large samples are required for most phenotypes to compensate for the background "noise" produced by other, non-genetic factors (32). Thirdly, the risk of a disease or trait associated with a particular genetic variant calculated from an unrepresentative

population, selected perhaps for its high incidence or risk of this trait, cannot be extrapolated to the general population. Large, population-based samples are therefore required to calculate the absolute risks associated with any genetic variant.

Starting from this, the next chapter will present an introduction to Mendelian randomization: a method for using genetic data to estimate causal associations of modifiable (non-genetic) risk factors using observational data.

# CHAPTER 2

## Mendelian randomization study

## 2.1 What is Mendelian randomization?

Mendelian randomization (MR) is commonly defined as the use of non-experimental studies to determine the causal effect of a phenotype on an outcome by making use of genetic variation. The word "phenotype" refers to the putative causal risk factor, which can be thought of as an exposure, a biomarker, or any other risk factor which may affect the outcome (33). Usually, the outcome is a disease, although there is no methodological restriction as to what outcomes can be considered. Non-experimental studies encompass all observational studies, including cross-sectional, cohort, and case-control designs, where there is no intervention imposed by the researcher. These are contrasted with clinical trials.

### 2.1.1 Motivation

A foundational aim of epidemiological studies is the estimation of the effect of changing one risk factor on an outcome (25). This is known as the causal effect of the phenotype on the outcome and typically differs from the observational association between phenotype and outcome, due to endogeneity of the phenotype. Endogeneity, literally "coming from within", of a variable in an equation means that there is a correlation between the variable and the error term, and occurs when the variable is predicted by the terms in the model in which it appears (34). For example, those who regularly take headache tablets are likely to have more headaches than those who do not, but taking headache tablets is unlikely to be a cause of the increased incidence of headaches. Taking tablets is an endogenous variable in this context, and so the causal effect of taking tablets on headaches cannot be estimated from this observational setting. The opposite of endogenous is exogenous; an exogenous variable comes from

outside of the model and is not explained by the terms in the model. The idea of MR is to find an exogenous genetic variant (or variants) which is associated with the phenotype, but is not associated with any other risk factor which affects the outcome, and is not directly associated with the outcome, in that any impact of the genetic variant on the outcome must come via its association with the phenotype (35). These assumptions define an instrumental variable (IV) (36, 37).

### 2.1.2 Instrumental variables

An alternative definition of MR is "instrumental variable analysis using genetic instruments" (38). The use of genetic variants as IVs is at the core of MR. An IV is an exogenous variable associated with an endogenous exposure which is used to estimate the causal effect of changing the exposure while keeping all other factors equal (39). The choice of the genetic IV is essential to a successful MR study. To allow unbiased estimation of the causal effect of the exposure on the outcome, a valid genetic IV fulfils three core assumptions (**Figure 3** panel A) (36, 40):

1) it must be reproducibly and strongly associated with the exposure (**Figure 3** panel B);
2) it must not be associated with confounders, i.e., factors that confound the relationship between exposure and outcome (**Figure 3** panel C);
3) it is only associated with the outcome through the exposure, i.e., it is independent of the outcome given the exposure (**Figure 3** panel D).

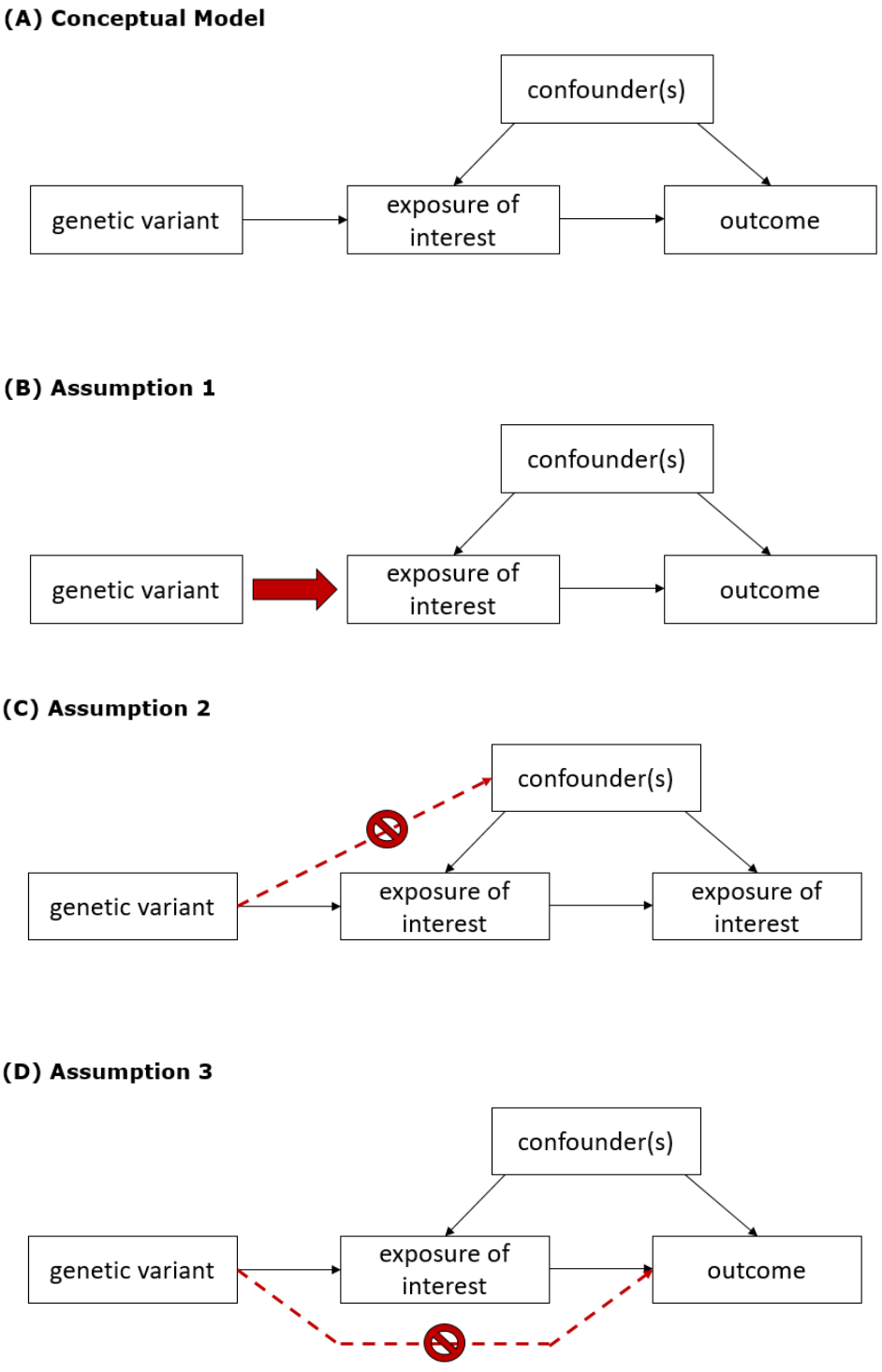



**Figure 3.** Conceptual illustration of the MR method and its three underlying core assumptions as directed acyclic graphs. (A) Conceptual model; (B) Assumption 1; (C) Assumption 2; (D) Assumption 3.

Although MR analyses often involve a single genetic variant, multiple variants can be used either as separate IVs or combined into a single IV. The first assumption guarantees that genetic subgroups defined by the variant will have different average levels of the exposure. This ensures that there is a systematic difference between the subgroups. If the genetic variant is not strongly associated with the exposure (in the sense of its statistical strength of association), then it is referred to as a weak instrument. A weak instrument differs from an invalid instrument in that a weak instrument can be made stronger by collecting more data. If a single genetic variant is a weak instrument, then it will still give a valid test of the null hypothesis of no causal effect, but the power to detect a true causal effect may be low. However, combining multiple weak instruments in an analysis model to obtain a single effect estimate can lead to misleading inferences. The second assumption can be understood as ensuring that the comparison between the genetic subgroups is a fair test, that is, all other variables are distributed equally between the subgroups. The third assumption is often expressed using the concept of conditional independence as "the genetic variant is not associated with the outcome conditional on the value of the exposure and confounders of the exposure–outcome association". It ensures that the only causal pathway(s) from the genetic variant to the outcome are via the exposure. This means that the genetic variant is not directly associated with the outcome, nor is there any alternative pathway by which the variant is associated with the outcome other than through the exposure.

### 2.1.3 Analogy with randomized controlled trials

MR has been defined as analogous to a RCT. As reported before, a RCT, considered the "gold standard" of medical evidence, involves

dividing a target population into two or more subgroups in a random way. These subgroups are each given different treatment programmes. Randomization is preferred over any other assignment to subgroups as all possible confounders, known and unknown, are on average balanced. However, in many situations, for ethical or practical reasons, it is not possible to intervene on the factor of interest to estimate the causal effect by direct experiment. In MR, we use the IV to form subgroups analogous to those in a RCT, as shown in **Figure 4**.

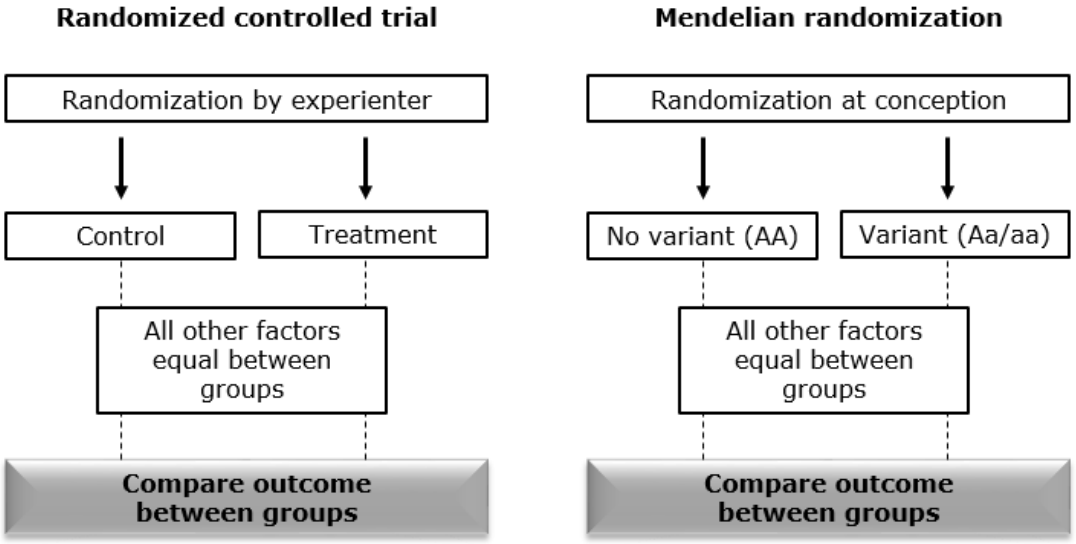



**Figure 4.** Comparison of randomized controlled trial and Mendelian randomization design.

From the IV assumptions, these subgroups differ systematically in the exposure, but not in any other factor except for those causally "downstream" of the exposure. A difference in outcomes between these subgroups would therefore indicate a causal effect of the exposure on the outcome (41). Inferring a causal effect of the exposure on the outcome from an association between the genetic variant and the outcome is analogous to inferring an intention-to-

treat effect from an association between randomization and the outcome in an RCT (that is, assignment to the treatment group affects the outcome). Genetic variants for an individual are inherited from their parents, and so are not randomly assigned. For example, if neither of an individual's parents carry a particular genetic mutation, there is no way that the individual will carry that mutation. Nonetheless, under fairly realistic conditions, the distribution of genetic variants in the population can be thought of as random with respect to environmental and social factors which may be important confounders. The necessary assumptions for a variant to be randomly distributed are random mating and lack of selection effects relating to the variant of interest. Considerable departures from the random mating assumptions which may invalidate the use of a genetic variant can be assessed by performing a test of Hardy–Weinberg equilibrium, to see if the frequency of heterozygotes and homozygotes in the population is in line with what is expected. A variable which is distributed as if being randomly assigned despite the lack of true randomness in the assignment is known as quasi-randomized. Most natural experiments rely on quasi-randomization rather than the strict randomization of experimental units.

However, MR differs from a RCT in another respect. The aim of MR is not to estimate the size of a genetic effect, but the causal effect of the exposure on the outcome. The average change in the outcome associated with a genetic variant may differ in magnitude from that resulting from an intervention in the exposure. When the proportion of variation in the phenotype associated with the genetic variant is not large or is imprecisely estimated, studies will require large sample sizes, such as 10,000 or even 30,000 cases (42), as the risk ratio from the difference in phenotype due to the genetic variant may be low. However, the population attributable risk of the phenotype is

not necessarily low. Although the variation in phenotype attributable to the gene may be small, it can be similar to that attributable to treatment in a RCT (43).

### 2.1.4 Violations of IV assumptions

Sometimes, the use of a particular genetic variant as an IV is controversial as the assumptions presented above cannot be fully tested and may be violated for various epidemiological and biological reasons. Overall, justification of the assumptions relies on biological knowledge about the genetic markers in question. Among reasons why a genetic variant may not be a valid IV there are issues of biological mechanism, genetic coinheritance, and population effects (44). Invalid IVs lead to unreliable inferences for the causal effect of an exposure.

The first category of issues resulting in violations of the IV assumptions is the underlying biological mechanism.

- **Pleiotropy.** Pleiotropy refers to a genetic variant being associated with multiple risk factors. If a genetic variant used as an IV is also associated with another risk factor for the same outcome, then either the second or the third IV assumption is violated (depending on whether the risk factor is a confounder of the exposure–outcome association or not), and the variant is not a valid IV (45). If the genetic variant is associated with an additional variable solely due to mediation of the genetic association via the exposure of interest (sometimes called vertical pleiotropy), this is not regarded as pleiotropy. Concerns about pleiotropy can be alleviated by using genetic variants located in genes, the biological function of which are well-understood (**Figure 5**).

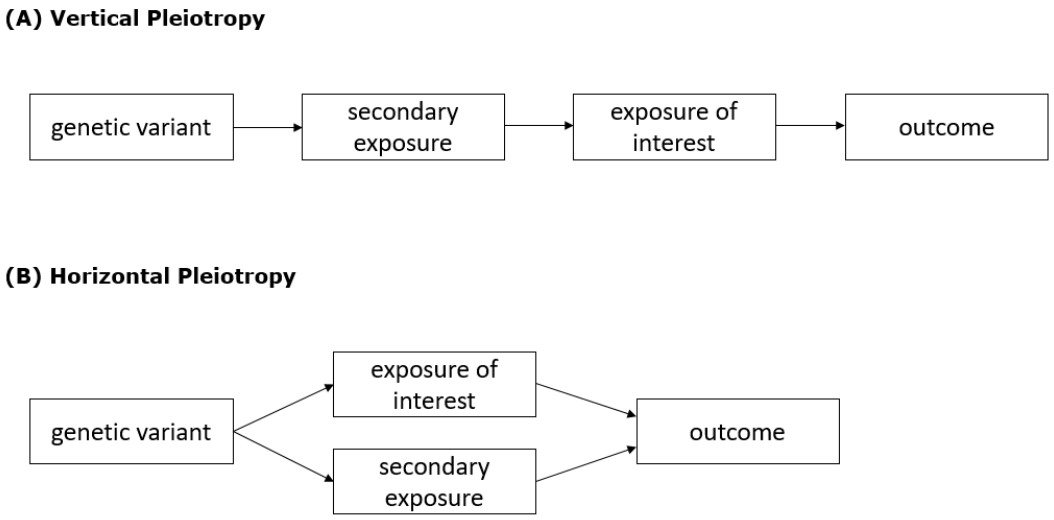



**Figure 5.** Explanations for the pleiotropy. An example of vertical pleiotropy (A), and an example of horizontal pleiotropy (B).

- **Canalization.** Canalization, or developmental compensation, is the phenomenon by which an individual adapts in response to genetic change in such a way that the expected effect of the change is reduced or absent (46). Often the organism develops a compensatory mechanism to allow for the missing gene such that the functionality of the gene is expressed via a different biological pathway. This buffering of the genetic effect may have downstream effects on other variables. Canalization may be a problem in MR if groups with different levels of the genetic variants differ with respect not only to the exposure of interest, but also to other risk factors via a canalization mechanism. In a sense, canalization is not a violation of the IV assumptions, but merely an (often unwanted) consequence. Canalization is the same process as that assessed by MR, as any change in other risk factors from canalization occurs as a causal effect of the genetic variant. However, the aim of MR is not simply to describe the effects of genetic change, but to assess the causal

effect of the (non-genetic) exposure. If there is substantial canalization, MR estimates may be unrepresentative of clinical interventions on the exposure performed in a cohort.

The second category of issues resulting in violations of the IV assumptions is the non-Mendelian inheritance. Although Mendelian principles state that separate characteristics are inherited separately, this is not always true. Non-Mendelian inheritance refers to patterns of inheritance which do not correspond to Mendel's laws, specifically the law of independent assortment.

- **Linkage disequilibrium.** One particular reason for genetic variants to be inherited together is the physical proximity of the variants on the same chromosome (47). Variants whose distributions are correlated are said to be in linkage disequilibrium (LD). LD has both desirable and undesirable consequences. If genetic variants were truly independently distributed, then only the genetic variant which was causally responsible for variation in the exposure could be used as an IV, as all other genetic variants would not be associated with the exposure. An undesirable consequence of LD is that genetic variants correlated with the variant used in the analysis may have effects on competing risk factors (**Figure 6**). This would lead to the violation of the second or the third IV assumption (similar to violations due to pleiotropy). Concerns about invalid inferences due to LD can be alleviated by empirical testing of the association of known potential confounders with the measured variant.

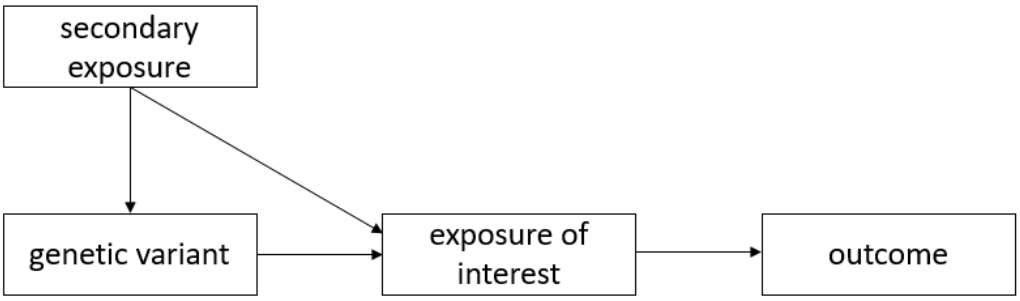



**Figure 6.** Explanations for linkage disequilibrium.

- **Effect modification.** Unlike confounding, effect modification relates to a statistical interaction between the effect of a variable (usually an effect of the exposure) and the value of a covariate, leading to the causal effect of the exposure varying across strata defined by the covariate. Factors that may lead to effect modification include issues of non-Mendelian inheritance, such epigenetic variation (48) and parent-of-origin effects (49). Effect modification alone is unlikely to represent a violation of the IV assumptions; however, it may lead to difficulties in interpreting MR investigations.

The third category of issues resulting in violations of the IV assumptions is about population effects.

- **Population stratification.** Population stratification occurs when the population under investigation can be divided into distinct subpopulations (50). This may occur, for example, when the population is a mixture of individuals of different ethnic origins. If the frequency of the genetic variant and the distribution of the exposure are different in the different subpopulations, a spurious association between the variant and the exposure will be induced which is due to subpopulation differences, not the effect of the genetic variant. Concerns

about population stratification can be alleviated by restricting the study population to those with the same ethnic background (although there may still be differences associated with ancestry in broadly-defined ethnic groups). In a GWAS, genomic control approaches, such as adjustment for genetic principal components, are possible.

- **Ascertainment effects.** If the genetic variant is associated with recruitment into the study, then the relative proportions of individuals in each genetic subgroup are not the same as those in the population, and so a genetic association with the outcome in the sample may not be present in the original population (51). If the study cohort is taken from the general population, ascertainment effects are unlikely to be a major problem in practice. However, if, for example, the study cohort is composed by pregnant mothers, and the genetic variant is associated with fertility, then the distributions of the covariates in the genetic subgroups will differ and not be the same as those in the general population. This may introduce bias in the estimation of causal effects, as there is a pathway opened up from the genetic variant to the outcome by conditioning on a common cause of the variant and the outcome (sometimes called collider bias). This would also be a problem in studies looking at genetic associations in populations of individuals with pathological conditions, such as clinical trials of secondary disease prevention. Individuals with greater genetically determined disease risk are less likely to survive to study recruitment, and so the randomization of individuals into genetic subgroups at conception would not hold in the study population, leading to biased genetic associations.

Although it is not possible to demonstrate conclusively the validity of

the IV assumptions, several tests and assessments are possible to increase or decrease confidence in the use of genetic variants as IVs. The simplest assessment of instrument validity is to test the association between the genetic variant and known confounders. Association of the variant with a covariate associated with the outcome which is not on the causal pathway between the exposure and outcome would violate the second IV assumption. However, there is no definitive way to tell whether the association with the covariate is due to violation of the IV assumptions (such as by pleiotropy or linkage disequilibrium) or due to mediation through the exposure of interest. Additionally, there is no way of testing whether or not the variant is associated with an unmeasured confounder. Other mathematical results for testing IV validity are available (52), but these are only likely to detect gross violations of the IV assumptions. Biological knowledge rather than statistical testing should form the backbone of any justification of the use of a particular genetic variant as an IV in MR.

## 2.2 Genetic markers

Generally, in MR, genetic markers used as IVs are single nucleotide polymorphisms (SNPs) (53). A SNP is defined as a modification in the DNA of an individual compared to the population at a single point (or locus), where one nucleotide, either A, C, G or T, has been replaced with another. These different variants in the genetic code are called alleles. Where there are two possible alleles at a particular locus (a diallelic SNP), the more common allele (the major allele or wild type) is reported as "A" and the less common allele (the minor allele or variant) as "a". The proportion of minor alleles in a population is called the "minor allele frequency". An arbitrary threshold of the

minor allele frequency is set at 1%, below which a SNP is considered a mutation rather than a polymorphism. As people have two copies of each DNA sequence, individuals can be categorized for each diallelic SNP into three possible subgroups corresponding to three combinations of alleles. These subgroups are named major homozygotes (AA), heterozygotes (Aa) and minor homozygotes (aa). We shall denote these subgroups as 0, 1 and 2, corresponding to the number of minor alleles for that SNP. For this reason, a diallelic SNP is usually considered to be a discrete random variable taking values from 0, 1, 2.

## 2.3 Multiple instruments

Although IV methods give estimates which are consistent for the causal effect, their variance is typically much larger than the variance of the estimate from an observational analysis (42). This is because the variation in the exposure explained by the IV is usually small. If there are multiple IVs available, a more precise causal effect estimate can be obtained by incorporating data on all the IVs simultaneously to estimate a single causal effect (54). However, a problem arising from including multiple IVs in an analysis is weak instruments. When there are large numbers of genetic variants, several IV methods give estimates which are biased in the direction of the observational estimate with incorrectly sized confidence intervals. Allele scores are a convenient way of summarizing a large number of genetic variants associated with an exposure. Using a univariate allele score as a single IV rather than each genetic variant as a separate IV helps resolve problems in IV estimation resulting from weak instruments. Using IVs which explain a greater proportion of the variance in the exposure (which can be achieved by including more genetic variants

in an analysis) leads to the gain in statistical power.

More specifically, an allele score (also called a genetic risk score, gene score, or genotype score) is a single variable summarizing multiple genetic variants in a univariate score. An unweighted allele score is constructed as the total number of exposure-increasing alleles present in the genotype of an individual. A weighted allele score can also be considered, where each allele contributes a weight reflecting the effect of the corresponding genetic variant on the exposure. These weights can be derived internally from the data under analysis, or externally from prior knowledge or an independent data source (55). The use of an allele score in MR requires the score to satisfy the assumptions for being an IV. This means that each variant which contributes to the allele score must satisfy the assumptions, except that it is not necessary for all the variants to be associated with the exposure (a variant not strongly associated with the exposure but satisfying the second and third IV assumptions will not invalidate the score, but neither will it add any information to the score) (36).

The choice of variants to be included in an allele score should be made prior to analysis, or on the basis of external (independent) data. This is particularly important if there are several candidate variants with similar magnitudes of association with the exposure. Additionally, the inclusion of variants which are highly correlated with each other (in high linkage disequilibrium) will not give extra information compared to including any one of these variants, and may lead to inefficiency if the correlation is not taken into account in determining the weights.

## 2.4 Types of Mendelian randomization studies

MR studies can be performed using several different strategies, with combinations of either one or two study samples to gain information on gene-risk factor and gene-outcome associations and level of detail on study participants with individual-level, study-level, and summary-level data (56).

The standard study design is a one-sample MR study using individual-level data. This corresponds to the study design shown in **Figure 7** (black arrows) and is often carried out using data from one population study. Advantages of this design are that (i) detailed information on potential confounding and mediating factors may be available and can be examined and accounted for; (ii) the assumptions necessary for the validity of the genetic variants can be tested, including testing for potential pleiotropy; (iii) the use of a population with known ethnicity reduces the risk of population stratification; (iv) it can be tested whether an additive or multiplicative per allele model fits the data best, resulting in more precise causal estimates; and (v) valuable information on the observational association between the risk factor and the outcome may be included (**Figure 7**, black arrow #1). Two variants of the classic one-sample MR study using individual-level data are the two-step MR study, testing whether the effect of the biomarker or lifestyle factor under examination is mediated through other measured factors on a causal pathway, and the bi-directional MR, aimed to assess the direction of causation (57, 58). It is often claimed that a potential limitation of the classic design is that the genotype-risk factor and genotype-outcome associations are correlated since they are obtained using the same individuals, and that this may bias a causal effect of the biomarker in the same direction as the

observational estimate (**Figure 7**, arrows #2 and #3). However, use of the classic design in a large homogenous study cohort with its many other advantages, including low risk of chance findings, will out weight this minor issue.

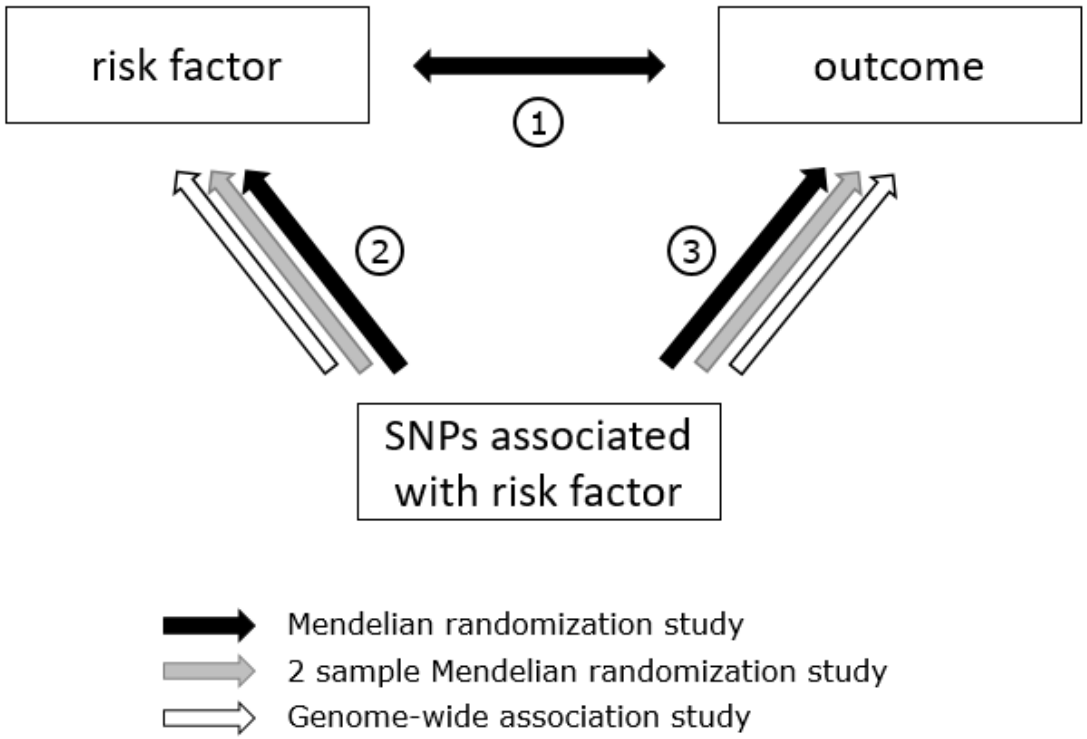



**Figure 7.** Types of Mendelian randomization designs.

MR can also be performed using study-level data where causal estimates from several studies are meta-analysed, increasing the statistical power of the combined study. Potential limitations are similar to the limitations of conventional meta-analyses, which are publication bias, inclusion of small studies which tend to show larger causal effects, and heterogeneity among studies included. Also, variation in the risk factor or lifestyle factor under examination and in the definition of endpoints between studies and affecting the causal estimate should be taken into account. Another study design is a two

sample MR study using two independent samples (**Figure 7**, grey arrows #2 and #3). This design can be used on individual level data, but is often carried out using summary level data, with a beta-coefficient and the standard error of the mean from the regression of the risk factor on the genotype from one study, and similar data for the regression of the outcome on the genotype from another study. Advantages of this study design are that (i) that causality can be inferred without information on the observational association (**Figure 7**, #1), (ii) the genotype-risk factor and genotype-outcome associations are not correlated and a bias will be in the direction of the null hypothesis, and (iii) data may be collected from very large GWAS where information on the genotype-outcome association comes from case–control studies with more cases than population studies, and thus a high statistical power. However, genetic variants identified in GWAS are common and often have small phenotypic effects, thus potentially introducing bias because of weak instruments. Requirements for the two-sample MR are that the samples included must not be overlapping, should be of similar age and gender distribution and ethnicity; and genetic variants should be completely independent and thus not in linkage disequilibrium (58). Limitations of the two sample MR study are that data from GWAS often use a case–control design, where selection of cases and controls may have introduced ascertainment bias, and inclusion of individuals from several populations may introduce population stratification.

Ideally, publications should present information both from individual-level data of own studies and if available, combined with summary-level data. This combination may provide information on confounding factors and mediating factors in biological pathways, with a high statistical power.

## 2.5 Why use Mendelian randomization?

Although the main reason to use MR is to avoid the problem of residual confounding, there are additional reasons for using MR in specific contexts: with case-control data and with exposures that are difficult to measure.

Reverse causation occurs when an association between the exposure and the outcome is not due to the exposure causing a change in the outcome, but the outcome causing a change in the exposure (42). This could happen if the exposure increased in response to pre-clinical disease, for example from cancer before it becomes clinically apparent or from atherosclerosis prior to clinical manifestations of CHD. As the genotype of an individual is determined at conception and cannot be changed, there is no possibility of reverse causation being responsible for an association between genotype and disease (59). For this reason, MR has great strengths in a retrospective setting where genetic variants are measured after the disease outcome, such as in a case-control study. Many exposures of interest cannot be reliably measured in cases, that is in individuals who have already experienced an outcome event, as the event may distort the measurement. In this case, the genetic variant can be used as a proxy for the exposure, and the genetic association with the outcome can be assessed retrospectively. As the genotype of an individual can be measured in diseased individuals, causal inferences can be obtained using MR in a case-control setting. MR can be also a useful technique when the exposure of interest is expensive or difficult to measure. For example, gold standard assays for biomarkers such as water-soluble vitamins may cost too much to be affordable for a large sample, or measurement of fasting blood glucose, which requires overnight fasting, may be impractical. If the genetic variant is

associated with the exposure and is a valid IV for the exposure, a causal relationship between the exposure and outcome can be inferred from an association between the genetic variant and the outcome even in the absence of measurement of the exposure. Additionally, instrumental variable estimates do not attenuate due to classical measurement error (including within-individual variation) in the exposure (60). This contrasts with observational studies, in which measurement error in the exposure usually leads to the attenuation of regression coefficients towards the null (known as regression dilution bias) (61). A further example is where the risk factor is not only difficult to measure, but also difficult to define. For example, a variant in the IL6R gene region that is associated with serum interleukin-6 (IL-6) concentrations (as well as levels of downstream inflammatory markers, including C-reactive protein [CRP] and fibrinogen) was shown to be associated with CHD risk (62). However, from knowledge about the functional role of the variant, the causal effect assessed is not thought to operate through elevated serum interleukin-6 concentrations, but rather through changes in signalling in interleukin-6 receptor pathways. This is a cellular phenotype which varies over time, and so a representative measurement for an individual is not straightforward to define. However, as the genetic variant can be measured, the causal role of interleukin-6 receptor-related pathways on CHD risk can be assessed by MR (63).

# CHAPTER 3

## Cardiovascular disease

## 3.1 A brief overview

Despite significant advances in the treatment of Cardiovascular Disease (CVD), it continues to be the leading cause of mortality and morbidity globally (**Figure 8**) (64). Typically, CVD represents a cluster of disorders that are associated with the heart, the vasculature of the brain, or blood vessels, and predominantly includes coronary and ischemic heart disease, deep vein or arterial thrombosis and cerebrovascular disease (65).

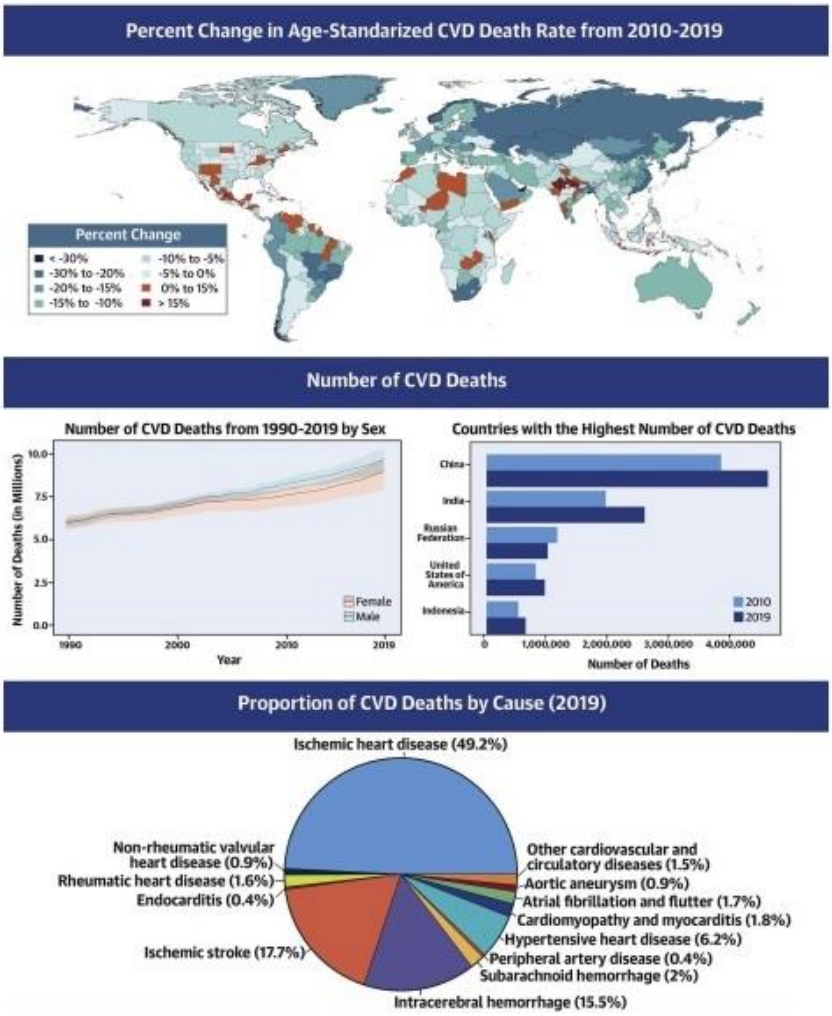



**Figure 8.** Cardiovascular disease burden across time, location, and cause. (66).

According to the World Health Organization (WHO), 17.9 million people die each year from CVDs, which accounts for an estimated 31% of all deaths worldwide. It has also been projected that approximately 23.3 million people will die annually from CVDs by 2030 (67). Additionally, in contrast to developed countries, where CVD mortality has decreased in the past few decades, the incidence of CVD mortality is on rise in low and middle-income countries, with an estimated account of more than 75% of the world's deaths. This is mainly attributed to the lack of preventive and treatment measures in these countries (68). Therefore, developing newer and better strategies to combat CVDs is of global importance and should be regarded as a priority by health care investigators. An improved and early diagnosis of CVDs may effectively shift the focus of the therapy from treatment to prevention and improve the overall survival.

Most of the CVDs are primarily of atherosclerotic origin (69). Atherosclerosis is a lipoprotein-driven disease that leads to plaque formation at specific sites of the arterial tree through intimal inflammation, necrosis, fibrosis, and calcification. After decades of indolent progression, such plaques may suddenly cause life-threatening coronary thrombosis presenting as an acute coronary syndrome. Atherosclerosis alone may obstruct coronary blood flow and cause stable angina pectoris, but this is rarely fatal in the absence of scarring of the myocardium, which can elicit an arrhythmia presenting as sudden cardiac death. Acute coronary syndrome is nearly always caused by a luminal thrombus or a sudden plaque haemorrhage imposed on an atherosclerotic plaque with or without concomitant vasospasm. Plaque rupture is the most frequent cause of thrombosis. In plaque rupture, a structural defect -a gap- in the fibrous cap exposes the highly thrombogenic core to the blood (70). In addition, a range of other factors also contributes towards

the progression of atherosclerosis, classified into non-modifiable and modifiable risk factors. The non-modifiable risk factors include age, sex and genetic causes, whereas, hypertension, diabetes, dyslipidaemia, smoking, poor diet, and obesity are important modifiable risk factors (71, 72).

### 3.1.1 Risk factors for cardiovascular disease

Traditional risk factors for CVD are well known (**Table 2**).

**Table 2.** Summary of risk factors for cardiovascular disease.

| **Category** | **Factors** |
|---|---|
| Non-modifiable risk factors | Advancing age |
| | Sex |
| | Family history |
| | Genetic predisposition |
| Metabolic risk factors | Hypertension |
| | Hyperlipidaemia |
| | Diabetes mellitus |
| | Metabolic syndrome |
| | Obesity/overweight |
| Lifestyle risk factors | Diet |
| | Smoking |
| | Physical activity |
| Novel risk factors | Elevated homocysteine level |
| | Elevated lipoprotein(a) level |
| | Small dense LDL-C |
| | Elevated inflammatory markers levels |
| | Elevated haemostatic factors levels |

Aging is the single most important factor that affects cardiovascular health. Its impact on CVD can be assessed from the fact that the risk

for cardiovascular morbidity increases by about 10-fold, between the age of 50 and 80 (73). Moreover, according to an estimate by the National Institute of Aging, by 2030, approximately 20% of the world population will be aged 65 or older, 40% of which would die as a consequence of CVD (74). Aging results in an increase in oxidative stress and a reduction in telomere length, causing DNA damage, along with impaired cell division and senescence of cardiovascular tissues (75). These factors greatly affect metabolic processes, the integrity of the vasculature, and cardiovascular repair mechanisms that render cardiovascular tissues vulnerable towards damage. In addition to this, aging has a remarkable effect on the arterial system and the heart, both of which play an important role in the development of CVDs. Aging results in thickening and stiffness of the arteries, which markedly affects their elasticity. It also causes endothelium dysfunction, which arises due to an imbalance between the production of vasodilators (such as nitric oxide, acetylcholine, and prostacyclin) and vasoconstrictors (such as angiotensin, nitric oxide synthases, leukotrienes, and thromboxanes). There is also a decrease in the affinity of vascular smooth muscle cells towards these regulators (76). Besides these, the heart also undergoes a series of aging-mediated changes in the structure and cellular composition. There is a substantial increase in oxidative stress, apoptosis, and necrosis that considerably reduce the absolute number of cardiomyocytes (77). It is now known that aging also affects the division and regeneration capability of cardiomyocytes, which prevents maintenance of an adequate number upon loss of these cells (78). Moreover, cardiomyocyte senescence, defined by the increased expression of senescence markers and decreased telomere length, also increases with age (74). Collectively, these outcomes of aging influences other cardiovascular risk factors that includes

atherosclerosis, hypertension, diabetes, which may lead to heart failure or stroke (79, 80).

Although CVD is the leading cause of death in both men and women it has been well established that men are more prone towards CVDs, while there is a delayed onset of CVDs in women (81, 82). The Framingham Study comprehensively investigated CVDs for over five decades in a mixed-gender general population sample and concluded that women demonstrate delayed inception of atherosclerosis associated CVD events by 10-20 years in comparison to men (83). This is primarily due to sex-based differences in hormones that regulate cardiovascular functions in both women and men. It has been studied that endogenous oestrogens in women exert diverse cardio-protective effects such as defending blood vessels from atherosclerotic lesion formation, reducing the levels of low-Density Lipoprotein cholesterol (LDL-C) while augmenting high-Density Lipoprotein cholesterol (HDL-C) concentration in plasma (84). Premenopausal women with normal oestrogenic levels are less prone to developing CVDs in comparison to men, while early menopause or bilateral oophorectomy increases lifetime cardiovascular risk (85). Additionally, other biological distinctions, such as a smaller artery dimension and different plaque composition in women, retard the progression of CVDs in comparison to men (86, 87). Diabetes mellitus, hypertension, smoking, hypercholesterolemia, and obesity are conventional risk factors that have a more profound contribution in women than men towards the development of CVDs (88). In contrast, men are generally more exposed to deleterious cardiac risk factors, such as the increased propensity of smoking and higher alcohol consumption, low-fibre diet, low vitamin C levels, and high blood viscosity (89). However, as discussed previously, all these risk factors affect women more vigorously after the age of 55 or 65, in

contrast to men, so that in women younger than 50 years, smoking is the leading cause of CVDs (87, 90). It is also important to note that diabetes reverses the protective effects of oestrogens in women and can result in a 3 to 7 fold increased CVD risk in women compared with men (91). CHD affects both men and women; however, it has been found that it is 2 to 5 times more prominent in men in comparison to women of the younger age group (92). Moreover, women are widely susceptible to dying from stroke and heart failure in contrast to men (93).

High blood pressure is regarded as one of the major underlying causes for the onset of almost all the CVDs, which include CHD, left ventricular hypertrophy, valvular heart diseases, atrial fibrillation, atherosclerosis, and cerebral stroke. Hypertension is an outcome of dysfunction in blood vessels that may be due to calcification or loss of elastin (causing reduced elasticity), in addition to the inability of smooth muscle cells to contract or relax. The incidence of high blood pressure increases with age and it has been estimated to affect 65% of the people aging more than 60 years (94). It is also estimated that globally 54% of strokes and 47% of CHDs are an outcome of hypertension (95). According to the Framingham Heart Study, an elevation of 20 mmHg systolic blood pressure can enhance the risk of heart failure by 50% (96). Numerous RCTs have shown that antihypertensive treatment reduces the risk of stroke and CHD by 40% and 16% respectively (97). Hypertension is also associated with the modification in the blood vessels, myocardial structure, coronary vasculature, and conduction system of the heart. This is referred to as remodelling and is regarded as an adaptive response against prolonged hemodynamic changes. These alterations subsequently impair myocardial performance and can lead to ventricular hypertrophy, CAD, cardiac arrhythmias (especially atrial fibrillation)

(98). Hypertension is also a major independent risk factor that can accelerate the progression of atherosclerotic CVD (99). The exact mechanism by which hypertension regulates the pathogenesis of atherosclerosis is not clearly understood. However, it is believed that high blood pressure exerts proinflammatory effects in the arteries through vasoactive peptides, such as angiotensin and endothelin-1, leading to the recruitment of monocytes into the intima, which is a prerequisite for the development of atherosclerosis (100). Furthermore, in comparison to normotensive people, the extent of fatty streaks and fibrous plaque build-up is heightened in the coronary arteries and aorta of hypertensive people. It has also been reported that plaques rarely develop solely on account of hypertension, which only promotes the progression of atherosclerosis, in the presence of hypercholesterolemia that has much more pronounced effect on atherosclerosis itself (101). It is also worth noting that hypertension generally co-exist with other risk factors such as obesity, dyslipidaemia and dysglycaemia, and only a small fraction of the population has hypertension as an isolated cause of CVDs. Moreover, when present concurrently, hypertension and other risk factors can substantially lead to a cardiovascular risk, which is higher than the sum of risk conferred by the individual factors (102).

Presence of an abnormally high level of lipids (such as cholesterol and triglycerides) in the blood is referred to as hyperlipidaemia. It can be an outcome of a sedentary lifestyle, smoking, or consumption of a diet rich in saturated fats and cholesterol. Diabetes, hypothyroidism, and obesity are also important causative factors. Furthermore, patients with familial hypercholesterolemia (a genetic disorder) are more prone to developing hyperlipidaemia at an early age (103). People with high cholesterol have approximately twice the

risk of heart disease as people with lower levels, but many are unaware of their condition because there are no symptoms (104). Physiologically desirable total cholesterol levels are up to 200 mg/dL and it has been estimated that a rise in total cholesterol in men from 200 to 240 mg/dL results in a 3-fold increase in deaths from cardiac diseases (105). Lipids are not soluble in plasma and are transported by specific kind of transporter proteins known as lipoproteins, which are classified, in order of increasing density, as: chylomicrons, Very-Low-Density Lipoprotein (VLDL), Low-Density Lipoprotein (LDL), and High-Density Lipoprotein (HDL). Of these, LDL and HDL are regarded as important markers of determining dyslipidaemia and have a profound contribution to the development of CVDs. LDL transport cholesterol from the liver to the cells, whereas HDL participates in the removal of excess cholesterol from different tissues and transferring it back to the liver (106). Up to 100 mg/dl for LDL and not less than 50 mg/dl of HDL is regarded to be a physiologically normal concentration (107). Elevated levels of LDL lipids in plasma tend to settle on the arterial walls, leading to their progressive hardening, along with the formation of an atherosclerotic plaque that can significantly affect the supply of oxygenated blood to the tissues, leading to ischemia (108). Besides this, hyperlipidaemia has also been shown to promote platelet activation through a variety of mechanisms, increasing the risk of thrombosis (109, 110). Such an event occurring in the heart or brain may lead to myocardial infarction or stroke respectively, which is life-threatening. Therefore, given the severity and level of threat associated with hyperlipidaemia, its early detection and treatment are imperative to prevent atherosclerotic and other related cardiovascular risks.

Type-2 Diabetes Mellitus (DM) is a metabolic disorder, characterized by insulin resistance and beta-cell impairment that leads to

hyperglycaemia (high blood glucose level). It is one of the most dominantly existing chronic diseases globally and its incidence is on a rapid and progressive rise. In the past three decades, the global burden of diabetes has increased markedly from 30 million to over 400 million presently (111). According to a recent estimate by international diabetes federation, there will be 592 million people suffering from type-2 DM by 2035, which is closely 1 in every 10 people (112). It has been widely reported that in contrast to non-diabetic patients, adults with diabetes exhibit a 2 to 4 times higher risk of developing CHD, making type-2 DM an important and independent risk factor for the development of CVDs (113). Mortality from stroke is also increased by almost 3-folds when patients with diabetes are compared with those without diabetes (114). Moreover, CVDs in diabetic patients exhibit a considerably poor prognosis for survival in comparison with CVD patients without diabetes (114). In general, patients with type-2 DM are also prone to other exiting classical cardiovascular risk factors such as hyperlipidaemia, hypertension, and obesity that can significantly augment the risk of developing CVDs.

Accumulating evidence suggests a strong connection between hyperglycaemia and atherosclerosis. For instance, diabetes can significantly enhance the likelihood of severe carotid atherosclerosis (115). This is because of several reasons. Firstly, diabetic blood is more likely to be rich in triglycerides due to impaired lipid flux, a function which is regulated by insulin (116). In addition to a high level of triglycerides and decreased HDL-C in the plasma, abnormalities have also been noticed in the structure of LDL particles of diabetic patients. There exists a smaller and dense form of LDL in diabetic blood, which is more atherogenic given its increased ability to penetrate the arterial wall along with higher susceptibility to

oxidation (117). Apart from oxidation, increased glycation of LDL particles has also been observed in patients with diabetes, which considerably lengthens their half-life (118). Besides these factors, high glucose levels also result in hypercoagulability of the blood and endothelial dysfunction, leading to platelet activation, leukocyte adhesion, thrombogenesis, and inflammation, which concomitantly accelerates atherosclerosis, making diabetic patient vulnerable to the risk of heart failure (119, 120). Diabetes also features a state of chronic and low-level inflammation (120). Numerous studies have confirmed a reduction in the secretion of potent vasodilator, nitric oxide, coupled with increased secretion of the vasoconstrictor and growth factor endothelin-1, in patients with diabetes. This condition not only leads to vasoconstriction but is also associated with the release of pro-inflammatory cytokines that arbitrates a strong relationship between diabetes, inflammation, and CVD (121). Furthermore, numerous other pathological factors such as increased oxidative stress and autonomic neuropathy have been observed in type-2 DM patients that may directly contribute towards the progression of CVDs (122).

Obesity is a chronic metabolic disorder, which has reached epidemic proportions globally not only in adults but in the paediatric population as well, in the past two decades. According to the WHO, 39% of the global population above 18 years of age are overweight, of which 13% are obese (123). Body mass index (BMI) is widely used to measure obesity and overweight. It is calculated by dividing the weight of an individual (Kg) by square of height ($m^2$). BMI between 18.5 to 25 $kg/m^2$ is considered as normal weight, overweight if the BMI ranges between 25.0 to 29.9 $kg/m^2$, and obese if the BMI is $\geq$30.0 $kg/m^2$ (124). Additionally, the waist circumference is regarded as a more accurate predictor of abdominal obesity in contrast to BMI.

Waist size over 40 inches (102 cm) in men and over 35 inches (88 cm) in women are at risk of heart disease (125). A broad range of clinical and epidemiological studies have identified obesity as an independent risk factor for a range of CVDs that include CHD, hypertension, cerebrovascular disease, atrial fibrillation, atherosclerosis, and sudden cardiac arrest (126). McGill et al. reported that excess body weight, especially fat accumulation in the abdominal region, can accelerate the progression of atherosclerosis that can go unnoticed for decades before the first clinical manifestation of CHD appears (127). Chief factors that contribute to obesity include dietary habits, physical inactivity, certain medical conditions, and medications. It is important to identify the contributing factors of weight gain as the intervention generally involves treatment plans that are tailored to the individual patient. Diet or caloric modification, physical activity, and behavioural therapy are three vital steps that form the basis to effectively manage obesity. Additionally, pharmacotherapy and bariatric surgery are often used as a resort in patients who are unable to achieve targeted weight loss with lifestyle interventions (128).

### 3.1.2 Current therapies for the treatment of CVD

Despite the high prevalence of CVDs, its progression can be reversed by modifying or reducing these associated risks. Management of CVDs can be categorized broadly into three stages. Primary stage involves maintaining a healthy lifestyle by the individual, which includes consumption of a balanced diet, regular exercise, and smoking cessation, when present. Secondary stage majorly emphasizes on an early diagnosis of the disease, which would enhance the chances for a successful treatment and prevent any serious or long-term health effects associated with the CVD. Tertiary

stage mainly deals with the treatment of the CVD in a chronic state that has resulted in long term health effects. It majorly aims at managing pain, increasing life expectancy and the quality of life (129).

Identification of the risk factors and their subsequent control forms the foundation of the treatment strategy against CVDs. Unfortunately, therapies administered against CVDs does not offer a complete cure from the cardiovascular conditions. Its effectiveness is limited to an extent, which only prevents or reduces any further progression of an already persisting condition. Still, these therapies have principally resulted in an overall decrease in the mortality and morbidity caused by numerous CVDs, such as atherosclerosis, stroke, and heart failure. According to the current international guidelines, LDL-C is the primary target, while HDL-C and triglycerides constitute the secondary targets of lipid-lowering drugs. Important lipid-lowering drugs presently available or under investigation includes statins, fibrates, bile acid sequestrants, niacin, PCSK9 inhibitors, ezetimibe and omega-3 fatty acids (130). A summary of lipid-lowering drugs is reported in **Table 3**.

**Table 3.** Lipid-lowering drugs.

| **Class** | **Mechanism** | **LDL-C reduction** | **HDL-C increase** | **TG reduction** |
|---|---|---|---|---|
| **Statins** | Inhibition of HMG-CoA reductase activity | 20-55% | 0-15% | 0-30% |
| **Fibrates** | Stimulates β-oxidation of fatty acids *via* PPARα | 0-15% | 5-20% | 10-50% |

| | | | | |
|---|---|---|---|---|
| **Bile Acid Sequestrants** | Binding to bile acids in the intestine | 12-20% | 3-5% | 0-25% |
| **Nicotinic Acid** | Still unclear | 10-15% | 15-25% | 10-35% |
| **PCSK9 Inhibitors** | Blocking PCSK9 and therefore reducing intracellular degradation of LDL receptors | 50-60% | 4-7% | 6-20% |
| **Ezetimibe** | Down-regulation of intestinal absorption of cholesterol *via* NPC1L1 inhibition | 14-18% | 1-4% | 9% |
| **Omega-3 fatty acids** | Still unclear; hypothesis: upregulation of lipoprotein lipase activity | 6-25% | 5-7% | 25-35% |

LDL-C: low density lipoprotein cholesterol; HDL-C: high density lipoprotein cholesterol; TG: triglycerides

As reported before, beside LDL-lowering strategy, the treatment of CVDs is based on the modification of one or more of the risk factors. Depending on the identification of the risk factor underlying the development of CVDs, appropriate therapy can be administered. This might include also the use of antihypertensives, antiplatelet, and anticoagulants.

## 3.2 Use of MR in cardiovascular disease

Multiple GWAS on cardiovascular disease have identified common genetic variants in very large case–control studies. GWAS have also

examined genetic determinants of conventional risk factors for cardiovascular disease, including lipids and lipoproteins, markers of inflammation, haemostasis and thrombosis, arterial wall function, metabolism, antioxidants, and lifestyle factors. MR have provided novel information in cardiovascular medicine over the past more than 10 years, adding to our understanding of cardiovascular disease pathogenesis and disease pathways.

MR studies have the potential to reshape cardiovascular medicine by generating robust naturally randomized evidence that can help to fill evidence gaps when a RCT would be either impossible or impractical to conduct. Results from some of the MR studies are summarized below.

### 3.2.1 Lipids and lipoproteins

MR has been extensively used to examine the causal role of lipids and lipoproteins. LDL-C (131-133), triglyceride-rich lipoproteins (134, 135), and lipoprotein(a) (136, 137) concentrations have been shown to be causally associated with higher CVD risk, while MR studies on HDL-C have failed to show a causal role for HDL-C in CVD risk (138, 139). High HDL-C concentration is associated with reduced risk of CVD in observational studies, but this may be due to confounding by other factors (i.e. physical activity, obesity, or diabetes) or due to low concentration of triglyceride-rich lipoproteins inversely correlated with HDL-C concentration. This have contributed to discard this lipoprotein fraction as a therapeutic target. The MR study design has also been used to predict potential effects of pharmacological intervention on other lipid and lipoprotein on CVD risk (for example, MR did predict the effect of inhibiting the PCSK9 protein), to examine side effects of LDL-lowering therapies, such as increased risk of new onset diabetes (140, 141), and to rebut other

potential side effects of LDL lowering, like cancer development (142), Alzheimer's disease, dementia, or Parkinson's disease.

More recently, MR studies revealed some of the most important scientific evidence in the cardiovascular field. In 2015, a large individual-level MR study found that the effect of lowering LDL-C with ezetimibe, a statin, or combination therapy with both ezetimibe and a statin should each reduce the risk of CHD by approximately the same amount per unit lower LDL-C, and the magnitude of the observed clinical benefit should be proportional to the absolute magnitude of the reduction in LDL-C, regardless of which treatment is used (143). More generally, these results suggest that the effect of lower LDL-C on the risk of CHD appears to be determined by the absolute magnitude of exposure to lower LDL-C, independently of the mechanism by which LDL-C is lowered. Therefore, it may be time to consider changing the notion of "lower is better" to "lower is better, and earlier is better" to maximize the potential lifetime benefit of exposure to lower LDL-C; and to reconsider the focus on "high intensity statins" and instead focus on "high intensity LDL-C lowering" as the preferred strategy to reduce the risk of cardiovascular events, while at the same time minimize the potential for dose-dependent statin-induced side-effects. In addition, very recently, MR analyses evaluating the associations of genetic scores composed of triglyceride-lowering variants in the LPL gene and LDL-C–lowering variants in the LDLR gene, respectively, with the risk of cardiovascular events among participants enrolled in 63 cohort or case-control studies, showed that triglyceride-lowering LPL variants and LDL-C–lowering LDLR variants were associated with similar lower CHD risk per unit lower level of apolipoprotein B (ApoB)-containing lipoproteins (144). The associations between lower triglyceride level and lower LDL-C level with risk of CHD due to these variants

appeared to be independent, additive, and proportional to the absolute change in ApoB. These results suggest that the clinical benefit of lowering triglyceride levels is similar to the clinical benefit of lowering LDL cholesterol levels per unit change in ApoB and is proportional to the net absolute reduction in ApoB-containing lipoproteins.

### 3.2.2 Inflammation

Recent advances in basic science have established a fundamental role for low degree chronic inflammation in mediating all stages of atherosclerosis, from initiation through progression and, ultimately, to the rupture of plaque and ensuing thrombotic complications of atherosclerosis. Interestingly, cholesterol accumulation in cells triggers the inflammasome response and results in the production of inflammatory mediators. High CRP concentration is associated with increased risk of cardiovascular disease in observational studies (145, 146), and treating high-risk patients with statins reduces CRP concentrations. However, several very large MR studies have failed to show a causal effect of CRP on CVD risk (146, 147), suggesting that pharmacological reduction of CPR may not results in a reduced cardiovascular risk. Thus, recent clinical investigations have sequentially moved upstream, first to IL-6 and then to interleukin-1, seeking more promising targets for anti-inflammatory atheroprotection. The recent CANTOS trial (148) showed that reducing vascular inflammation, through the inhibition of interleukin-1β (IL-1B), in the absence of concomitant lipid-lowering effect, reduces the rates of cardiovascular events, and also that individuals with a reduction of plasma IL-6 level on anti-IL-1B therapy experienced a greater cardiovascular risk reduction. The stimulation of the downstream IL-6-receptor signalling pathway, mediated by IL-

1B, represents only one of many potential anti-inflammatory pathways that might serve as targets for atheroprotection. In the blood plasma, a soluble fraction of the IL-6 receptor (sIL6R) is able to form an inhibitory complex that acts as a decoy receptor and negatively regulates IL-6 signalling. Since a clinical investigation of sIL6R pathways can be difficult because they are prone to fluctuation in the circulation, a study of the genetic determinants of these factors might be informative to evaluate the relevance of proximal inflammatory mediators to coronary heart disease.

### 3.2.3 Other risk factors

Several GWAS have identified genetic variants associated with both systolic and diastolic blood pressure (149). The variants identified only explain a small proportion of the variation in blood pressure, but despite this, they have shown large effects on cardiovascular disease risk (150). Type 2 diabetes has consistently been associated with cardiovascular disease in observational and in MR studies (151), and plasma glucose concentrations have also been shown to causally contribute to this risk (152, 153). MR studies have confirmed that genetic variation increasing smoking amount and extent is a cause of cardiovascular disease risk (154) and that genetically low alcohol intake is causally associated with less coronary heart events (155). Despite very strong epidemiological evidence for an association between low plasma vitamin D levels and increased cardiovascular disease risk, MR studies of genetically low vitamin D could not support a causal relationship (156). Interestingly, however, genetically low vitamin D did appear to be causally related to hypertension (157, 158) and to high all-cause mortality including cancer and other mortality, but not cardiovascular mortality (159). Also, shorter telomeres were associated with higher risk of ischaemic

heart disease, both observationally and genetically in MR studies (160, 161). In **Figure 9** more detailed information of biomarkers and lifestyle factors examined for a causal effect on risk of CVD using the MR design are represented. Biomarkers marked with a red "+" have been shown to have a causal effect with higher cardiovascular disease risk; biomarkers marked with a green "-" have been examined, but did not show causal effects on risk; and biomarkers marked with yellow "×" have been examined, but results have been conflicting.

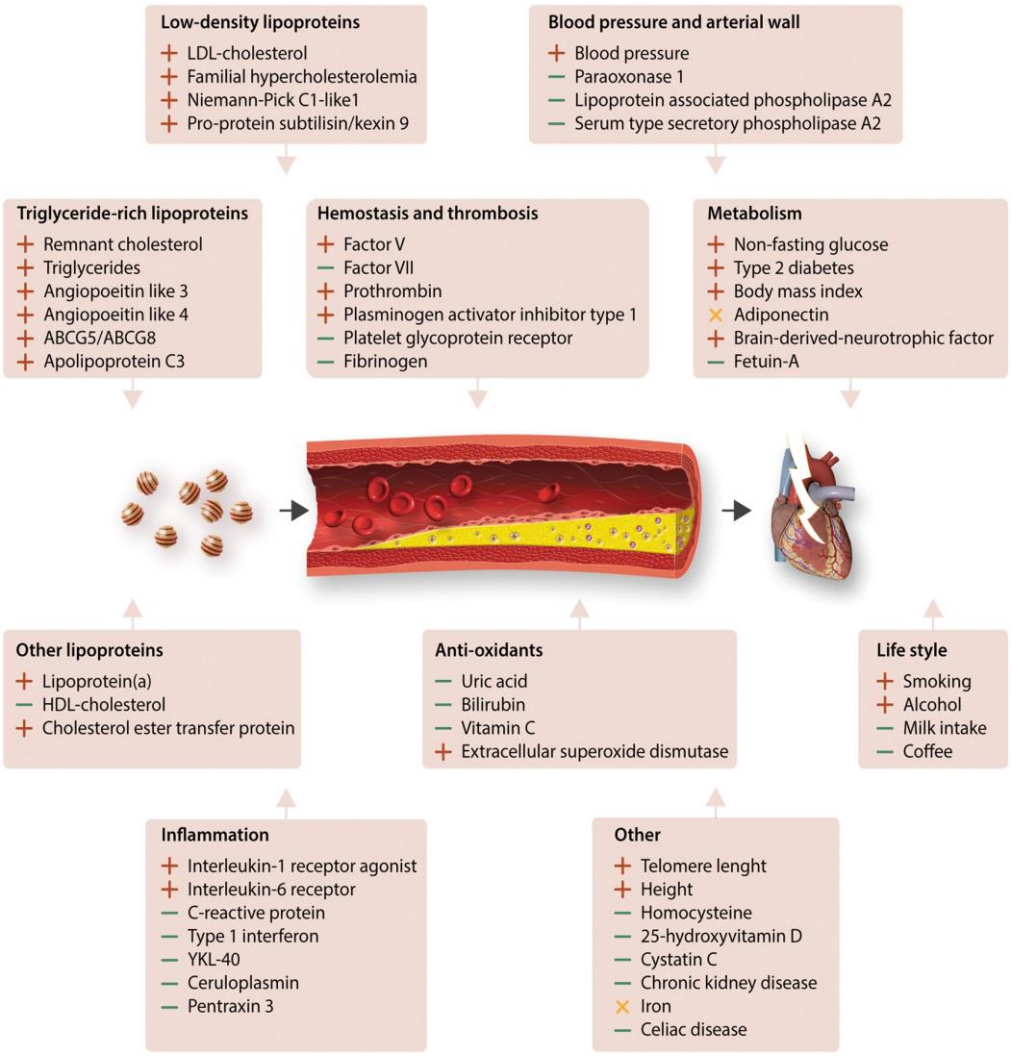



**Figure 9.** Overview of biomarkers and lifestyle factors examined for a causal effect on the risk of cardiovascular disease (56).

# CHAPTER 4

## Overview of dissertation

## 4.1 Aim

Having discussed several of the statistical issues regarding Mendelian randomization analyses, in the following chapter, many studies carried out using Mendelian randomization design will be presented in order to clarify the methodology and improve our knowledge about cardiovascular disease.

Thanks to an important collaboration established with Professor Brian A. Ference from the Cardiovascular Epidemiology Unit of the Department of Public Health and Primary Care (University of Cambridge, UK), many projects have been conducted through the access to the UK Biobank (see **Appendix I** for further details):

- "*Body mass index, polygenic predisposition, and risk of type 2 diabetes*"

  This project aimed at assessing whether measurements of body mass index and polygenic predisposition to have high values of this factor can be combined to better estimate the risk of developing type 2 diabetes. In addition, we also evaluated whether body mass index has a cumulative effect over time to make inference about the optimal timing of interventions to prevent diabetes.

- "*Combining family history of coronary heart disease and individual polygenic predisposition to provide risk estimation and guide therapy decision making*"

  This project aimed at understanding both the role of parental family history of coronary heart disease in predicting the individual lifetime risk of major coronary events in combination with the polygenic predisposition and lifetime exposure to LDL cholesterol, and whether these factors can be evaluated jointly to identify people who may benefit the most from lowering

cholesterol early in life.

- "*Does the risk of atherosclerotic cardiovascular disease vary based on measured or genetically determined lipoprotein(a)?*"
  This project aimed at comparing the cumulative lifetime risk of major coronary event among subjects with different Lp(a) genotype and Lp(a) measured concentrations.
- "*Does Lipoprotein(a) have a prothrombotic effect?*"
  This project aimed at clarifying the relation between Lp(a) levels and venous thromboembolism events, addressing whether Lp(a) has a genetically and clinically meaningful prothrombotic effect.
- "*A practical strategy to use measured lipoprotein(a) levels to guide clinical management*"
  This project aimed at addressing the amount of extra LDL cholesterol reduction needed to abolish the extra cardiovascular risk due to increased levels of Lp(a), specifically for different Lp(a) levels at baseline and depending on what age the treatment is started.

Details and analyses are reported in the following chapter, where each project is presented independently of each other and with the standard setting of a scientific article (introduction, methods, results, discussion, conclusions).

# CHAPTER 5

## Mendelian randomization projects

## 5.1 Body mass index, polygenic predisposition, and risk of type 2 diabetes

### Background

Type 2 diabetes (T2D) is chronic metabolic disorder characterized by dysglycemia leading to microvascular and macrovascular complications. It is a major cause of morbidity, mortality, and increased health care costs throughout the world (162). The world-wide prevalence of T2D has been rapidly increasing over the past decades and the International Diabetes Federation estimated that the number of people living with diabetes will rise to 700 million by 2045, if adequate prevention measures are not taken.

T2D mostly results from the interaction among genetic, environmental and other risk factors (including some unmodifiable factors such as family history, increasing age, or ethnicity). While people may have a strong genetic disposition towards T2D, the risk is greatly increased if people display a number of modifiable lifestyle factors including high blood pressure, overweight or obesity, insufficient physical activity, poor diet and if extra weight is carried around the waist (163).

Among those, certainly obesity plays a key role. Most patients with type 2 diabetes are obese, and the global epidemic of obesity largely explains the dramatic increase in the incidence and prevalence of T2D over the past 20 years (164). Taking into account also the inherited polygenic risk, these two elements are the two strongest risk factors for developing T2D (165).

As a result, there is an urgent public health need to develop better ways to identify persons who are at risk for developing T2D and more effective ways to prevent the development of diabetes among at-risk persons.

Whether information about obesity, as measured by increased body mass index (BMI), and a polygenic score (PGS) estimating inherited risk (polygenic form of obesity) can be combined to better identify persons at risk for developing diabetes or provide information about the optimal timing of interventions to prevent the development of diabetes among at-risk persons is unknown.

To address this issue, we sought to evaluate the separate and combined effects of BMI and a PGS on the risk of developing type 2 diabetes. In addition, we sought to compare the effect of lifelong exposure to increased BMI as compared to BMI changes later in life on plasma glycated haemoglobin (HbA1c) levels and the risk of T2D to assess whether BMI has a cumulative effect on the risk of diabetes over time and thus make inferences about the optimal timing of interventions to prevent diabetes.

**Methods**

Study population

A total of 445,765 participants enrolled in the UK Biobank with complete genetic and principal component data who self-identified as being of white ancestry were included in the study. Participants underwent genotyping with one of two closely related custom arrays (UK BiLEVE Axiom Array or the UK Biobank Axiom Array) consisting of over 800,000 genetic markers, with additional genotypes imputed using the Haplotype Reference Consortium resource, the UK10K panel, and the 1000 Genomes panel (more details in **Appendix I**).

Study outcomes

The primary outcome was type 2 diabetes, defined as the diagnosis of diabetes after the age of 35 years. Prevalent cases of T2D were defined as participants who reported a history of diabetes diagnosed

by a doctor at an age greater than or equal to age 35 years at the initial assessment visit (2006-2010) upon enrolment into UK Biobank (excluding women who reported gestational diabetes only). Incident cases of T2D were defined as participants who reported being diagnosed by a doctor with diabetes after the age of enrolment into UK biobank during one of three follow-up sub-study examinations and who reported no history of diabetes at the initial assessment visit, or participants who had other evidence of being diagnosed with "non-insulin dependent diabetes" after the age of enrolment into UK biobank and who reported no history of diabetes at the initial assessment visit.

Construction of PGS and Mendelian randomization instruments

The PGS for T2D was constructed using external weights for 6,917,436 variants evaluated by the DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) consortium that had an INFO score >0.3 in the UK Biobank. A PGS was calculated for each participant by summing the number risk-increasing alleles inherited at each variant weighted by the T2D effect size for each allele as reported in DIAGRAM. To optimize the PGS, variants in high linkage disequilibrium were progressively pruned using the clumping procedure in PLINK2 until there was no further improvement in the magnitude of the association between the PGS and the risk of diabetes per standard deviation (SD) change in the PGS.

A BMI instrumental variable genetic score for Mendelian randomization analysis was constructed by combining 96 independently inherited (linkage disequilibrium $r^2<0.001$) variants associated with BMI at genome-wide level of significance ($p<5x10^{-8}$) as reported in Genetic Investigation of ANthropometric Traits (GIANT) consortium (see **Appendix II** for the list of the single

nucleotide polymorphisms included in the score). The BMI genetic score was calculated for each participant by summing the number of BMI-increasing alleles inherited at each variant included in the BMI instrumental variable genetic score weighted by the BMI effect size of each allele.

Statistical analysis

Logistic regression was used to evaluate the associations between type 2 diabetes and the PGS (lifelong exposure to BMI using the BMI instrumental variable genetic score) and measured BMI in separate analyses. All analyses were adjusted for age at baseline, sex, and the first 10 principal components of ancestry. In sensitivity analyses, Cox proportional hazards models with age as the time scale were used to evaluate the associations with risk of being diagnosed with T2D. In these analyses, each participant was censored at the age they were first diagnosed with diabetes, died (treated as a competing risk), or the age at last reported follow-up. Cumulative lifetime risk of T2D was estimated using Kaplan-Meier curves with age as the time scale for participants within each quintile of PGS, and within each quintile of measured BMI. To evaluate the combined effect of measured BMI and polygenic predisposition to have high values of BMI on the risk of T2D, the proportion of participants diagnosed with diabetes was compared after ordering participants by quintiles of BMI and PGS, respectively. To assess whether BMI increases plasma HbA1c levels and the risk of diabetes with increasing duration of exposure, the effect of a one-unit increase in BMI measured in middle life in observational analyses was compared with the effect of a one-unit increase in genetically determined lifelong exposure to BMI using the BMI instrumental variable genetic score in Mendelian randomization analyses for both plasma HbA1c levels and the risk of type 2 diabetes.

In a test of external replication, 2-sample Mendelian randomization analyses were performed to assess the effect of lifelong exposure to a one unit increase in BMI instrumented by a genetic score consisting of 494 independently inherited ($r^2$<0.001) variants associated with BMI at GWAS level of significance in a combined analysis including 681,275 participants of European descent enrolled in either the GIANT consortium or UK Biobank and plasma HbA1c levels measured among 123,665 participants of European descent without diabetes as reported by the MAGIC (the Meta-Analyses of Glucose and Insulin-related traits Consortium) investigators; and the risk of type 2 diabetes among 159,208 participants of European descent (26,676 cases of type 2 diabetes) as reported by the DIAGRAM (DIAbetes Genetics Replication and Meta-analysis) consortium.

All analyses were performed using Stata (version 16; StataCorp), R (version 3.3.3). A 2-tailed p-value less than 0.05 was considered statistically significant.

## Results

Participant characteristics

Baseline characteristics are presented as means and SD for normally distributed variables, median and interquartile ranges (IQR) for non-normally distributed variables (triglycerides and C-reactive protein), or percentages for dichotomous variables. The mean age of participants at enrolment was 57.3 years (range: 38.9-73.7), the median follow-up time was 8.1 years; and 54.3% were women (**Table 1**). A total of 28,563 participants (6.4%) were diagnosed with T2D after the age of 35 years, including 10,711 (4.4%) women and 17,852 (8.8%) men. A total of 18,278 cases of T2D were prevalent at the time of enrolment, and 10,285 incident cases were diagnosed during follow-up.

Associations of polygenic score and BMI with diabetes

A one standard deviation increase in the PGS containing 2,037,596 variants was associated with an odds ratio (OR) of 1.50 (95%CI: 1.49-1.52). A one SD increase in the BMI (4.77 units) measured at the time of enrolment into UK Biobank was associated with an OR for diabetes of 2.26 (95%CI: 2.23-2.28). Both increasing quintiles of PGS, and increasing quintiles of BMI, were associated with increasingly steeper trajectories of lifetime risk for diabetes (**Figure 1**). There was a step wise increase in the risk of diabetes with each increasing quintile of PGS (**Table 1**). Participants in the highest PGS quintile had an OR for diabetes of 3.09 (95%CI: 2.96-3.22) as compared to participants in the lowest PGS quintile. Similarly, there was a step wise increase in the risk of diabetes with each increasing quintile of BMI (**Table 1**). Participants in the highest quintile of BMI (mean BMI 34.7 kg/$m^2$) had an OR of 12.57 (95%CI: 11.83-13.34) as compared to participants in the lowest BMI quintile (mean BMI 21.8 kg/$m^2$).

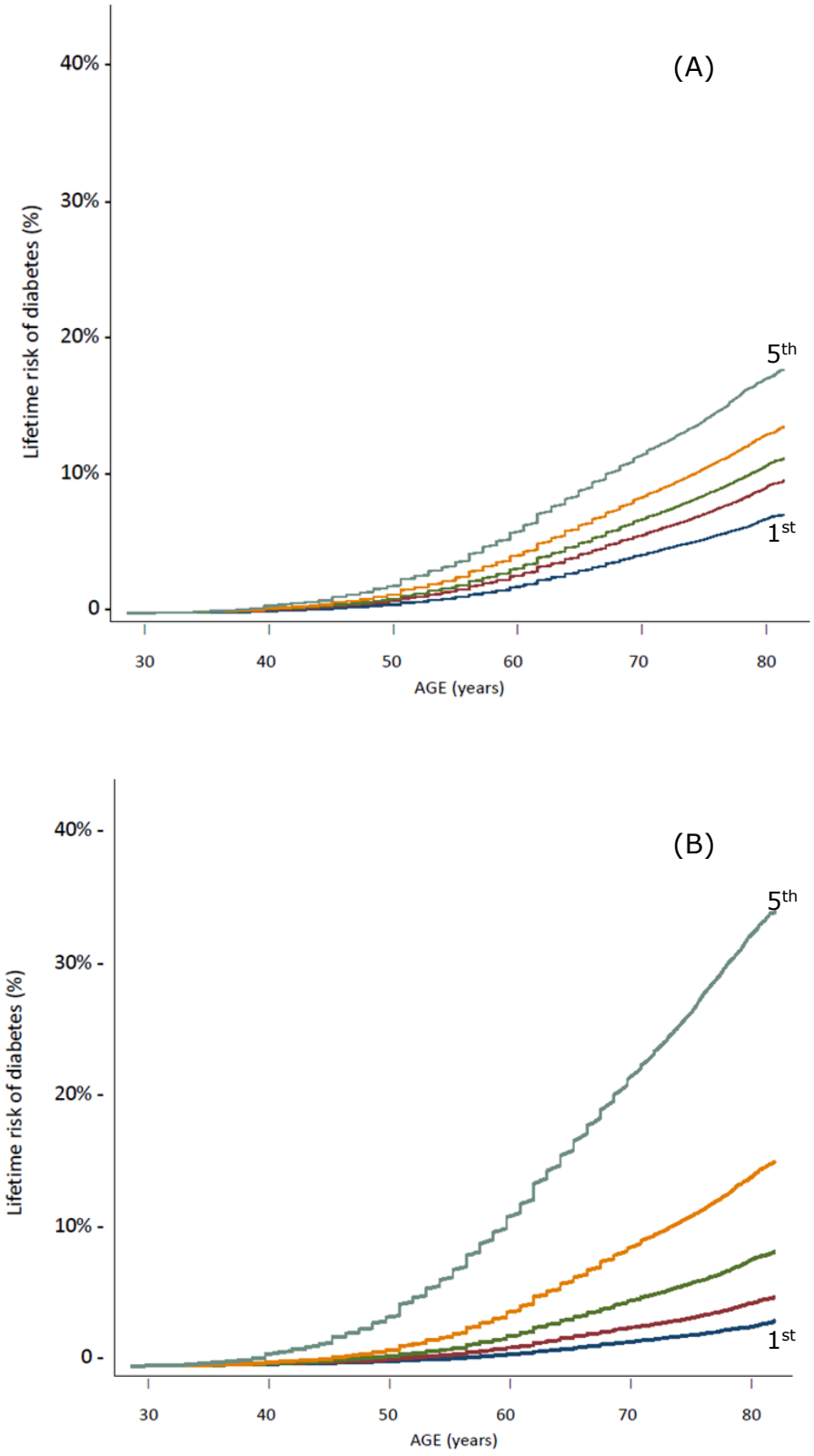



**Figure 1.** Trajectories of lifetime risk for type 2 diabetes by polygenic score (panel A) and BMI (panel B) quintiles.

**Table 1.** Baseline characteristics by quintiles of polygenic score and body mass index. *Odds ratio (OR) per standard deviation (SD) increase in polygenic score (PGS) for diabetes, or per SD change in body mass index.

| | | **Polygenic Score (PGS) for Type 2 Diabetes** | | | | |
|---|---|---|---|---|---|---|
| **Characteristic** | **All** | **PGS Q1** | **PGS Q2** | **PGS Q3** | **PGS Q4** | **PGS Q5** |
| Age, y (SD) | 65.3 (8.0) | 65.4 (8.0) | 65.3 (8.0) | 65.3 (8.0) | 65.2 (8.0) | 65.1 (8.0) |
| Women (%) | 54.3 | 54.3 | 53.8 | 54.4 | 54.1 | 54.9 |
| LDL-C, mg/dL (SD) | 138.0 (33.6) | 137.9 (33.1) | 138.0 (33.2) | 138.1 (33.5) | 138.0 (33.9) | 138.0 (34.4) |
| HDL-C, mg/dL (SD) | 56.2 (14.8) | 57.4 (14.9) | 56.7 (14.9) | 56.2 (14.8) | 55.7 (14.7) | 55.0 (14.6) |
| Triglycerides, mg/dL (IQR) | 131.9 (93.1-190.7) | 125.1 (88.8-179.8) | 129.5 (91.7-187.1) | 132.0 (93.4-190.4) | 134.6 (94.9-195.3) | 138.5 (97.6-200.5) |
| SBP, mmHg (SD) | 137.8 (18.6) | 136.9 (18.6) | 137.5 (18.5) | 137.9 (18.6) | 138.1 (18.9) | 138.6 (18.6) |
| Current smoker (%) | 7.2 | 6.7 | 7.1 | 7.2 | 7.4 | 7.7 |
| CRP, mg/L (IQR) | 1.3 (0.7-2.8) | 1.2 (0.6-2.6) | 1.3 (0.6-2.7) | 1.3 (0.7-2.8) | 1.4 (0.7-2.9) | 1.4 (0.7-3.0) |
| BMI (SD) | 27.4 (4.8) | 26.9 (4.5) | 27.2 (4.7) | 27.4 (4.7) | 27.6 (4.8) | 27.9 (4.9) |
| No. with diabetes (%) | 28,563 (6.4) | 3,372 (3.8) | 4,542 (5.1) | 5,361 (6.0) | 6,569 (7.4) | 8,719 (9.8) |
| OR (95%CI) | 1.50 (1.49-1.52)* | reference | 1.40 (1.33-1.46) | 1.71 (1.63-1.79) | 2.16 (2.07-2.26) | 3.09 (2.96-3.22) |
| | | **Body Mass Index (BMI)** | | | | |
| | **All** | **BMI Q1** | **BMI Q2** | **BMI Q3** | **BMI Q4** | **BMI Q5** |
| BMI (range) | 27.4 (17.5-50.0) | 21.8 (17.5-23.5) | 24.6 (23.5-25.7) | 26.7 (25.7-27.8) | 29.2 (27.8-30.8) | 34.7 (30.8-50.0) |
| No. with diabetes (%) | 28,563 (6.4) | 1,193 (1.3) | 2,089 (2.4) | 3,679 (4.1) | 6,638 (7.5) | 14,776 (16.6) |
| OR (95%CI) | 2.26 (2.23-2.28)* | reference | 1.50 (1.39-1.61) | 2.48 (2.32–2.66) | 4.58 (4.30–4.88) | 12.57 (11.83-13.34) |

In stratified analyses, the risk of diabetes varied by at least 10-fold within each quintile of PGS depending on differences in BMI (panel A). By contrast, ordering by BMI quintile stratified by PGS quintile (panel B) appeared to separate participants into distinct categories of risk without similar overlapping risk (**Figure 2**).

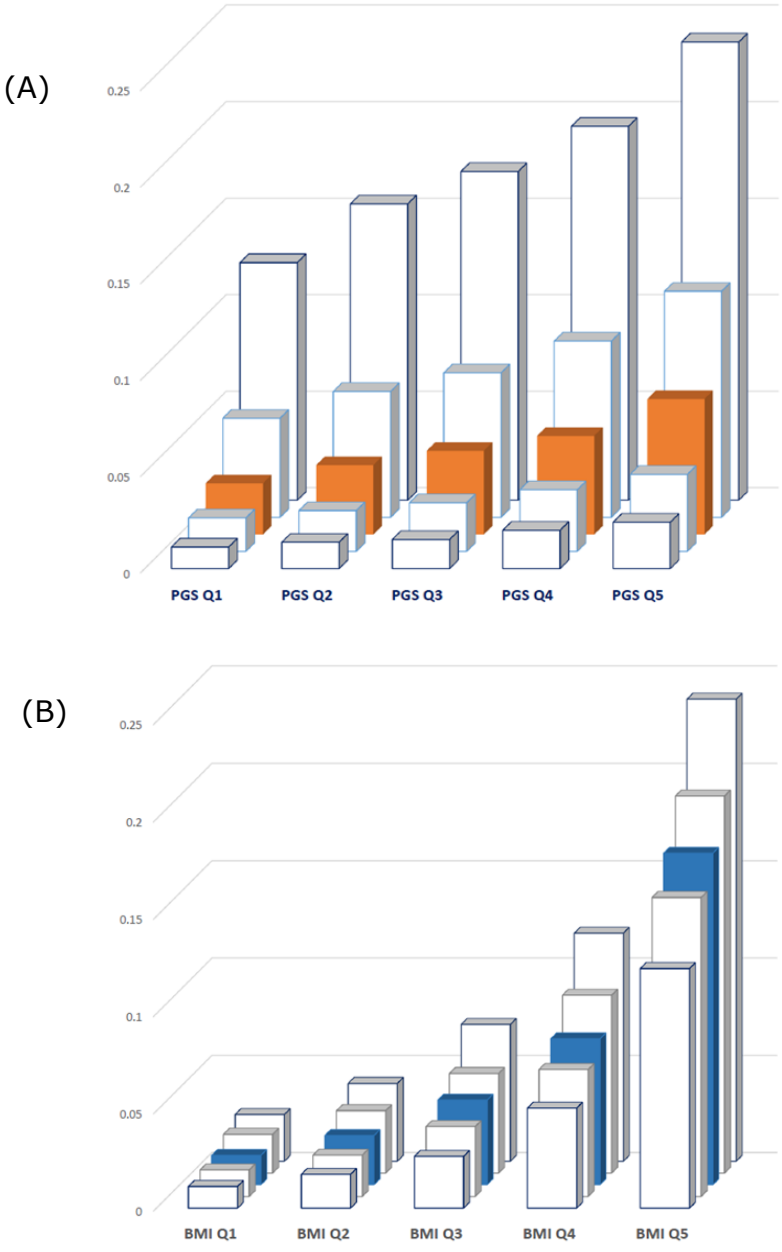



**Figure 2.** Combined effect of BMI and polygenic score on lifetime risk of diabetes by quintiles of polygenic score (panel A) and BMI (panel B)

Indeed, participants in the lowest PGS quintile with BMI >30 had a 6-fold greater risk of type 2 diabetes (OR: 6.14, 95%CI: 5.43-6.95) as compared to participants in the highest PGS quintile with BMI <25 (**Figure 3**).

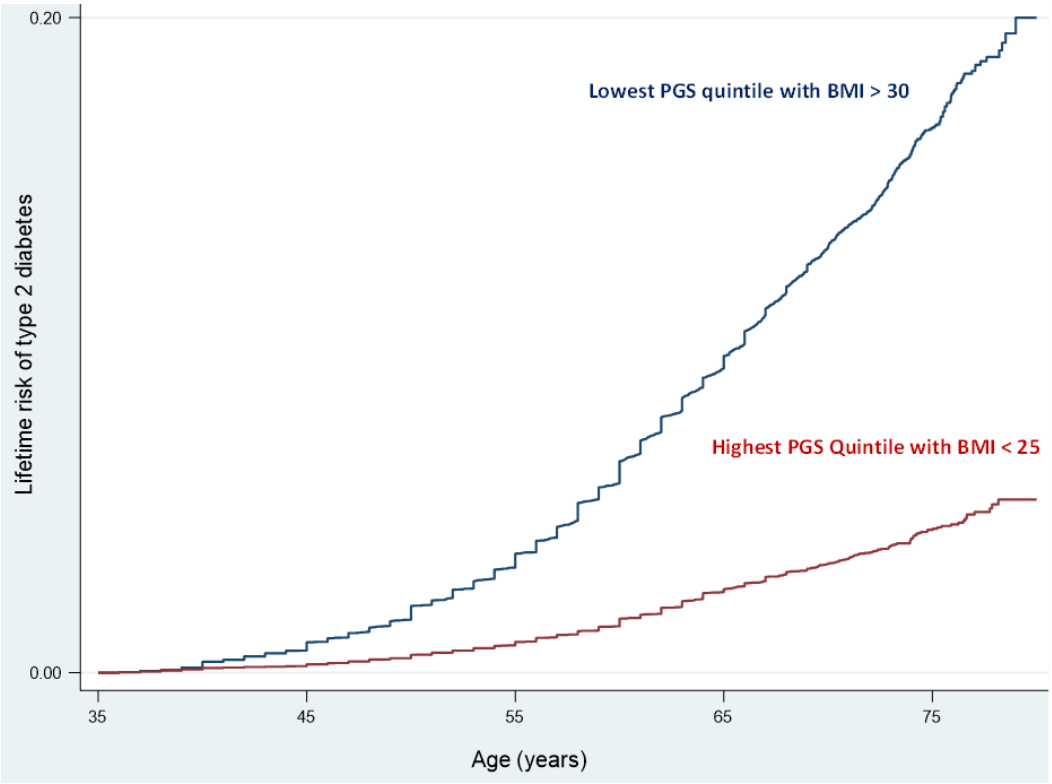



**Figure 3.** Kaplan Meier estimates of lifetime risk of being diagnosed with T2D after the age of 35 years up to age 80 years for participants in the lowest PGS quintile with obesity (BMI >30), and participants in the highest PGS quintile with normal body weight (BMI <25). PGS is polygenic score, BMI is body mass index.

<u>Comparison of lifelong and middle life changes in BMI on T2D</u>

In observational analyses (**Figure 4**), a one-unit increase in BMI measured in middle life was associated with an OR for diabetes of 1.21 (95%CI: 1.21-1.22), corrected for regression dilution bias. In Mendelian randomization analyses, a one-unit increase in genetically

determined lifetime exposure to BMI was associated with an OR of 1.25 (95%CI: 1.23-1.28). In external replication analyses using data from the DIAGRAM consortium, a one-unit increase in lifetime exposure to BMI was associated with an OR of 1.20 (95%CI: 1.17-1.23). In a combined analysis including a total of 604,973 participants and 55,239 cases of type 2 diabetes, a one-unit increase in genetically determined lifetime exposure to BMI was associated with an OR of 1.22 (95%CI: 1.21-1.25). Similarly, in observational analyses, a one-unit increase in BMI measured in middle life was associated with a 0.013% change in plasma HbA1c (95%CI:0.013-0.014). In Mendelian randomization analyses, a one-unit increase in genetically determined lifetime exposure to BMI was associated with a 0.011% change plasma HbA1c (95%CI: 0.008-0.013). In external replication analyses using data from the MAGIC consortium, a one-unit increase in lifetime exposure to BMI was associated with a similar 0.011% change in HbA1c (95%CI: 0.009-0.014).

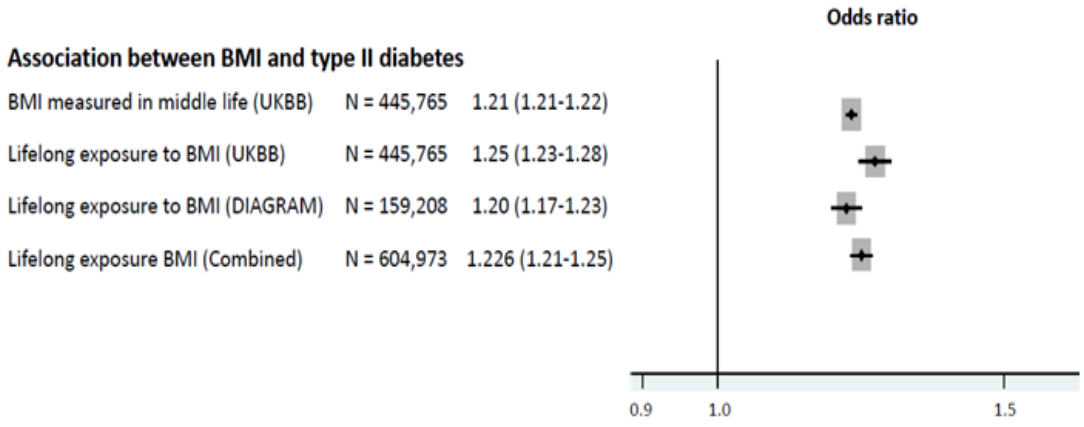



**Figure 4.** Comparison of the effect lifelong exposure and shorter-term exposure to increased BMI on the risk of type 2 diabetes

In sensitivity analyses, the results of all analyses were similar among men and women, in analyses restricted to prevalent or incident cases of diabetes, and when using proportional hazard regression rather than logistic regression.

**Discussion**

We found that BMI is a much stronger risk factor for type 2 diabetes than polygenic predisposition. The risk of T2D was high among participants with high BMI and low among participants with low BMI regardless of PGS. By contrast, the risk of T2D varied by at least 10-fold among persons with the same PGS depending on differences in BMI. In addition, we found that lifelong exposure to increased BMI and increases in BMI that occur later in life appear to have similar effects on both plasma HbA1c levels and the risk of T2D, thus suggesting that the effect of BMI on dysglycemia and the risk of T2D does not increase with increasing duration of exposure. These findings may have practice changing implications for screening, preventing, and treating T2D.

The finding that BMI is a much stronger risk factor for T2D than inherited polygenic predisposition implies that the risk of type 2 diabetes is largely modifiable if clinical interventions are put in place in the early stages of the disease. Furthermore, the finding that the effect of BMI does not increase with increasing duration of exposure suggests that increased BMI does not cause irreversible structural changes but instead leads to reversible metabolic changes. Together, these two findings suggest that most cases of T2D can be either prevented or reversed, acting in the early stages of the disease.

Specifically, increasing BMI appears to lead to a corresponding increase in HbA1c levels. At a certain BMI level, the corresponding HbA1c level will exceed the threshold for the diagnosis of

prediabetes. This BMI is the BMI threshold for prediabetes for that person. Further increases in BMI will lead to further increases in HbA1c. The BMI at which the corresponding HbA1c level exceeds the threshold for the diagnosis of T2D is that person's BMI threshold for diabetes. The results of this study suggest that most cases of T2D can be prevented by keeping BMI below each person's BMI threshold for diabetes. Furthermore, because BMI does not appear to lead to irreversible structural changes, the results of this study suggest that if a person's BMI exceeds his/her diabetes threshold, then losing weight to reduce BMI below the threshold for diabetes (or prediabetes) should lead to a corresponding reduction in HbA1c large enough to potentially "reverse" the diagnosis of diabetes.

These conclusions are consistent with the results of several randomized trials. In the Diabetes Prevention Program trial (166), both an intensive lifestyle intervention program and treatment with metformin led to weight loss and a lower risk of incident T2D as compared to placebo among persons with impaired fasting glucose. The intensive lifestyle intervention program led to greater weight loss than treatment with metformin, and a corresponding greater reduction in the incidence of diabetes. Similarly, in the randomized Finnish Diabetes Prevention Study, weight loss led to a lower incidence of T2D among overweight persons with impaired glucose tolerance (167). Furthermore, in the DiRECT trial (168), an aggressive weight-loss program with whole diet replacement resulted in remission of diabetes in almost one-half of participants allocated to the active intervention group as compared to the best practices control group. In addition, several trials evaluating mechanical interventions have demonstrated that substantial weight loss can reverse the diagnosis of T2D by re-establishing more normal glycaemic control (169).

Importantly, the BMI threshold for T2D is likely to be different for each person. Indeed, the effect of BMI on HbA1c may be modified by the distribution of fat (waist-to-hip ratio), adipocyte activity, or ancestral background. This may explain why some persons with normal BMI develop T2D while most persons with morbid obesity do not. It may also explain why some ethnic groups develop diabetes at lower BMI thresholds than others (170). Therefore, perhaps the best way to screen for T2D currently would be to serially measure each person's BMI and HbA1c over time. The slope of the change in HbA1c with increasing BMI can be plotted for each person to estimate his/her individual BMI threshold for both prediabetes and diabetes.

Finally, the results of this study combined with the results of randomized trials suggests that the treatment of diabetes should be refocused to have dual primary goals. First, consistent with current practice, plasma glucose should be controlled to prevent complications caused by elevated glucose levels. Second, however, greater emphasis should be placed on therapeutic weight loss as a strategy to lower BMI below a person's threshold for diabetes (or prediabetes), in an explicit attempt to reverse diabetes. Both goals can potentially be accomplished using newer hypoglycemic agents including GLP1 receptor antagonists which control plasma glucose and lead to substantial weight loss (171).

Limitations

This study has limitations. The analysis was restricted to participants who self-identified as being of white European ancestry, and therefore may not apply to persons of other ethnicities. As a result, this study should be repeated in other populations, particularly those that include participants who may be more vulnerable to developing diabetes at lower BMI thresholds. In addition, BMI was measured at

the time of enrolment into UK biobank and therefore may not reflect each participant's BMI at the time diabetes was diagnosed. However, the effect of a one-unit increase in BMI on the risk of diabetes was broadly similar for both prevalent cases of diabetes that occurred prior to the measurement of BMI and for incident cases that occurred after BMI was measured at enrolment. Additional research is needed to identify factors that influence each person's change in HbA1c in response to increasing BMI.

**Conclusion**

In conclusion, we found that BMI is a much stronger risk factor for type 2 diabetes than polygenic predisposition which leads to reversible metabolic changes that do not appear to accumulate over time. Therefore, most cases of diabetes can potentially be prevented or reversed if preventive interventions are taken promptly.

## 5.2 Combining family history of coronary heart disease and individual polygenic predisposition to provide risk estimation and guide therapy decision making

### Background

Despite remarkable successes in the treatment and prevention in the past decades, coronary heart disease (CHD) is still the leading cause of death and premature disability in developed countries (172).

Understanding the genetic basis of CHD can improve management and prevention. Family and twin studies, animal models, and gene association studies suggest a genetic basis for CHD, supporting the hypothesis that genes contribute to CHD development and progression, and response to risk factor modification and lifestyle choices (173). For this reason, individuals with genetic predisposition to atherosclerosis are at the greatest risk for developing CHD, especially at early ages, and they have the most to gain from timely preventive interventions.

There are mainly two ways to conceptualize inherited risk of CHD: family history and polygenic predisposition. In the cardiovascular field, a positive family history is associated with a significant doubling in cardiovascular risk (174). Researchers from the Framingham Study reported that having cardiovascular disease in at least one parent doubled the 8-year risk among men, and increased the risk among women by 70% (174), independently from the other risk factors. Family history captures inherited genetic predisposition as well as shared environments and behaviours. Despite this, it has been shown to be partially independent from genome-wide polygenic scores (PGSs) in diseases such as heart disease (175, 176). Other

studies have also shown that genome-wide PGSs associated with incident coronary artery disease (CAD) are independent of family history (177).

The systematic collection and interpretation of family history information is the most appropriate initial screening approach to identify individuals with genetic susceptibility to CHD. Indeed, the simplicity of data collection about family history, which can be easily and systematically queried in the clinical setting, grants for inexpensive and easy-to-obtain predictive information, potentially allowing for intervention before prolonged exposure to clinical risk factors, such as smoking or elevated lipid levels. On the other hand, despite PGSs are more expensive and onerous to obtain than a standard lipid panel or family history, they can provide important information regarding an exposure present from birth that could be ascertained early in life as part of a broad set of risk evaluations. Together, family history and PGSs have the potential to enhance risk prediction in cardiovascular diseases.

Several studies have evaluated the inclusion of self-reported family history alongside genetics in risk-prediction models for complex diseases such as CAD (178, 179). For example, the use of six conventional risk factors for CAD, including family history of heart disease, was shown to improve the prediction of incident CAD when used in combination with PGS compared to prediction based on PGS alone or conventional risk factors alone (180). Current prevention guidelines recommend that premature family history should be incorporated into the risk estimation process that guides treatment decisions. However, family history is incorporated in some, but not all, short-term risk prediction equations. This probably because it is not fully understood whether family history of heart disease provides independent and/or additional information to the prediction of

individual cardiovascular risk beyond classical cardiovascular risk factors and genetic predisposition.

Given this background, we aimed at clarifying the role of family history in predicting the individual lifetime risk of CHD in combination with polygenic predisposition and lifetime exposure to high LDL cholesterol (LDL-C) level, and understanding whether the use of these factors jointly can identify people with the highest lifetime risk who may benefit the most from lowering LDL-C early in life.

## Methods

### Study population

A total of 445,744 participants enrolled in the UK Biobank with complete genetic and principal component data who self-identified as being of white ancestry were included in the study. The UK Biobank is a prospective observational study of approximately 500,000 volunteer adults aged 40 to 69 years recruited from 22 sites across the United Kingdom between 2006 and 2010, with follow-up ongoing. Biochemical measurements, physical examination measurements, and medical histories were assessed at the time of study enrolment. Participants underwent genotyping with one of two closely related custom arrays (UK BiLEVE Axiom Array or the UK Biobank Axiom Array) consisting of over 800,000 genetic markers, with additional genotypes imputed using the Haplotype Reference Consortium resource, the UK10K panel, and the 1000 Genomes panel. The KING toolset was used to identify up to third-degree relatedness based on kinship coefficients >0.044 (more details in **Appendix I**).

### Construction of the LDL instrumental variable

To construct the LDL instrumental variable, all variants associated with LDL-C at genome-wide level of significance ($p<5x10^{-8}$) as

reported in external consortia were included in the polygenic score (181). The LDL variants were then pruned by excluding all variants with a linkage disequilibrium (LD) $r^2$>0.1 to select independently inherited variants for inclusion in the instrumental variable genetic scores. An LDL score was calculated for each participant by summing the number of LDL-increasing alleles inherited at each variant included in the LDL score weighted by the LDL effect size of each allele (**Appendix III**).

Construction of the PGS score for CAD

Polygenic risk scores were created following an additive model for CAD, atrial fibrillation (AF), stroke, hypertension, and diabetes separately, as described elsewhere (182). Briefly, the number of alleles (0, 1 or 2) for each individual was summed after multiplication with the effect size between the single nucleotide polymorphism (SNP) and disease of interest. Effect sizes of SNP–disease associations were based on previously published genome-wide association studies. For CAD, 169 SNPs were used; for AF, 25 SNPs; for stroke, 11 SNPs; for hypertension, 107 SNPs; and for diabetes, 38 SNPs. If multiple effect sizes were reported in a study, those estimated in the largest sample size were used. Effect sizes were not considered for the polygenic score if estimated with UK Biobank data to avoid potential overestimation. SNPs were excluded if they were missing in UK Biobank data. As some studies reported multiple correlated variants in the same locus, independent SNPs were selected based on the highest reported p-value by using the LD clumping procedure (at $r^2$< 0.01).

Definition of family history

Self-reported information on family history of cardiovascular disease

was collected at the time of the enrolment in the UK Biobank. For family history of CHD, we considered history in any first-degree relative (father or mother; fields #20107, and 20110, respectively).

Study outcomes and statistical analysis

The primary outcome was major coronary events (MCE), defined as the first occurrence of either a fatal or non-fatal myocardial infarction (MI), or coronary revascularization. The analysis used Cox proportional hazards models adjusted for age and the first 10 principal components of ancestry, with age as the time scale. Each participant was censored either at the age primary outcome event was experienced, death due to a cause other than MI (treated as a competing risk), or at the age of last reported follow-up. The dates of all incident events were recorded from hospital episode statistics, while the dates of events that were prevalent at the time of enrolment into UK Biobank were recorded either from hospital episode statistics or self-reported. Lifetime risk of MCE was plotted using Kaplan-Meier curves by the presence of parental family history of CHD, and by polygenic predisposition (deciles of the PGS). The combined effect of parental family history, polygenic predisposition and the lifetime exposure to high LDL-C in predicting the individual lifetime risk of MCE was estimated carrying out adjusted Cox models. Since the events rate disease for heart attack, stroke or coronary revascularization among women enrolled in the UK Biobank is only one third of the events rate in the UK general practice database (while the events rate among men is more equivalent), all the analyses were focused on male cohort to make reliable conclusions.
All analyses were performed using Stata (version 16; StataCorp). A 2-tailed p-value less than 0.05 was considered statistically significant.

## Results

Participant characteristics

The mean age of participants at enrolment was similar between subjects with and without family history of CAD (approximately 58 vs 57 years, respectively) (**Table 1**). In addition, no remarkable differences were observed for the lipid profile based on family history of CHD, neither for the distribution of obesity indicators, such as body mass index and waist-to-hip ratio. However, the percentage of patients on lipid-lowering or anti-hypertensive therapies was greater among subjects with a positive family history (17.9% vs 27.2% vs 37.2, and 20.7% vs 27.4% vs 35.1, for subjects with no family history of CHD, and with one parent or both parents with a history of heart disease, respectively). Same distributions were observed among individuals with different polygenic predisposition (**Table 2**).

**Table 1.** Baseline characteristics, measured at the time of enrolment in the UK Biobank, are presented by presence of family history of coronary heart disease.

| | **Family history** | | |
|---|---|---|---|
| **Characteristics** | **No** | **One parent** | **Both parents** |
| No. participants | 108,677 | 60,613 | 11,592 |
| Age, y (SD) | 56.6 (8.4) | 57.8 (7.7) | 58.8 (7.0) |
| TC, mg/dL (SD) | 222.2 (39.6) | 224.4 (39.2) | 224.8 (39.7) |
| LDL-C, mg/dL (SD) | 142.4 (30.3) | 144.4 (30.0) | 145.0 (30.3) |
| ApoB, mg/dL (SD) | 106.9 (22.7) | 108.5 (22.6) | 109.4 (22.7) |
| HDL-C, mg/dL (SD) | 50.7 (12.0) | 50.4 (12.0) | 49.7 (11.9) |
| TG, mg/dL (IQR) | 147.1 (102.3-213.1) | 150.4 (105.1-216.9) | 153.0 (105.9-222.1) |
| SBP, mmHg (SD) | 138.8 (16.9) | 140.1 (17.2) | 140.8 (17.5) |
| CRP, mg/L (IQR) | 1.3 (0.6-2.5) | 1.3 (0.7-2.5) | 1.3 (0.7-2.6) |
| BMI, $Kg/m^2$ (SD) | 27.7 (4.2) | 27.9 (4.3) | 28.3 (4.4) |

| | | | |
|---|---|---|---|
| WHR (SD) | 0.9 (0.1) | 0.9 (0.1) | 0.9 (0.1) |
| Lipid-lowering therapy, % | 17.9 | 27.2 | 37.2 |
| Anti-hypertensive therapy, % | 20.7 | 27.4 | 35.1 |
| Diabetes,% | 5.3 | 6.6 | 8.5 |
| Hypertension, % | 55.5 | 61.7 | 67.0 |
| Ever smoker (%) | 34.4 | 35.1 | 36.6 |

Characteristics are presented as means and standard deviations (SD) for normally distributed variables, median and interquartile ranges (IQR) for non-normally distributed variables (TG and CRP), or percentages for dichotomous variables. TC: total cholesterol; LDL-C: low-density lipoprotein cholesterol; ApoB: apolipoprotein B; HDL-C: high-density lipoprotein cholesterol; TG: triglycerides; SBP: systolic blood pressure; CRP: c-reactive protein; BMI: body mass index; WHR: waist to hip ratio.

**Table 2.** Baseline characteristics, measured at the time of enrolment in the UK Biobank, are presented by levels of the polygenic risk score (PGS) for coronary artery disease.

| | **PGS** | | |
|---|---|---|---|
| **Characteristics** | **Low level** | **Average level** | **High level** |
| No. participants | 20,132 | 122,653 | 20,235 |
| Age, y (SD) | 57.6 (8.1) | 57.5 (8.1) | 57.3 (8.1) |
| TC, mg/dL (SD) | 220.9 (38.7) | 223.1 (39.5) | 224.4 (40.6) |
| LDL-C, mg/dL (SD) | 141.2 (29.6) | 143.2 (30.2) | 144.6 (31.0) |
| ApoB, mg/dL (SD) | 105.9 (22.2) | 107.6 (22.7) | 109.0 (23.1) |
| HDL-C, mg/dL (SD) | 51.2 (12.2) | 50.5 (12.0) | 49.6 (12.0) |
| TG, mg/dL (IQR) | 142.8 (98.9-207.1) | 149.3 (103.7-215.8) | 154.8 (108.0-222.8) |
| SBP, mmHg (SD) | 138.0 (16.8) | 139.7 (17.2) | 140.8 (17.3) |
| CRP, mg/L (IQR) | 1.2 (0.6-2.4) | 1.3 (0.7-2.5) | 1.4 (0.7-2.7) |
| BMI, $Kg/m^2$ (SD) | 27.4 (4.1) | 27.9 (4.2) | 28.4 (4.4) |
| WHR (SD) | 0.9 (0.1) | 0.9 (0.1) | 0.9 (0.1) |
| Lipid-lowering therapy, % | 17.3 | 22.6 | 29.3 |

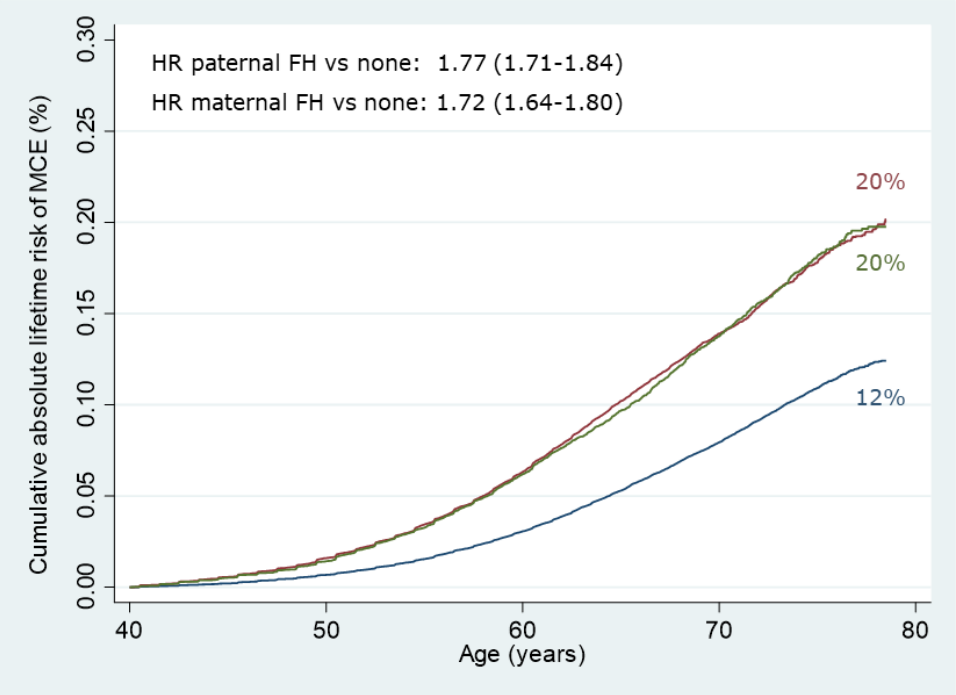
| | | | |
|---|---|---|---|
| Anti-hypertensive therapy, % | 18.9 | 24.3 | 31.0 |
| Diabetes,% | 4.6 | 6.0 | 8.4 |
| Hypertension, % | 53.1 | 59.1 | 64.7 |
| Ever smoker (%) | 32.3 | 35.6 | 38.6 |

Characteristics are presented as means and standard deviations (SD) for normally distributed variables, median and interquartile ranges (IQR) for non-normally distributed variables (TG and CRP), or percentages for dichotomous variables. TC: total cholesterol; LDL-C: low-density lipoprotein cholesterol; ApoB: apolipoprotein B; HDL-C: high-density lipoprotein cholesterol; TG: triglycerides; SBP: systolic blood pressure; CRP: c-reactive protein; BMI: body mass index; WHR: waist to hip ratio. Low PGS level is defined as the lowest decile. Average PGS is defined as the combination of deciles 3,4,5,6,7, and 8; high PGS level is defined as the highest decile.

Family history of heart disease and lifetime risk of MCE

In **Figure 1** panel A, the blue line (reference) represents people with no family history of CHD, while the red line and the green lines represent people with paternal or maternal family history of CHD, respectively. Either having the mother or the father with a history of heart disease increases the lifetime risk of MCE by about the same amount (comparable absolute rate of the disease, 20%), suggesting that maternal and paternal family history of CHD contributes roughly the same in term of risk prediction. In **Figure 1** panel B, is tested the hypothesis of a dose response relationship between parental family history of heart disease and lifetime risk of MCE. The blue line (reference) still represents people with no parental family history of CHD, while the red and the green lines represent people with one parent (either the mother or the father) or both parents with a history of heart disease, respectively. A dose dependent response in the association with MCE appears to be present: having both parents with a history of CHD roughly double the risk (hazard ratio [HR]: 2.78, 95%CI: 2.64-2.92), compared to having only one parent with a history of heart disease (HR: 1.75, 95%CI: 1.70-1.82).

(A) Comparing maternal and paternal family history




(B) Family history dose response

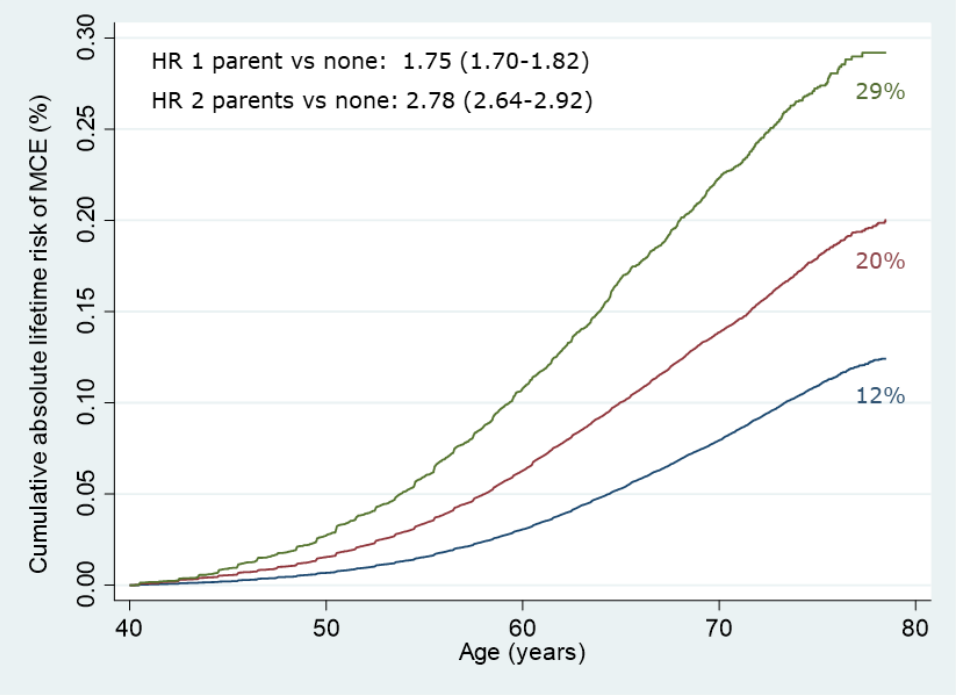



**Figure 1.** Association between family history of heart disease and lifetime risk of major coronary event (MCE)

Polygenic predisposition and lifetime risk of MCE

In **Figure 2**, the blue line represents people in the lowest decile of the polygenic score, the red curve individuals in the second lowest PGS decile, the green line (reference) subjects with an average PGS (defined as the combination of deciles 3, 4, 5, 6, 7, and 8), the orange curve individuals in the second highest PGS decile, while the grey-blue line represents people in the highest PGS decile. This analysis confirmed that the inherited risk that has captured by the polygenic score has an impact on the lifetime risk of coronary heart disease. As it has been observed for family history, also polygenic predisposition has an impact on the lifetime risk of MCE with a dose dependent response.

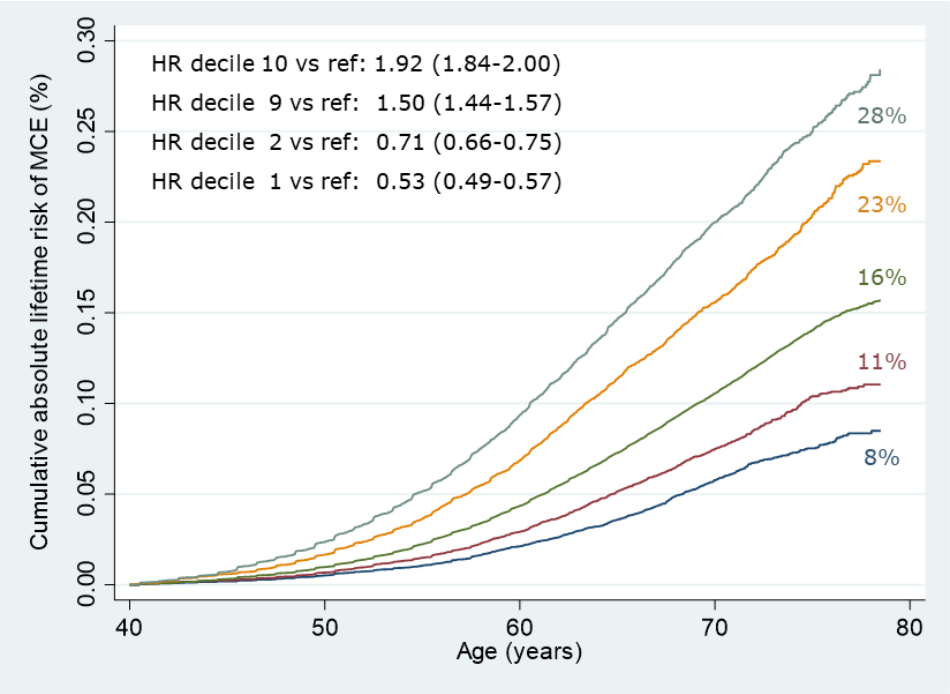



**Figure 2.** Association between polygenic risk score and lifetime risk of major coronary event (MCE)

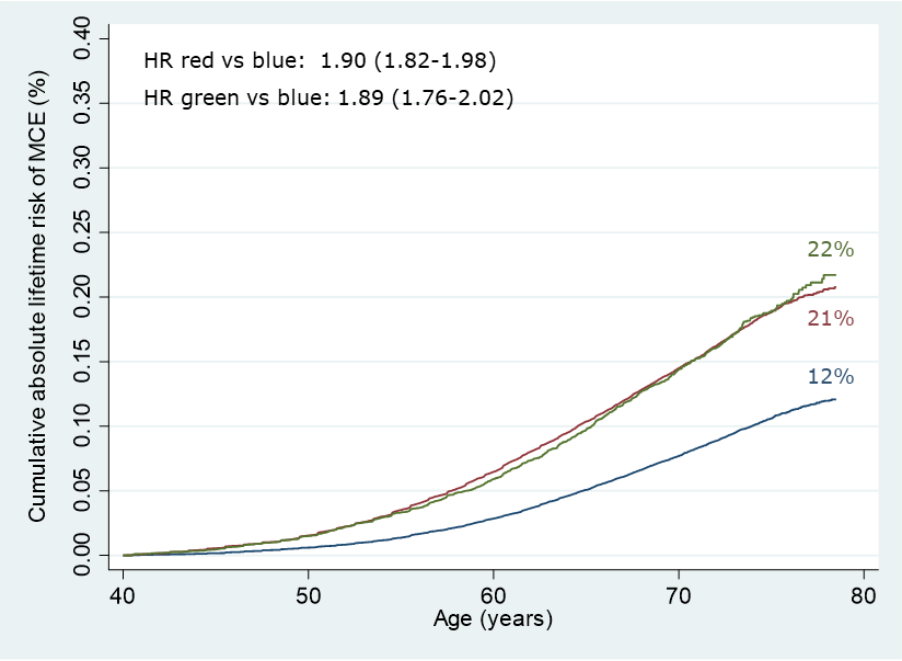
Evaluating the combined effect of family history and polygenic predisposition on the lifetime risk of MCE

In **Figure 3** panel A, the blue line (reference) represents people with no family history of CHD with an average PGS, the red curve represents subjects with one or more parents with a history of heart disease and with an average PGS, while the green line represents people with no family history of CHD but belonging to the highest decile of the polygenic score. It appears clear that having a parental history of CHD (either or both parents) is equivalent, in term of lifetime risk of MCE, as having a very high polygenic predisposition (belonging to the highest PGS decile). **Figure 3** panel B tried to explore whether these effects are independent and additive or largely redundant. The blue (reference) and the red lines represent people with no family history of CHD with an average PGS, and subjects with one or two parents with a history of heart disease and with an average PGS, respectively, while, the green line represents people with parental family history of CHD and belonging to the highest decile of the polygenic score. Having one or two parents with a history of heart disease determined a risk of MCE of 90%, (HR: 1.90, 95%CI: 1.82-1.98), but if in addition it is present also a very high polygenic predisposition, the risk is almost doubled (HR: 3.54, 95%CI: 3.34-3.75). This result shows a clear dose response relationship among this independent information, which is confirmed also evaluating the impact of family history in the lowest decile of the polygenic score (**Figure 3** panel C). The blue (reference) and the red lines still represent people with no family history of CHD with an average PGS, and subjects with one or more parents with a history of heart disease and with an average PGS, respectively, while, the green line represents people with parental family history of CHD and belonging to the lowest decile of the polygenic score. Again, if a

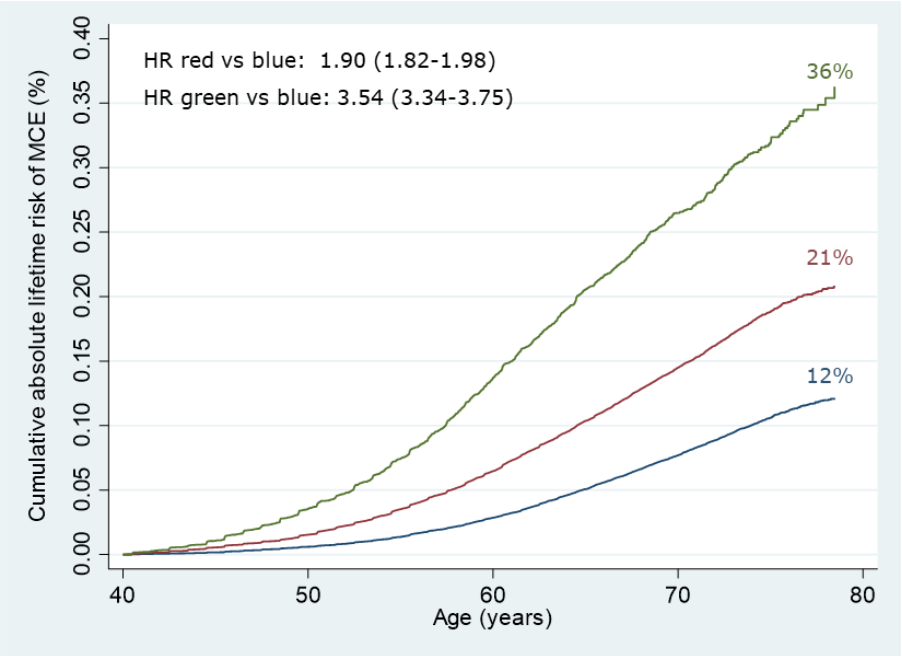
subject has a family history of CHD is exposed to an increased risk; however, if in the same condition the subject belongs also to the lowest decile of the polygenic score, its contribution to the risk disappears. This further implies that parental family history and PGS really provide additional and complementary information.

In summary: (i) the effect of parental family history of CHD and of having very high polygenic predisposition is essentially the same; (ii) the combined effect on lifetime risk of MCE of these two factors is additive; (iii) however, in subjects with a positive family history of CHD (one or two parents) the lifetime risk of MCE is comparable to that of the average population, if they are characterized by a very low polygenic predisposition. This evidence emphasizes that inherited risk has to be characterized jointly by both family history and polygenic predisposition, because they contribute with independent and additive information to the characterization of the lifetime risk.

(A) Comparing the effect of parental family history and highest decile of PGS on lifetime risk of MCE



(B) Combined effect of family history and highest decile of PGS on lifetime risk of MCE




(C) Combined effect of family history and lowest decile of PGS on lifetime risk of MCE

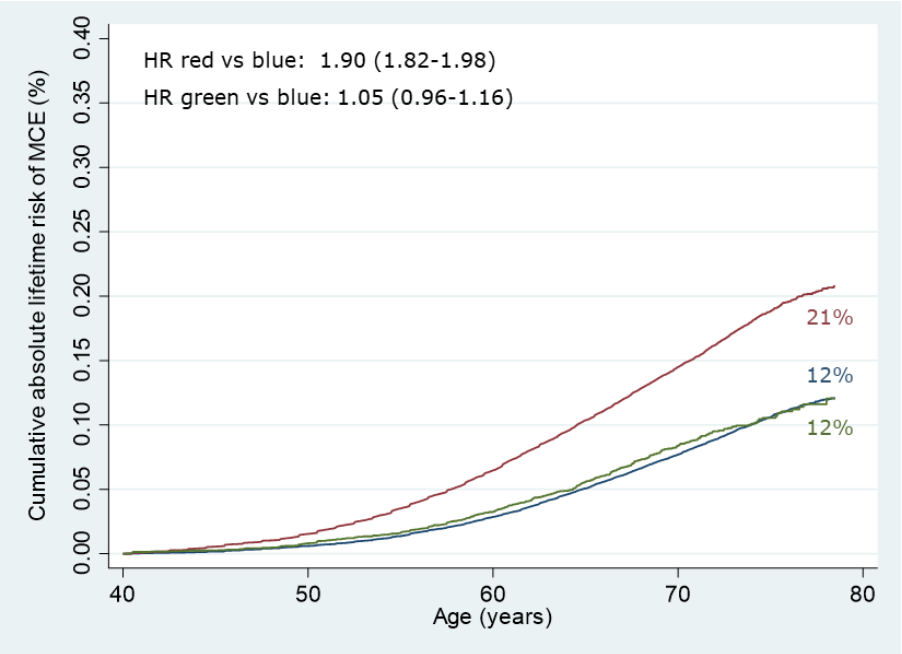
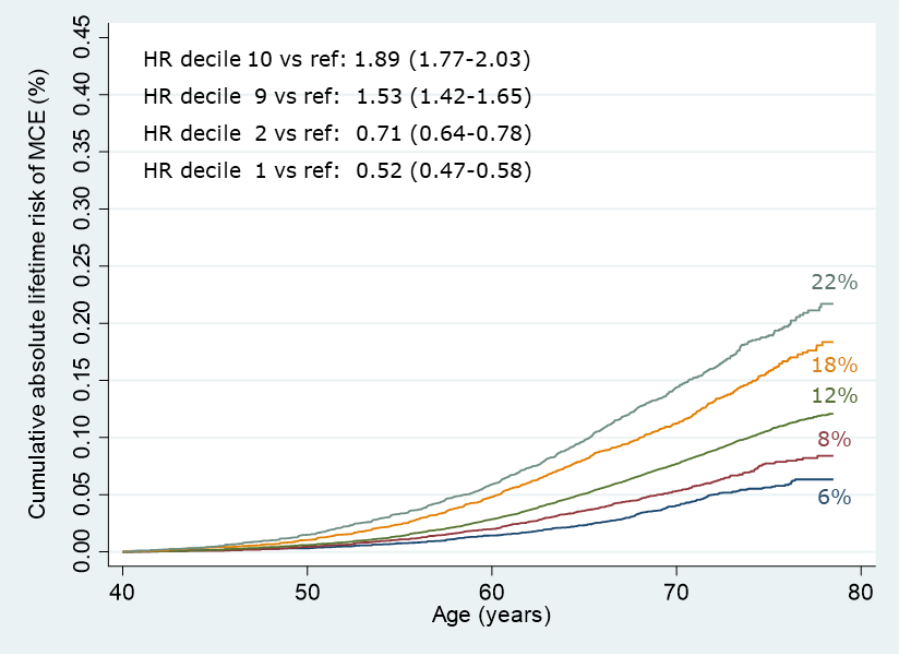



**Figure 3.** Combined effect of family history and PGS on lifetime risk of major coronary event (MCE).

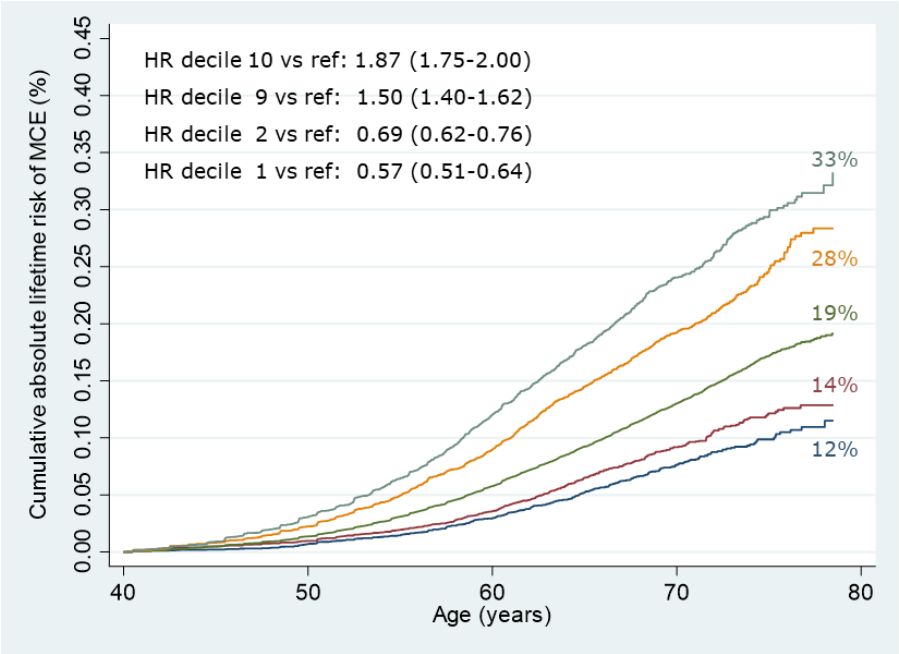
Effect of PGS on lifetime risk of MCE by family history

Regardless the presence of family history (**Figure 4**), the lifetime risk of MCE depends on the polygenic predisposition (the risk across PGS classes varied by roughly the same amount in each scenario).

(A) No family history of heart disease




(B) One parent with a history of heart disease



(C) Both parents with a history of heart disease

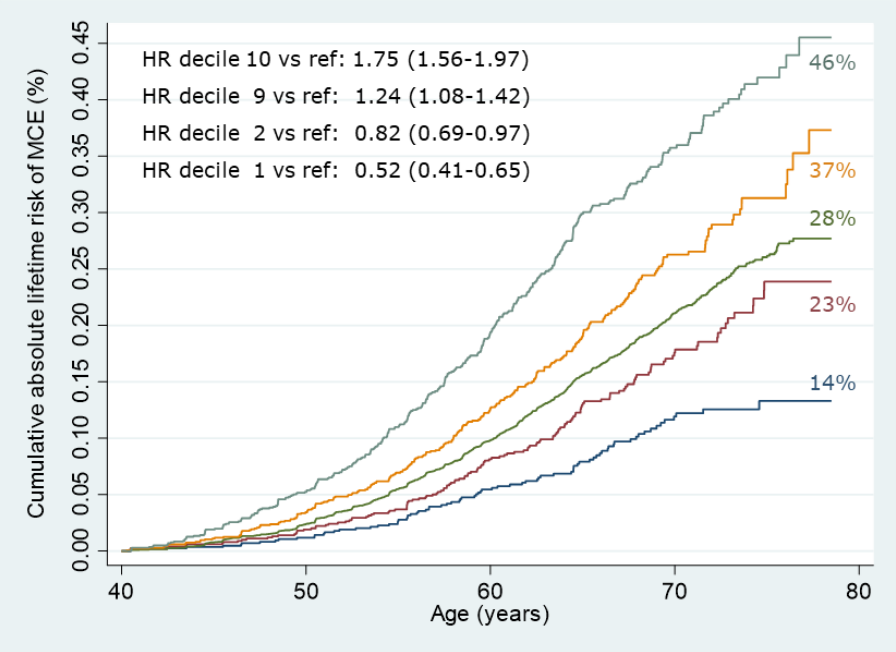


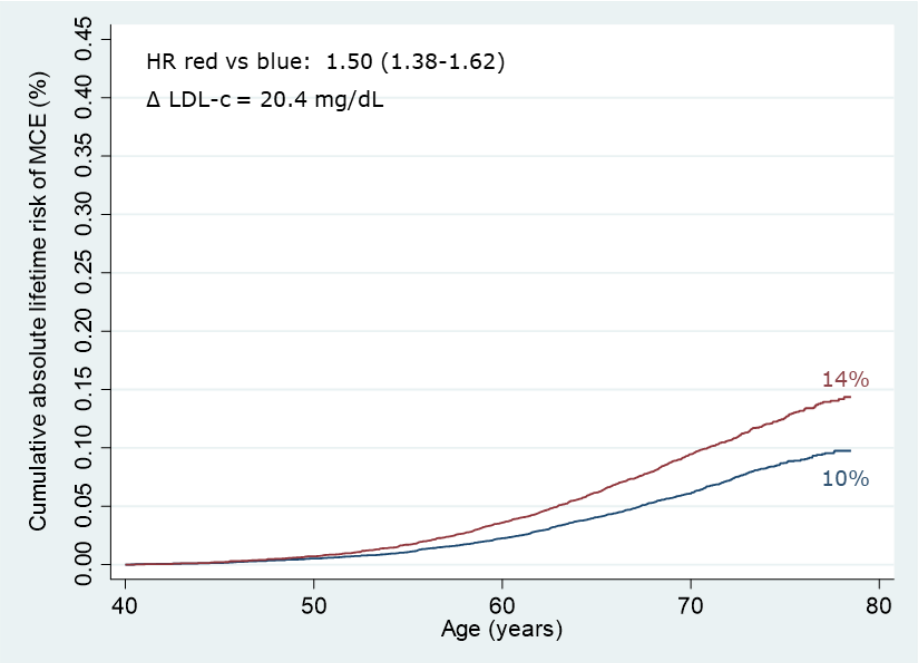

**Figure 4.** Effect of PGS on lifetime risk of major coronary event (MCE) by parental family history

Effect of LDL-C on lifetime risk of MCE by family history, PGS or both

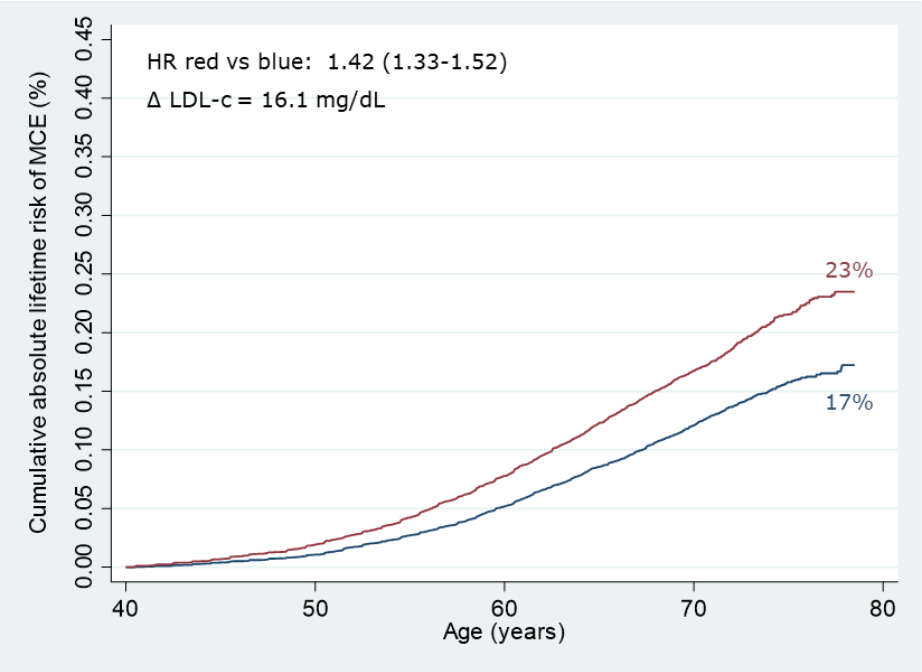
Before trying to use these two factors together to identify people at the highest risk who may benefit the most from lowering LDL-C early in life, we first assessed if the lifetime exposure to half mmol higher LDL-C has the same effect depending on whether or not the family history is present, on PGS level, or both. In **Figure 5** panel A, both the blue (reference) and red lines represent people with no family history of CHD with an average PGS, but the second group is also characterized by half mmol higher LDL-C compared to the reference group. In this scenario each half mmol LDL-C increases the risk of 50%. In panel B, the same analysis was performed but in this case only among people with one or two parents with heart disease. Again, LDL-C increases the risk by roughly the same amount (adjusting by the observed difference in cholesterol between the groups). The same was observed also among people with the highest decile of the

polygenic score as compared with the average decile. In **Figure 5** panel C, both the blue (reference) and red lines represent people with no family history of CHD in the highest PGS decile, but the second group is also characterized by half mmol higher LDL-C compared to the reference group. As previously, the estimates are roughly the same: the effect of LDL-C increases the risk of MCE regardless the presence of both parental family history or high polygenic predisposition. In panel C it is addressed what happens if both these conditions are present. The blue (reference) and red lines represent people with family history of CHD in the highest PGS decile, but the second group is also characterized by half mmol higher LDL-C compared to the reference group. LDL-C increases the risk again by roughly the same amount, adjusting by the observed difference in cholesterol between the groups.

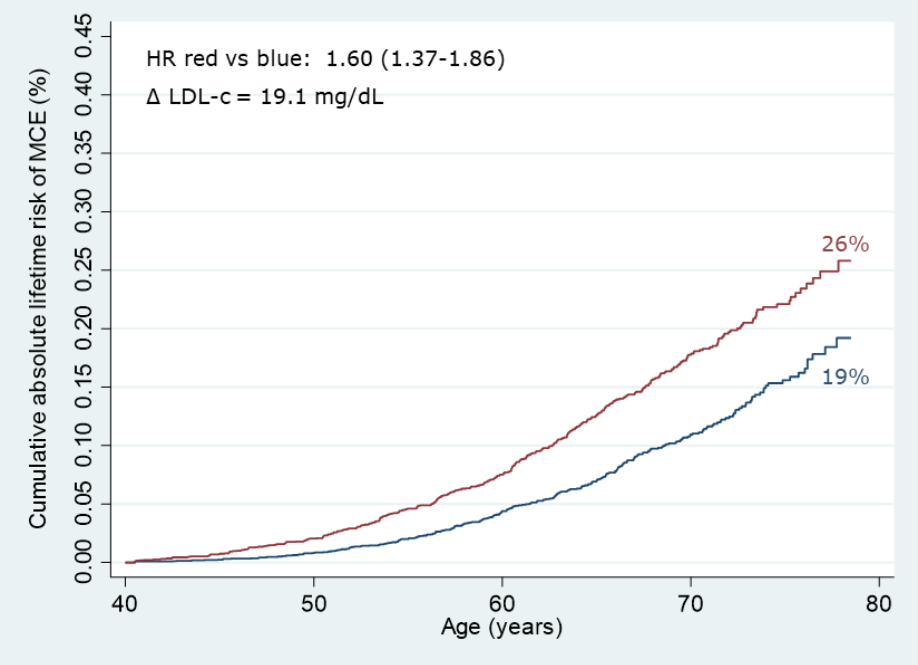
(A) Effect of LDL-C on lifetime risk among participants with no family history of heart disease and average PGS



(B) Effect of LDL-C on lifetime risk among participants with family history of heart disease and average PGS




(C) Effect of LDL-C on lifetime risk among participants with no family history of heart disease and highest decile of PGS



(D) Effect of LDL-C on lifetime risk among participants with a family history of heart disease and highest decile PGS

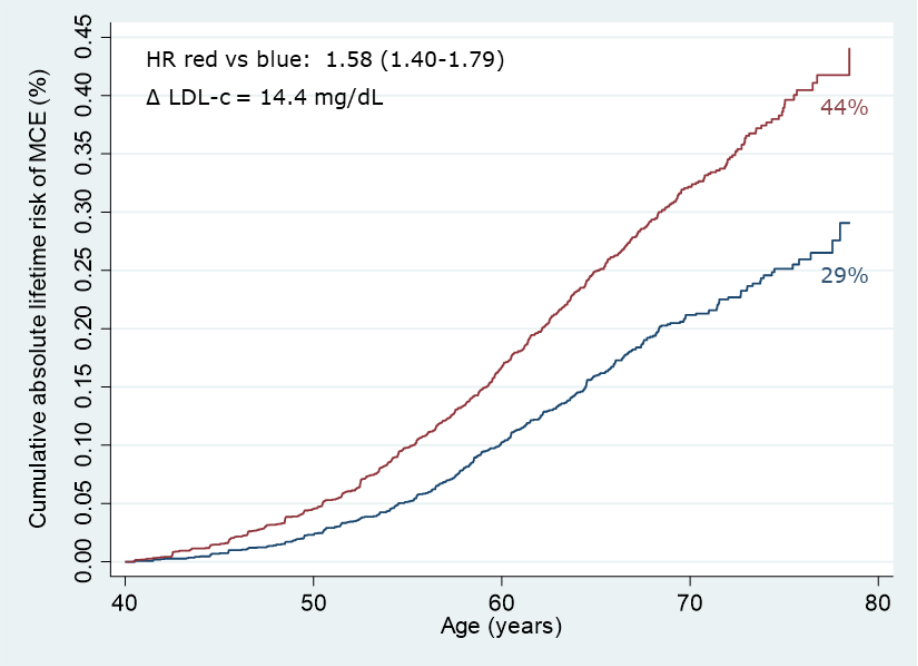



**Figure 5.** Effect of LDL-C on lifetime risk of major coronary event (MCE) by family history, polygenic predisposition (PGS) or both.

Identifying people who may benefit from an early intervention

**Table 3** identifies individuals in the population who have the most to gain from preventive interventions based not only on the evaluation of LDL-C levels but also on the presence of parental family history of CHD and on polygenic predisposition. Overall, the lifetime risk of MCE associated with the average LDL-C level in the population is 16% (estimated carried out a survival model). Increasing or decreasing LDL-C of half and one mmol compared with the median, determines the increase or decrease of the lifetime risk accordingly (23%, 33%, 11%, 7%, respectively). If we assume to treat everybody with a lifetime risk equal or greater than 25%, in this scenario, only subjects with very high LDL-C would be treated. The scenarios presented in the second and third part of the table illustrate what would happen if

the assessment of who should be treated is made not only considering the value of LDL-C but also the parental family history of CHD and the individual polygenic predisposition.

Considering LDL-C lifetime exposure and family history, everybody who has LDL-C one mmol higher than the median is still treated. However, in this case, even subjects with median LDL-C level but with two parents with a history of heart disease are eligible for treatment (lifetime risk of MCE equal to 29% that is higher than the established threshold), and they would have been completely missed just relying on LDL-C. Similarly, those people who have half mmol higher LDL-C than the average, they now need to be treated if they have either or both parents with a history of CHD. In this case, parental family history information has been useful to identify additional people with high lifetime risk of CHD who deserved to be selected for treatment that they would have been missed just relying on cholesterol.

Considering LDL-C lifetime exposure and polygenic predisposition, everybody who has LDL-C one mmol higher than the median is still treated, except people in the lowest PGS group, because they have such a low genetic predisposition which probably overcome the risk due to high LDL-C. In addition to that, even subjects with half nmol higher and either the highest nine or ten deciles are now eligible for treatment because their lifetime risk of MCE exceeds the established threshold, as well as individuals in the highest PGS decile even if they have average LDL-C level. Once again, this analysis proved that family history and polygenic predisposition can capture few additional groups of patients with high lifetime risk (and not only people with very high LDL-C) that they would have been missed otherwise.

**Table 4** presents what happen if lifetime exposure to LDL-C, parental family history, and polygenic predisposition are considered jointly to

identify subjects at higher risk instead of using only cholesterol level evaluation alone.

**Table 3.** Lifetime risk (%) of major coronary events by exposure to LDL cholesterol, parental family history, and polygenic predisposition (PGS). A lifetime risk greater than 25% is assumed as the threshold to identify people to treat.

| | **LDL cholesterol level** | | | | |
|---|---|---|---|---|---|
| | **-1.0 mmol/L** | **-0.5 mmol/L** | **median mmol/L** | **+0.5 mmol/L** | **+1.0 mmol/L** |
| | | | | | |
| All | 7 | 11 | 16 | 23 | 33 |
| | | | | | |
| Both parents | 13 | 19.7 | 29 | 41 | 57 |
| One parents | 10 | 14.3 | 20 | 28 | 37 |
| None | 6 | 8.2 | 12 | 17 | 25 |
| | | | | | |
| PGS decile 10 | 11 | 18 | 28 | 42 | 60 |
| PGS decile 9 | 10 | 15 | 23 | 34 | 48 |
| PGS average | 8 | 11 | 16 | 23 | 31 |
| PGS decile 2 | 5 | 7 | 11 | 17 | 25 |
| PGS decile 1 | 5 | 6 | 8 | 11 | 14 |

The lifetime risk of CHD is now evaluated using both the information of LDL-C lifetime exposure and polygenic predisposition among people without a parental family history of CHD (**Table 4** panel A), and people with one (panel B) or both (panel C) parents with heart disease. If a subject has no parents with heart disease but very high LDL-C, he/she appears to be at higher risk of MCE only if is present also a very high polygenic predisposition ($9^{th}$ or $10^{th}$ deciles). The high polygenic predisposition is responsible for the increased lifetime risk (>25%) even among subjects without a parental history of CHD

and having half mmol higher LDL-C than the median in the population. Amon people with just one parent with heart disease we would treat almost everybody with high LDL-C, except people in the very lowest decile of the polygenic score. In addition to that, with this new evaluation, we can capture also almost half of the people with one parent with heart disease and having also only half mmol higher LDL-C, but surprisingly also those with average LDL-C level who are in the highest deciles of the polygenic score. Finally, everyone with both parents with heart disease and very high LDL-C are at higher lifetime risk, and eligible for treatment, exactly how it would have been. However, a further selection can be made among subjects having only half mmol higher LDL-C or average LDL-C level based on PGS level. Remarkably, even an individual with both parents with heart disease who has half mmol lower LDL-C than the average cholesterol level in the population would be selected for treatment if exposed to a very high polygenic predisposition.

**Table 4.** Lifetime risk (%) of major coronary events by exposure to LDL cholesterol and polygenic predisposition (PGS) stratified by parental family history. A lifetime risk greater than 25% is assumed as the threshold to identify people to treat.

(A) No family history of heart disease

| | **LDL cholesterol level** | | | | |
|---|---|---|---|---|---|
| | **-1.0 mmol/L** | **-0.5 mmol/L** | **median mmol/L** | **+0.5 mmol/L** | **+1.0 mmol/L** |
| | | | | | |
| All | 7 | 11 | 16 | 23 | 33 |
| | | | | | |
| PGS decile 10 | 9 | 15 | 22 | 33 | 47 |
| PGS decile 9 | 7 | 11 | 18 | 28 | 42 |

| | | | | | |
|---|---|---|---|---|---|
| PGS average | 6 | 9 | 12 | 17 | 23 |
| PGS decile 2 | 3 | 5 | 8 | 12 | 19 |
| PGS decile 1 | 3 | 4 | 6 | 9 | 14 |

(B) One parent with heart disease

| | **LDL cholesterol level** | | | | |
|---|---|---|---|---|---|
| | **-1.0 mmol/L** | **-0.5 mmol/L** | **median mmol/L** | **+0.5 mmol/L** | **+1.0 mmol/L** |
| | | | | | |
| All | 7 | 11 | 16 | 23 | 33 |
| | | | | | |
| PGS decile 10 | 15 | 22 | 33 | 47 | 64 |
| PGS decile 9 | 16 | 21 | 28 | 37 | 47 |
| PGS average | 10 | 14 | 19 | 26 | 36 |
| PGS decile 2 | 7 | 10 | 14 | 20 | 27 |
| PGS decile 1 | 7 | 9 | 12 | 16 | 21 |

(C) Both parents with heart disease

| | **LDL cholesterol level** | | | | |
|---|---|---|---|---|---|
| | **-1.0 mmol/L** | **-0.5 mmol/L** | **median mmol/L** | **+0.5 mmol/L** | **+1.0 mmol/L** |
| | | | | | |
| All | 7 | 11 | 16 | 23 | 33 |
| | | | | | |
| PGS decile 10 | 12 | 25 | 46 | 74 | 95 |
| PGS decile 9 | 12 | 21 | 37 | 59 | 82 |
| PGS average | 14 | 20 | 28 | 38 | 51 |
| PGS decile 2 | 13 | 17 | 23 | 30 | 39 |
| PGS decile 1 | 8 | 10 | 14 | 19 | 25 |

## Discussion

Based on these results, it appears clear that: (i) maternal and paternal family history of CHD have the same effect on the lifetime

risk of MCE, but if both the parents were affected by heart disease then the risk doubles (dose response relationship); (ii) also the polygenic predisposition affects the lifetime risk of MCE, in a dose response way; (iii) parental family history not only captures inherited genetic variation, but likely represents also lifestyle and social determinants of health, and therefore, providing independent, complementary and additive information to the polygenic predisposition, but also to the lifetime exposure to LDL-C. Consequently, in order to more accurately identify those who are at risk of having an event early in life, is essential to evaluate LDL-C exposure, parental family history of CHD, and the individual polygenic predisposition jointly. If it is true that LDL-C level (because it is the target of the therapy) mainly drives the definition of the lifetime cardiovascular risk, a further improvement in the estimation of the risk can be obtained evaluating this exposure among subjects with a family history of CHD and taking into account the level of the genetic predisposition (even individuals with low level of LDL-C may benefit from lowering cholesterol).

Atherosclerotic cardiovascular disease (ASCVD) has many risk factors, some of which cannot be changed, and some of which are more easily modifiable. One of the major non-modifiable risk factors for ASCVD is family history (183). Several studies in the past have suggested that, especially when other important risk factors are accounted for, even a history of a single first-degree relative of any age with a history of CHD identifies the patient as having an increased risk of CHD (184). Nevertheless, family history is not part of all CHD risk assessment algorithms or calculators. This may have occurred for several reasons, such as bias in recalling family history, as well as its variable predictive value, depending on whether the prediction is for premature or later-onset events (183). Despite this,

family history (especially if it includes premature ASCVD), in combination with the evaluation of LDL-C exposure and polygenic predisposition, represents a key factor in risk stratification for ASCVD in general and specifically in identifying subjects eligible for lipid-lowering therapy. Before embracing family history as a public health screening strategy, a number of critical issues should be addressed. First, family history must be consistently recorded in the electronic health record to be impactful in advanced risk estimation algorithms. For example, a binary predictor describing the presence or absence of family history is less informative than more precise family history records such as: age at time of family history report, the number of affected relatives, relationship to relatives with disease, severity of disease in the family member, or age of disease onset/diagnosis in these relatives. Differentiating between first-degree relative (mother, father, sibling) and second-degree relative (sisters and brothers, grandparents, aunts, uncles) will yield specificity as to the degree of shared genetic liability. Even more useful is a grid of diseases and relationships to allow for higher resolution family history variables. The age at time of reporting family history should be recorded and regular updates to family history information will improve prediction based on family history (185).

Limitations

This study has several limitations. First, participants in the UK biobank are a self-selected group who tend to be at lower risk of cardiovascular events than members of other populations. As a result, all analyses should be repeated in populations at higher risk. Second, we conducted the analysis just among men, because in the UK Biobank the events rate in this cohort is 3-4-fold time higher than those in women. In addition, the analysis was restricted to

participants who self-identified as being of white European ancestry, and therefore results may not apply to persons of other ethnicities. Therefore, this study should be repeated in other populations.

**Conclusion**

Parental family history of CHD provides complementary and additive information to the individual polygenic predisposition in the definition of the inherited genetic variation as well as to LDL-C levels exposure in the estimation of the lifetime cardiovascular risk. In order to develop a simple, but powerful, algorithm to really identify subjects at higher risk of having an early event, especially if they are young, it is essential to retrieve information about parental family history of heart disease and individual polygenic predisposition to CAD, in addition to the measurement of all the other well-known cardiovascular risk factors, especially LDL-C levels. Only considered together, these three factors are able to contextualize the frame of who will need to be treated.

## 5.3 Does the risk of atherosclerotic cardiovascular disease vary based on measured or genetically determined lipoprotein(a)?

### Background

Apolipoprotein(a), which is encoded by the LPA gene, covalently binds to a cholesterol-rich low-density lipoprotein (LDL) particle to form lipoprotein(a) [Lp(a)].

Meta-analyses of prospective observational studies have reported that higher plasma Lp(a) concentration is associated with dose-dependent higher risk of atherosclerotic cardiovascular disease (ASCVD) (186). These findings are confirmed by Mendelian randomization analyses which have provided strong evidence that Lp(a) is a causal contributor to ASCVD (187, 188).

Lp(a) levels are 75% to 95% heritable and predominately determined by single-nucleotide variants at the LPA gene and copy number variants specifically in the kringle IV type 2 domain (189, 190). Genetic association studies have identified genetic variants explaining approximately 60% of the variability in Lp(a) levels in European populations (191). Elevated Lp(a), defined as Lp(a) levels of more than 120 nmol/L or approximately 50 mg/dL, is relatively common, though its prevalence varies in prevalence by ancestry (affects 1 in 5 European individuals) (186).

Both the diagnostic yield and clinical value of genetic testing of LPA are not well understood. Specifically, it remains unclear whether genetic factors related to Lp(a) that provide information regarding lifetime exposure relevant to ASCVD risk prediction may provide different or additional features compared with the measurement of Lp(a) concentrations in clinical practice.

Therefore, we aimed at comparing the cumulative lifetime risk of major coronary events (MCE) among subjects with different Lp(a) genotype and measured Lp(a) concentrations.

## Methods

Study population

The UK Biobank is a prospective observational study of approximately 500,000 volunteer adults aged 40 to 69 years recruited from 22 sites across the United Kingdom between 2006 and 2010, with follow-up ongoing. Biochemical measurements, physical examination measurements, and medical histories were assessed at the time of study enrolment. Participants underwent genotyping with one of two closely related custom arrays (UK BiLEVE Axiom Array or the UK Biobank Axiom Array) consisting of over 800,000 genetic markers, with additional genotypes imputed using the Haplotype Reference Consortium resource, the UK10K panel, and the 1000 Genomes panel. The KING toolset was used to identify up to third-degree relatedness based on kinship coefficients >0.044.

The UK Biobank protocol was approved by the Northwest Multi-Center Research Ethics Committee, and all study participants provided written informed consent (more details in **Appendix I**).

Lipoprotein(a) measurement

Lp(a) was measured in nanomoles per litre at study enrolment using an immunoturbidimetric method on the Beckman Coulter AU5800 platform (Randox Bioscience, UK), which is isoform insensitive (192). To convert Lp(a) values to milligrams per decilitre, divide by 2.15.

Lp(a) polygenic score

An Lp(a) polygenic score was calculated for each UK Biobank

participant by summing the number risk-increasing alleles inherited at rs3798220 and rs10455872 variants, accounting for at least 40% of Lp(a) concentrations variation, weighted by the effect size for each allele (**Appendix IV**). Because these two variants are strongly associated with plasma Lp(a) concentrations, they can serve as surrogate markers for Lp(a) levels.

Study outcomes and statistical analysis

The primary outcome was MCE, defined as the first occurrence of either a fatal or non-fatal myocardial infarction (MI), or coronary revascularization. The analysis used Cox proportional hazards models adjusted for age, sex, and the first 10 principal components of ancestry, with age as the time scale. Each participant was censored at the age he/she experienced either a primary outcome event, death due to a cause other than MI (treated as a competing risk), or at the age of last reported follow-up. The dates of all incident events were recorded from hospital episode statistics, while the dates of events that were prevalent at the time of enrolment into UK Biobank were recorded either from hospital episode statistics or self-reported. Cumulative lifetime risk of MCE was plotted, using Kaplan-Meier curves, for participants within each class of the Lp(a) genetic score. All analyses were performed using Stata (version 16; StataCorp). A 2-tailed p-value less than 0.05 was considered statistically significant.

## Results

Participant characteristics

A total of 445,744 participants enrolled in the UK Biobank with complete genetic and principal component data who self-identified as being of white ancestry were included in the study.

The median (SD) age at study enrolment was about 57 (8) years, with no difference observed among Lp(a) genetic score genotype (**Table 1**). No differences were also observed in the distribution of all the other covariates evaluated, across the different classes, confirming a random allocation of subjects. The median [IQR] level of Lp(a) [nmol/L] increased with increasing number of the genetic score copies (13.6 [6.2-35.0], 146.3 [104.8-200.2], 261.8 [190.2-336.0]), as expected. Descriptive analyses stratified by rs10455872 and rs3798220 variants confirmed the same evidence (**Appendix V**).

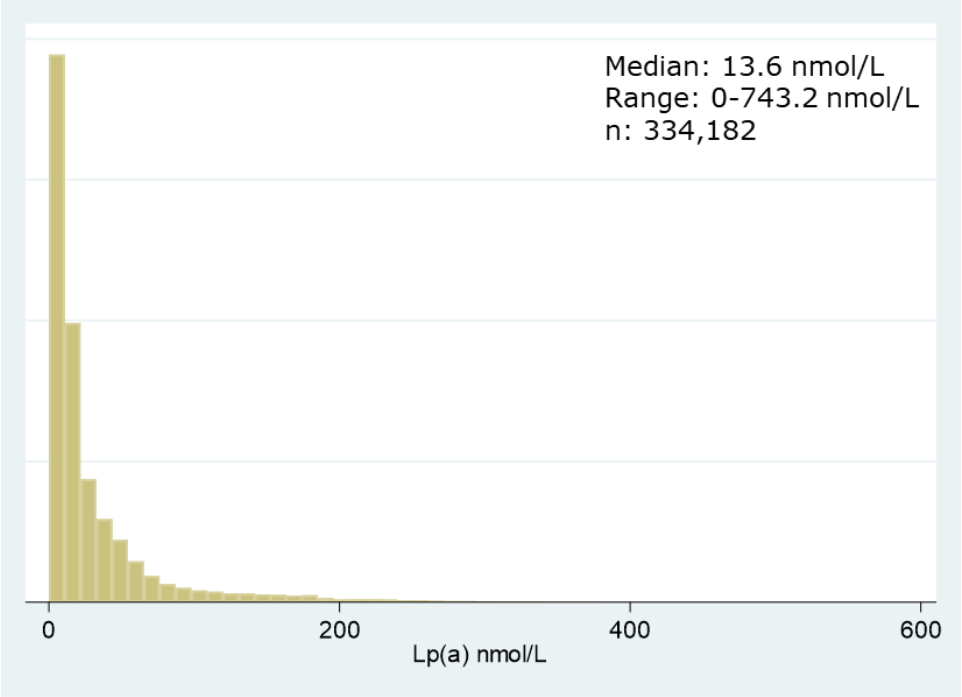
**Table 1.** Baseline characteristics, measured at the time of enrolment in the UK Biobank, are presented for the entire cohort and by Lp(a) instrument variable (IV) genetic score copies.

| **Characteristics** | **Overall** | **Lp(a) IV genetic score copies** | | |
|---|---|---|---|---|
| | | **0** | **1** | **2** |
| No. participants | 445,774 | 358,469 | 77,658 | 4,250 |
| Age, y (SD) | 57.3 (8.0) | 57.2 (8.0) | 57.3 (8.0) | 57.2 (8.0) |
| Female Sex (%) | 54.3 | 54.2 | 54.3 | 54.3 |
| LDL-C, mg/dL (SD) | 138.0 (33.6) | 137.4 (33.5) | 140.4 (34.1) | 142.7 (34.7) |
| ApoB, mg/L (SD) | 103.4 (23.8) | 103.0 (23.8) | 105.2 (23.9) | 106.8 (24.2) |
| TG, mg/dL (IQR) | 131.9 [93.1-190.7] | 132.7 [93.7-191.8] | 128.5 [91-186.9] | 124 [87.7-180.7] |
| HDL-C, mg/dL (SD) | 56.2 (14.8) | 56.1 (14.8) | 56.4 (14.9) | 56.5 (15.3) |
| Lp(a), nmol/L (IQR) | 18.7 [7.4-72.9] | 13.6 [6.2-35.0] | 146.3 [104.8-200.2] | 261.8 [190.2-336] |
| CRP, mg/L (IQR) | 1.33 [0.66-2.75] | 1.3 [0.7-2.8] | 1.3 [0.7-2.8] | 1.3 [0.7-2.7] |
| SBP, mmHg (SD) | 137.8 (18.6) | 137.8 (18.6) | 137.9 (18.6) | 137.8 (18.6) |
| BMI, kg/m$^2$ (SD) | 27.4 (4.8) | 27.4 (4.8) | 27.4 (4.8) | 27.4 (4.9) |
| No. MCE | 23,032 | 17,110 | 5,313 | 365 |

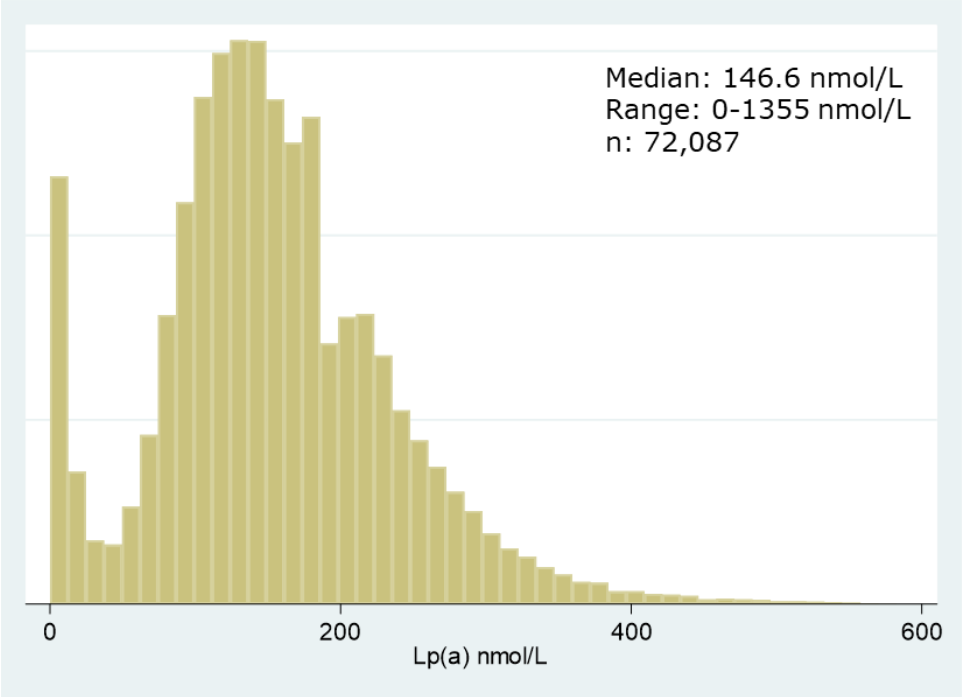
Characteristics are presented as means and standard deviations (SD) for normally distributed variables, median and interquartile ranges (IQR) for non-normally distributed variables (triglycerides, Lp(a), and CRP), or percentages for dichotomous variables. LDL-C: low-density lipoprotein cholesterol; ApoB: apolipoprotein B; TG: triglycerides; HDL-C: high-density lipoprotein cholesterol; Lp(a): lipoprotein(a); SBP: systolic blood pressure; CRP: c-reactive protein; BMI: body mass index; MCE: major coronary events.

Among subjects with the same genotype, the distribution of measured Lp(a) was characterized by high variability, as shown in **Figure 1** (panel A for individuals with zero copies of the score, panel B for subjects with one copy of either rs10455872 or rs3798220, and panel C for individuals with two copies of either rs10455872 or rs3798220).

(A) Lp(a) score equal to 0



(B) Lp(a) score equal to 1




(C) Lp(a) score equal to 2

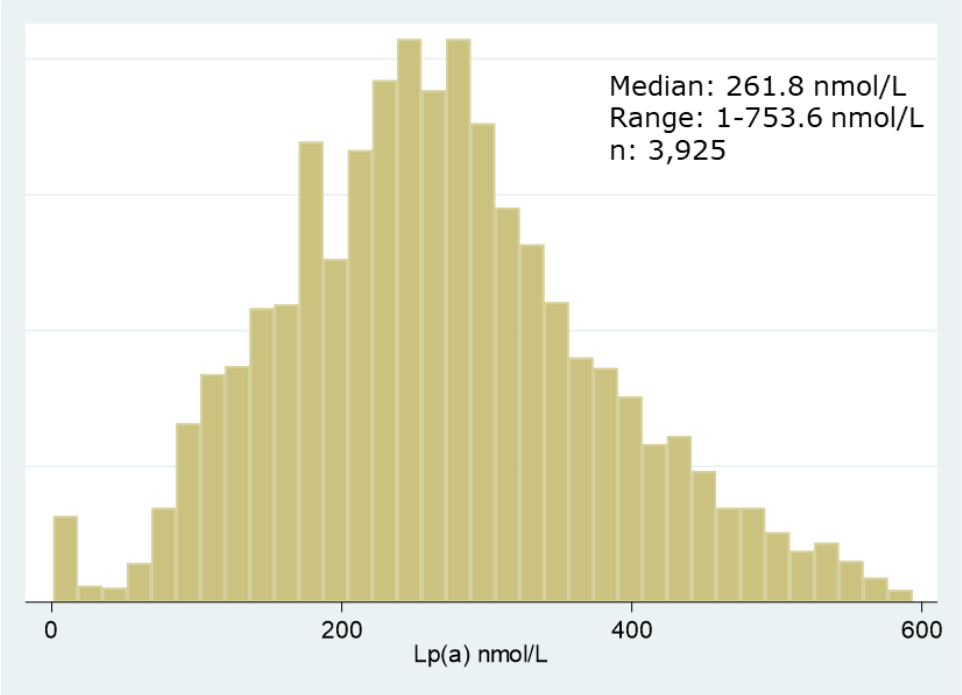
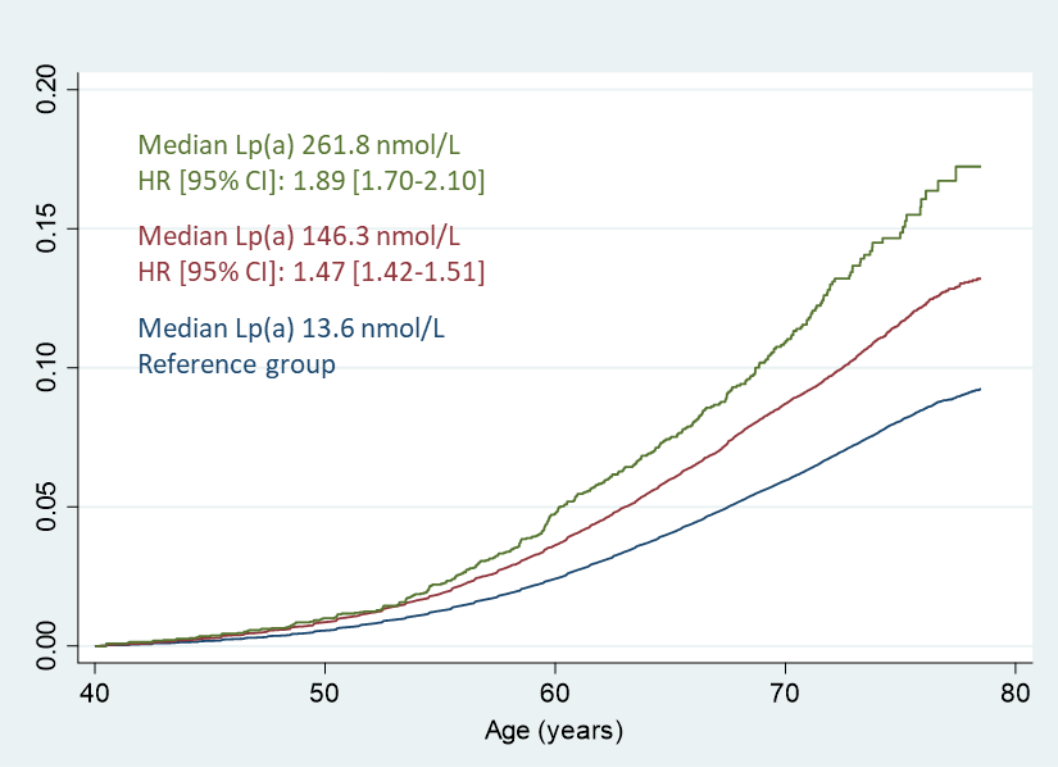



**Figure 1.** Distribution of measured Lp(a) concentrations by Lp(a) genetic score.

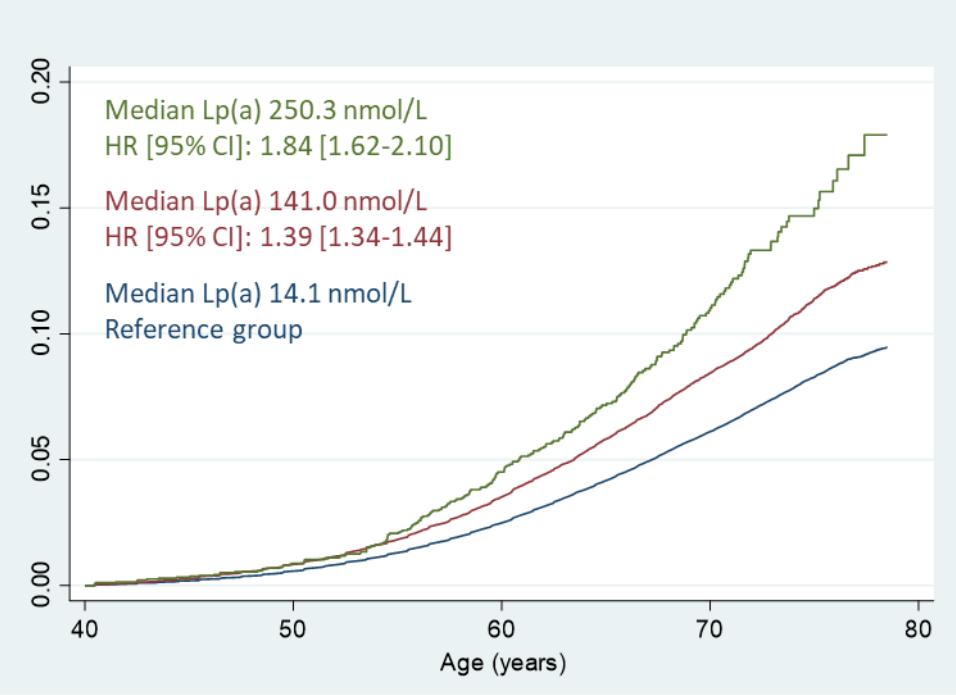
Association of Lp(a) genetic score with MCE

In **Figure 2**, it can be observed that the increased number of copies of the Lp(a) score was associated with raising trajectories of lifetime risk of MCE (panel A). Accordingly, we found similar rate of incident MCE with increasing number of copies of rs10455872 and rs3798220 variants (panel B and C, respectively).

(A)



(B)




(C)

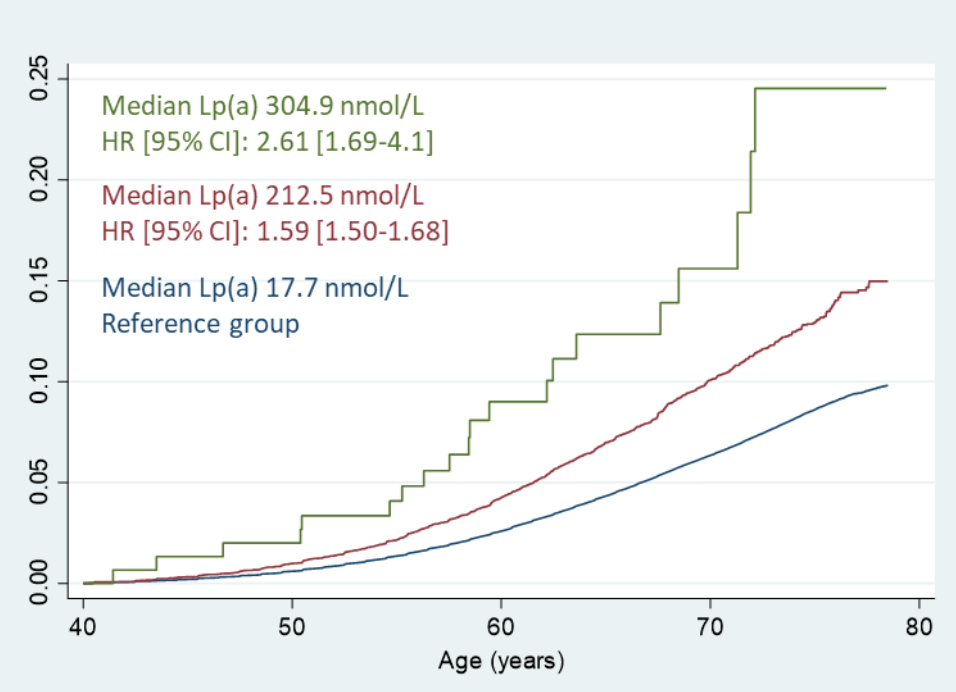
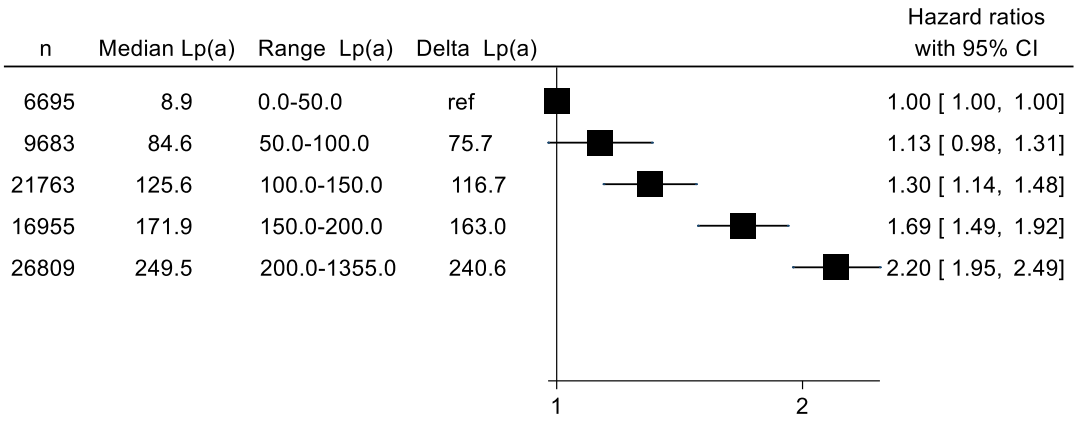


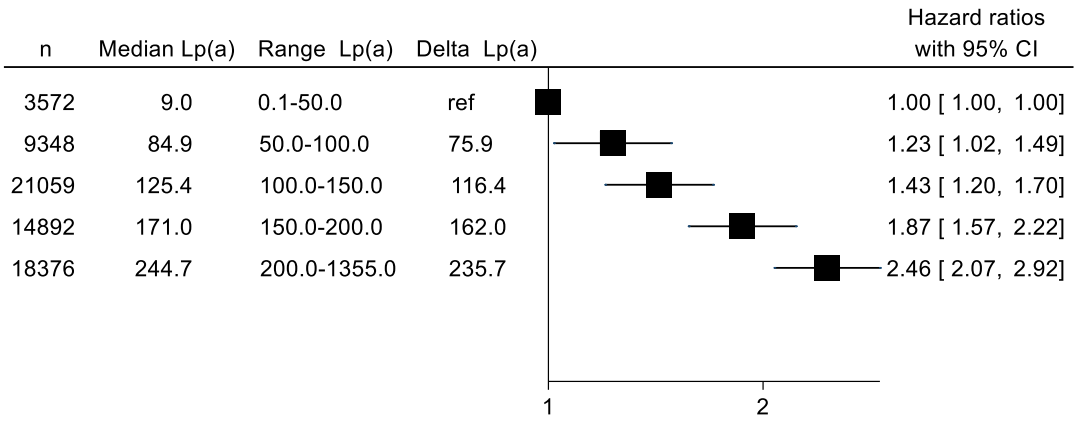

**Figure 2.** Event curves for lifetime risk of major coronary events (MCE) by Lp(a) genetic score, rs10455872, and rs3798220 copies. Effects of genetically predicted Lp(a) on MCE are reported as Hazard ratios (HR) and 95% confidence intervals (95%CI).

Even among subjects with the same genotype (Lp(a) genetic score equal to 1 or 2 vs 0; rs10455872 equal to 1 or 2 vs 0; rs3798220 equal to 1 or 2 vs 0), the risk is driven by genetically predicted Lp(a) levels (**Figure 3**). Increasing quintiles of Lp(a) concentrations, were associated with a step-wise increase in the risk of MCE in all the scenarios: participants in the highest Lp(a) range had a hazard ratio (HR) for MCE of 2.20 (95%CI: 1.95-2.49), 2.46 (95%CI: 2.07-2.92), and of 1.96 (95%CI: 1.64-2.34), respectively, as compared to participants in the lowest range. This analysis also suggests that the risk of ASCVD is proportional to the absolute change in Lp(a) concentrations.

(A) Lp(a) score equal to 1




(B) rs10455872 equal to 1



(C) rs3798220 equal to 1

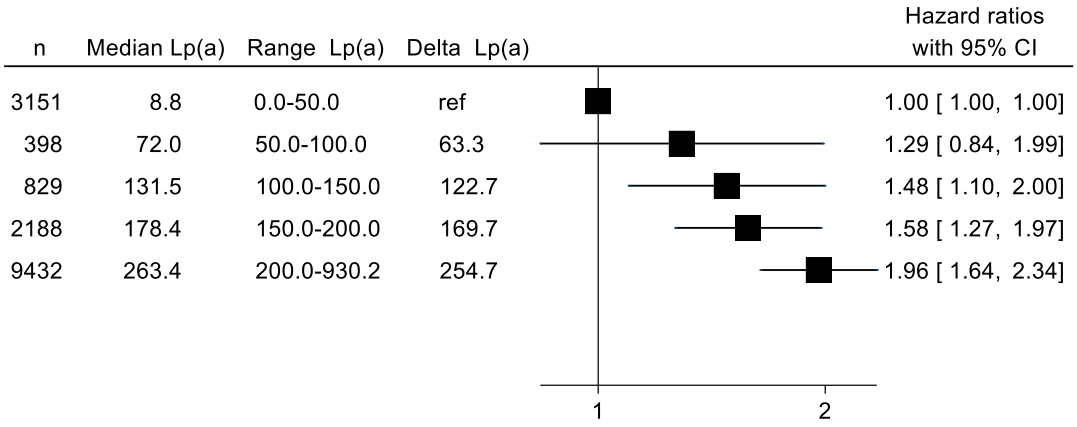
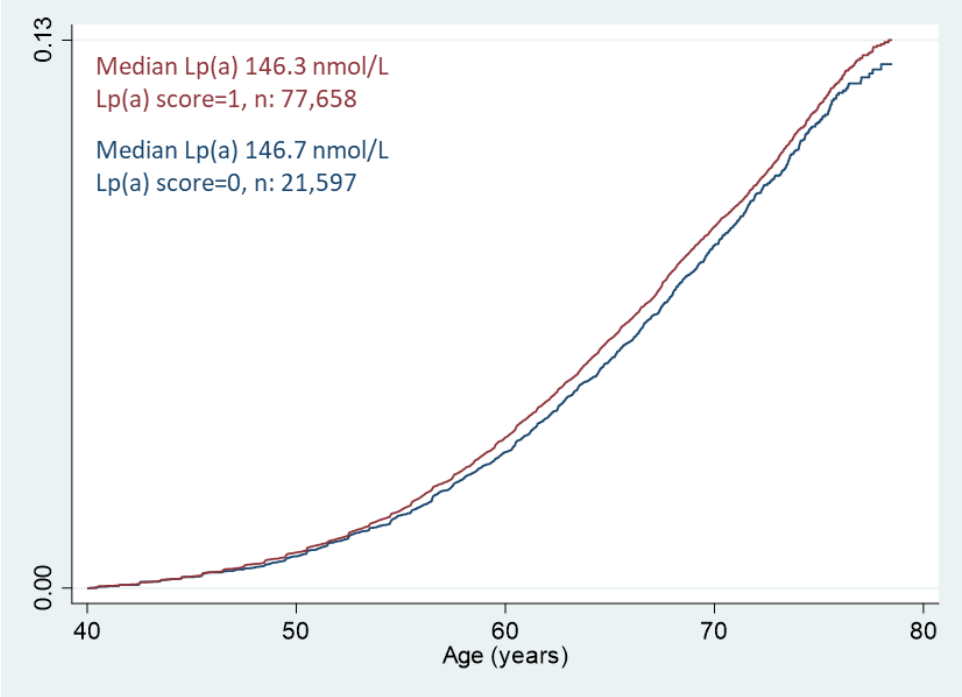



**Figure 3.** Cumulative hazard estimates of major coronary events by defined ranges of measured Lp(a) values among people with the same genotype of the Lp(a) score, and of the two variants evaluated, separately.

In addition, even among individuals with different genotype, the lifetime risk of MCE was comparable regardless the number of Lp(a) genetic score copies, when subjects have been selecting for having similar median Lp(a) concentrations (**Figure 4**), suggesting that the lifetime risk can be accurately predicted using measured values when the genotype is unknown.

(A) Lp(a) score equal to 0 and 1




(B) Lp(a) score equal to 0 and 2

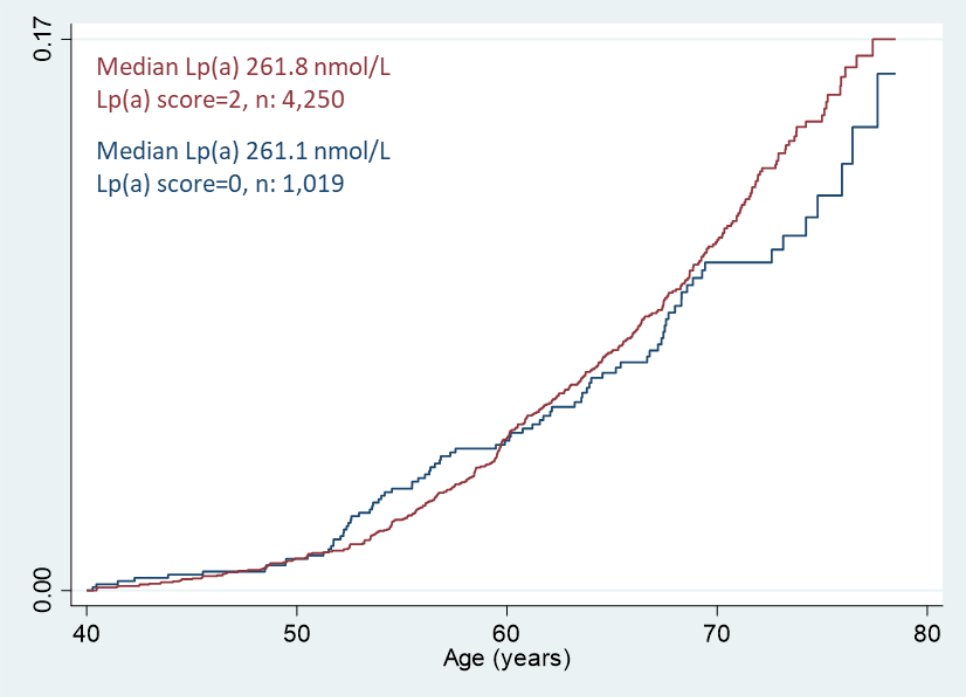



**Figure 3.** Event curves for lifetime risk of major coronary events by Lp(a) score among participants with comparable median Lp(a).

## Discussion

The results of this study highlighted that profiling the genetic determinants of plasma Lp(a) provides comparable value to measured levels of Lp(a) for ASCVD risk prediction. However, since even among individuals with the same genotype the risk changed accordingly with measured Lp(a) concentrations, our data emphasizes the importance of measuring Lp(a) level in clinical practice, even on top of the genetic background, to better identify patients at risk.

Our findings are in accordance with a very recent observational study, published in 2020 and using a cohort of 283,540 adults recruited by the UK Biobank, demonstrating that an LPA genetic risk score offered comparable ASCVD risk prediction to directly measured lipoprotein(a) (193).

An isolated Lp(a) measurement could be potentially an easier and cost saving approach to quantify the exposure to Lp(a) than the Lp(a) score. This can be explained by two key factors: (1) Lp(a) displays higher heritability than LDL cholesterol levels, and (2) Lp(a) levels are generally much more stable throughout life compared with other circulating lipoproteins (ie, minimal influence of age, sex, genetic factors outside the LPA gene, environmental factors, or currently available medicines) (194). Despite that, in some cases, the genetic determinants of elevated Lp(a) may help discern the familial risk of lipoprotein(a)-associated ASCVD that is not always conclusive from Lp(a) measurement alone. For instance, a single copy of the rs10455872-G allele, which is common in European populations, is known to associate with extremely elevated Lp(a) that can result in a phenotype similar to familial hypercholesteremia and thus may be amenable to cascade screening (195, 196).

While some guidelines support broad population-based screening

with Lp(a) (197), the small improvement in powerful clinical risk scores has led to think that measured Lp(a) or genetic score could be an inefficient approach for refinement of ASCVD risk among asymptomatic middle-aged adults broadly. However, because Lp(a) levels display an extremely right-skewed distribution in the general population (potentially varying more than 1000-fold between individuals, approximately 0.2 to ≥200 mg/dL), only individuals with extreme Lp(a) levels greater than 200 mg/dL could have a 3- to 4-fold increased lifetime risk of ASCVD (198, 199). In such cases, the modest improvement in ASCVD risk discrimination that was observed in this study, when continuous levels of measured Lp(a) or Lp(a) genetic score were added to clinical risk scores, may underestimate cardiovascular risk. This is probably the main reason for the limited predictive power of all the reclassification models that included the entire range of Lp(a) levels in ASCVD risk discrimination and failed to accurately quantify the cardiovascular risk associated with the extremes of elevated Lp(a).

Overall, evidence seems to suggest a role for Lp(a) genotype/measurement in refining cardiovascular risk prediction, although limitedly to specific circumstances.

Limitations

First, participants in the UK biobank are a self-selected group who tend to be at lower risk of cardiovascular events than members of other populations. As a result, all analyses should be repeated in populations at higher risk. In addition, our analysis was restricted to participants who self-identified as being of white European ancestry, and therefore results may not apply to persons of other ethnicities. As a result, this study should be repeated in other populations.

**Conclusion**

Lp(a) genetic score provides comparable risk prediction for incident ASCVD compared with measured Lp(a). Since the distribution of measured Lp(a) is quite wide even among people with the same genotype, it may be preferable to rely on measured values for the management of patients in clinical practice. Our evidence supports the role of Lp(a) as a risk-enhancing factor, but further researches are needed to improve guideline-supported risk scores.

## 5.4 Does Lipoprotein(a) have a prothrombotic effect?

### Background

Since its discovery, lipoprotein(a) [Lp(a)] has been the subject of controversy and debate about its physiological functions and roles in atherogenesis, thrombogenesis and development of cardiovascular diseases.

In recent years, a number of prospective epidemiological and clinical studies have shown that elevated Lp(a) is an independent risk factor for development of coronary and cerebral atherosclerosis (200). This has been also confirmed by Mendelian randomization studies, showing a causal dose-response effect of exposure to alleles associated with higher Lp(a) on the risk of coronary disease (188).

The exact physiological role of Lp(a) has not been fully elucidated, however it has been speculated a possible implication of Lp(a) in inhibiting the activation of transforming growth factor and contributing to the progress of arterial atherosclerotic lesions by promoting proliferation of vascular smooth muscle cells and migration of smooth muscle cells to endothelial cells, or by acting as a proinflammatory mediator, increasing the lesion formation in atherosclerotic plaques (201).

Due to structural homology with plasminogen, Lp(a) may also compete with plasminogen for its receptors on endothelial cells, thus leading to diminished plasmin formation, delaying clot lysis, and favouring venous thrombosis (202). However, evidence on its role as a risk factor for venous thromboembolic events (VTE) remains controversial. Indeed, although convincing epidemiological evidences were brought forward to propose a causal role of Lp(a) in the development, progression, and complication of occlusive arterial disease, data supporting a positive association between elevated

Lp(a) values and VTE are less consistent. A previous systematic review and meta-analysis of the literature, conducted a decade ago and including a limited number of studies, found a statistically significant, albeit modest, association between high Lp(a) (>300 mg/L) and VTE (odds ratio [OR]: 1.87, 95%CI: 1.51–2.30) (203). On the other hand, a more recent study reported an association between the two variants of the LPA gene (rs10455872 and rs3798220 polymorphisms) and systemic and coronary atherosclerosis, but not with VTE (204).

Therefore, to better clarify the relation between Lp(a) levels and VTE, we conducted a study to address whether Lp(a) has a genetically and clinically meaningful venous or arterial prothrombotic effect.

## Methods

Study population

A total of 445,744 participants enrolled in the UK Biobank with complete genetic and principal component data who self-identified as being of white ancestry were included in the study. Participants underwent genotyping with one of two closely related custom arrays (UK BiLEVE Axiom Array or the UK Biobank Axiom Array) consisting of over 800,000 genetic markers, with additional genotypes imputed using the Haplotype Reference Consortium resource, the UK10K panel, and the 1000 Genomes panel. The KING toolset was used to identify up to third-degree relatedness based on kinship coefficients >0.044 (more details in **Appendix I**).

Construction of Lp(a) polygenic score

It has been recognized that more than 90% of variation of plasma Lp(a) concentration is genetically regulated, with Lp(a) gene (LPA) being a major determinant. To date, several genetic variants in the

LPA gene have been shown to influence Lp(a) plasma values, with rs3798220 (Ile4399→Met) and rs10455872 (intronic A/G polymorphism) polymorphisms accounting in particular for at least 40% of such variation (136). Because these two variants are strongly associated with plasma Lp(a) concentrations, they can serve as surrogate markers for Lp(a) levels, reflecting also lifelong elevation in plasma Lp(a) and eliminating the artefact of "reverse causality". In this study, an Lp(a) polygenic score was calculated for each UK Biobank participant by summing the number of risk-increasing alleles inherited at rs3798220 and rs10455872 variants weighted by the effect size for each allele (**Appendix IV**).

Lipoprotein(a) measurement

Lipoprotein(a) was measured in nanomoles per litre at study enrolment using an immunoturbidimetric method on the Beckman Coulter AU5800 platform (Randox Bioscience, UK).

Study outcomes and statistical analysis

The primary outcome was VTE (data Fields: #20002, #4012, #4022, #131308, #131308), a composite of deep vein thrombosis (DVT) and pulmonary embolism (PE).

First, we evaluated the effect of increase Lp(a) on major coronary events (MCE), defined as the first occurrence of either a fatal or non-fatal myocardial infarction, or coronary revascularization, for a positive control. The analysis used Cox proportional hazards models adjusted for age, sex, and the first 10 principal components of ancestry, with age as the time scale. The risk was estimated using the Lp(a) genetic score and concentrations of measured Lp(a) (both in continuous [each 100 nmol/L increase in measured level] and dividing the distribution into deciles [where the first six deciles were

grouped together and used as reference]) in sensitivity analyses. Then, adjusted logistic regressions were used to estimate the effect of measured and genetically determined Lp(a) (as defined above) on VTE, DVT, and PE. For further validation, the effect of 100 nmol/L increase in genetically determined Lp(a) level on MCE were also assessed stratifying by GUCY1A3 score and by Factor II and V score (**Appendix VI**), that mimic the effect of antiplatelet and antithrombin (anticoagulant) therapies respectively, to assess whether these treatments are likely to influence cardiovascular risk associated with high Lp(a) levels.

All analyses were performed using Stata (version 16; StataCorp). A 2-tailed p-value less than 0.05 was considered statistically significant.

## Results

A total of 445,744 participants were included in the study. The median (SD) age at study enrolment was about 57 (8) years, with no difference observed among Lp(a) genetic score genotype (**Table 1**). No differences were also observed in the distribution of all the other covariates evaluated, across the different classes, confirming a random allocation of subjects. The median [IQR] level of Lp(a) [nmol/L] increased with increasing number of the genetic score copies (13.6 [6.2-35.0], 146.3 [104.8-200.2], 261.8 [190.2-336.0]).

**Table 1.** Baseline characteristics, measured at the time of enrolment in the UK Biobank, are presented for the entire cohort and by Lp(a) instrument variable (IV) genetic score copies.

| **Characteristics** | **Overall** | **Lp(a) IV genetic score copies** | | |
|---|---|---|---|---|
| | | **0** | **1** | **2** |
| No. participants | 445,774 | 358,469 | 77,658 | 4,250 |

| | | | | |
|---|---|---|---|---|
| Age, y (SD) | 57.3 (8.0) | 57.2 (8.0) | 57.3 (8.0) | 57.2 (8.0) |
| Female Sex (%) | 54.3 | 54.2 | 54.3 | 54.3 |
| LDL-C, mg/dL (SD) | 138.0 (33.6) | 137.4 (33.5) | 140.4 (34.1) | 142.7 (34.7) |
| ApoB, mg/L (SD) | 103.4 (23.8) | 103.0 (23.8) | 105.2 (23.9) | 106.8 (24.2) |
| TG, mg/dL (IQR) | 131.9 [93.1-190.7] | 132.7 [93.7-191.8] | 128.5 [91-186.9] | 124 [87.7-180.7] |
| HDL-C, mg/dL (SD) | 56.2 (14.8) | 56.1 (14.8) | 56.4 (14.9) | 56.5 (15.3) |
| Lp(a), nmol/L (IQR) | 18.7 [7.4-72.9] | 13.6 [6.2-35.0] | 146.3 [104.8-200.2] | 261.8 [190.2-336] |
| CRP, mg/L (IQR) | 1.3 [0.7-2.8] | 1.3 [0.7-2.8] | 1.3 [0.7-2.8] | 1.3 [0.7-2.7] |
| SBP, mmHg (SD) | 137.8 (18.6) | 137.8 (18.6) | 137.9 (18.6) | 137.8 (18.6) |
| BMI, kg/m$^2$ (SD) | 27.4 (4.8) | 27.4 (4.8) | 27.4 (4.8) | 27.4 (4.9) |
| Current smoker (%) | 7.2 | 7.2 | 7.2 | 6.8 |
| No. MCE | 23,032 | 17,110 | 5,313 | 365 |
| No. VTE events | 15,974 | 12,866 | 2,762 | 153 |
| No. DVT events | 11,079 | 8,934 | 1,906 | 106 |
| No. PE events | 6,602 | 5,320 | 1,144 | 58 |

Characteristics are presented as means and standard deviations (SD) for normally distributed variables, median and interquartile ranges (IQR) for non-normally distributed variables (triglycerides, Lp(a), and CRP), or percentages for dichotomous variables. LDL-C: low-density lipoprotein cholesterol; ApoB: apolipoprotein B; TG: triglycerides; HDL-C: high-density lipoprotein cholesterol; Lp(a): lipoprotein(a); SBP: systolic blood pressure; CRP: c-reactive protein; BMI: body mass index; WHR: waist to hip ratio. MCE: major coronary events; VTE: venous thromboembolism; DVT: deep vein thrombosis, PE: pulmonary embolism.

Associations of lipoprotein(a) levels with major coronary events

Each 100 nmol/L increase in genetically predicted Lp(a) levels was associated with a 35% higher risk of MCE (hazard ratio [HR]: 1.35, 95%CI: 1.32-1.38). There was also a step-wise increase in the risk of MCE with increasing number of copies of the genetic score (**Figure 1**): participants with both copies of rs10455872 and rs3798220

variants had a HR for MCE of 1.89 (95%CI: 1.70-2.10), while the risk of MCE was 47% higher (HR: 1.47, 95%CI: 1.43-1.52) for subjects with one copies, compared with the reference group (delta in median Lp(a) equal to 262 nmol/L and about 146 nmol/L, respectively).
In the sensitivity analysis, using measured Lp(a), increasing deciles of Lp(a) concentration were still associated with a step-wise increase in the risk of MCE (**Appendix VII**): participants in the highest decile had a HR for MCE of 2.14 (95%CI: 2.06-2.22) as compared to participants in the reference group (delta in median Lp(a) level equal to 212 nmol/L).

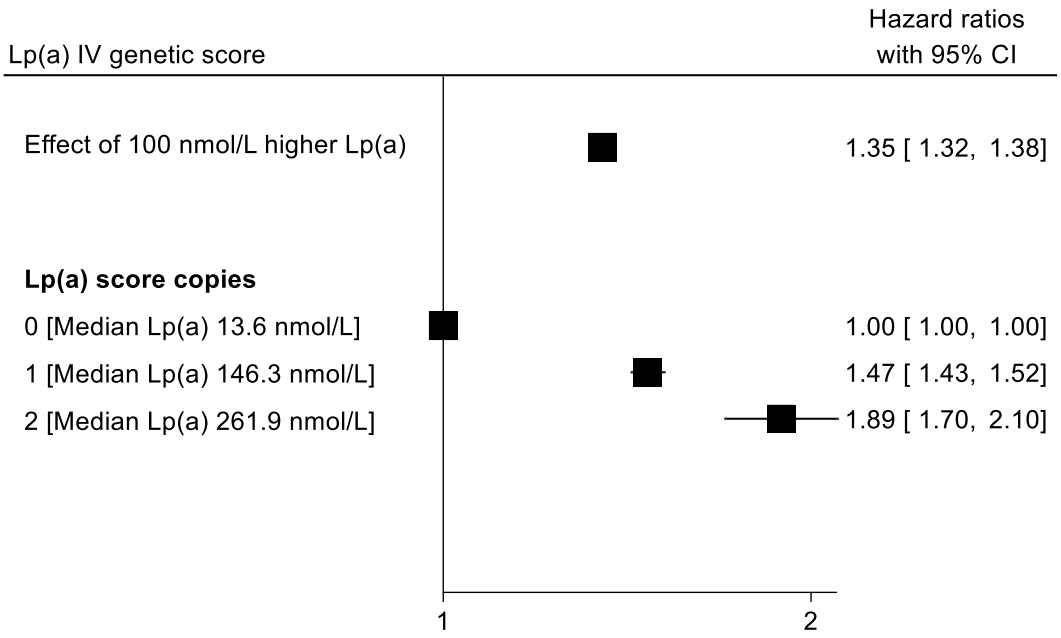


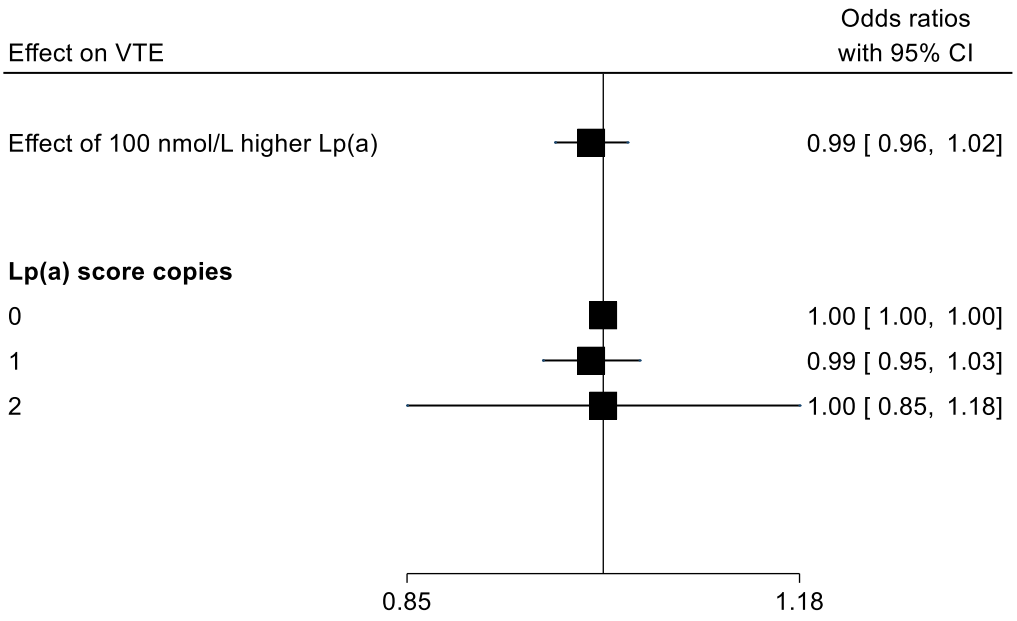

**Figure 1.** Effect of Lp(a) on major coronary events.
*IV: instrumental variable; CI: confidence interval.

<u>Associations of Lp(a) with venous thromboembolism, deep vein thrombosis, and pulmonary embolism</u>

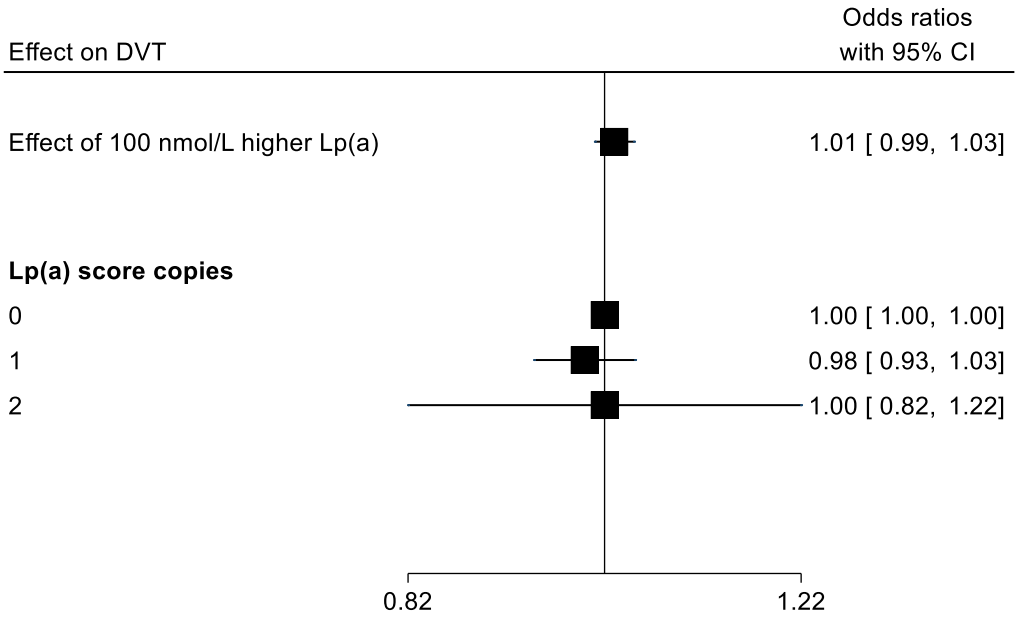
There was no statistically significant evidence of an association of genetically predicted Lp(a) concentrations (both for 100 nmol/L

higher level and increasing numbers of genetic score copies) with VTE (**Figure 2,** panel A), DVT (**Figure 2,** panel B), and PE (**Figure 2,** panel C). Same evidence was observed using measured Lp(a) as exposure (**Appendix VIII**).

(A)




(B)



(C)

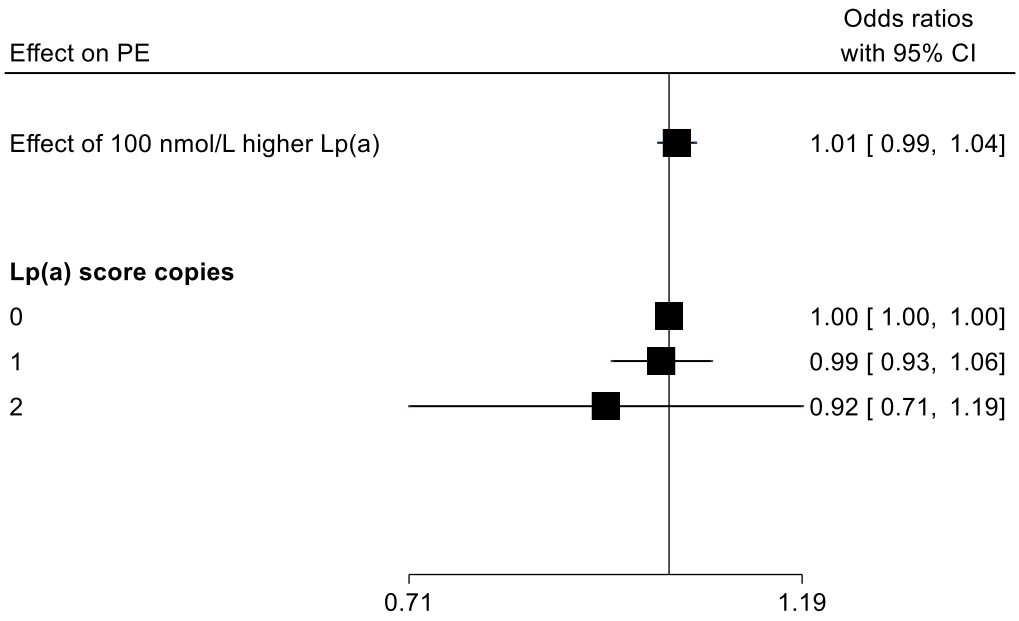


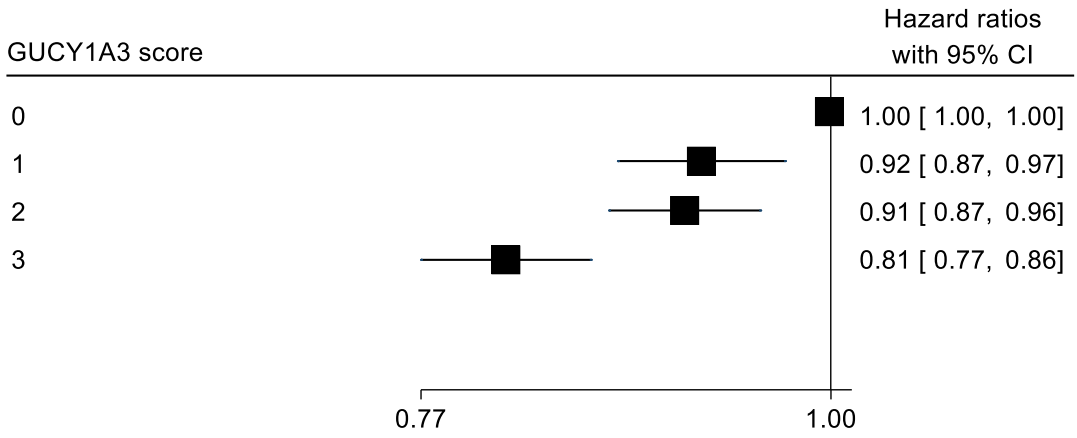

**Figure 2.** Effect of Lp(a) on venous thromboembolism (VTE, panel A), deep vein thrombosis (DVT, panel B), and pulmonary embolism (PE, panel C). * CI: confidence interval.

Stratified analyses by GUCY1A3 and Factor II and V scores

Despite the increase of the number of copies of GUCY1A3 score produced an increase in antiplatelet inhibition that corresponds to a step-wise clinically significant reduction of cardiovascular events (**Figure 3**, panel A), no effect was observed for increasing Lp(a) levels. In fact, no differences were observed in the association between genetically predicted Lp(a) concentrations (each 100 nmol/L increase concentrations) and the risk of MCE stratifying by GUCY1A3 score copies (**Figure 3**, panel B). Similarly, in **Figure 4** (panel A) there is a clear step-wise increase in greater antithrombotic (anticoagulant) effect that produced a significant step-wise decrease in the risk of VTE, which in turn does not impact the effect of Lp(a) on MCE, as can be observed in **Figure 4** panel B. This evidence suggests that platelet inhibition and anticoagulant effect are not

mediating the association between Lp(a) levels and MCE risk, since the trend persisted in the cohort regardless of whether they carried gene variants for platelet activation, or for Factor II and Factor V genes for prothrombin disorders.

(A) Effect of GUCY1A3 instrumental variable score on MCE




(B) Effect of Lp(a) on MCE stratified by GUCY1A3 score

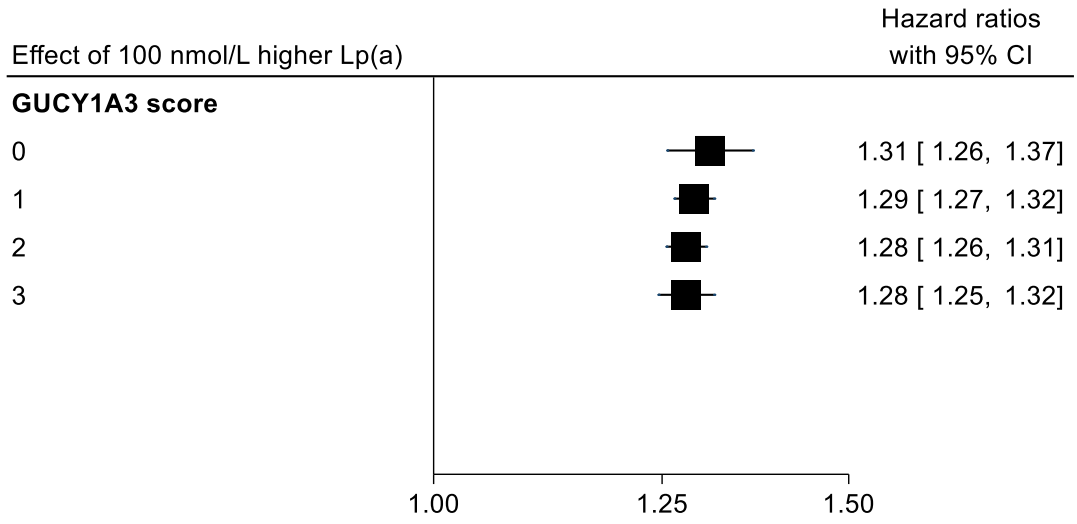
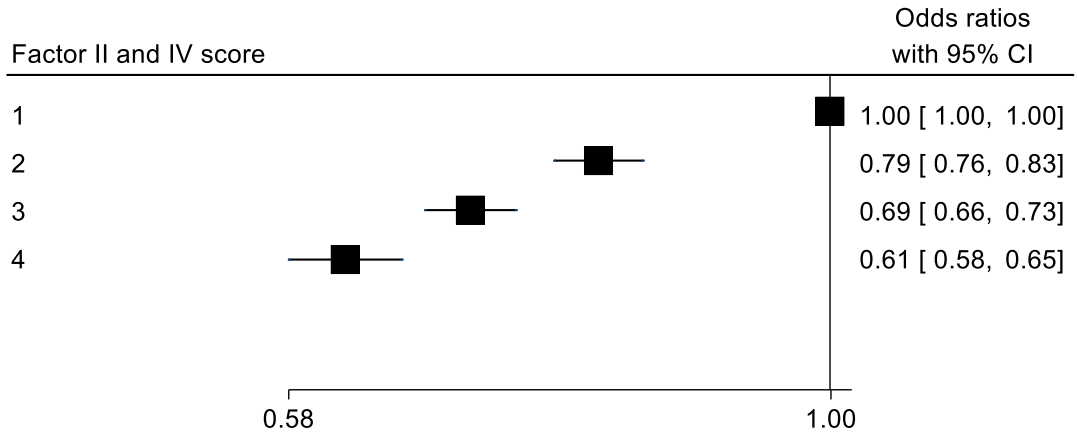



**Figure 3.** Effect of Lp(a) and GUCY1A3 score on major coronary events (MCE). * CI: confidence interval.

(A) Effect of Factor II and V instrumental variable score on VTE




(B) Effect of Lp(a) on MCE stratified by Factor II and V score

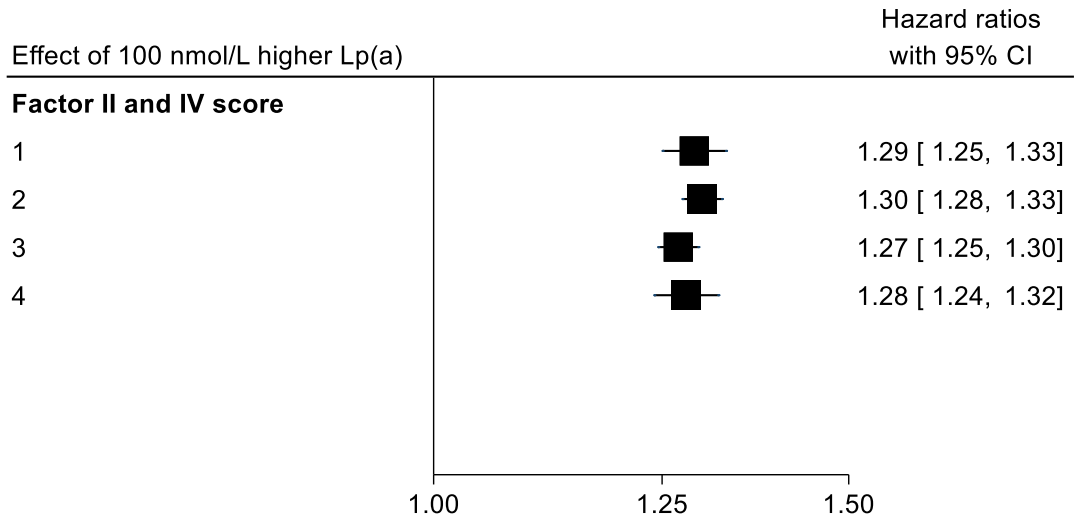



**Figure 4.** Effect of Lp(a) and Factor II and V score on major coronary events (MCE) and venous thromboembolism (VTE), respectively.

* CI: confidence interval.

## Discussion

We found that Lp(a) does not appear to have a clinically significant venous or arterial prothrombotic effect. Indeed, genetically predicted and measured Lp(a) concentrations were not associated with clinically significant thrombotic events, neither the effect of increased

Lp(a) levels on MCE was attenuated by either genetically determined lifelong platelet or thrombin inhibition. This is extremely important for designing trials enrolling people with a previous cardiovascular event, and accordingly are on antiplatelet or anticoagulant therapy, because this means that the benefit obtainable from the reduction of high Lp(a) levels is not attenuate by these therapies (an elevated Lp(a) value is still leading to high cardiovascular risk, despite patients are taking these medications). Likewise, high Lp(a) concentrations in individuals in primary prevention cannot probably be significantly reduced with these therapies.

Apolipoprotein (a) moiety has sequence homology to plasminogen by the presence of kringle IV and kringle V domains as well as a protease domain, which however is catalytically inactive. The similarity to plasminogen as well as the presence of oxidized phospholipids have been made responsible for the thrombogenic properties of Lp(a) which include the inhibition of fibrinolysis, the induction of plasminogen activator inhibitor type 1 (PAI-1) expression in endothelial cells, as well as the increasing of activity of tissue factor pathway inhibitor and platelet responsiveness (205). These properties have led to the hypothesis that elevated plasma concentrations of Lp(a) are a risk factor for purely thrombotic disorders, such as venous thromboembolism.

In the past literature, the majority of studies which investigated the association of Lp(a) with VTE were cross-sectional rather than prospective, resulting in the inability to make causal inference. The most recent meta-analysis analysed the data of ten case-control and cohort studies encompassing 13,541 subjects of whom 5660 had a history of deep vein thrombosis and/or pulmonary embolism (206). As the cut-off defining elevated Lp(a), the authors used the upper limit of the manufacturer’s product reference range (usually

30 mg/dL) in nine studies and the 75th percentile of Lp(a) value in the control group in one study. With these definitions, elevated Lp(a) was associated with the presence of VTE at a significant odds ratio of 1.56 (95%CI: 1.36-1.79). Much stronger risk associations were found in patients who have a very high risk of VTE, for example patients immobilized due to paraplegia.

In contrast to case–control studies, the data of prospective studies are more controversial. Among 8960 participants of the Copenhagen City Heart Study, of whom 735 experienced a VTE event during 15 to 18 years of follow-up, Kamstrup et al. did not find any significant association of Lp(a) levels with the incidence of VTE. Adjusted hazard ratios for second (median, IQR: 17, 12–27 mg/dL) and third tertiles (median, IQR: 59, 40–94 mg/dL) vs. first tertile of Lp(a) (median, IQR: 3,1–5 mg/dL) were 1.1 (95%CI 0.8–1.4) and 0.8 (95%CI 0.6–1.1), respectively (207). Likewise, in a 12-center study of 510 patients with first unprovoked VTE treated for 5–7 months with anticoagulants and followed up for 16.9 ± 11.2 months, Rodgers et al. did not find any significant association of Lp(a) levels >300 mg/L with risk of recurrent VTE events (relative risk 1.4; 95%CI 0.7–2.6) (208). Conversely, in a study of 467 patients with first VTE followed up for one year, Marcucci et al. found a 5-fold increased risk of recurrent VTE for Lp(a) >30 mg/dL (OR 5.1; 95%CI 3.1–8.4) which was similar to that for hyperhomocysteinemia and even higher than that for factor V Leiden or the factor II 20210GA polymorphism (209).

Mendelian randomization studies also made controversial findings on the association of LPA polymorphisms with risk of VTE. Kamstrup et al. (207) excluded any contribution of the kringle IV repeat polymorphism to VTE in the Copenhagen City Heart Study (N = 9190, 443 events) and the Copenhagen General Population study

(N = 28,538; 926 with history of VTE). Of note, in the same study, the authors found genetically causal associations of Lp(a) levels and kringle IV repeats with coronary, carotid and femoral atherosclerosis as well as of factor V Leiden with VTE. A more recent but smaller study of 516 patients with a history of VTE and 1117 controls found significant inverse and dose-dependent associations of kringle IV repeat numbers with venous thrombosis (210). More recently, however, genetic approaches have provided strong evidence against a role for Lp(a) in venous thrombosis. When rs3798220 and rs10455872 variants have been used as proxy to assess associations between genetically elevated Lp(a) levels and different forms of venous thrombosis, the results have been consistently negative (207, 211). These findings, together with our analysis, quite definitely rule out a role for elevated Lp(a) in the aetiology of venous thrombosis and a pro-coagulant effect of the latter.

Limitations

Participants in the UK biobank are a self-selected group who tend to be at lower risk of cardiovascular events than members of other populations. Thus, all analyses should be repeated in populations at higher risk. In addition, the analysis was restricted to participants who self-identified as being of white European ancestry, and therefore results may not apply to persons of other ethnicities.

## Conclusion

Taken together the results of this study seem to argue against a clinically significant venous or arterial prothrombotic effect for Lp(a). Moreover, the increased risk of MCE caused by elevated Lp(a) is unlikely to be reduced by either an antiplatelet or antithrombin therapy.

## 5.5 A practical strategy to use measured lipoprotein(a) levels to guide clinical management

### Background

Despite significant advances in the diagnosis and therapy of cardiovascular disease (CVD), patients continue to experience myocardial infarction, stroke, peripheral arterial disease, and need for revascularization. The advances in identifying modifiable risk factors for CVD, including smoking, hypertension, dyslipidemias, diabetes mellitus, and obesity, have allowed the development of practice patterns and evidence-based guidelines in medical therapy and revascularization that have contributed to the reduction of CVD mortality. However, ~40% of all deaths can be still attributed to CVD (212). These observations suggest that probably the presence of additional modifiable risk factors contributes to CVD risk.

Lipid disorders can be broadly divided into 4 "clinical" categories: elevated low-density lipoprotein cholesterol (LDL-C), low high-density lipoprotein cholesterol (HDL-C), elevated triglycerides, and elevated lipoprotein(a) [Lp(a)]. In the current genetic era, it has become apparent that only elevated levels of apolipoprotein B-100 (apoB)−containing lipoproteins (intermediate-density lipoprotein, LDL, Lp(a), very low-density lipoprotein) are causally associated with increased cardiovascular risk.

At the clinical level, elevated Lp(a) has been the least studied of these disorders. However, this is rapidly changing with the growing awareness of its role in CVD and calcific aortic valve stenosis, and the potential of novel therapies to substantially lower Lp(a) levels.

Mendelian randomization studies indeed showed a dose-response effect of exposure to alleles associated with higher Lp(a) on the risk of coronary disease, suggesting a causal relation. The main issue

remains to understand and estimate how much Lp(a) should be decreased to achieve a clinically meaningful effect.

Lp(a) levels show a highly skewed distribution in the population, such that the vast majority of people (about 80%) have levels below 50 mg/dL. If it is true that only very large reductions of Lp(a) are likely to yield clinically meaningful risk reductions, this means that when looking at the distribution of Lp(a) levels in the population, only a small minority of patients have the potential to benefit from lowering Lp(a) concentration: specifically, only those with very high levels (191).

This finding likely explains why therapies that reduce Lp(a) concentration by 20% to 35% have failed to provide clear evidence that decreasing Lp(a) concentration reduces the risk of cardiovascular events in previous randomized trials. The median Lp(a) concentration among participants enrolled in these trials was approximately 12 to 20 mg/dL (213, 214). Therefore, a 30% reduction in Lp(a) concentrations would translate into only a 3- to 6-mg/dL absolute reduction in circulating plasma Lp(a) concentrations, a small absolute reduction that was likely far too modest to appreciably reduce the risk of cardiovascular events in a short-term randomized trial. To date, since evidence from randomized trials are inconsistent, guideline suggests to focus on reducing LDL-C and other apoB-containing lipoproteins as first line to most effectively reduce cardiovascular risk, even in case of elevated Lp(a). However, if it has been estimated how much Lp(a) concentration must be lowered pharmacologically to produce the same change as lowering LDL-C level by 38.67 mg/dL (i.e., 1 mmol/L) (191), it is unknown how much LDL-C must be further lowered in order to overcome the extra cardiovascular risk due to increased levels of Lp(a).

Thus, we sought to evaluate the amount of further LDL-C reduction

needed to abolish the extra risk due to increased levels of Lp(a), in order to provide specific clinical guideline that may help the therapy management in clinical practice.

## Methods

Study population

A total of 445,744 participants enrolled in the UK Biobank with complete genetic and principal component data who self-identified as being of white ancestry were included in the study. Participants underwent genotyping with one of two closely related custom arrays (UK BiLEVE Axiom Array or the UK Biobank Axiom Array) consisting of over 800,000 genetic markers, with additional genotypes imputed using the Haplotype Reference Consortium resource, the UK10K panel, and the 1000 Genomes panel. The KING toolset was used to identify up to third-degree relatedness based on kinship coefficients >0.044 (more details in **Appendix I**).

Lipoprotein(a) measurement

Lipoprotein(a) was measured in nanomoles per litre at study enrolment using an immunoturbidimetric method on the Beckman Coulter AU5800 platform (Randox Bioscience, UK).

Construction of Lp(a) polygenic score

To date, two specific genetic variants in the LPA gene have been shown to influence Lp(a) plasma values: rs3798220 (Ile4399→Met) and rs10455872 (intronic A/G polymorphism) (136). An Lp(a) polygenic score was calculated for each UK Biobank participants by summing the number risk-increasing alleles inherited at rs3798220 and rs10455872 variants weighted by the effect size for each allele (**Appendix IV**).

Construction of the LDL instrumental variable

To construct the LDL instrumental variable, all variants associated with LDL at genome-wide level of significance ($p<5x10^{-8}$) as reported in external consortia were included in the polygenic score (181). The LDL variants were then pruned by excluding all variants with a linkage disequilibrium (LD) $r^2>0.1$ to select independently inherited variants for inclusion in the instrumental variable genetic score. An LDL score was calculated for each participant by summing the number of LDL-increasing alleles inherited at each variant included in the LDL score weighted by the LDL effect size of each allele (**Appendix III**).

Study outcomes and statistical analysis

The primary outcome was major coronary events (MCE), defined as the first occurrence of either a fatal or non-fatal myocardial infarction (MI), or coronary revascularization. The analysis used Cox proportional hazards models adjusted for age, sex, and the first 10 principal components of ancestry, with age as the time scale. Each participant was censored at the age they experienced either a primary outcome event, death due to a cause other than MI (treated as a competing risk), or at the age of last reported follow-up. The dates of all incident events were recorded from hospital episode statistics, while the dates of events that were prevalent at the time of enrolment into UK Biobank were recorded either from hospital episode statistics or self-reported. Cumulative lifetime risk of MCE was plotted using Kaplan-Meier curves.

All analyses were performed using Stata (version 16; StataCorp). A 2-tailed p-value less than 0.05 was considered statistically significant.

## Results

Participants characteristics

A total of 445,744 participants were included in the study. The median (SD) age at study enrolment was about 57 (8) years, with no difference observed among Lp(a) genetic score genotype (**Table 1**). No differences were also observed in the distribution of all the other covariates, across different classes, confirming a random allocation of subjects. The median [IQR] level of Lp(a) [nmol/L] increased with increasing number of the genetic score copies (13.6 [6.2-35.0], 146.3 [104.8-200.2], 261.8 [190.2-336.0]). Baseline characteristics stratified by sex are presented in **Appendix IX**.

**Table 1.** Baseline characteristics, measured at the time of enrolment in the UK Biobank, are presented for the entire cohort and by Lp(a) instrument variable (IV) genetic score copies.

| **Characteristics** | **Overall** | **Lp(a) IV genetic score copies** | | |
|---|---|---|---|---|
| | | **0** | **1** | **2** |
| No. participants | 445,774 | 358,469 | 77,658 | 4,250 |
| Age, y (SD) | 57.3 (8.0) | 57.2 (8.0) | 57.3 (8.0) | 57.2 (8.0) |
| Female Sex (%) | 54.3 | 54.2 | 54.3 | 54.3 |
| LDL-C, mg/dL (SD) | 138.0 (33.6) | 137.4 (33.5) | 140.4 (34.1) | 142.7 (34.7) |
| ApoB, mg/L (SD) | 103.4 (23.8) | 103.0 (23.8) | 105.2 (23.9) | 106.8 (24.2) |
| TG, mg/dL (IQR) | 131.9 [93.1-190.7] | 132.7 [93.7-191.8] | 128.5 [91-186.9] | 124 [87.7-180.7] |
| HDL-C, mg/dL (SD) | 56.2 (14.8) | 56.1 (14.8) | 56.4 (14.9) | 56.5 (15.3) |
| Lp(a), nmol/L (IQR) | 18.7 [7.4-72.9] | 13.6 [6.2-35.0] | 146.3 [104.8-200.2] | 261.8 [190.2-336] |
| CRP, mg/L (IQR) | 1.33 [0.66-2.75] | 1.3 [0.7-2.8] | 1.3 [0.7-2.8] | 1.3 [0.7-2.7] |

| | | | | |
|---|---|---|---|---|
| SBP, mmHg (SD) | 137.8 (18.6) | 137.8 (18.6) | 137.9 (18.6) | 137.8 (18.6) |
| BMI, kg/m$^2$ (SD) | 27.4 (4.8) | 27.4 (4.8) | 27.4 (4.8) | 27.4 (4.9) |
| Current smoker (%) | 7.2 | 7.2 | 7.2 | 6.8 |
| No. MCE | 23,032 | 17,110 | 5,313 | 365 |

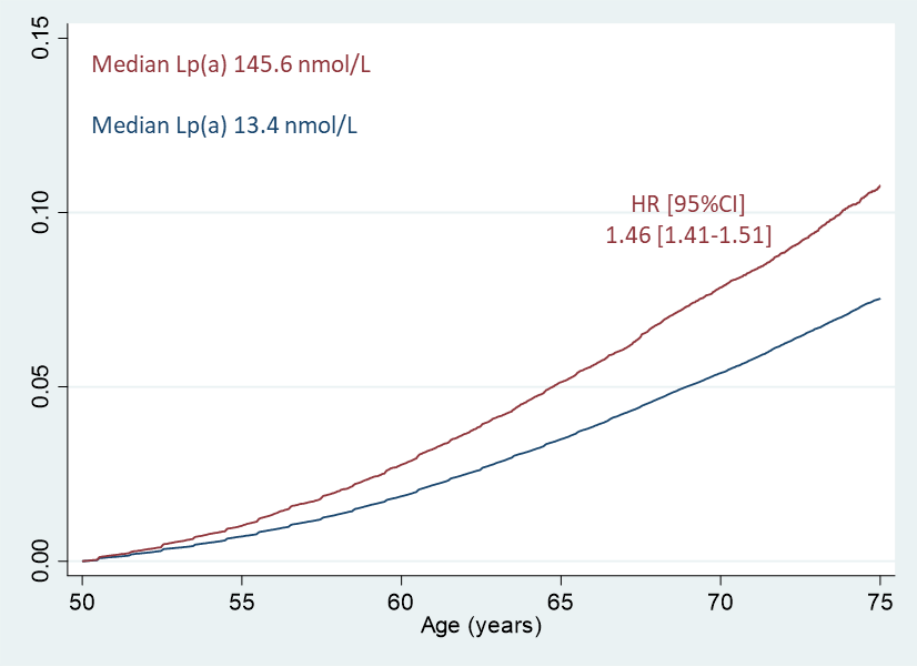
Characteristics are presented as means and standard deviations (SD) for normally distributed variables, median and interquartile ranges (IQR) for non-normally distributed variables (triglycerides, Lp(a), and CRP), or percentages for dichotomous variables. LDL-C: low-density lipoprotein cholesterol; ApoB: apolipoprotein B; TG: triglycerides; HDL-C: high-density lipoprotein cholesterol; Lp(a): lipoprotein(a); SBP: systolic blood pressure; CRP: c-reactive protein; BMI: body mass index; WHR: waist to hip ratio. MCE: major coronary events.

## Association of Lp(a) with major coronary events

Participants with one copy of either rs10455872 or rs3798220 had a hazard ratio (HR) for MCE of 1.47 (95%CI: 1.42-1.51), compared with the reference group (no copies), respectively, attributable to a difference in Lp(a) levels of 132 nmol/L (**Figure 1** panel A). The lifetime risk of MCE was constant over time, considering different enter time in the survival analysis (**Figure 1** panel B and C).

(A)

(B)

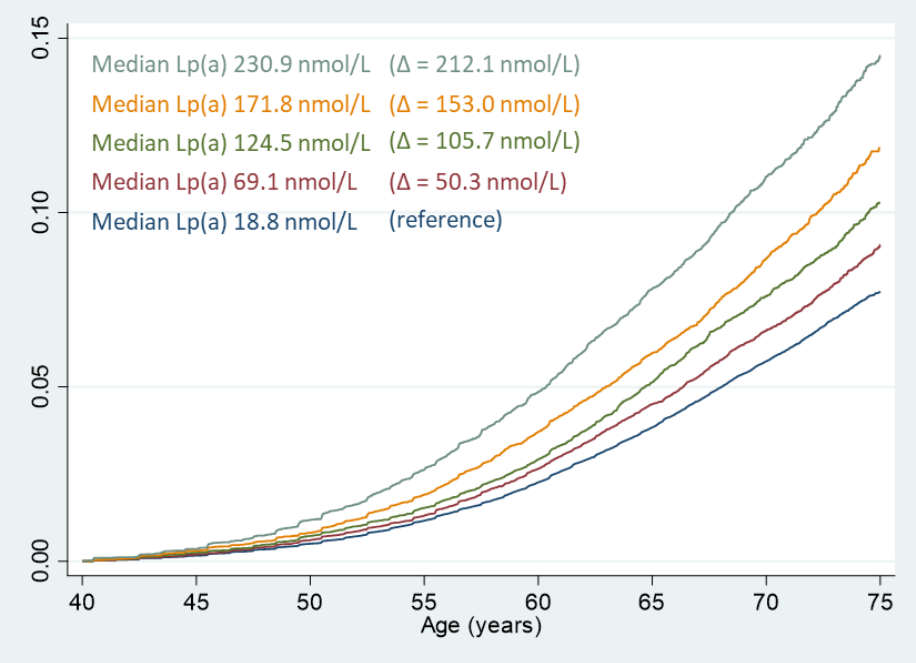



(C)

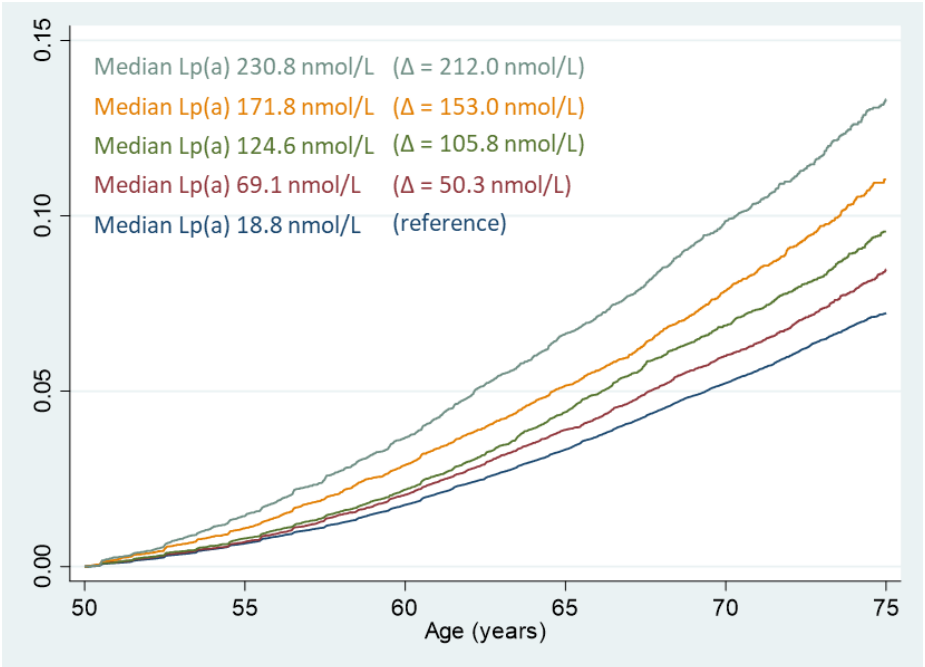

**Figure 1.** Event curves for lifetime risk of major coronary events (MCE) by Lp(a) genetic score copies (one or zero) at different enter time (40 years panel A, 50 years panel B, 60 years panel C). Effects of genetically predicted Lp(a) on MCE are reported as Hazard ratios (HR) and 95% confidence intervals (95%CI).

In the same way, increasing median concentrations of measured Lp(a) levels, were associated with increasingly steeper trajectories of lifetime risk for MCE, with the risk remaining stable even considering different enter time (**Figure 2**).

(A)




(B)



(C)

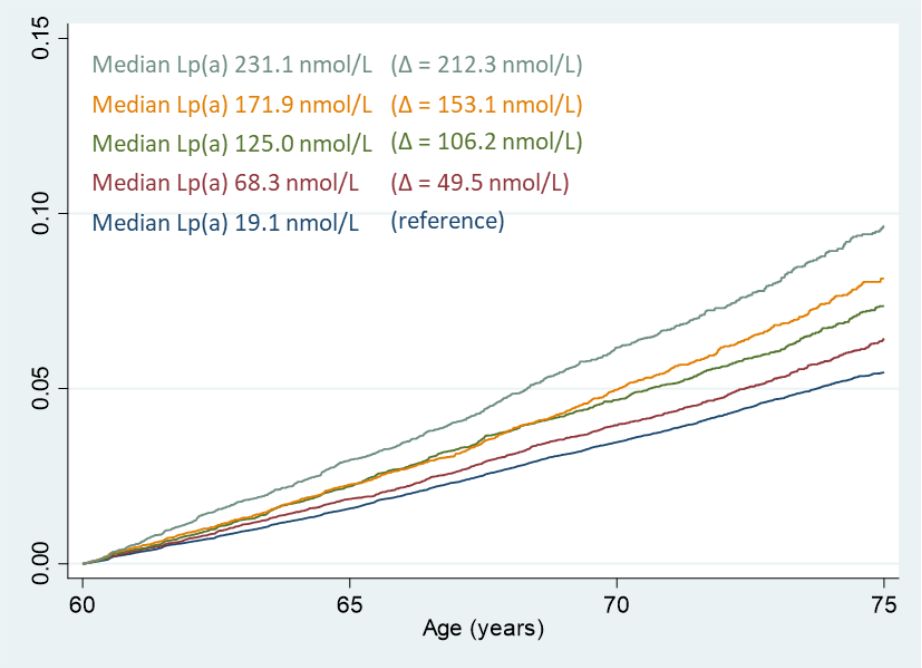


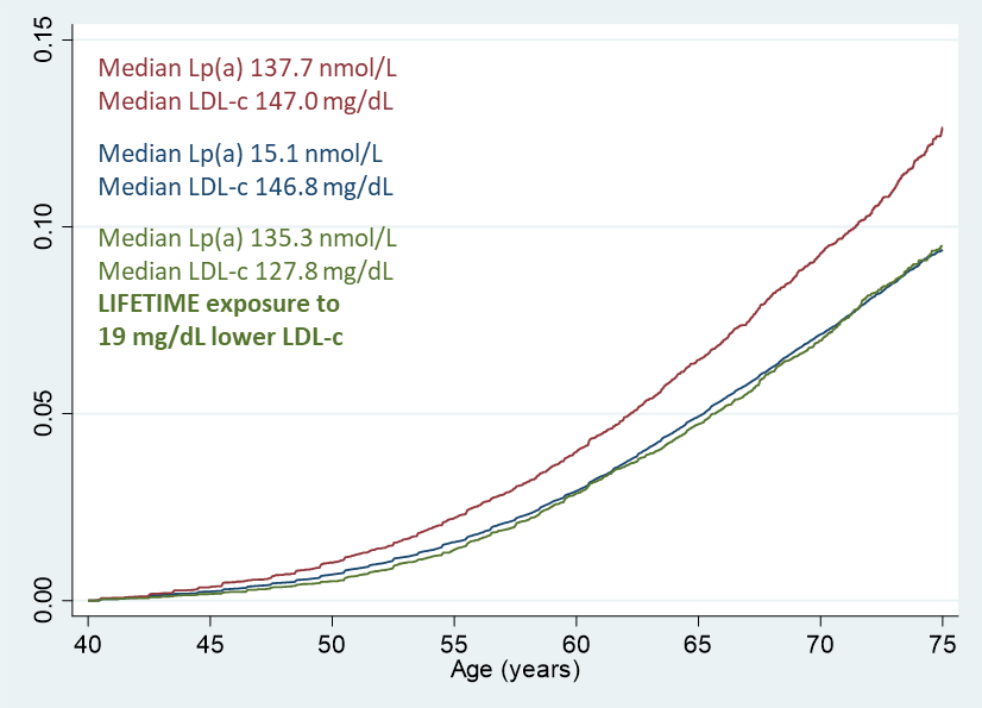

**Figure 2.** Event curves for lifetime risk of major coronary events by measured Lp(a) groups at different enter time (40 years panel A, 50 years panel B, 60 years panel C).

Overcoming the risk caused by high Lp(a) by reducing LDL cholesterol levels

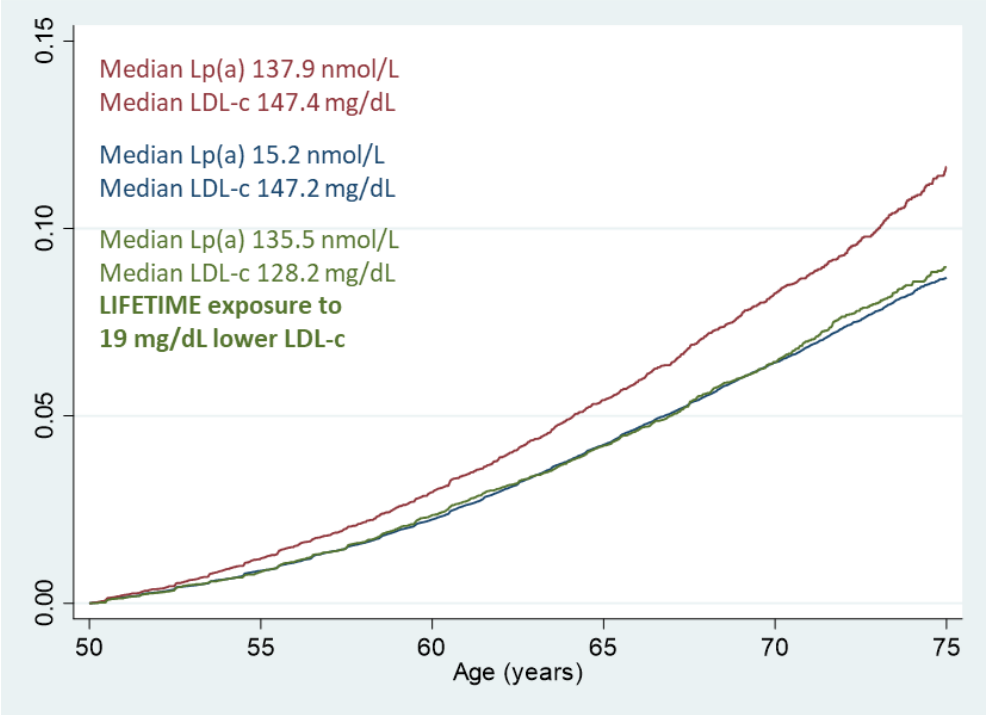
In **Figure 3** (panel A) the trajectories of the lifetime risk of major coronary events are plotted for subjects with one copy of either rs10455872 or rs3798220 variant (red line) and for individuals belonging to the reference group (no copies, blue line). In both groups subjects are characterized by relatively high and comparable LDL-C values (selected from the third tertile of the LDL instrumental variable genetic score; median values around 147 mg/dL), while they mainly differ for median Lp(a) concentrations (about 148 nmol/L vs 14 nmol/L, respectively), which are responsible for the observed increase MCE risk in the first group. The green curve, added to the graph, represents subjects having similar median Lp(a) value

compared with those in the red curve (not statistically significant difference) but exposed to 19 mg/dL lower LDL-C lifetime. In other words, the lifetime exposure to about half mmol/L lower LDL-C, without a clinically significant change in Lp(a) concentration, is able to abolish the extra risk caused by high Lp(a) levels, reducing the risk to exactly the same level observed in subjects belonging to the reference group and characterized by low Lp(a) concentration (the green curve is superimposable at every single age with the blue one). This extra reduction of LDL-C, is able to overcome the risk caused by elevated Lp(a) level, also considering different enter time (panel B and C). In sensitivity analyses stratified by sex, we confirmed all these results both in male and female sex cohorts (**Appendix X** and **Appendix XI**), and also using measured Lp(a) concentrations as exposure (**Appendix XII**).

(A)



(B)




(C)

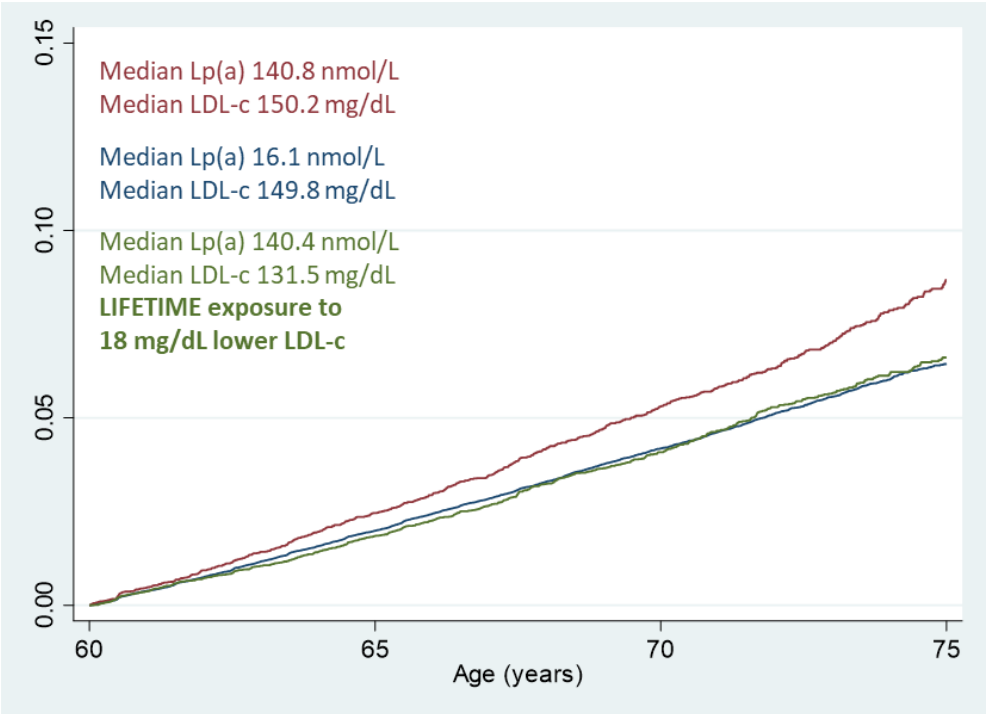



**Figure 3.** Event curves for lifetime risk of major coronary events by Lp(a) genetic score, LDL instrumental variable, at different enter time (40 years panel A, 50 years panel B, 60 years panel C).

Extra LDL-C reduction needed to abolish the risk due to high Lp(a)

**Table 2** reports the amount of extra LDL-C reduction (in mmol/L) needed to abolish the risk due to increased levels of Lp(a), specifically for different Lp(a) values at baseline (expressed in nmol/L) and depending on what age the treatment is started.

To obtain the amount of extra LDL-C reduction needed to overcome the risk (Y), the following formula was applied:

$$Y = \ln(RR) \,/\, [X * (-0.0126) - 0.1601\,]$$

For example, if a subject is in the 90$^{th}$ percentile and has an Lp(a) level of 170 nmol/L (150 nmol higher than the median level in the population, approximated to 20 nmol/L), which increases the lifetime risk of MCE of 60% (for each 100 nmol/L increment), in order to overcome this extra risk, he/she has to lower LDL-C by an extra amount of 0.71 mmol/L (27.46 mg/dL) starting at age 40. However, if this subject, instead of starting the treatment at 40 years, decides to start reducing cholesterol at 60 years, LDL-C has to be additionally lowered by more than 1 mmol in order to abolish the same risk to which the subject is exposed due to elevate levels of Lp(a). Same interpretation can be made for different starting points of Lp(a) concentration or age at which the treatment is initiated.

Overall, for people below the 90$^{th}$ percentile, the LDL-C has to be lowered just by a quarter of a 1 mmol if the lipid-lowering treatment is started at age 30, to overcome the risk caused by high Lp(a). If they survive to 40, 50, 60 years, and the risk is still high, each decade the treatment is started later, a less relative risk reduction is produced. Since LDL-C has a cumulative effect over time, the longer is waited to start the treatment, the more aggressive the LDL-C has to be decreased to reduce the risk due to high Lp(a) concentrations.

**Table 2.** Extra LDL cholesterol reduction needed to abolish the risk due to high Lp(a), compared with the median value in the population.

| | | | | Age initiate treatment (years) | | | |
|---|---|---|---|---|---|---|---|
| **Lp(a) nmol/L** | **Percentile** | **Δ nmol/L** | **HR for MCE** | **30** | **40** | **50** | **60** |
| 320 | 99 | 300 | 2.56 | 1.19 | 1.41 | 1.74 | 2.28 |
| 270 | 97.5 | 250 | 2.19 | 0.99 | 1.18 | 1.45 | 1.90 |
| 220 | 93.5 | 200 | 1.87 | 0.79 | 0.94 | 1.16 | 1.52 |
| 170 | 90 | 150 | 1.60 | 0.59 | 0.71 | 0.87 | 1.14 |
| 120 | 82.5 | 100 | 1.37 | 0.40 | 0.47 | 0.58 | 0.76 |
| 70 | 75 | 50 | 1.17 | 0.20 | 0.24 | 0.29 | 0.38 |
| 20 | 50 | ref | ref | ref | ref | ref | ref |

HR: hazard ratio. Estimates have been standardized for 100 nmol/L increment in Lp(a) level; MCE: major coronary events; the extra LDL-C reduction needed to abolish the risk due to high Lp(a) is reported in grey stratified by different enter time and by different Lp(a) concentrations. LDL cholesterol is expressed as mmol/L, to covert LDL cholesterol to milligrams per decilitre, multiply by 38.67. A step by step example of how to calculate the extra LDL cholesterol reduction needed to abolish the risk due to high Lp(a), compared with the median value in the population, if the treatment is started at 40 years, is presented in **Appendix XIII**.

## Discussion

The results of this study show how to use Lp(a) clinically. Specifically, we quantify how much LDL-C has to be lowered to further reduce the risk and overcame the increased risk caused by high Lp(a) concentrations. However, because LDL-C has a cumulative effect over time, the earlier we start treating patients, the less LDL-C has to be reduced in order to overcome that risk. By contrast, the later we start the lipid-lowering treatment, the more aggressive LDL-C has to be lowered in order to abolish the risk. This hepatizes that it is important to screen for Lp(a) level earlier, so LDL-C can be lowered by a modest amount to overcome the risk, but also that it is important to screen for Lp(a) also later in life, because people with high Lp(a) concentrations are exposed to a really high risk and to

overcome it the LDL-C has to be decreased more aggressively.

Although genetic and epidemiological data strongly support the prognostic causality of Lp(a), patients with elevated Lp(a) are significantly under-diagnosed and the screening is frequently opportunistic rather than systematic. Lp(a) concentrations remain quite stable through the life course; this means that it is possible to measure Lp(a) once in a person's lifetime to ascertain risk.

Because most patients are not aware of their Lp(a)-mediated risk, there is a rationale to add Lp(a) measurement to the lipid panel of a patient in whom lipids are measured for the first time. If Lp(a) is in the normal range, then subsequent measurements are not needed, irrespective of any change in the patient's medical therapy. If Lp(a) level is elevated, we offered a strategy to estimate the cardiovascular risk to which the subject is exposed and a solution to overcome this risk, quantifying the additional amount of LDL-C that has to be lowered in order to abolish this risk, without acting through Lp(a)-lowering drugs.

So far, it was still unclear if Lp(a) remains a risk factor when LDL-C is controlled. Previous data from angiographic progression studies suggested that Lp(a) is no longer a risk factor when LDL-C is under control. For this reason, many clinicians have practiced with the assumption that when an elevated Lp(a) level is discovered, the most appropriate course of action was to treat the LDL-C and not the Lp(a). Recent studies have suggested that this is a false assumption, and that elevated Lp(a) remains a risk factor even in patients who achieve LDL-C <70 mg/dL (215-217). Furthermore, the concept of diminishing returns is now apparent in outcomes trials of LDL-C lowering, in which the starting LDL-C is often <100 mg/dL, but the absolute risk reduction is small. For example, in IMPROVE-IT (Improved Reduction of Outcomes: Vytorin Efficacy International

Trial) (218), after a median follow-up of 6 years, the major adverse cardiac event (MACE) rate was 32.7% in the simvastatin/ezetimibe group, which achieved LDL-C of 54 mg/dL, and 34.7% in the simvastatin-alone arm, which achieved LDL-C of 70 mg/dL. Although this was a laudatory achievement, a 32.7% recurrent hard MACE rate in the setting of an LDL-C of 54 mg/dL suggests that LDL-C–directed risk reduction might not reduce events optimally, even with very effective therapies such as PCSK9 inhibitors. Recent reports from the AIM-HIGH (Atherothrombosis Intervention in Metabolic Syndrome with Low HDL/High Triglyceride and Impact on Global Health Outcomes) (215), JUPITER (Justification for the Use of Statins in Prevention: an Intervention Trial Evaluating Rosuvastatin) (216), and LIPID (Long-Term Intervention with Pravastatin in Ischaemic Disease) (217) trials have suggested that a portion of this "residual risk" is due to highly elevated Lp(a) in the setting of controlled LDL-C levels. For example, in AIM-HIGH, patients who achieved LDL-C levels of 65.2 mg/dL and had Lp(a) >125 nmol/L (~50 mg/dL), which was ≥75th percentile of Lp(a) levels, had an 89% higher risk of MACEs compared with those who had a similar LDL-C, but low Lp(a) level. In JUPITER, patients who achieved LDL-C of 55.0 mg/dL and Lp(a) >54 nmol/L (~21 mg/dL) had a 71% higher risk of MACEs. In LIPID, in patients who achieved LDL-C of ~112 mg/dL and Lp(a) >73.7 mg/dL, a 23% increase in MACEs was found. The overall data encompassing 13,167 statin-treated patients shows a weighed hazard ratio of 1.61 in the setting of LDL-C of 89.1 mg/dL and Lp(a) of 54.9 mg/dL. These data strongly support the independent role of Lp(a) in mediating CVD events that may explain some of the residual risk in patients on well-established statin therapy. Our results offer a practical strategy to reduce, at list in part, this residual risk when Lp(a) concentrations are high, even if LDL-C level are controlled.

Limitations

First, participants in the UK biobank are a self-selected group who tend to be at lower risk of cardiovascular events than members of other populations. As a result, all analyses should be repeated in populations at higher risk. In addition, the analysis was restricted to participants who self-identified as being of white European ancestry, and therefore results may not apply to persons of other ethnicities.

## Conclusion

Based on different starting points of Lp(a) levels, we quantitatively estimated the amount of additional LDL-C reduction needed to overcome the extra risk due to increased levels of Lp(a). This evidence encourages even more to add Lp(a) measurement to the lipid panel in clinical practice in order to accurately assess individuals' cardiovascular risk and establish the most appropriate therapy.

# CHAPTER 6

## Discussion and conclusions

## 6.1 New insights into improving CVD management

The methodology of Mendelian randomization presented in this dissertation have been applied to several studies. This investigation has shed light on new important evidence, has improved our knowledge about cardiovascular disease, and has showed how Mendelian randomization can be used to inform clinical practice.
In particular, we found that:

- body mass index is a much stronger risk factor for type 2 diabetes compared to genetic predisposition to have high values of body mass index and that the effect of body mass index does not increase with increasing duration of exposure suggesting that increased body mass index leads to reversible metabolic changes. From a clinical point of view this implies that the risk of type 2 diabetes is largely modifiable if clinical interventions are put in place in the early stages of the disease, not only through the pharmacological control of plasma glucose but also suggesting weight loss as a strategy to lower body mass index. A graphical summary is presented in **Figure 10**.
- both parental family history of coronary heart disease and individual polygenic predisposition affect the lifetime risk of major coronary events through a dose response relationship, and that these two factors provide complementary and additive information to the definition of the inherited genetic variation as well as to LDL cholesterol exposure in the estimation of the lifetime cardiovascular risk. Consequently, only considered together, they are able to identify people who will need to be treated because exposed to a very high cardiovascular risk. A graphical summary is presented in **Figure 11**.
- profiling the genetic determinants of plasma lipoprotein(a)

provides comparable value to measured levels in terms of atherosclerotic cardiovascular disease risk prediction. However, since the distribution of measured Lp(a) levels was quite wide even among people with the same genotype (with the risk being affected accordingly), our evidence emphasizes the importance of measuring Lp(a) levels in clinical practice to better identify patients at higher risk for a correct management of them. A graphical summary is presented in **Figure 12**.

- Lp(a) does not appear to have a venous or arterial prothrombotic effect. Indeed, genetically predicted and measured Lp(a) concentrations were not associated with clinically significant thrombotic events, neither the effect of increased Lp(a) levels on major coronary events was attenuated by either an antiplatelet or antithrombin therapy, leading to an important evidence for the design of clinical trials enrolling patients in secondary prevention. A graphical summary is presented in **Figure 13**.
- an extra reduction of LDL cholesterol can overcome the increased risk due to exposure to high Lp(a) levels. More precisely, we exactly quantify how much LDL cholesterol has to be lowered to abolish this risk. From a clinical point of view this evidence encourages even more to add Lp(a) measurement to the lipid panel in order to establish the most appropriate therapy. A graphical summary is presented in **Figure 14**.

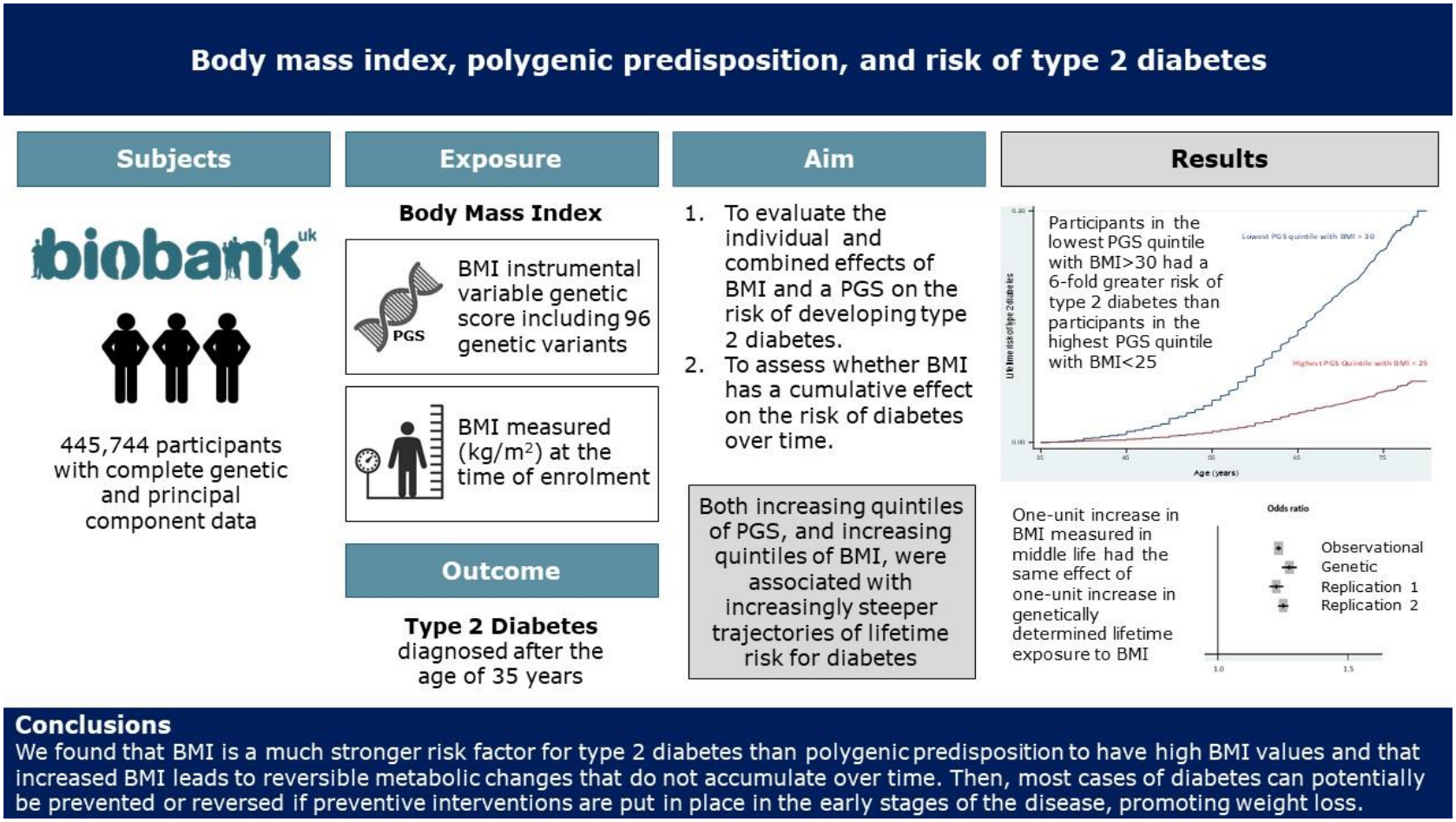



**Figure 10.** A graphical summary to briefly present the take-home messages of the analysis performed.

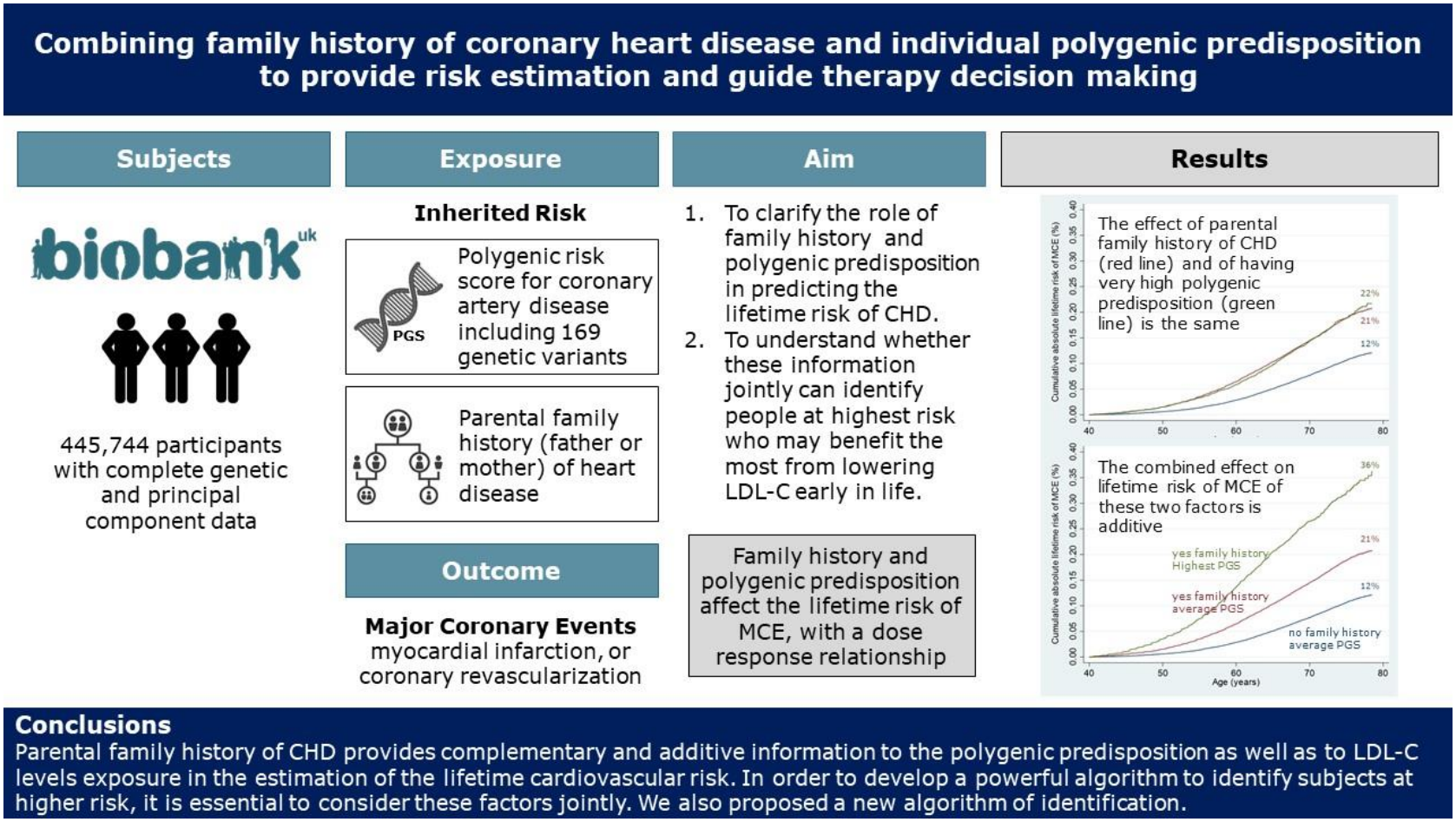



**Figure 11.** A graphical summary to briefly present the take-home messages of the analysis performed.

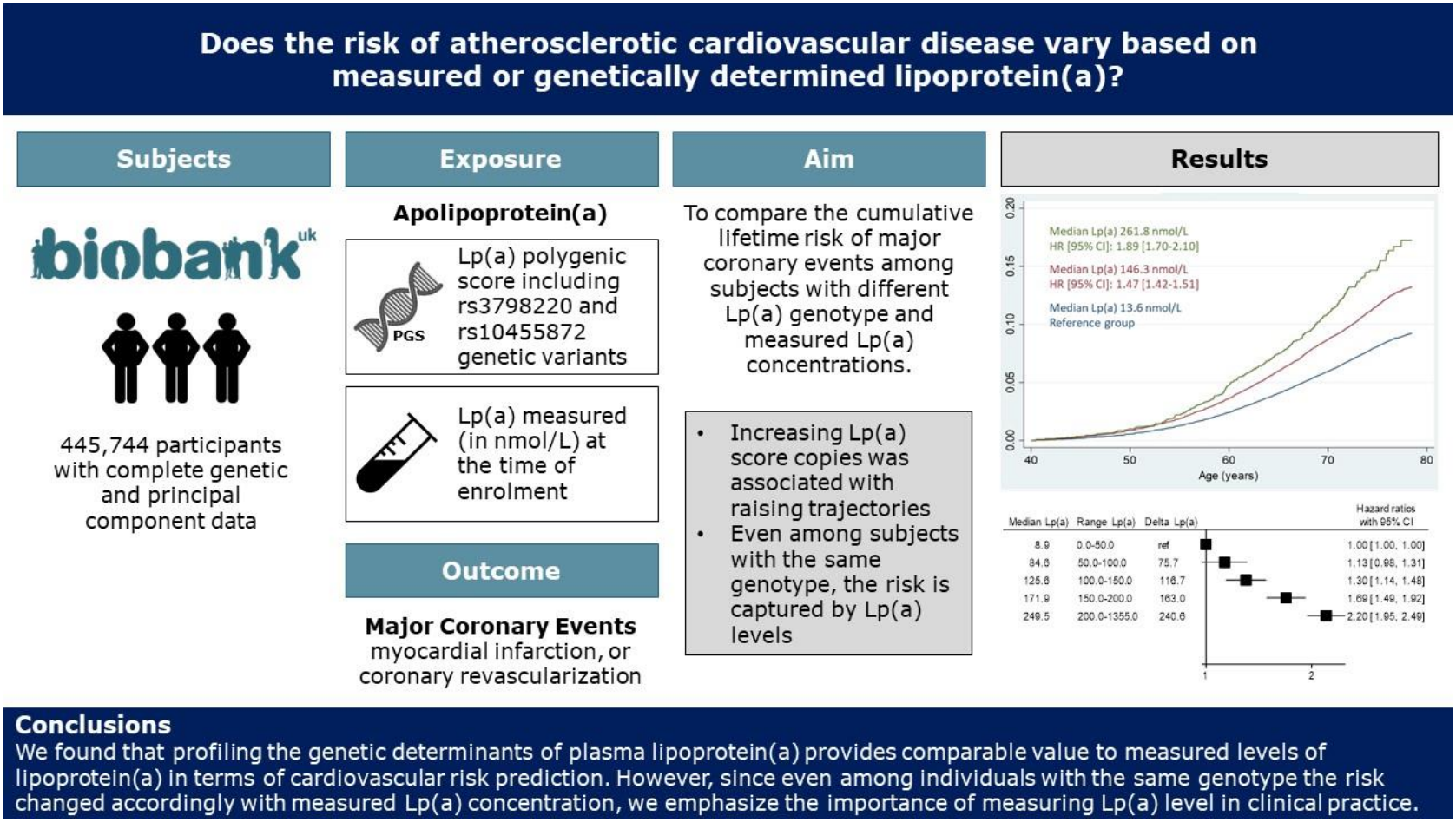



**Figure 12.** A graphical summary to briefly present the take-home messages of the analysis performed.

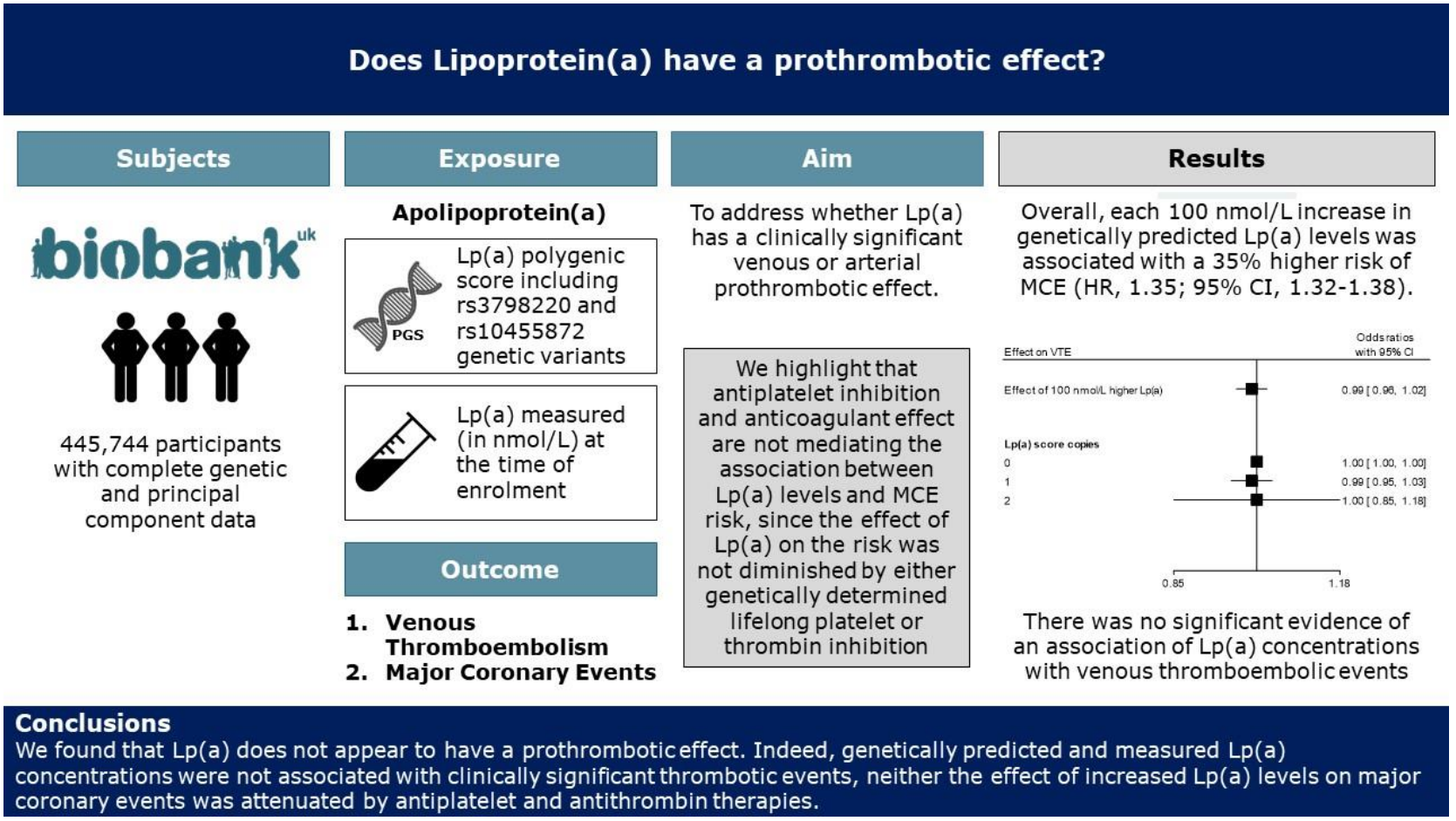



**Figure 13.** A graphical summary to briefly present the take-home messages of the analysis performed.

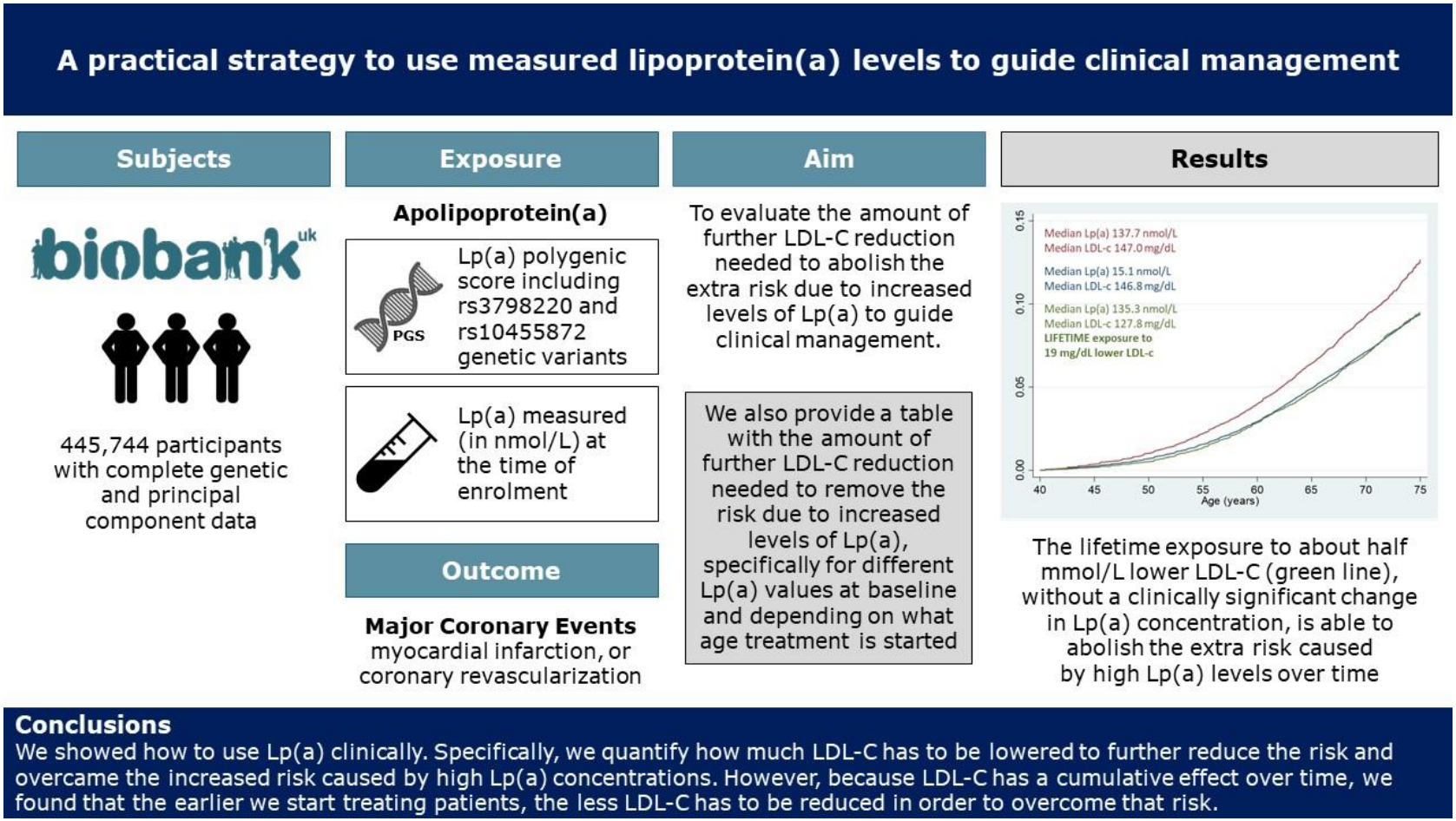



**Figure 14.** A graphical summary to briefly present the take-home messages of the analysis performed.

## 6.2 Biomarkers and therapeutic targets

Projects presented in this dissertation (chapter 5), are an example of how genetic evidence can be used to support causation, and to discover modifiable risk factors that are causal and amenable to therapeutic modification. Indeed, these biomarkers can potentially be modified pharmacologically by manipulating a drug target. Although important differences exist between genetic investigations of a quantitative trait and a drug target (219), MR has proven to be a valid approach to support both these types of research.

The primary question and motivation of a MR study of a biomarker is typically whether the biomarker has a role in disease. This knowledge is of interest from a biological perspective (because it can highlight new aetiological pathways, which enriches our understanding of the biological processes that underlie a disease and can stimulate new avenues of investigation, permitting also the development of therapies that act on risk factors along the causal pathway to ameliorate the consequences of harmful exposures) and from a public health perspective (because lifestyle modification-mediated alterations in the biomarker might yield large benefits in public health).

By contrast, the primary question and motivation of a MR study of a drug target is whether modifying the drug target alters the risk of disease and the primary goal is to estimate the likelihood that a specific therapeutic agent (typically designed to modify a complex biomarker, which might differ from the effects of a protein drug target that is under investigation on the basis that the protein modifies the same complex biomarker) will show efficacy in a final phase of a clinical trial. MR analysis of drug targets is useful because this approach uses genetic variants that closely mimic the action of

the therapeutic target of interest, and for this reason, this analysis can potentially provide some of the most reliable evidence of the effects of modifying the biomarker, via a specific therapeutic target, on long-term health outcomes.

Anyway, in both scenarios, human genetics can provide an invaluable tool to elucidate the causal roles in disease aetiology and, through careful application, these studies can guide the design and interpretation of clinical trials of medicines.

## 6.3 Potential use of MR to inform the design of RCTs

To be licensed and available to patients, drugs must undergo rigorous investigations, including initial studies of tolerability, followed by phase I trials (typically dose-ranging in nature), phase II trials (exploring efficacy), and then randomized, controlled, phase III outcome trials, which investigate clinical efficacy for the intended outcome, while showing a lack of clinically meaningful adverse events. Phase III trials are typically the final hurdle before applying for marketing authorization. If the drug is found to be efficacious for the primary outcome and safe, the regulators are likely to grant the applicant a licence to market the drug. For interventions for CVD, the process from date of first testing in humans (clinical entry) through to marketing authorization takes a median of 10 years (220), and often the development process is interrupted due to lack of efficacy or safety issues. Indeed, one of the crucial issues in drug development is failure due to lack of efficacy (accounting for 52% of drug failures in phase II and phase III trials) (221), with therapies targeting risk factors that are probably non-causal being a major roadblock to innovation (222). Another major hurdle to the development of new drugs is safety (accounting for 24% of drug

failures (221)), with therapies abandoned owing to both dose-related and idiosyncratic adverse events. Mostly for these reasons, fewer than 10% of drugs move from phase I through to marketing authorization, with the success rate being disease-specific. It is here that MR can find one of its most valuable application. Beyond simply providing quantitative evidence of whether a therapy acting on a given drug target is likely to be efficacious and safe, MR studies of drug targets can provide a variety of additional information that can feed into multiple facets of clinical trial design.

For example, in the trial design phase, MR can:

- provide information when a trial might be unethical to conduct (for example, potentially harmful intervention, potential ethical issues of withholding treatment or placebo comparator);
- inform on whether a trial should be conducted in particular phenotypically defined or genetically defined subgroups;
- predict the outcome of a factorial trial design, for example, by predicting drug–drug interactions.

In addition, MR can help to define the characteristics of the intervention when no therapies exist, or to deprioritize the development of therapies acting on non-causal targets, or to explore the nature of the causal relationship across the physiological distribution of the exposure. This genetic approach can be also useful to explain mechanisms of effect, or to clarify pleiotropic effects, such as the off-target effects of a drug class (for example, whether the drug has effects on an outcome beyond that which is considered to be the primary exposure) or the off-target effects of an individual drug, or to explore the full repertoire of target-mediated effects (both beneficial and potentially harmful)

Finally, MR can be used to inform on potential alterations to eligibility criteria to enrol participants who will experience:

- greater benefit, which might allow the trial to recruit fewer participants or have a shorter duration and thereby lower the cost of drug development;
- fewer adverse effects: although it might not be possible to separate out the mechanism-based adverse effects of a therapeutic using genetic variants if an adverse event arises owing to a mechanistic pathway that is separate from the mechanistic pathway leading to clinical benefit, it might be possible to identify genetic variants that modify the propensity of the pharmacological agent to modify the pathway leading to adverse events and, by doing so, disentangle benefit from harm.

## 6.4 Potential use of MR to anticipate RCTs results

MR studies have become more prevalent in the literature as they are quicker and cheaper to conduct and can utilise existing data from large genetic consortia to provide qualitative information of treatment efficacy, target-mediated adverse effects and opportunities for drug repurposing, which can be very informative to prioritise biomarkers to take forward into phase II/III clinical trials in humans.

### 6.4.1 Predicting efficacy

There is considerable interest in predicting the efficacy of potential therapeutic targets as early as possible in the drug development process, as genetic support for a drug target can substantially enhance the probability for a RCT of a therapy targeting such a drug target to have success (223). Very often a lot of moneys are spent on what then turn out to be failed clinical trials. While the negative

outcomes cannot be fully anticipated when the trials are set up, if genetic investigations are pursued prior to embarking on the phase III trials, the drug development could be sidelined with unpromising results, and more promising targets prioritised in its place.

MR studies evaluate the effect of genetically determined lifelong changes in an exposure on an outcome, whereas randomized trials evaluate the effect of therapeutically induced short-term changes in the same exposure. The challenge when attempting to use MR to anticipate the results of randomized trials is therefore to translate the causal effect of lifelong changes in an exposure on an outcome to the effect that can be expected to occur in response to short-term therapeutically induced changes in that exposure on the outcome of interest. Unfortunately, the results of a MR study cannot be used to estimate directly the expected effect of a therapy in a short-term trial. This is because the effect of most causal exposures appears to have cumulative effects on the associated outcome over time, and the use of MR would lead to overestimate the expected effect size. Moreover, a common misconception is that a therapy directed against any causal exposure will probably improve the associated outcome. Unfortunately, finding that an exposure is causally associated with an outcome tells us almost nothing about whether a therapy directed against that exposure will improve the outcome in a randomized trial. To anticipate the results of a randomized trial, the critical question that must be answered is by how much the causal exposure must be changed to improve the associated outcome in a short-term randomized trial. Careful application of the algorithm presented in **Figure 15** should help to answer this critical question (224).

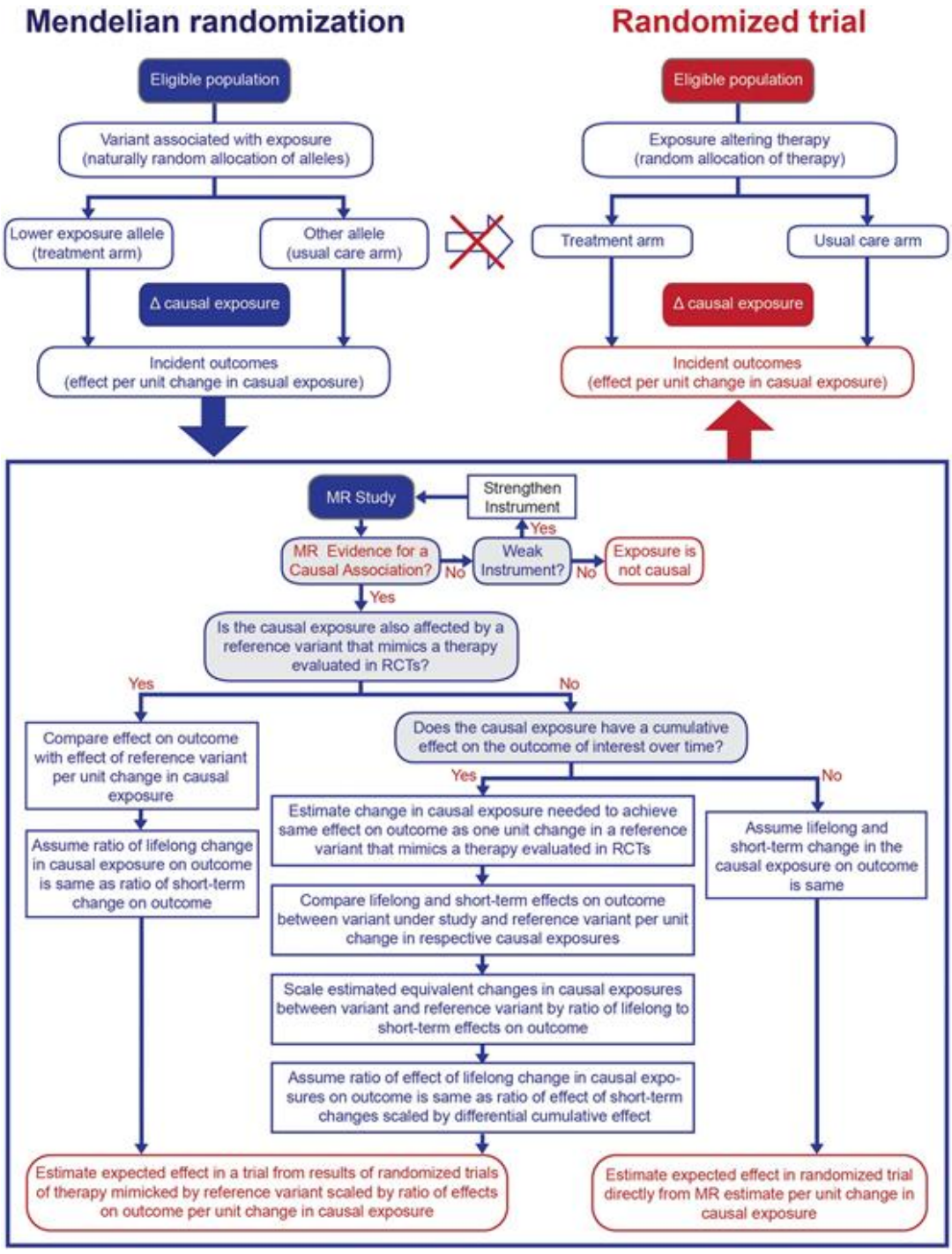



**Figure 15.** Algorithm for using Mendelian randomization studies to anticipate results of randomized trials.

### 6.4.2 Predicting target-mediated adverse effects

MR studies could also be a valuable tool to suggest potential target-mediated adverse events. Let's clarify with an example. Statins are

one of the most widely prescribed medications for lowering LDL-C for primary and secondary prevention of CVD, and several large-scale trials, and individual patient data meta-analyses of large-scale RCTs have clearly demonstrated that they are effective, compared with placebo, at lowering risk of CVD (225). Even though statins are very safe, they are not without side effects, and there is controversy about the frequency of these adverse effects and how they should be reliably investigated, such as long-term follow-up of trials or using observational data (226). One of the adverse effects of statins is type 2 diabetes (T2D), and meta-analysis of RCTs and MR studies have shown this to be an on-target effect of hydroxymethylglutaryl coenzyme A (HMG-CoA) reductase (141). The question naturally arises whether other LDL-C lowering drugs (such as ezetimibe or PCSK9 inhibitors) will have the same diabetogenic effects. Lotta et al. (227) used data from a meta-analysis of 50 000 T2D cases and 270 000 controls and found that the LDL-C SNPs in the NPC1L1 gene (a genetic proxy for ezetimibe) and HMGCR gene (a genetic proxy for statins) were associated with an OR of T2D of 2.42 (95%CI: 1.70 to −3.43) and 1.39 (1.12 to −1.73) per genetically predicted 1 mmol/L lower LDL-C, respectively. Two recent MR studies (140, 228) have provided orthogonal evidence that LDL-C lowering through PCSK9 inhibition is, as with statins, likely to lead to increased risk of T2D, and more generally, a recent study using SNPs across the genome provides evidence that an increased risk of T2D may arise as a general consequence of LDL-C lowering (133).

MR does not require the exposure of patients to the drug, this means it can be implemented at any point during the drug development process and beyond. This can allow pre-specification of likely adverse outcomes in trials; and reduce the possibility of exposing patients to unnecessary risks and harm.

### 6.4.3 Repurposing drugs

MR can be used to inform on potential repurposing of drugs, for example through use of a "phenome-wide scan", which is facilitated by the availability of large-scale prospective biobanks with incident diagnoses procured through electronic health records, such as the UK Biobank. MR can also be used to investigate for potential pharmacogenetic associations (whereby one or more genetic variants are used to identify patients more likely to respond to a drug and/or patients less likely to suffer an adverse drug reaction): by stratifying a MR analysis by these genetic variants, clarity can be provided as to whether subpopulations are likely to derive greater benefit from a drug. This may be helpful prior to embarking on de novo RCTs targeting such subgroups (229).

Developments in genotyping and availability of large-scale genetic consortia with hypothesis-free reliable discovery of new genetic variants for biomarkers that may play causal roles in disease development and application of such genetic variants in MR analyses present a wide range of opportunities to identify potentially modifiable exposures that, if shown to be causal, may be tested in future intervention studies.

## 6.5 Conclusions

Studies based on MR are increasingly being used to distinguish causal relationships from observational associations in epidemiology and to prioritize potential targets for pharmaceutical intervention, especially in recent years, where -omics technologies, including genomics, transcriptomics, proteomics, metabolomics, and more recently epigenomics, have developed rapidly. The application of these technologies in observational studies has generated a very large

number of novel exposures/intermediate phenotypes that researchers can use to assess associations with clinical endpoints. These scans are so called "hypothesis free" approaches, because they do not rely on underlying biologic assumptions and are, therefore, suited to unravel unknown biology. The results of such association studies represent a vast amount of unbiased information on potentially (modifiable) exposures and instrumental variables, which can subsequently be used to assess novel causal relationships or verify those examined in RCTs. To address this new situation, several extensions to MR approaches have been developed in recent years to allow for more complex questions to be answered under a MR framework.

Although MR appears to be a perfect epidemiological approach to directly estimating the causal effect, there are still limitations and assumptions in its application which must be taken into account. First, genetic association studies often investigate only common genetic variants or combine the effect of rare genetic variants. This results in a situation where individual genetic variants may explain very little of the observed variation. Careful consideration must therefore be given to the choice of genetic variant when conducting an MR study. Second, MR estimates indicate lifelong perturbations in an exposure, which may not be equivalent to interventions given at a specific point in time and for a shorter time period. Therefore, careful consideration of the exposure and its timing must be made to avoid misinterpretation of results. For example, some exposures are cumulative whereby repeated exposure, over a sustained period, results in the outcome. MR analyses of such exposures are likely to overestimate the effect observed in other study designs, including RCTs, as these designs consider much shorter periods of exposure with lower compliance. A further example is time-dependent

exposures. MR analyses of this type of exposure can provide misleading evidence about the effect of manipulating an exposure after the critical period. This is because the MR estimate will, by definition, include any critical periods in its assessment of lifelong exposure. Finally, a large proportion of the genetic variants that have been identified to date are concerned with the incidence of disease. In order to predict unintended drug effects that relate to the treatment of that disease, genetic variants relating to progression will need to be identified. Indeed, so far, only a small proportion of GWAS studies (about 8% of associations curated in the GWAS Catalog) have attempted to identify variants associated with disease progression or severity (230). All of these technical issues can strongly influence the conclusions of MR analyses and thus highlight the importance of cautious interpretation of findings.

However, despite all these limitations, MR can still be used to evaluate whether a causality exists between an exposure and an outcome. MR uses germline genetic variants that are less likely to be confounded by environmental, lifestyle or disease-related factors operating later in life. Consequently, if a genetic variant is associated with an outcome only through its association with a drug effect, it is likely to be because the genetic variant causes the outcome. Moreover, as reported before, MR does not require the exposure of patients to the drug, this means it can be implemented at any point during the drug development process and beyond. This can increase the efficiency of drug development by identifying unsuitable targets, allow pre-specification of likely adverse outcomes in trials, reduce the possibility of exposing patients to unnecessary risks and harms, and can be highly useful to inform drug development and repurposing (230).

In conclusion, MR offers a novel and appealing approach to assess the causality of observed exposure-outcome associations through genetic instrumental variables, which can address some of the limitations associated with existing methods in this field. The use of Mendelian randomization will likely become increasingly popular in medical research in the development of drug profiles to help prevent adverse drug events and identify novel indications for existing drugs.

# CHAPTER 7

## Bibliography

1.Parascandola M, Weed DL. Causation in epidemiology. J Epidemiol Community Health. 2001;55(12):905-12.

2.Rubin DB. For Objective Causal Inference, Design Trumps Analysis. Ann Appl Stat. 2008;2(3):808-40.

3.Taubes G. Epidemiology faces its limits. Science. 1995;269(5221):164-9.

4.Meldrum ML. A brief history of the randomized controlled trial. From oranges and lemons to the gold standard. Hematol Oncol Clin North Am. 2000;14(4):745-60, vii.

5.Hariton E, Locascio JJ. Randomised controlled trials - the gold standard for effectiveness research: Study design: randomised controlled trials. BJOG. 2018;125(13):1716.

6.Saturni S, Bellini F, Braido F, Paggiaro P, Sanduzzi A, Scichilone N, et al. Randomized Controlled Trials and real life studies. Approaches and methodologies: a clinical point of view. Pulm Pharmacol Ther. 2014;27(2):129-38.

7.Price D, Chisholm A, van der Molen T, Roche N, Hillyer EV, Bousquet J. Reassessing the evidence hierarchy in asthma: evaluating comparative effectiveness. Curr Allergy Asthma Rep. 2011;11(6):526-38.

8.Ware JH, Hamel MB. Pragmatic trials--guides to better patient care? N Engl J Med. 2011;364(18):1685-7.

9.van Deudekom FJ, Postmus I, van der Ham DJ, Pothof AB, Broekhuizen K, Blauw GJ, et al. External validity of randomized controlled trials in older adults, a systematic review. PLoS One. 2017;12(3):e0174053.

10.Nallamothu BK, Hayward RA, Bates ER. Beyond the randomized clinical trial: the role of effectiveness studies in evaluating cardiovascular therapies. Circulation. 2008;118(12):1294-303.

11.Pauwels RA, Lofdahl CG, Postma DS, Tattersfield AE, O'Byrne P, Barnes PJ, et al. Effect of inhaled formoterol and budesonide on exacerbations of asthma. Formoterol and Corticosteroids

Establishing Therapy (FACET) International Study Group. N Engl J Med. 1997;337(20):1405-11.

12.Price D, Hillyer EV, van der Molen T. Efficacy versus effectiveness trials: informing guidelines for asthma management. Curr Opin Allergy Clin Immunol. 2013;13(1):50-7.

13.Perez-Gomez A, Mejia-Trujillo J, Mejia A. How useful are randomized controlled trials in a rapidly changing world? Glob Ment Health (Camb). 2016;3:e6.

14.Thiese MS. Observational and interventional study design types; an overview. Biochem Med (Zagreb). 2014;24(2):199-210.

15.Mahajan R. Real world data: Additional source for making clinical decisions. Int J Appl Basic Med Res. 2015;5(2):82.

16.Mann CJ. Observational research methods. Research design II: cohort, cross sectional, and case-control studies. Emerg Med J. 2003;20(1):54-60.

17.Samet JM. Measuring the effectiveness of inhaled corticosteroids for COPD is not easy! Am J Respir Crit Care Med. 2003;168(1):1-2.

18.Jarvinen TL, Sievanen H, Kannus P, Jokihaara J, Khan KM. The true cost of pharmacological disease prevention. BMJ. 2011;342:d2175.

19.Rizzo JD, Powe NR. Methodological hurdles in conducting pharmacoeconomic analyses. Pharmacoeconomics. 1999;15(4):339-55.

20.Tunis SR, Stryer DB, Clancy CM. Practical clinical trials: increasing the value of clinical research for decision making in clinical and health policy. JAMA. 2003;290(12):1624-32.

21.Albert RK. "Lies, damned lies ..." and observational studies in comparative effectiveness research. Am J Respir Crit Care Med. 2013;187(11):1173-7.

22.Salas M, Hofman A, Stricker BH. Confounding by indication: an example of variation in the use of epidemiologic terminology. Am J Epidemiol. 1999;149(11):981-3.

23.Kunz R, Oxman AD. The unpredictability paradox: review of empirical comparisons of randomised and non-randomised clinical trials. BMJ. 1998;317(7167):1185-90.

24.Boyko EJ. Observational research--opportunities and limitations. J Diabetes Complications. 2013;27(6):642-8.

25.Smith GD, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? Int J Epidemiol. 2003;32(1):1-22.

26.Pauling L, Itano HA, et al. Sickle cell anemia a molecular disease. Science. 1949;110(2865):543-8.

27.McPherson JD, Marra M, Hillier L, Waterston RH, Chinwalla A, Wallis J, et al. A physical map of the human genome. Nature. 2001;409(6822):934-41.

28.Roberts L, Davenport RJ, Pennisi E, Marshall E. A history of the Human Genome Project. Science. 2001;291(5507):1195.

29.Khoury MJ, Beaty TH, Cohen BH. Fundamentals of genetic epidemiology. New York: Oxford University Press; 1993. vi, 383 p. p.

30.Smith BH, Watt GC, Campbell H, Sheikh A. Genetic epidemiology and primary care. Br J Gen Pract. 2006;56(524):214-21.

31.Ioannidis JP, Trikalinos TA, Ntzani EE, Contopoulos-Ioannidis DG. Genetic associations in large versus small studies: an empirical assessment. Lancet. 2003;361(9357):567-71.

32.Wright AF, Carothers AD, Campbell H. Gene-environment interactions--the BioBank UK study. Pharmacogenomics J. 2002;2(2):75-82.

33.Ebrahim S, Davey Smith G. Mendelian randomization: can genetic epidemiology help redress the failures of observational epidemiology? Hum Genet. 2008;123(1):15-33.

34.Davidson R, MacKinnon JG. Estimation and inference in econometrics. New York: Oxford University Press; 1993. xx, 874 p. p.

35.Lawlor DA, Harbord RM, Sterne JA, Timpson N, Davey Smith G. Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. Stat Med. 2008;27(8):1133-63.

36.Greenland S. An introduction to instrumental variables for epidemiologists. Int J Epidemiol. 2000;29(4):722-9.

37.Sussman JB, Hayward RA. An IV for the RCT: using instrumental variables to adjust for treatment contamination in randomised controlled trials. BMJ. 2010;340:c2073.

38.Wehby GL, Ohsfeldt RL, Murray JC. 'Mendelian randomization' equals instrumental variable analysis with genetic instruments. Stat Med. 2008;27(15):2745-9.

39.Martens EP, Pestman WR, de Boer A, Belitser SV, Klungel OH. Instrumental variables: application and limitations. Epidemiology. 2006;17(3):260-7.

40.Didelez V, Sheehan N. Mendelian randomization as an instrumental variable approach to causal inference. Stat Methods Med Res. 2007;16(4):309-30.

41.Hernan MA, Robins JM. Instruments for causal inference: an epidemiologist's dream? Epidemiology. 2006;17(4):360-72.

42.Smith GD, Ebrahim S. Mendelian randomization: prospects, potentials, and limitations. Int J Epidemiol. 2004;33(1):30-42.

43.Pai JK, Mukamal KJ, Rexrode KM, Rimm EB. C-reactive protein (CRP) gene polymorphisms, CRP levels, and risk of incident coronary heart disease in two nested case-control studies. PLoS One. 2008;3(1):e1395.

44.Thanassoulis G, O'Donnell CJ. Mendelian randomization: nature's randomized trial in the post-genome era. JAMA. 2009;301(22):2386-8.

45.Verduijn M, Siegerink B, Jager KJ, Zoccali C, Dekker FW. Mendelian randomization: use of genetics to enable causal inference in observational studies. Nephrol Dial Transplant. 2010;25(5):1394-8.

46.Debat V, David P. Mapping phenotypes: canalization, plasticity and developmental stability. Trends Ecol Evol. 2001;16(10):555-61.

47.National Research Council (U.S.). Committee on Advances in Collecting and Utilizing Biological Indicators and Genetic Information in Social Science Surveys., Weinstein M, Vaupel JW, Wachter KW, National Research Council (U.S.). Committee on Population. Biosocial surveys. Washington, D.C.: National Academies Press; 2008. xiii, 414 p. p.

48.Ogbuanu IU, Zhang H, Karmaus W. Can we apply the Mendelian randomization methodology without considering epigenetic effects? Emerg Themes Epidemiol. 2009;6:3.

49.Bochud M, Chiolero A, Elston RC, Paccaud F. A cautionary note on the use of Mendelian randomization to infer causation in observational epidemiology. Int J Epidemiol. 2008;37(2):414-6; author reply 6-7.

50.Thomas DC, Conti DV. Commentary: the concept of 'Mendelian Randomization'. Int J Epidemiol. 2004;33(1):21-5.

51.Lee K, Lim CY. Mendelian Randomization Analysis in Observational Epidemiology. J Lipid Atheroscler. 2019;8(2):67-77.

52.Glymour MM, Tchetgen Tchetgen EJ, Robins JM. Credible Mendelian randomization studies: approaches for evaluating the instrumental variable assumptions. Am J Epidemiol. 2012;175(4):332-9.

53.Sleiman PM, Grant SF. Mendelian randomization in the era of genomewide association studies. Clin Chem. 2010;56(5):723-8.

54.Palmer TM, Lawlor DA, Harbord RM, Sheehan NA, Tobias JH, Timpson NJ, et al. Using multiple genetic variants as instrumental variables for modifiable risk factors. Stat Methods Med Res. 2012;21(3):223-42.

55.Burgess S, Thompson SG. Use of allele scores as instrumental variables for Mendelian randomization. Int J Epidemiol. 2013;42(4):1134-44.

56.Benn M, Nordestgaard BG. From genome-wide association studies to Mendelian randomization: novel opportunities for understanding cardiovascular disease causality, pathogenesis, prevention, and treatment. Cardiovasc Res. 2018;114(9):1192-208.

57.Burgess S, Thompson SG. Mendelian Randomization: Methods for Using Genetic Variants in Causal Estimation: CRC Press; 2015.

58.Haycock PC, Burgess S, Wade KH, Bowden J, Relton C, Davey Smith G. Best (but oft-forgotten) practices: the design, analysis, and interpretation of Mendelian randomization studies. Am J Clin Nutr. 2016;103(4):965-78.

59.Keavney B. Commentary: Katan's remarkable foresight: genes and causality 18 years on. Int J Epidemiol. 2004;33(1):11-4.

60.Pierce BL, VanderWeele TJ. The effect of non-differential measurement error on bias, precision and power in Mendelian randomization studies. Int J Epidemiol. 2012;41(5):1383-93.

61.Frost C, Thompson SG. Correcting for regression dilution bias: comparison of methods for a single predictor variable. J R Stat Soc a Stat. 2000;163:173-89.

62.Interleukin-6 Receptor Mendelian Randomisation Analysis C, Swerdlow DI, Holmes MV, Kuchenbaecker KB, Engmann JE, Shah T, et al. The interleukin-6 receptor as a target for prevention of coronary heart disease: a mendelian randomisation analysis. Lancet. 2012;379(9822):1214-24.

63.Collaboration IRGCERF, Sarwar N, Butterworth AS, Freitag DF, Gregson J, Willeit P, et al. Interleukin-6 receptor pathways in coronary heart disease: a collaborative meta-analysis of 82 studies. Lancet. 2012;379(9822):1205-13.

64.Celermajer DS, Chow CK, Marijon E, Anstey NM, Woo KS. Cardiovascular disease in the developing world: prevalences, patterns, and the potential of early disease detection. J Am Coll Cardiol. 2012;60(14):1207-16.

65.Behera SS, Pramanik K, Nayak MK. Recent Advancement in the Treatment of Cardiovascular Diseases: Conventional Therapy to Nanotechnology. Curr Pharm Des. 2015;21(30):4479-97.

66.Roth GA, Mensah GA, Johnson CO, Addolorato G, Ammirati E, Baddour LM, et al. Global Burden of Cardiovascular Diseases and Risk Factors, 1990-2019: Update From the GBD 2019 Study. J Am Coll Cardiol. 2020;76(25):2982-3021.

67.Capewell S, Ford ES, Croft JB, Critchley JA, Greenlund KJ, Labarthe DR. Cardiovascular risk factor trends and potential for reducing coronary heart disease mortality in the United States of America. Bull World Health Organ. 2010;88(2):120-30.

68.Gaziano TA, Bitton A, Anand S, Abrahams-Gessel S, Murphy A. Growing epidemic of coronary heart disease in low- and middle-income countries. Curr Probl Cardiol. 2010;35(2):72-115.

69.Frostegard J. Immunity, atherosclerosis and cardiovascular disease. BMC Med. 2013;11:117.

70.Davies MJ. The pathophysiology of acute coronary syndromes. Heart. 2000;83(3):361-6.

71.Hobbs FD. Cardiovascular disease: different strategies for primary and secondary prevention? Heart. 2004;90(10):1217-23.

72.Hobbs FD. Prevention of cardiovascular diseases. BMC Med. 2015;13:261.

73.Ghebre YT, Yakubov E, Wong WT, Krishnamurthy P, Sayed N, Sikora AG, et al. Vascular Aging: Implications for Cardiovascular Disease and Therapy. Transl Med (Sunnyvale). 2016;6(4).

74.North BJ, Sinclair DA. The intersection between aging and cardiovascular disease. Circ Res. 2012;110(8):1097-108.

75.Minamino T, Komuro I. Vascular aging: insights from studies on cellular senescence, stem cell aging, and progeroid syndromes. Nat Clin Pract Cardiovasc Med. 2008;5(10):637-48.

76.Brandes RP, Fleming I, Busse R. Endothelial aging. Cardiovasc Res. 2005;66(2):286-94.

77.Goldspink DF, Burniston JG, Tan LB. Cardiomyocyte death and the ageing and failing heart. Exp Physiol. 2003;88(3):447-58.

78.Anversa P, Kajstura J, Leri A, Bolli R. Life and death of cardiac stem cells: a paradigm shift in cardiac biology. Circulation. 2006;113(11):1451-63.

79.Hamilton PK, Lockhart CJ, Quinn CE, McVeigh GE. Arterial stiffness: clinical relevance, measurement and treatment. Clin Sci (Lond). 2007;113(4):157-70.

80.Lakatta EG, Levy D. Arterial and cardiac aging: major shareholders in cardiovascular disease enterprises: Part II: the aging heart in health: links to heart disease. Circulation. 2003;107(2):346-54.

81.Baker L, Meldrum KK, Wang M, Sankula R, Vanam R, Raiesdana A, et al. The role of estrogen in cardiovascular disease. J Surg Res. 2003;115(2):325-44.

82.Maas AH, Appelman YE. Gender differences in coronary heart disease. Neth Heart J. 2010;18(12):598-602.

83.Kannel WB. The Framingham Study: historical insight on the impact of cardiovascular risk factors in men versus women. J Gend Specif Med. 2002;5(2):27-37.

84.Meyer MR, Haas E, Barton M. Gender differences of cardiovascular disease: new perspectives for estrogen receptor signaling. Hypertension. 2006;47(6):1019-26.

85.Merz AA, Cheng S. Sex differences in cardiovascular ageing. Heart. 2016;102(11):825-31.

86.Krejza J, Arkuszewski M, Kasner SE, Weigele J, Ustymowicz A, Hurst RW, et al. Carotid artery diameter in men and women and the relation to body and neck size. Stroke. 2006;37(4):1103-5.

87.Moller-Leimkuhler AM. Gender differences in cardiovascular disease and comorbid depression. Dialogues Clin Neurosci. 2007;9(1):71-83.

88.Barrett-Connor E. Sex differences in coronary heart disease. Why are women so superior? The 1995 Ancel Keys Lecture. Circulation. 1997;95(1):252-64.

89.Spence JD, Pilote L. Importance of sex and gender in atherosclerosis and cardiovascular disease. Atherosclerosis. 2015;241(1):208-10.

90.Bolego C, Poli A, Paoletti R. Smoking and gender. Cardiovasc Res. 2002;53(3):568-76.

91.Eastwood JA, Doering LV. Gender differences in coronary artery disease. J Cardiovasc Nurs. 2005;20(5):340-51; quiz 52-3.

92.Lloyd-Jones DM, Larson MG, Beiser A, Levy D. Lifetime risk of developing coronary heart disease. Lancet. 1999;353(9147):89-92.

93.Roger VL, Go AS, Lloyd-Jones DM, Adams RJ, Berry JD, Brown TM, et al. Heart disease and stroke statistics--2011 update: a report from the American Heart Association. Circulation. 2011;123(4):e18-e209.

94.Hajjar I, Kotchen TA. Trends in prevalence, awareness, treatment, and control of hypertension in the United States, 1988-2000. JAMA. 2003;290(2):199-206.

95.Lawes CM, Vander Hoorn S, Rodgers A, International Society of H. Global burden of blood-pressure-related disease, 2001. Lancet. 2008;371(9623):1513-8.

96.Haider AW, Larson MG, Franklin SS, Levy D, Framingham Heart S. Systolic blood pressure, diastolic blood pressure, and pulse pressure as predictors of risk for congestive heart failure in the Framingham Heart Study. Ann Intern Med. 2003;138(1):10-6.

97.Collins R, Peto R, MacMahon S, Hebert P, Fiebach NH, Eberlein KA, et al. Blood pressure, stroke, and coronary heart disease. Part 2, Short-term reductions in blood pressure: overview of randomised drug trials in their epidemiological context. Lancet. 1990;335(8693):827-38.

98.Messerli FH, Rimoldi SF, Bangalore S. The Transition From Hypertension to Heart Failure: Contemporary Update. JACC Heart Fail. 2017;5(8):543-51.

99.Rafieian-Kopaei M, Setorki M, Doudi M, Baradaran A, Nasri H. Atherosclerosis: process, indicators, risk factors and new hopes. Int J Prev Med. 2014;5(8):927-46.

100.Li JJ, Chen JL. Inflammation may be a bridge connecting hypertension and atherosclerosis. Med Hypotheses. 2005;64(5):925-9.

101.Chobanian AV, Alexander RW. Exacerbation of atherosclerosis by hypertension. Potential mechanisms and clinical implications. Arch Intern Med. 1996;156(17):1952-6.

102.Kjeldsen SE. Hypertension and cardiovascular risk: General aspects. Pharmacol Res. 2018;129:95-9.

103.Vogt A. The genetics of familial hypercholesterolemia and emerging therapies. Appl Clin Genet. 2015;8:27-36.

104.Roy S. Atherosclerotic Cardiovascular Disease Risk and Evidence-based Management of Cholesterol. N Am J Med Sci. 2014;6(5):191-8.

105.Stamler J, Daviglus ML, Garside DB, Dyer AR, Greenland P, Neaton JD. Relationship of baseline serum cholesterol levels in 3 large cohorts of younger men to long-term coronary, cardiovascular, and all-cause mortality and to longevity. JAMA. 2000;284(3):311-8.

106.Daniels TF, Killinger KM, Michal JJ, Wright RW, Jr., Jiang Z. Lipoproteins, cholesterol homeostasis and cardiac health. Int J Biol Sci. 2009;5(5):474-88.

107.Zodda D, Giammona R, Schifilliti S. Treatment Strategy for Dyslipidemia in Cardiovascular Disease Prevention: Focus on Old and New Drugs. Pharmacy (Basel). 2018;6(1).

108.Zarate A, Manuel-Apolinar L, Saucedo R, Hernandez-Valencia M, Basurto L. Hypercholesterolemia As a Risk Factor for

Cardiovascular Disease: Current Controversial Therapeutic Management. Arch Med Res. 2016;47(7):491-5.

109.Otsuka F, Yasuda S, Noguchi T, Ishibashi-Ueda H. Pathology of coronary atherosclerosis and thrombosis. Cardiovasc Diagn Ther. 2016;6(4):396-408.

110.Podrez EA, Byzova TV, Febbraio M, Salomon RG, Ma Y, Valiyaveettil M, et al. Platelet CD36 links hyperlipidemia, oxidant stress and a prothrombotic phenotype. Nat Med. 2007;13(9):1086-95.

111.Leon BM, Maddox TM. Diabetes and cardiovascular disease: Epidemiology, biological mechanisms, treatment recommendations and future research. World J Diabetes. 2015;6(13):1246-58.

112.Guariguata L, Whiting DR, Hambleton I, Beagley J, Linnenkamp U, Shaw JE. Global estimates of diabetes prevalence for 2013 and projections for 2035. Diabetes Res Clin Pract. 2014;103(2):137-49.

113.De Rosa S, Arcidiacono B, Chiefari E, Brunetti A, Indolfi C, Foti DP. Type 2 Diabetes Mellitus and Cardiovascular Disease: Genetic and Epigenetic Links. Front Endocrinol (Lausanne). 2018;9:2.

114.Grundy SM, Benjamin IJ, Burke GL, Chait A, Eckel RH, Howard BV, et al. Diabetes and cardiovascular disease: a statement for healthcare professionals from the American Heart Association. Circulation. 1999;100(10):1134-46.

115.Folsom AR, Eckfeldt JH, Weitzman S, Ma J, Chambless LE, Barnes RW, et al. Relation of carotid artery wall thickness to diabetes mellitus, fasting glucose and insulin, body size, and physical activity. Atherosclerosis Risk in Communities (ARIC) Study Investigators. Stroke. 1994;25(1):66-73.

116.Fagot-Campagna A, Rolka DB, Beckles GL, Gregg EW, Narayan KM. Prevalence of lipid abnormalities, awareness, and treatment in US adults with diabetes. Diabetes. 2000;49:A78-A9.

117.Festa A, Williams K, Hanley AJ, Otvos JD, Goff DC, Wagenknecht LE, et al. Nuclear magnetic resonance lipoprotein

abnormalities in prediabetic subjects in the Insulin Resistance Atherosclerosis Study. Circulation. 2005;111(25):3465-72.

118.Verges B. Pathophysiology of diabetic dyslipidaemia: where are we? Diabetologia. 2015;58(5):886-99.

119.Chait A, Bornfeldt KE. Diabetes and atherosclerosis: is there a role for hyperglycemia? J Lipid Res. 2009;50 Suppl:S335-9.

120.Pereira SS, Alvarez-Leite JI. Low-Grade Inflammation, Obesity, and Diabetes. Curr Obes Rep. 2014;3(4):422-31.

121.Dokken BB. The pathophysiology of cardiovascular disease and diabetes: beyond blood pressure and lipids. Diabetes Spectrum. 2008;21:160-5.

122.Matheus AS, Tannus LR, Cobas RA, Palma CC, Negrato CA, Gomes MB. Impact of diabetes on cardiovascular disease: an update. Int J Hypertens. 2013;2013:653789.

123.Csige I, Ujvarosy D, Szabo Z, Lorincz I, Paragh G, Harangi M, et al. The Impact of Obesity on the Cardiovascular System. J Diabetes Res. 2018;2018:3407306.

124.Mandviwala T, Khalid U, Deswal A. Obesity and Cardiovascular Disease: a Risk Factor or a Risk Marker? Curr Atheroscler Rep. 2016;18(5):21.

125.Janssen I, Katzmarzyk PT, Ross R. Waist circumference and not body mass index explains obesity-related health risk. Am J Clin Nutr. 2004;79(3):379-84.

126.Poirier P, Giles TD, Bray GA, Hong Y, Stern JS, Pi-Sunyer FX, et al. Obesity and cardiovascular disease: pathophysiology, evaluation, and effect of weight loss. Arterioscler Thromb Vasc Biol. 2006;26(5):968-76.

127.McGill HC, Jr., McMahan CA, Malcom GT, Oalmann MC, Strong JP. Relation of glycohemoglobin and adiposity to atherosclerosis in youth. Pathobiological Determinants of Atherosclerosis in Youth (PDAY) Research Group. Arterioscler Thromb Vasc Biol. 1995;15(4):431-40.

128.Shukla AP, Buniak WI, Aronne LJ. Treatment of obesity in 2015. J Cardiopulm Rehabil Prev. 2015;35(2):81-92.

129.Karunathilake SP, Ganegoda GU. Secondary Prevention of Cardiovascular Diseases and Application of Technology for Early Diagnosis. Biomed Res Int. 2018;2018:5767864.

130.Flora GD, Nayak MK. A Brief Review of Cardiovascular Diseases, Associated Risk Factors and Current Treatment Regimes. Curr Pharm Des. 2019;25(38):4063-84.

131.Cohen JC, Boerwinkle E, Mosley TH, Jr., Hobbs HH. Sequence variations in PCSK9, low LDL, and protection against coronary heart disease. N Engl J Med. 2006;354(12):1264-72.

132.Holmes MV, Asselbergs FW, Palmer TM, Drenos F, Lanktree MB, Nelson CP, et al. Mendelian randomization of blood lipids for coronary heart disease. Eur Heart J. 2015;36(9):539-50.

133.White J, Swerdlow DI, Preiss D, Fairhurst-Hunter Z, Keating BJ, Asselbergs FW, et al. Association of Lipid Fractions With Risks for Coronary Artery Disease and Diabetes. JAMA Cardiol. 2016;1(6):692-9.

134.Jorgensen AB, Frikke-Schmidt R, West AS, Grande P, Nordestgaard BG, Tybjaerg-Hansen A. Genetically elevated non-fasting triglycerides and calculated remnant cholesterol as causal risk factors for myocardial infarction. Eur Heart J. 2013;34(24):1826-33.

135.Varbo A, Benn M, Tybjaerg-Hansen A, Jorgensen AB, Frikke-Schmidt R, Nordestgaard BG. Remnant cholesterol as a causal risk factor for ischemic heart disease. J Am Coll Cardiol. 2013;61(4):427-36.

136.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(26):2518-28.

137.Kamstrup PR, Benn M, Tybjaerg-Hansen A, Nordestgaard BG. Extreme lipoprotein(a) levels and risk of myocardial infarction in the

general population: the Copenhagen City Heart Study. Circulation. 2008;117(2):176-84.

138.Haase CL, Tybjaerg-Hansen A, Grande P, Frikke-Schmidt R. Genetically elevated apolipoprotein A-I, high-density lipoprotein cholesterol levels, and risk of ischemic heart disease. J Clin Endocrinol Metab. 2010;95(12):E500-10.

139.Haase CL, Tybjaerg-Hansen A, Qayyum AA, Schou J, Nordestgaard BG, Frikke-Schmidt R. LCAT, HDL cholesterol and ischemic cardiovascular disease: a Mendelian randomization study of HDL cholesterol in 54,500 individuals. J Clin Endocrinol Metab. 2012;97(2):E248-56.

140.Ference BA, Robinson JG, Brook RD, Catapano AL, Chapman MJ, Neff DR, et al. Variation in PCSK9 and HMGCR and Risk of Cardiovascular Disease and Diabetes. N Engl J Med. 2016;375(22):2144-53.

141.Swerdlow DI, Preiss D, Kuchenbaecker KB, Holmes MV, Engmann JE, Shah T, et al. HMG-coenzyme A reductase inhibition, type 2 diabetes, and bodyweight: evidence from genetic analysis and randomised trials. Lancet. 2015;385(9965):351-61.

142.Benn M, Tybjaerg-Hansen A, Stender S, Frikke-Schmidt R, Nordestgaard BG. Low-density lipoprotein cholesterol and the risk of cancer: a mendelian randomization study. J Natl Cancer Inst. 2011;103(6):508-19.

143.Ference BA, Majeed F, Penumetcha R, Flack JM, Brook RD. Effect of naturally random allocation to lower low-density lipoprotein cholesterol on the risk of coronary heart disease mediated by polymorphisms in NPC1L1, HMGCR, or both: a 2 x 2 factorial Mendelian randomization study. J Am Coll Cardiol. 2015;65(15):1552-61.

144.Ference BA, Kastelein JJP, Ray KK, Ginsberg HN, Chapman MJ, Packard CJ, et al. Association of Triglyceride-Lowering LPL Variants and LDL-C-Lowering LDLR Variants With Risk of Coronary Heart Disease. JAMA. 2019;321(4):364-73.

145.Emerging Risk Factors C, Kaptoge S, Di Angelantonio E, Pennells L, Wood AM, White IR, et al. C-reactive protein, fibrinogen, and cardiovascular disease prediction. N Engl J Med. 2012;367(14):1310-20.

146.Zacho J, Tybjaerg-Hansen A, Jensen JS, Grande P, Sillesen H, Nordestgaard BG. Genetically elevated C-reactive protein and ischemic vascular disease. N Engl J Med. 2008;359(18):1897-908.

147.Collaboration CRPCHDG, Wensley F, Gao P, Burgess S, Kaptoge S, Di Angelantonio E, et al. Association between C reactive protein and coronary heart disease: mendelian randomisation analysis based on individual participant data. BMJ. 2011;342:d548.

148.Ridker PM, Everett BM, Thuren T, MacFadyen JG, Chang WH, Ballantyne C, et al. Antiinflammatory Therapy with Canakinumab for Atherosclerotic Disease. N Engl J Med. 2017;377(12):1119-31.

149.Levy D, Ehret GB, Rice K, Verwoert GC, Launer LJ, Dehghan A, et al. Genome-wide association study of blood pressure and hypertension. Nat Genet. 2009;41(6):677-87.

150.Lieb W, Jansen H, Loley C, Pencina MJ, Nelson CP, Newton-Cheh C, et al. Genetic predisposition to higher blood pressure increases coronary artery disease risk. Hypertension. 2013;61(5):995-1001.

151.Jansen H, Loley C, Lieb W, Pencina MJ, Nelson CP, Kathiresan S, et al. Genetic variants primarily associated with type 2 diabetes are related to coronary artery disease risk. Atherosclerosis. 2015;241(2):419-26.

152.Benn M, Tybjaerg-Hansen A, McCarthy MI, Jensen GB, Grande P, Nordestgaard BG. Nonfasting glucose, ischemic heart disease, and myocardial infarction: a Mendelian randomization study. J Am Coll Cardiol. 2012;59(25):2356-65.

153.Ross S, Gerstein HC, Eikelboom J, Anand SS, Yusuf S, Pare G. Mendelian randomization analysis supports the causal role of dysglycaemia and diabetes in the risk of coronary artery disease. Eur Heart J. 2015;36(23):1454-62.

154.Kaur-Knudsen D, Bojesen SE, Tybjaerg-Hansen A, Nordestgaard BG. Nicotinic acetylcholine receptor polymorphism, smoking behavior, and tobacco-related cancer and lung and cardiovascular diseases: a cohort study. J Clin Oncol. 2011;29(21):2875-82.

155.Holmes MV, Dale CE, Zuccolo L, Silverwood RJ, Guo Y, Ye Z, et al. Association between alcohol and cardiovascular disease: Mendelian randomisation analysis based on individual participant data. BMJ. 2014;349:g4164.

156.Brondum-Jacobsen P, Benn M, Afzal S, Nordestgaard BG. No evidence that genetically reduced 25-hydroxyvitamin D is associated with increased risk of ischaemic heart disease or myocardial infarction: a Mendelian randomization study. Int J Epidemiol. 2015;44(2):651-61.

157.Afzal S, Nordestgaard BG. Vitamin D, Hypertension, and Ischemic Stroke in 116 655 Individuals From the General Population: A Genetic Study. Hypertension. 2017.

158.Vimaleswaran KS, Cavadino A, Berry DJ, LifeLines Cohort Study i, Jorde R, Dieffenbach AK, et al. Association of vitamin D status with arterial blood pressure and hypertension risk: a mendelian randomisation study. Lancet Diabetes Endocrinol. 2014;2(9):719-29.

159.Afzal S, Brondum-Jacobsen P, Bojesen SE, Nordestgaard BG. Genetically low vitamin D concentrations and increased mortality: Mendelian randomisation analysis in three large cohorts. BMJ. 2014;349:g6330.

160.Brouilette S, Singh RK, Thompson JR, Goodall AH, Samani NJ. White cell telomere length and risk of premature myocardial infarction. Arterioscler Thromb Vasc Biol. 2003;23(5):842-6.

161.Scheller Madrid A, Rode L, Nordestgaard BG, Bojesen SE. Short Telomere Length and Ischemic Heart Disease: Observational and Genetic Studies in 290 022 Individuals. Clin Chem. 2016;62(8):1140-9.

162.Hostalek U. Global epidemiology of prediabetes - present and future perspectives. Clin Diabetes Endocrinol. 2019;5:5.

163.Kolb H, Martin S. Environmental/lifestyle factors in the pathogenesis and prevention of type 2 diabetes. BMC Med. 2017;15(1):131.

164.Bhupathiraju SN, Hu FB. Epidemiology of Obesity and Diabetes and Their Cardiovascular Complications. Circ Res. 2016;118(11):1723-35.

165.Khera AV, Chaffin M, Wade KH, Zahid S, Brancale J, Xia R, et al. Polygenic Prediction of Weight and Obesity Trajectories from Birth to Adulthood. Cell. 2019;177(3):587-96 e9.

166.Diabetes Prevention Program Research G. The Diabetes Prevention Program (DPP): description of lifestyle intervention. Diabetes Care. 2002;25(12):2165-71.

167.Lindstrom J, Peltonen M, Eriksson JG, Ilanne-Parikka P, Aunola S, Keinanen-Kiukaanniemi S, et al. Improved lifestyle and decreased diabetes risk over 13 years: long-term follow-up of the randomised Finnish Diabetes Prevention Study (DPS). Diabetologia. 2013;56(2):284-93.

168.Leslie WS, Ford I, Sattar N, Hollingsworth KG, Adamson A, Sniehotta FF, et al. The Diabetes Remission Clinical Trial (DiRECT): protocol for a cluster randomised trial. BMC Fam Pract. 2016;17:20.

169.Hallberg SJ, Gershuni VM, Hazbun TL, Athinarayanan SJ. Reversing Type 2 Diabetes: A Narrative Review of the Evidence. Nutrients. 2019;11(4).

170.Eckel RH, Kahn SE, Ferrannini E, Goldfine AB, Nathan DM, Schwartz MW, et al. Obesity and type 2 diabetes: what can be unified and what needs to be individualized? J Clin Endocrinol Metab. 2011;96(6):1654-63.

171.Prasad-Reddy L, Isaacs D. A clinical review of GLP-1 receptor agonists: efficacy and safety in diabetes and beyond. Drugs Context. 2015;4:212283.

172.Jeemon P, Harikrishnan S, Ganapathi S, Sivasankaran S, Binukumar B, Padmanabhan S, et al. Efficacy of a family-based cardiovascular risk reduction intervention in individuals with a family history of premature coronary heart disease in India (PROLIFIC): an open-label, single-centre, cluster randomised controlled trial. Lancet Glob Health. 2021;9(10):e1442-e50.

173.Scheuner MT. Genetic predisposition to coronary artery disease. Curr Opin Cardiol. 2001;16(4):251-60.

174.Lloyd-Jones DM, Nam BH, D'Agostino RB, Sr., Levy D, Murabito JM, Wang TJ, et al. Parental cardiovascular disease as a risk factor for cardiovascular disease in middle-aged adults: a prospective study of parents and offspring. JAMA. 2004;291(18):2204-11.

175.Ripatti S, Tikkanen E, Orho-Melander M, Havulinna AS, Silander K, Sharma A, et al. A multilocus genetic risk score for coronary heart disease: case-control and prospective cohort analyses. Lancet. 2010;376(9750):1393-400.

176.Tikkanen E, Havulinna AS, Palotie A, Salomaa V, Ripatti S. Genetic risk prediction and a 2-stage risk screening strategy for coronary heart disease. Arterioscler Thromb Vasc Biol. 2013;33(9):2261-6.

177.Abraham G, Havulinna AS, Bhalala OG, Byars SG, De Livera AM, Yetukuri L, et al. Genomic prediction of coronary heart disease. Eur Heart J. 2016;37(43):3267-78.

178.Chatterjee N, Wheeler B, Sampson J, Hartge P, Chanock SJ, Park JH. Projecting the performance of risk prediction based on polygenic analyses of genome-wide association studies. Nat Genet. 2013;45(4):400-5, 5e1-3.

179.Do CB, Hinds DA, Francke U, Eriksson N. Comparison of family history and SNPs for predicting risk of complex disease. PLoS Genet. 2012;8(10):e1002973.

180.Inouye M, Abraham G, Nelson CP, Wood AM, Sweeting MJ, Dudbridge F, et al. Genomic Risk Prediction of Coronary Artery Disease in 480,000 Adults: Implications for Primary Prevention. J Am Coll Cardiol. 2018;72(16):1883-93.

181.Liu DJ, Peloso GM, Yu H, Butterworth AS, Wang X, Mahajan A, et al. Exome-wide association study of plasma lipids in >300,000 individuals. Nat Genet. 2017;49(12):1758-66.

182.Verweij N, Eppinga RN, Hagemeijer Y, van der Harst P. Identification of 15 novel risk loci for coronary artery disease and genetic risk of recurrent events, atrial fibrillation and heart failure. Sci Rep. 2017;7(1):2761.

183.Sharif M, Al Rifai M, Stone NJ, Blumenthal RS, Shapiro MD, Patel J. Power of the pedigree: the family history variable for ASCVD risk stratification. J Am Coll Cardiol. 2021.

184.Patel J, Al Rifai M, Scheuner MT, Shea S, Blumenthal RS, Nasir K, et al. Basic vs More Complex Definitions of Family History in the Prediction of Coronary Heart Disease: The Multi-Ethnic Study of Atherosclerosis. Mayo Clin Proc. 2018;93(9):1213-23.

185.Wolford BN, Surakka I, Graham SE, Nielsen JB, Zhou W, Gabrielsen ME, et al. Utility of family history in disease prediction in the era of polygenic scores. MedRxiv. 2021.

186.Emerging Risk Factors C, Erqou S, Kaptoge S, Perry PL, Di Angelantonio E, Thompson A, et al. Lipoprotein(a) concentration and the risk of coronary heart disease, stroke, and nonvascular mortality. JAMA. 2009;302(4):412-23.

187.Kamstrup PR, Tybjaerg-Hansen A, Steffensen R, Nordestgaard BG. Genetically elevated lipoprotein(a) and increased risk of myocardial infarction. JAMA. 2009;301(22):2331-9.

188.Nordestgaard BG, Langsted A. Lipoprotein (a) as a cause of cardiovascular disease: insights from epidemiology, genetics, and biology. J Lipid Res. 2016;57(11):1953-75.

189.Boerwinkle E, Leffert CC, Lin J, Lackner C, Chiesa G, Hobbs HH. Apolipoprotein(a) gene accounts for greater than 90% of the variation in plasma lipoprotein(a) concentrations. J Clin Invest. 1992;90(1):52-60.

190.Zekavat SM, Ruotsalainen S, Handsaker RE, Alver M, Bloom J, Poterba T, et al. Deep coverage whole genome sequences and

plasma lipoprotein(a) in individuals of European and African ancestries. Nat Commun. 2018;9(1):2606.

191.Burgess S, Ference BA, Staley JR, Freitag DF, Mason AM, Nielsen SF, et al. Association of LPA Variants With Risk of Coronary Disease and the Implications for Lipoprotein(a)-Lowering Therapies: A Mendelian Randomization Analysis. JAMA Cardiol. 2018;3(7):619-27.

192.Verbeek R, Boekholdt SM, Stoekenbroek RM, Hovingh GK, Witztum JL, Wareham NJ, et al. Population and assay thresholds for the predictive value of lipoprotein (a) for coronary artery disease: the EPIC-Norfolk Prospective Population Study. J Lipid Res. 2016;57(4):697-705.

193.Trinder M, Uddin MM, Finneran P, Aragam KG, Natarajan P. Clinical Utility of Lipoprotein(a) and LPA Genetic Risk Score in Risk Prediction of Incident Atherosclerotic Cardiovascular Disease. JAMA Cardiol. 2020.

194.Jenner JL, Ordovas JM, Lamon-Fava S, Schaefer MM, Wilson PW, Castelli WP, et al. Effects of age, sex, and menopausal status on plasma lipoprotein(a) levels. The Framingham Offspring Study. Circulation. 1993;87(4):1135-41.

195.Langsted A, Kamstrup PR, Benn M, Tybjaerg-Hansen A, Nordestgaard BG. High lipoprotein(a) as a possible cause of clinical familial hypercholesterolaemia: a prospective cohort study. Lancet Diabetes Endocrinol. 2016;4(7):577-87.

196.Trinder M, DeCastro ML, Azizi H, Cermakova L, Jackson LM, Frohlich J, et al. Ascertainment Bias in the Association Between Elevated Lipoprotein(a) and Familial Hypercholesterolemia. J Am Coll Cardiol. 2020;75(21):2682-93.

197.Mach F, Baigent C, Catapano AL, Koskinas KC, Casula M, Badimon L, et al. 2019 ESC/EAS Guidelines for the management of dyslipidaemias: lipid modification to reduce cardiovascular risk. Eur Heart J. 2020;41(1):111-88.

198.Kamstrup PR, Tybjaerg-Hansen A, Nordestgaard BG. Extreme lipoprotein(a) levels and improved cardiovascular risk prediction. J Am Coll Cardiol. 2013;61(11):1146-56.

199.Suk Danik J, Rifai N, Buring JE, Ridker PM. Lipoprotein(a), measured with an assay independent of apolipoprotein(a) isoform size, and risk of future cardiovascular events among initially healthy women. JAMA. 2006;296(11):1363-70.

200.Labudovic D, Kostovska I, Tosheska Trajkovska K, Cekovska S, Brezovska Kavrakova J, Topuzovska S. Lipoprotein(a) - Link between Atherogenesis and Thrombosis. Prague Med Rep. 2019;120(2-3):39-51.

201.Malaguarnera M, Vacante M, Russo C, Malaguarnera G, Antic T, Malaguarnera L, et al. Lipoprotein(a) in cardiovascular diseases. Biomed Res Int. 2013;2013:650989.

202.Hajjar KA, Nachman RL. The role of lipoprotein(a) in atherogenesis and thrombosis. Annu Rev Med. 1996;47:423-42.

203.Sofi F, Marcucci R, Abbate R, Gensini GF, Prisco D. Lipoprotein (a) and venous thromboembolism in adults: a meta-analysis. Am J Med. 2007;120(8):728-33.

204.Helgadottir A, Gretarsdottir S, Thorleifsson G, Holm H, Patel RS, Gudnason T, et al. Apolipoprotein(a) genetic sequence variants associated with systemic atherosclerosis and coronary atherosclerotic burden but not with venous thromboembolism. J Am Coll Cardiol. 2012;60(8):722-9.

205.Boffa MB, Koschinsky ML. Lipoprotein (a): truly a direct prothrombotic factor in cardiovascular disease? J Lipid Res. 2016;57(5):745-57.

206.Dentali F, Gessi V, Marcucci R, Gianni M, Grandi AM, Franchini M. Lipoprotein(a) as a Risk Factor for Venous Thromboembolism: A Systematic Review and Meta-analysis of the Literature. Semin Thromb Hemost. 2017;43(6):614-20.

207.Kamstrup PR, Tybjaerg-Hansen A, Nordestgaard BG. Genetic evidence that lipoprotein(a) associates with atherosclerotic stenosis

rather than venous thrombosis. Arterioscler Thromb Vasc Biol. 2012;32(7):1732-41.

208.Rodger MA, Le Gal G, Carrier M, Betancourt MT, Kahn SR, Wells PS, et al. Serum lipoprotein (a) levels in patients with first unprovoked venous thromboembolism is not associated with subsequent risk of recurrent VTE. Thromb Res. 2010;126(3):222-6.

209.Marcucci R, Liotta AA, Cellai AP, Rogolino A, Gori AM, Giusti B, et al. Increased plasma levels of lipoprotein(a) and the risk of idiopathic and recurrent venous thromboembolism. Am J Med. 2003;115(8):601-5.

210.Sticchi E, Magi A, Kamstrup PR, Marcucci R, Prisco D, Martinelli I, et al. Apolipoprotein(a) Kringle-IV Type 2 Copy Number Variation Is Associated with Venous Thromboembolism. PLoS One. 2016;11(2):e0149427.

211.Danik JS, Buring JE, Chasman DI, Zee RY, Ridker PM, Glynn RJ. Lipoprotein(a), polymorphisms in the LPA gene, and incident venous thromboembolism among 21483 women. J Thromb Haemost. 2013;11(1):205-8.

212.Tsimikas S. A Test in Context: Lipoprotein(a): Diagnosis, Prognosis, Controversies, and Emerging Therapies. J Am Coll Cardiol. 2017;69(6):692-711.

213.Group HTC, Landray MJ, Haynes R, Hopewell JC, Parish S, Aung T, et al. Effects of extended-release niacin with laropiprant in high-risk patients. N Engl J Med. 2014;371(3):203-12.

214.Sabatine MS, Giugliano RP, Keech AC, Honarpour N, Wiviott SD, Murphy SA, et al. Evolocumab and Clinical Outcomes in Patients with Cardiovascular Disease. N Engl J Med. 2017;376(18):1713-22.

215.Albers JJ, Slee A, O'Brien KD, Robinson JG, Kashyap ML, Kwiterovich PO, Jr., et al. Relationship of apolipoproteins A-1 and B, and lipoprotein(a) to cardiovascular outcomes: the AIM-HIGH trial (Atherothrombosis Intervention in Metabolic Syndrome with Low HDL/High Triglyceride and Impact on Global Health Outcomes). J Am Coll Cardiol. 2013;62(17):1575-9.

216.Khera AV, Everett BM, Caulfield MP, Hantash FM, Wohlgemuth J, Ridker PM, et al. Lipoprotein(a) concentrations, rosuvastatin therapy, and residual vascular risk: an analysis from the JUPITER Trial (Justification for the Use of Statins in Prevention: an Intervention Trial Evaluating Rosuvastatin). Circulation. 2014;129(6):635-42.

217.Nestel PJ, Barnes EH, Tonkin AM, Simes J, Fournier M, White HD, et al. Plasma lipoprotein(a) concentration predicts future coronary and cardiovascular events in patients with stable coronary heart disease. Arterioscler Thromb Vasc Biol. 2013;33(12):2902-8.

218.Cannon CP, Blazing MA, Giugliano RP, McCagg A, White JA, Theroux P, et al. Ezetimibe Added to Statin Therapy after Acute Coronary Syndromes. N Engl J Med. 2015;372(25):2387-97.

219.Holmes MV, Richardson TG, Ference BA, Davies NM, Davey Smith G. Integrating genomics with biomarkers and therapeutic targets to invigorate cardiovascular drug development. Nat Rev Cardiol. 2021;18(6):435-53.

220.Beierlein JM, McNamee LM, Walsh MJ, Kaitin KI, DiMasi JA, Ledley FD. Landscape of Innovation for Cardiovascular Pharmaceuticals: From Basic Science to New Molecular Entities. Clin Ther. 2017;39(7):1409-25 e20.

221.Harrison RK. Phase II and phase III failures: 2013-2015. Nat Rev Drug Discov. 2016;15(12):817-8.

222.Fordyce CB, Roe MT, Ahmad T, Libby P, Borer JS, Hiatt WR, et al. Cardiovascular drug development: is it dead or just hibernating? J Am Coll Cardiol. 2015;65(15):1567-82.

223.Nelson MR, Tipney H, Painter JL, Shen J, Nicoletti P, Shen Y, et al. The support of human genetic evidence for approved drug indications. Nat Genet. 2015;47(8):856-60.

224.Ference BA. How to use Mendelian randomization to anticipate the results of randomized trials. Eur Heart J. 2018;39(5):360-2.

225.Baigent C, Keech A, Kearney PM, Blackwell L, Buck G, Pollicino C, et al. Efficacy and safety of cholesterol-lowering treatment:

prospective meta-analysis of data from 90,056 participants in 14 randomised trials of statins. Lancet. 2005;366(9493):1267-78.

226.Collins R, Reith C, Emberson J, Armitage J, Baigent C, Blackwell L, et al. Interpretation of the evidence for the efficacy and safety of statin therapy. Lancet. 2016;388(10059):2532-61.

227.Lotta LA, Sharp SJ, Burgess S, Perry JRB, Stewart ID, Willems SM, et al. Association Between Low-Density Lipoprotein Cholesterol-Lowering Genetic Variants and Risk of Type 2 Diabetes: A Meta-analysis. JAMA. 2016;316(13):1383-91.

228.Schmidt AF, Swerdlow DI, Holmes MV, Patel RS, Fairhurst-Hunter Z, Lyall DM, et al. PCSK9 genetic variants and risk of type 2 diabetes: a mendelian randomisation study. Lancet Diabetes Endocrinol. 2017;5(2):97-105.

229.Tardif JC, Rheaume E, Lemieux Perreault LP, Gregoire JC, Feroz Zada Y, Asselin G, et al. Pharmacogenomic determinants of the cardiovascular effects of dalcetrapib. Circ Cardiovasc Genet. 2015;8(2):372-82.

230.Walker VM, Davey Smith G, Davies NM, Martin RM. Mendelian randomization: a novel approach for the prediction of adverse drug events and drug repurposing opportunities. Int J Epidemiol. 2017;46(6):2078-89.

231.Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank resource with deep phenotyping and genomic data. Nature. 2018;562(7726):203-9.

# APPENDIX

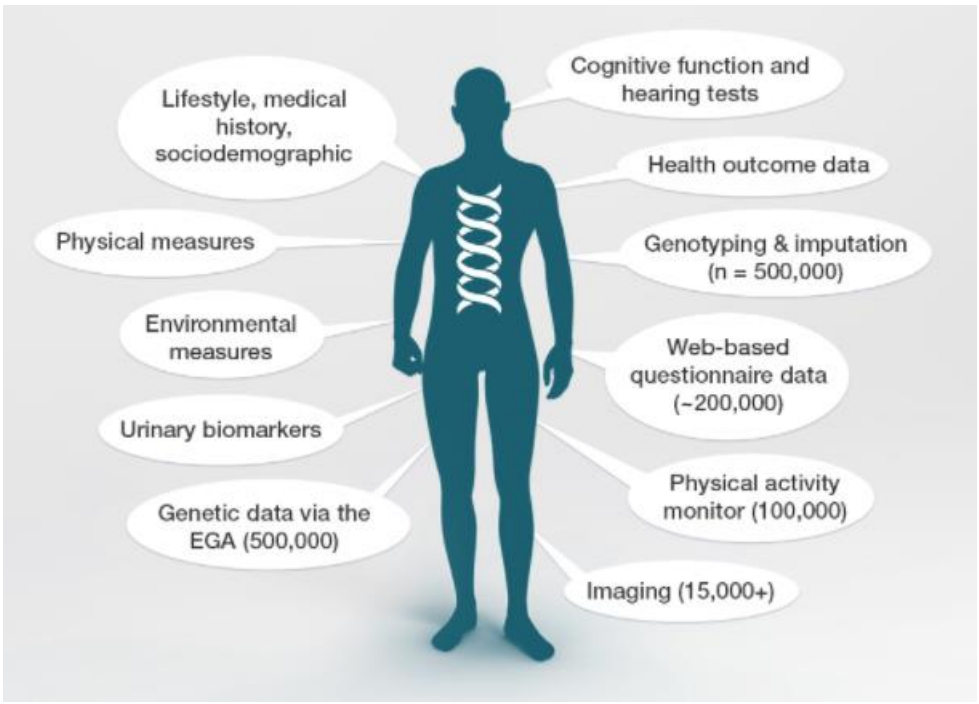
**Appendix I.** The UK Biobank resource

The UK Biobank is a prospective cohort study with deep genetic, physical and health data collected on ~500,000 volunteer participants aged 40-69 years (with a roughly even number of men and women) recruited between 2006 and 2010 throughout the UK and followed for at least 10 years thereafter.

Prospective participants were invited to visit an assessment centre, at which they completed an automated questionnaire and were interviewed about lifestyle, medical history and nutritional habits; basic variables such weight, height, blood pressure etc. were measured; and blood and urine samples were taken. These samples were preserved so that it was possible to later extract DNA and measure other biologically important substances. During the whole duration of the study, it was intended that all disease events, drug prescriptions and deaths of the participants are recorded in a database, taking advantage of the centralized UK National Health Service.

The UK Biobank database is regularly augmented with additional data and is globally accessible to approved researchers undertaking vital research into the most common and life-threatening diseases. It is a major contributor to the advancement of modern medicine and treatment and has enabled several scientific discoveries that improve human health.

The UK Biobank is a powerful example of the immense value that can be achieved from large population scale studies that combine genetics with extensive and deep phenotyping and linkage to health records coupled with a strong data sharing policy, to drive and enhance understanding of human biology and disease. Deep information can be found elsewhere (231).




Data on UK Biobank participants.

**Appendix II.** Variants included in the body mass index (BMI) instrumental variable genetic score.

| SNP | beta | se | effect allele | eaf |
|---|---|---|---|---|
| rs1000940 | 0.0192 | 0.0034 | G | 0.225 |
| rs10132280 | -0.023 | 0.0034 | A | 0.3333 |
| rs1016287 | 0.0229 | 0.0034 | T | 0.325 |
| rs10182181 | -0.0307 | 0.0031 | A | 0.5 |
| rs10733682 | 0.0174 | 0.0031 | A | 0.425 |
| rs10938397 | -0.0402 | 0.0031 | A | 0.5667 |
| rs10968576 | 0.0249 | 0.0033 | G | 0.2917 |
| rs11030104 | 0.0414 | 0.0038 | A | 0.8 |
| rs11057405 | -0.0307 | 0.0055 | A | 0.0917 |
| rs11126666 | -0.0207 | 0.0034 | G | 0.6917 |
| rs11165643 | -0.0218 | 0.0031 | C | 0.425 |
| rs11191560 | -0.0308 | 0.0053 | T | 0.9417 |
| rs11583200 | 0.0177 | 0.0031 | C | 0.375 |
| rs1167827 | -0.0202 | 0.0033 | A | 0.4583 |
| rs11688816 | -0.0172 | 0.0031 | A | 0.5417 |
| rs11727676 | -0.0358 | 0.0064 | C | 0.075 |
| rs11847697 | 0.0492 | 0.0084 | T | 0.0417 |
| rs9581854 | -0.0298 | 0.0047 | C | 0.7667 |
| rs12286929 | 0.0217 | 0.0031 | G | 0.4333 |

| | | | | |
|---|---|---|---|---|
| rs12401738 | 0.0211 | 0.0033 | A | 0.425 |
| rs12429545 | -0.0334 | 0.0047 | G | 0.9 |
| rs12446632 | -0.0403 | 0.0046 | A | 0.1333 |
| rs12566985 | 0.0242 | 0.0031 | G | 0.425 |
| rs12885454 | 0.0207 | 0.0033 | C | 0.6333 |
| rs12940622 | -0.0182 | 0.0031 | A | 0.4583 |
| rs13021737 | -0.0601 | 0.004 | A | 0.125 |
| rs13078960 | -0.0297 | 0.0039 | T | 0.8167 |
| rs13107325 | -0.0477 | 0.0068 | C | 0.8833 |
| rs13191362 | 0.0277 | 0.0048 | A | 0.8 |
| rs13201877 | -0.0233 | 0.0045 | A | 0.9167 |
| rs1441264 | 0.0175 | 0.0032 | A | 0.55 |
| rs1460676 | -0.0197 | 0.004 | T | 0.7833 |
| rs1516725 | -0.0451 | 0.0046 | T | 0.0917 |
| rs1528435 | 0.0178 | 0.0031 | T | 0.5833 |
| rs1558902 | 0.0818 | 0.0031 | A | 0.45 |
| rs16851483 | -0.0483 | 0.0077 | G | 0.9083 |
| rs16907751 | 0.035 | 0.0066 | C | 0.9583 |
| rs16951275 | -0.0311 | 0.0037 | C | 0.225 |
| rs17001654 | -0.0306 | 0.0053 | C | 0.8417 |
| rs17024393 | 0.0658 | 0.0088 | C | 0.04167 |
| rs17094222 | 0.0249 | 0.0038 | C | 0.2083 |

| | | | | |
|---|---|---|---|---|
| rs17203016 | 0.021 | 0.0039 | G | 0.2 |
| rs17405819 | -0.0224 | 0.0033 | C | 0.3667 |
| rs17724992 | 0.0194 | 0.0035 | A | 0.6917 |
| rs1808579 | -0.0167 | 0.0031 | T | 0.475 |
| rs1928295 | -0.0188 | 0.0031 | C | 0.425 |
| rs2033529 | 0.019 | 0.0033 | G | 0.2583 |
| rs2033732 | 0.0192 | 0.0035 | C | 0.7583 |
| rs205262 | -0.0221 | 0.0035 | A | 0.7333 |
| rs2075650 | 0.0258 | 0.0045 | A | 0.8583 |
| rs2080454 | -0.0168 | 0.0031 | A | 0.6083 |
| rs2112347 | -0.0261 | 0.0031 | G | 0.375 |
| rs2121279 | 0.0245 | 0.0044 | T | 0.1167 |
| rs2176040 | -0.0141 | 0.0031 | G | 0.6083 |
| rs2176598 | 0.0198 | 0.0036 | T | 0.2 |
| rs2207139 | 0.0447 | 0.004 | G | 0.1 |
| rs2245368 | -0.0317 | 0.0057 | T | 0.7583 |
| rs2287019 | 0.036 | 0.0042 | C | 0.85 |
| rs2365389 | 0.02 | 0.0031 | C | 0.6583 |
| rs2650492 | 0.0207 | 0.0035 | A | 0.3083 |
| rs2820292 | -0.0195 | 0.0031 | A | 0.4917 |
| rs2836754 | 0.0164 | 0.0032 | C | 0.65 |
| rs29941 | -0.0182 | 0.0033 | A | 0.3333 |

| | | | | |
|---|---|---|---|---|
| rs3101336 | -0.0334 | 0.0031 | T | 0.3509 |
| rs3736485 | 0.0176 | 0.0031 | A | 0.425 |
| rs3810291 | 0.0283 | 0.0036 | A | 0.625 |
| rs3817334 | -0.0262 | 0.0031 | C | 0.55 |
| rs3849570 | 0.0188 | 0.0034 | A | 0.3667 |
| rs3888190 | 0.0309 | 0.0031 | A | 0.3583 |
| rs4256980 | 0.0209 | 0.0031 | G | 0.725 |
| rs4740619 | 0.0179 | 0.0031 | T | 0.5333 |
| rs4787491 | -0.0159 | 0.0034 | A | 0.386 |
| rs492400 | -0.0158 | 0.0031 | T | 0.675 |
| rs543874 | 0.0482 | 0.0039 | G | 0.2667 |
| rs6091540 | 0.0188 | 0.0035 | C | 0.725 |
| rs6465468 | -0.0166 | 0.0035 | G | 0.675 |
| rs6477694 | 0.0174 | 0.0031 | C | 0.3583 |
| rs6567160 | 0.0556 | 0.0036 | C | 0.2833 |
| rs657452 | 0.0227 | 0.0031 | A | 0.4167 |
| rs6804842 | -0.0185 | 0.0031 | A | 0.425 |
| rs7138803 | -0.0315 | 0.0031 | G | 0.5583 |
| rs7141420 | 0.0235 | 0.0031 | T | 0.6167 |
| rs7164727 | 0.018 | 0.0033 | T | 0.775 |
| rs7239883 | 0.0164 | 0.0031 | G | 0.3167 |
| rs7243357 | -0.0217 | 0.004 | G | 0.1333 |

| | | | | |
|---|---|---|---|---|
| rs758747 | -0.0225 | 0.0037 | C | 0.7333 |
| rs7599312 | 0.022 | 0.0034 | G | 0.7083 |
| rs7715256 | 0.0163 | 0.0031 | G | 0.45 |
| rs7899106 | -0.0395 | 0.0071 | A | 0.95 |
| rs9374842 | 0.0187 | 0.0035 | T | 0.7417 |
| rs9400239 | 0.0188 | 0.0033 | C | 0.7 |
| rs9540493 | -0.0172 | 0.0033 | G | 0.55 |
| rs9641123 | -0.0191 | 0.0038 | G | 0.6083 |
| rs977747 | 0.0167 | 0.0031 | T | 0.4667 |
| rs9914578 | 0.0201 | 0.0038 | G | 0.1667 |
| rs9925964 | -0.0192 | 0.0031 | G | 0.3917 |

*SNP: single nucleotide polymorphism; eaf: expected allele frequency.

**Appendix III.** Variants included in the LDL cholesterol (LDL-C) instrumental variable genetic score. Effect allele and frequency referred to the allele that is associated with higher LDL-C.

| SNP | beta | se | effect allele | eaf |
|---|---|---|---|---|
| rs646776 | 5.002 | 0.083 | T | 0.77 |
| rs267733 | 0.75 | 0.094 | A | 0.84 |
| rs20558 | 0.43 | 0.07 | C | 0.57 |
| rs2738755 | 0.574 | 0.073 | C | 0.67 |
| rs10903129 | 1.099 | 0.069 | G | 0.56 |
| rs12748152 | 1.021 | 0.128 | T | 0.08 |
| rs11206510 | 0.664 | 0.094 | T | 0.81 |
| rs2479409 | 0.992 | 0.075 | G | 0.35 |
| rs11591147 | 13.572 | 0.263 | G | 0.98 |
| rs505151 | 2.792 | 0.197 | G | 0.03 |
| rs10889353 | 1.81 | 0.072 | A | 0.65 |
| rs17526895 | 1.478 | 0.125 | A | 0.92 |
| rs1801702 | 2.408 | 0.25 | C | 0.98 |
| rs533617 | 4.979 | 0.179 | T | 0.96 |
| rs1367117 | 2.604 | 0.077 | A | 0.33 |
| rs541041 | 3.748 | 0.093 | A | 0.82 |
| rs887829 | 0.289 | 0.074 | C | 0.68 |
| rs1260326 | 1.467 | 0.071 | T | 0.39 |

| | | | | |
|---|---|---|---|---|
| rs11556157 | 0.605 | 0.081 | T | 0.24 |
| rs4245791 | 2.449 | 0.074 | C | 0.32 |
| rs11125936 | 1.073 | 0.119 | T | 0.91 |
| rs7640978 | 1.197 | 0.121 | C | 0.91 |
| rs2251219 | 0.313 | 0.071 | C | 0.39 |
| rs3816873 | 0.444 | 0.079 | T | 0.74 |
| rs976002 | 0.839 | 0.08 | G | 0.24 |
| rs4530754 | 0.684 | 0.069 | A | 0.54 |
| rs1016988 | 0.714 | 0.088 | T | 0.81 |
| rs6882076 | 1.523 | 0.072 | C | 0.63 |
| rs351855 | 0.214 | 0.076 | G | 0.7 |
| rs12654264 | 2.593 | 0.072 | T | 0.37 |
| rs1999930 | 0.529 | 0.075 | C | 0.7 |
| rs12208357 | 2.288 | 0.124 | T | 0.08 |
| rs1564348 | 1.896 | 0.093 | C | 0.17 |
| rs3798220 | 5.332 | 0.265 | C | 0.02 |
| rs7770628 | 1.165 | 0.07 | C | 0.47 |
| rs9370867 | 0.984 | 0.069 | A | 0.54 |
| rs1800562 | 2.246 | 0.129 | G | 0.92 |
| rs1051794 | 0.514 | 0.081 | A | 0.26 |
| rs13192471 | 1.462 | 0.102 | C | 0.14 |
| rs2239619 | 0.352 | 0.071 | A | 0.61 |

| | | | | |
|---|---|---|---|---|
| rs12670798 | 1.295 | 0.08 | C | 0.25 |
| rs4722551 | 1.068 | 0.095 | C | 0.16 |
| rs11550029 | 0.758 | 0.088 | A | 0.19 |
| rs2737229 | 0.824 | 0.075 | A | 0.7 |
| rs2954029 | 2.433 | 0.069 | A | 0.53 |
| rs11136343 | 0.823 | 0.071 | G | 0.38 |
| rs4921914 | 1.091 | 0.084 | C | 0.22 |
| rs10102164 | 0.938 | 0.085 | A | 0.21 |
| rs2081687 | 1.247 | 0.073 | T | 0.34 |
| rs4841132 | 1.883 | 0.12 | G | 0.91 |
| rs1883025 | 1.096 | 0.079 | C | 0.74 |
| rs635634 | 2.495 | 0.09 | T | 0.18 |
| rs3812594 | 0.562 | 0.077 | G | 0.73 |
| rs67710536 | 0.86 | 0.139 | C | 0.07 |
| rs3780181 | 1.202 | 0.138 | A | 0.93 |
| rs2255141 | 0.729 | 0.077 | A | 0.28 |
| rs1891110 | 0.662 | 0.07 | A | 0.55 |
| rs1935 | 0.328 | 0.069 | G | 0.47 |
| rs2068888 | 0.96 | 0.069 | G | 0.55 |
| rs964184 | 3.208 | 0.103 | G | 0.13 |
| rs10128711 | 0.783 | 0.079 | C | 0.74 |
| rs174550 | 1.558 | 0.073 | T | 0.65 |

| | | | | |
|---|---|---|---|---|
| rs3816492 | 0.656 | 0.081 | C | 0.76 |
| rs11057830 | 0.505 | 0.082 | A | 0.18 |
| rs4942486 | 0.813 | 0.069 | T | 0.48 |
| rs8017377 | 0.801 | 0.069 | A | 0.47 |
| rs9646133 | 0.56 | 0.075 | G | 0.69 |
| rs13379043 | 0.653 | 0.078 | T | 0.72 |
| rs28929474 | 0.657 | 0.247 | T | 0.02 |
| rs173539 | 1.232 | 0.076 | C | 0.67 |
| rs7499892 | 0.856 | 0.094 | T | 0.18 |
| rs2000999 | 2.176 | 0.089 | A | 0.19 |
| rs704 | 0.533 | 0.069 | A | 0.47 |
| rs11080150 | 0.494 | 0.075 | A | 0.7 |
| rs11871606 | 1.077 | 0.069 | C | 0.5 |
| rs1801689 | 2.742 | 0.202 | C | 0.03 |
| rs77542162 | 5.829 | 0.237 | G | 0.02 |
| rs314253 | 0.629 | 0.072 | T | 0.65 |
| rs2125345 | 0.561 | 0.076 | T | 0.7 |
| rs4129767 | 0.592 | 0.069 | A | 0.49 |
| rs77960347 | 2.829 | 0.303 | G | 0.01 |
| rs7241918 | 0.666 | 0.091 | T | 0.82 |
| rs6511720 | 7.171 | 0.107 | G | 0.88 |
| rs11669576 | 0.904 | 0.152 | A | 0.05 |

| | | | | |
|---|---|---|---|---|
| rs7188 | 1.266 | 0.075 | C | 0.33 |
| rs58542926 | 4.442 | 0.13 | C | 0.92 |
| rs28399654 | 0.704 | 0.204 | G | 0.97 |
| rs157580 | 1.357 | 0.075 | A | 0.6 |
| rs769449 | 5.598 | 0.111 | A | 0.12 |
| rs7412 | 16.891 | 0.134 | C | 0.92 |
| rs492602 | 1.423 | 0.069 | G | 0.51 |
| rs364585 | 0.537 | 0.071 | G | 0.61 |
| rs1132274 | 0.453 | 0.096 | A | 0.15 |
| rs7261862 | 0.912 | 0.096 | T | 0.85 |
| rs6016373 | 1.141 | 0.072 | A | 0.59 |
| rs6029526 | 1.266 | 0.069 | A | 0.48 |
| rs1800961 | 2.22 | 0.199 | C | 0.97 |
| rs2076674 | 0.504 | 0.073 | C | 0.34 |
| rs738409 | 0.336 | 0.084 | C | 0.78 |
| rs13268 | 1.271 | 0.219 | A | 0.97 |

*SNP: single nucleotide polymorphism; eaf: expected allele frequency.

**Appendix IV.** Variants included in the LPA instrumental variable genetic score. Effect allele and frequency referred to the allele that is associated with higher Lp(a).

| SNP | Gene | Position | Effect allele | eaf |
|---|---|---|---|---|
| rs10455872 | LPA | 6:161010118 | G | 0.0736 |
| rs3798220 | LPA | 6:160961137 | C | 0.0099 |

*SNP: single nucleotide polymorphism; eaf: effect allele frequency.

**Appendix V.** Baseline characteristics, measured at the time of enrolment in the UK Biobank, by rs10455872 and rs3798220 variants.

| | **rs10455872** | | |
|---|---|---|---|
| **Characteristics** | **0 n=373,129** | **1 n=64,487** | **2 n=2,761** |
| Age, y (SD) | 57.2 (8) | 57.3 (8) | 57.3 (7.9) |
| Female sex (%) | 54.2 | 54.2 | 54.6 |
| LDL-C, mg/dL (SD) | 137.5 (33.5) | 140.4 (34.1) | 142.4 (34.8) |
| ApoB, mg/L (SD) | 103.1 (23.8) | 105.2 (23.9) | 106.6 (24.1) |
| TG, mg/dL (IQR) | 132.4 [93.4-191.5] | 129 [91.5-187.1] | 125.1 [88.3-181.9] |
| HDL-C, mg/dL (SD) | 56.1 (14.8) | 56.4 (14.9) | 56.6 (15.7) |
| Lp(a), nmol/L (IQR) | 14.1 [6.3-38.6] | 141 [105.6-183] | 250.3 [190.4-312.3] |
| CRP, mg/L (IQR) | 1.3 [0.7-2.8] | 1.3 [0.7-2.8] | 1.3 [0.7-2.7] |
| SBP, mmHg (SD) | 137.8 (18.6) | 137.8 (18.6) | 137.7 (18.6) |
| BMI, $kg/m^2$ (SD) | 27.4 (4.8) | 27.4 (4.8) | 27.4 (4.9) |
| | **rs3798220** | | |
| **Characteristics** | **0 n=424,377** | **1 n=15,849** | **2 n=151** |
| Age, y (SD) | 57.3 (8) | 57.2 (8) | 57.4 (7.4) |
| Female sex (%) | 54.2 | 54.6 | 45.0 |
| LDL-C, mg/dL (SD) | 137.9 (33.6) | 140.7 (34.5) | 146.3 (36.8) |
| ApoB, mg/L (SD) | 103.3 (23.8) | 105.4 (24.2) | 109.5 (25.1) |
| TG, mg/dL (IQR) | 132.4 [93.4-191.5] | 129 [91.5-187.1] | 125.1 [88.3-181.9] |
| HDL-C, mg/dL (SD) | 56.2 (14.8) | 56.4 (14.9) | 54.6 (13.1) |
| Lp(a), nmol/L (IQR) | 14.1 [6.3-38.6] | 141 [105.6-183] | 250.3 [190.4-312.3] |

| | | | |
|---|---|---|---|
| CRP, mg/L (IQR) | 1.3 [0.7-2.8] | 1.3 [0.7-2.8] | 1.3 [0.7-2.7] |
| SBP, mmHg (SD) | 137.8 (18.6) | 138.1 (18.7) | 138.5 (16.6) |
| BMI, $kg/m^2$ (SD) | 27.4 (4.8) | 27.4 (4.8) | 27.4 (4.5) |

Characteristics are presented as means and standard deviations (SD) for normally distributed variables, median and interquartile ranges (IQR) for non-normally distributed variables (triglycerides, Lp(a), and CRP), or percentages for dichotomous variables. LDL-C: low-density lipoprotein cholesterol; ApoB: apolipoprotein B; TG: triglycerides; HDL-C: high-density lipoprotein cholesterol; Lp(a): lipoprotein(a); SBP: systolic blood pressure; CRP: c-reactive protein; BMI: body mass index.

**Appendix VI.** Variants included in the antiplatelet instrumental variable genetic score (GUCY1A3 score, panel A), and in the anticoagulant/antithrombotic instrumental variable genetic score (Factor II + Factor V score, panel B). In the first case, effect allele and frequency referred to the allele that is associated with lower risk of major coronary events. In the second one, referred to the allele that is associated with lower risk of deep vein thrombosis.

A)

| **SNP** | **Gene** | **Position** | **Effect allele** | **eaf** |
|---|---|---|---|---|
| rs4691707 | RP13-487K5.1 | 4:156441314 | A | 0.6640 |
| rs7692387 | GUCY1A3 | 4:156635309 | A | 0.2048 |

*SNP: single nucleotide polymorphism; eaf: effect allele frequency.

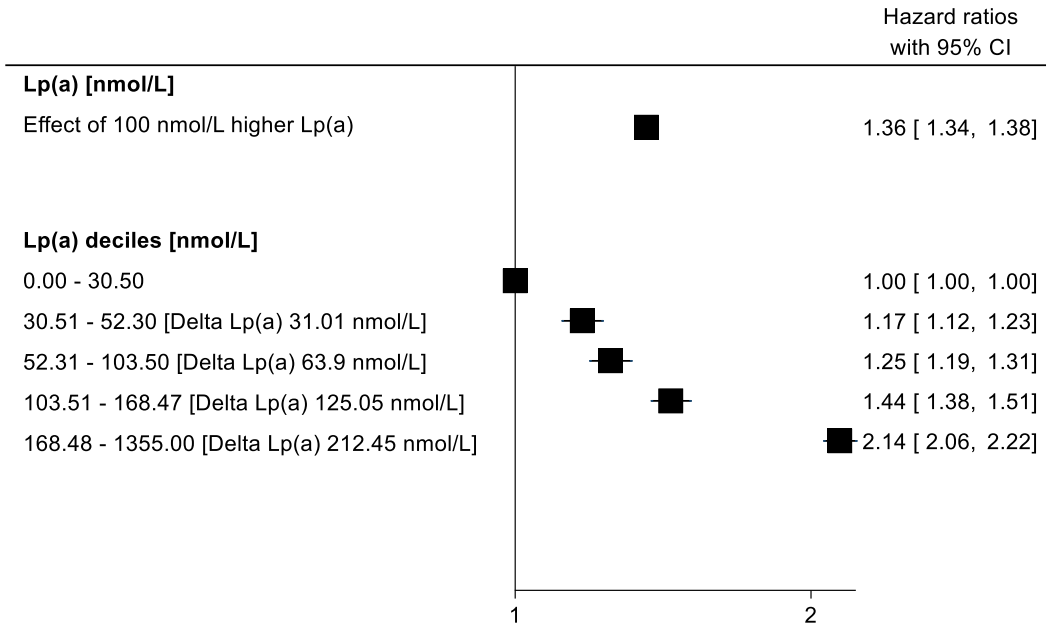
B)

| **SNP** | **Gene** | **Position** | **Effect allele** | **eaf** |
|---|---|---|---|---|
| rs6025 | F5 | 1:169519049 | C | 0.9881 |
| rs4524 | F5 | 1:169511755 | C | 0.2535 |
| rs3136516 | F2 | 11:46760756 | A | 0.5229 |
| rs1799963 | F2 | 11:46761055 | G | 0.992 |

*SNP: single nucleotide polymorphism; eaf: effect allele frequency.

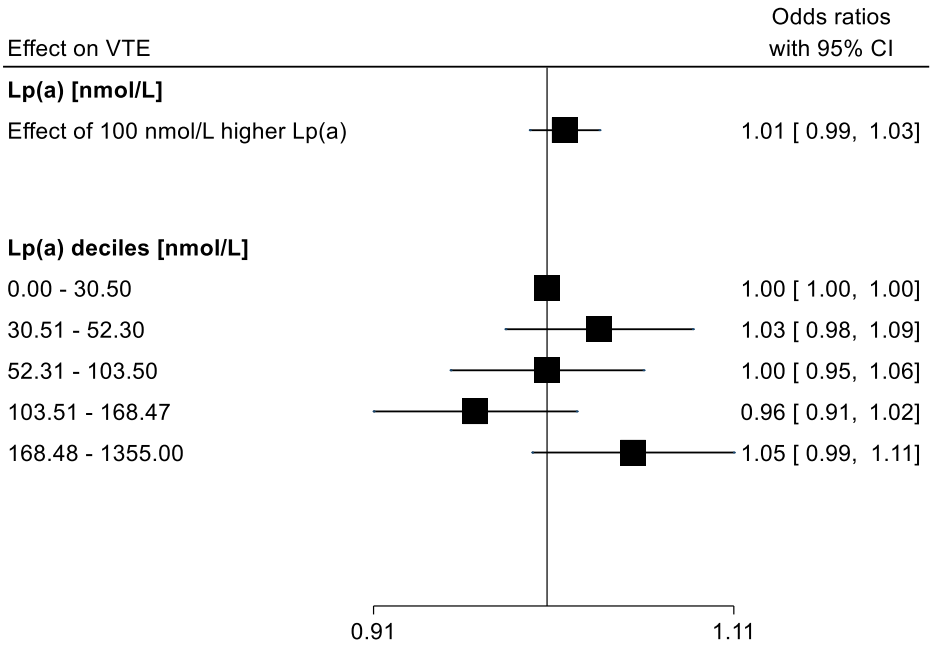
**Appendix VII.** Effect of measured Lp(a) on major coronary events.

* CI: confidence interval.

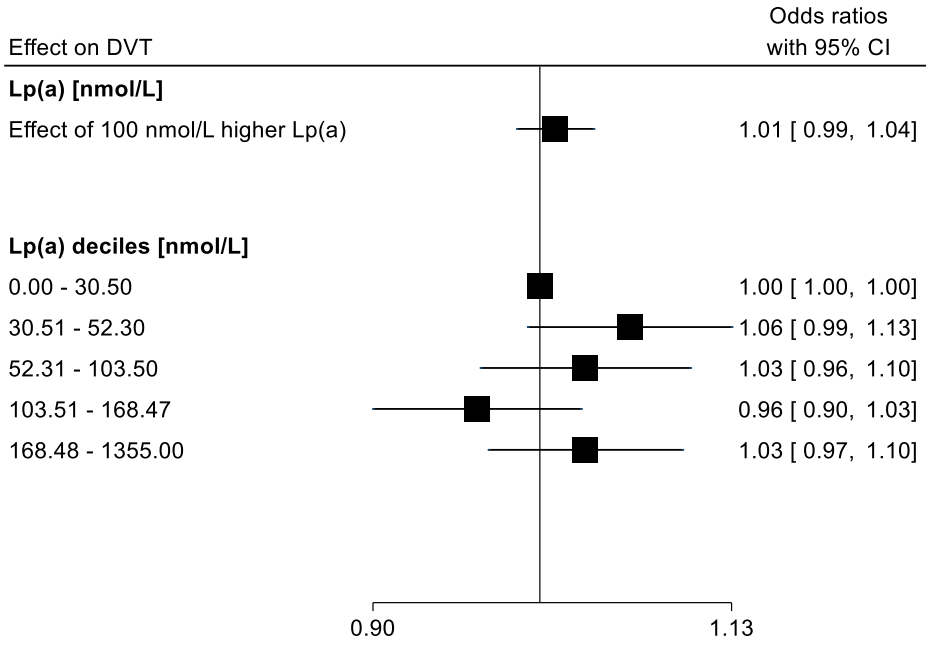


**Appendix VIII.** Effect of measured Lp(a) on venous thromboembolism (VTE, panel A), deep vein thrombosis (DVT, panel B), and pulmonary embolism (PE, panel C). * CI: confidence interval.

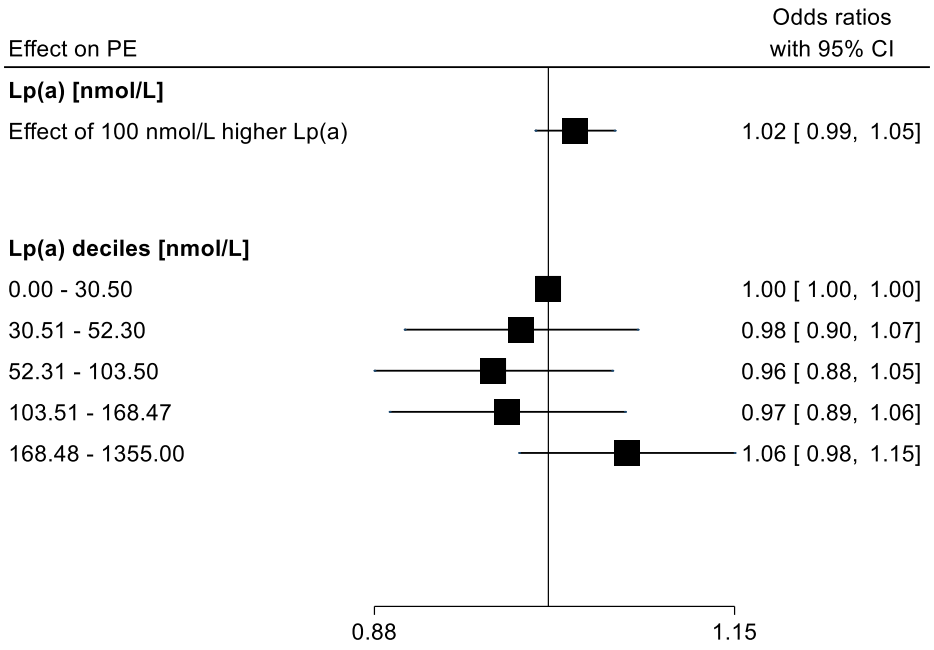
(A)




(B)



(C)


Effect on PE
Odds ratios with 95% CI
Lp(a) [nmol/L]
Effect of 100 nmol/L higher Lp(a)
1.02 [ 0.99, 1.05]
Lp(a) deciles [nmol/L]
0.00 - 30.50
1.00 [ 1.00, 1.00]
30.51 - 52.30
0.98 [ 0.90, 1.07]
52.31 - 103.50
0.96 [ 0.88, 1.05]
103.51 - 168.47
0.97 [ 0.89, 1.06]
168.48 - 1355.00
1.06 [ 0.98, 1.15]
0.88
1.15

**Appendix IX.** Baseline characteristics, measured at the time of enrolment in the UK Biobank, are presented stratified by sex and by Lp(a) instrument variable genetic score copies.

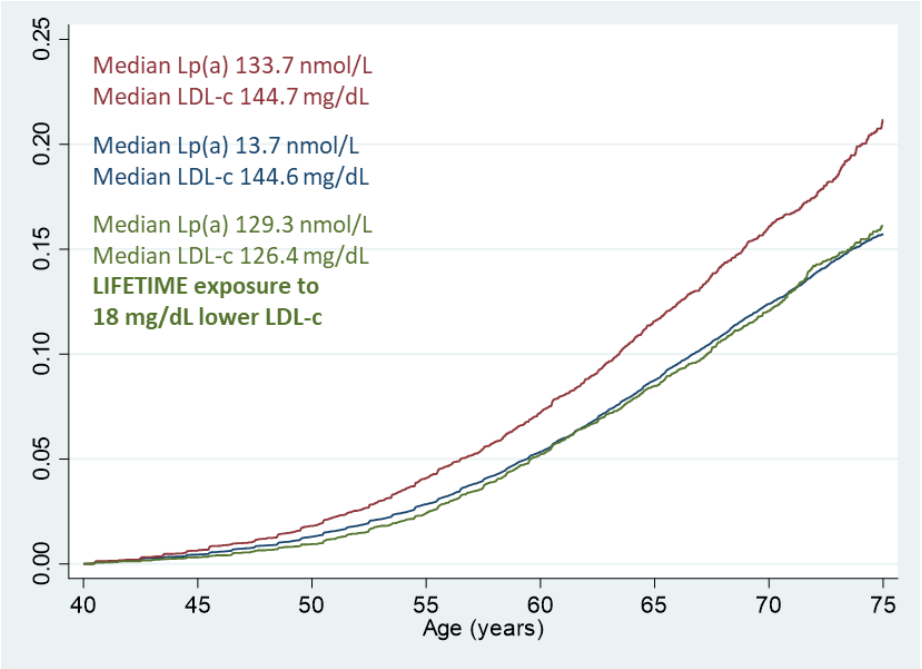
| **Characteristics male cohort** | **Overall** | **Lp(a) IV genetic score copies** | | |
|---|---|---|---|---|
| | | **0** | **1** | **2** |
| No. participants | 203,672 | 164,195 | 35,504 | 1,943 |
| Age, y (SD) | 57.5 (8.1) | 57.5 (8.1) | 57.5 (8.1) | 57.5 (7.9) |
| Lp(a), nmol/L (IQR) | 16.6 [2.4-6.7] | 12.2 [2.3-5.7] | 139 [9.1-99.9] | 248.1 [95.5-181.5] |
| LDL-C, mg/dL (SD) | 134.8 (33.3) | 134.4 (33.2) | 136.2 (33.6) | 138.6 (34.3) |
| ApoB, mg/L (SD) | 102.8 (23.8) | 102.5 (23.8) | 103.9 (23.8) | 106 (24.3) |
| TG, mg/dL (IQR) | 150.6 [65.8-105] | 151.5 [66.3-105.8] | 146.2 [64.2-101.9] | 141.8 [65.4-98.3] |
| HDL-C, mg/dL (SD) | 49.6 (12.1) | 49.6 (12) | 49.8 (12.2) | 49.3 (11.8) |
| CRP, mg/L (IQR) | 1.3 [0.3-0.7] | 1.3 [0.3-0.7] | 1.3 [0.3-0.7] | 1.3 [0.3-0.7] |
| SBP, mmHg (SD) | 140.9 (17.4) | 140.9 (17.4) | 140.9 (17.5) | 140.5 (17.3) |
| BMI, kg/m$^2$ (SD) | 27.8 (4.2) | 27.8 (4.2) | 27.9 (4.3) | 28.1 (4.5) |
| No. MCE | 17,613 | 13,254 | 4,089 | 270 |
| **Characteristics female cohort** | **Overall** | **Lp(a) IV genetic score copies** | | |
| | | **0** | **1** | **2** |
| No. participants | 242,102 | 194,274 | 42,154 | 2,307 |
| Age, y (SD) | 57.1 (7.9) | 57.1 (7.9) | 57.1 (7.9) | 57 (8) |
| Lp(a), nmol/L (IQR) | 20.7 [2.8-8] | 14.8 [2.6-6.7] | 153.1 [10.7-109.7] | 273.5 [105.1-203.3] |
| LDL-C, mg/dL (SD) | 140.7 (33.6) | 139.9 (33.4) | 144 (34.2) | 146.1 (34.7) |
| ApoB, mg/L (SD) | 104 (23.8) | 103.5 (23.7) | 106.3 (24) | 107.6 (24.1) |
| TG, mg/dL (IQR) | 118.7 [57.9-85.9] | 119.3 [58.2-86.4] | 116.1 [57.3-84.2] | 112.2 [55.5-81.1] |

| | | | | |
|---|---|---|---|---|
| HDL-C, mg/dL (SD) | 61.8 (14.6) | 61.7 (14.6) | 62.1 (14.7) | 62.6 (15.2) |
| CRP, mg/L (IQR) | 1.4 [0.3-0.7] | 1.4 [0.3-0.7] | 1.4 [0.3-0.7] | 1.3 [0.3-0.7] |
| SBP, mmHg (SD) | 135.2 (19.2) | 135.1 (19.2) | 135.3 (19.2) | 135.6 (19.4) |
| BMI, $kg/m^2$ (SD) | 27 (5.1) | 27 (5.1) | 27 (5.2) | 26.9 (5.1) |
| No. MCE | 5,175 | 3,856 | 1,224 | 95 |

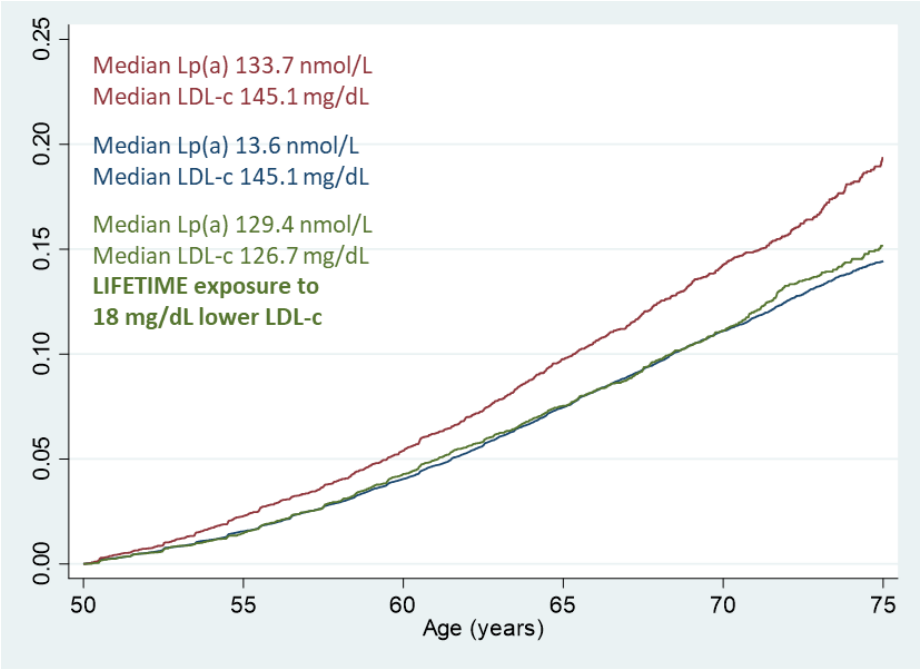
Characteristics are presented as means and standard deviations (SD) for normally distributed variables, median and interquartile ranges (IQR) for non-normally distributed variables (triglycerides, Lp(a), and CRP), or percentages for dichotomous variables. LDL-C: low-density lipoprotein cholesterol; ApoB: apolipoprotein B; TG: triglycerides; HDL-C: high-density lipoprotein cholesterol; Lp(a): lipoprotein(a); SBP: systolic blood pressure; CRP: c-reactive protein; BMI: body mass index; MCE: major coronary events.

**Appendix X.** Event curves for lifetime risk of major coronary events in male cohort by Lp(a) genetic score, LDL instrumental variable, at different enter time (40 years panel A, 50 years panel B, 60 years panel C).

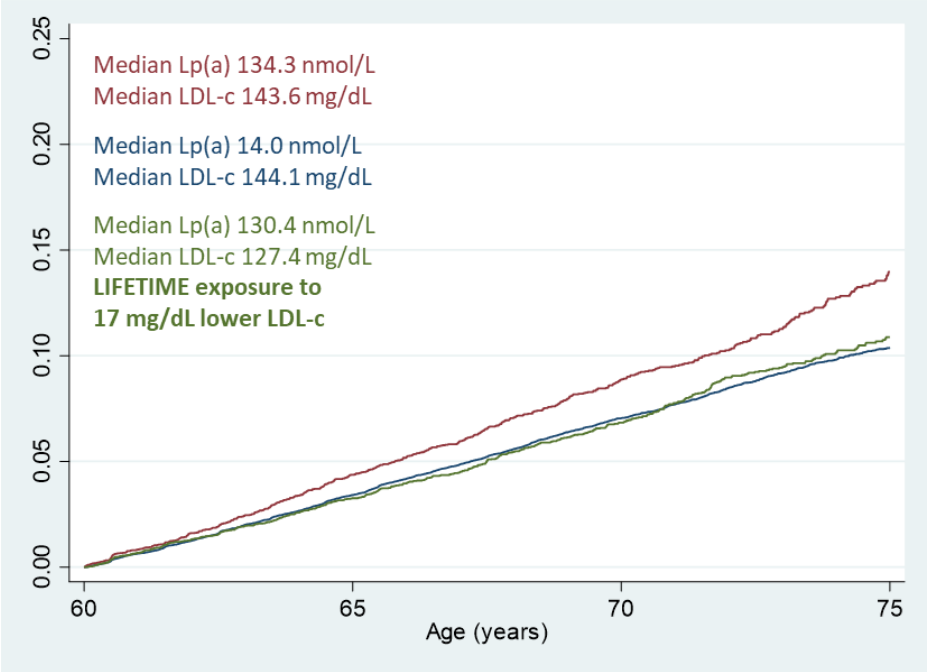
(A)




(B)

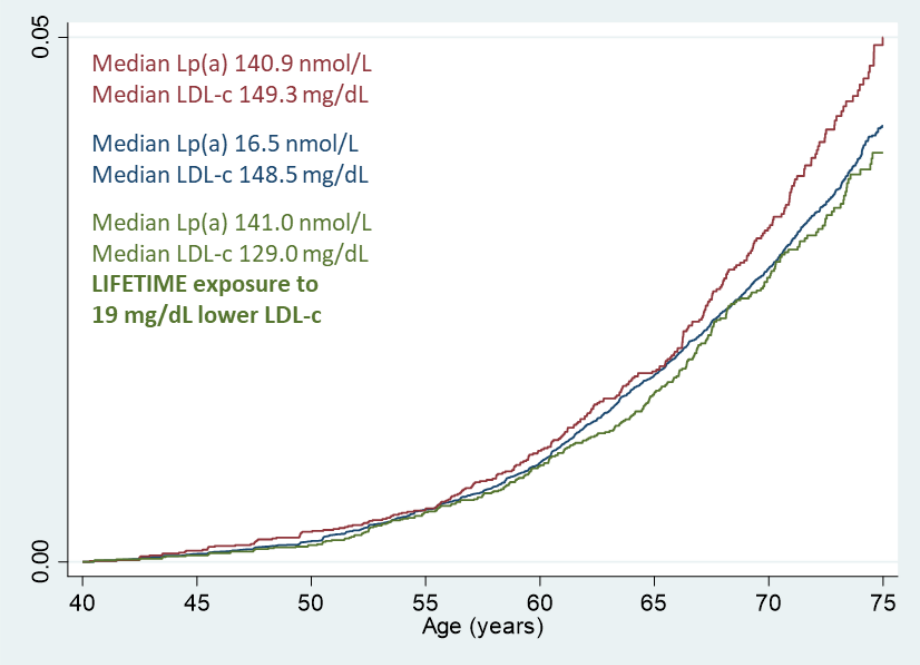


(C)

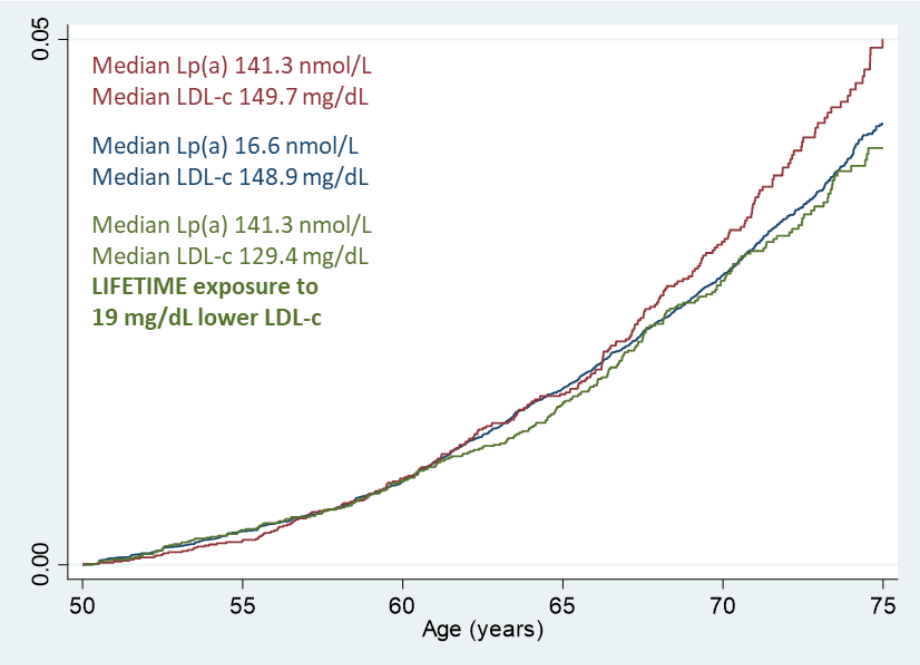

Median Lp(a) 134.3 nmol/L
Median LDL-c 143.6 mg/dL
Median Lp(a) 14.0 nmol/L
Median LDL-c 144.1 mg/dL
Median Lp(a) 130.4 nmol/L
Median LDL-c 127.4 mg/dL
LIFETIME exposure to
17 mg/dL lower LDL-c
0.00
0.05
0.10
0.15
0.20
0.25
60
65
70
75
Age (years)

**Appendix XI.** Event curves for lifetime risk of major coronary events in female cohort by Lp(a) genetic score, LDL instrumental variable, at different enter time (40 years panel A, 50 years panel B, 60 years panel C).

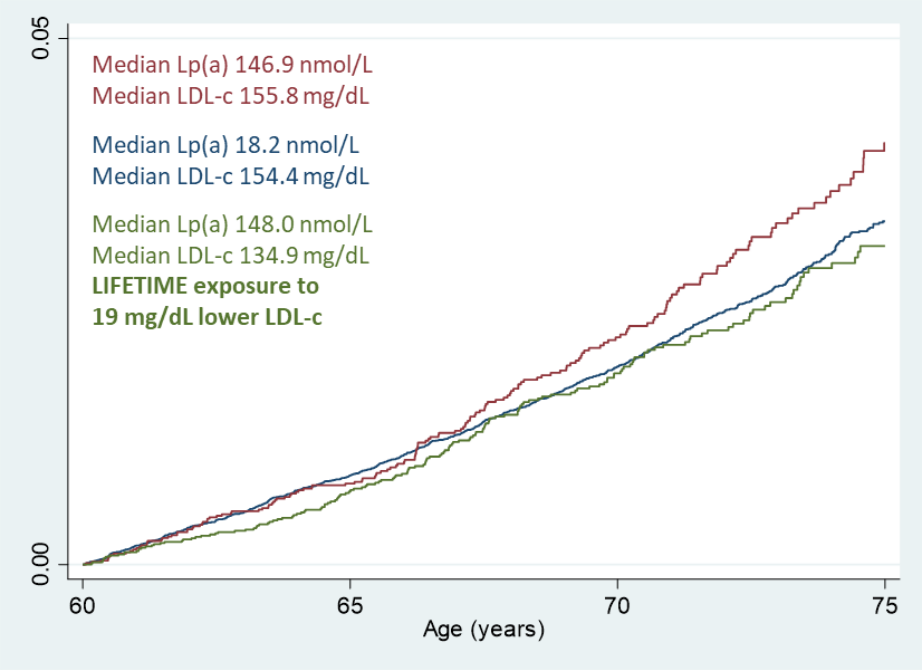
(A)




(B)

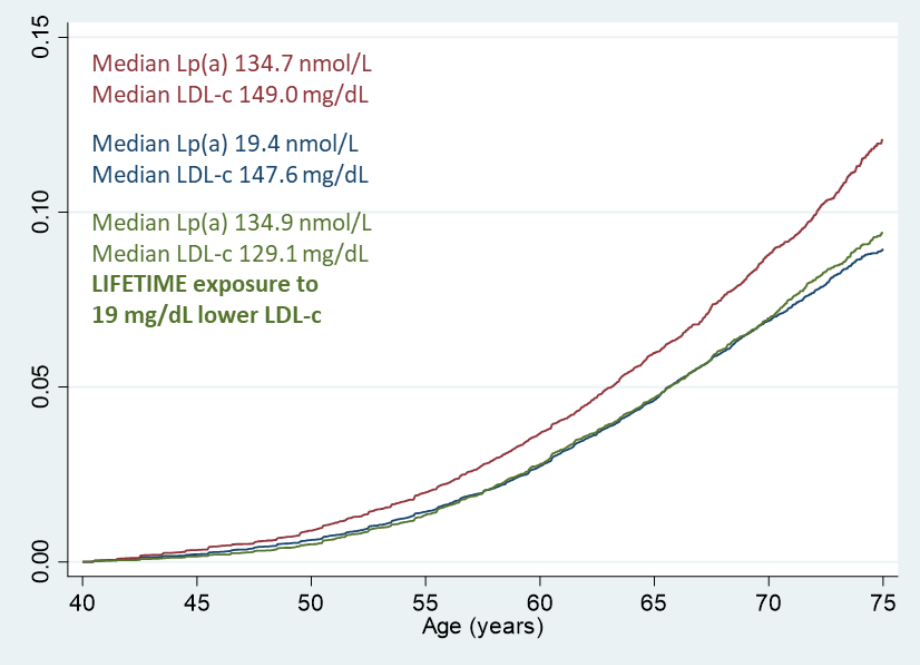


(C)

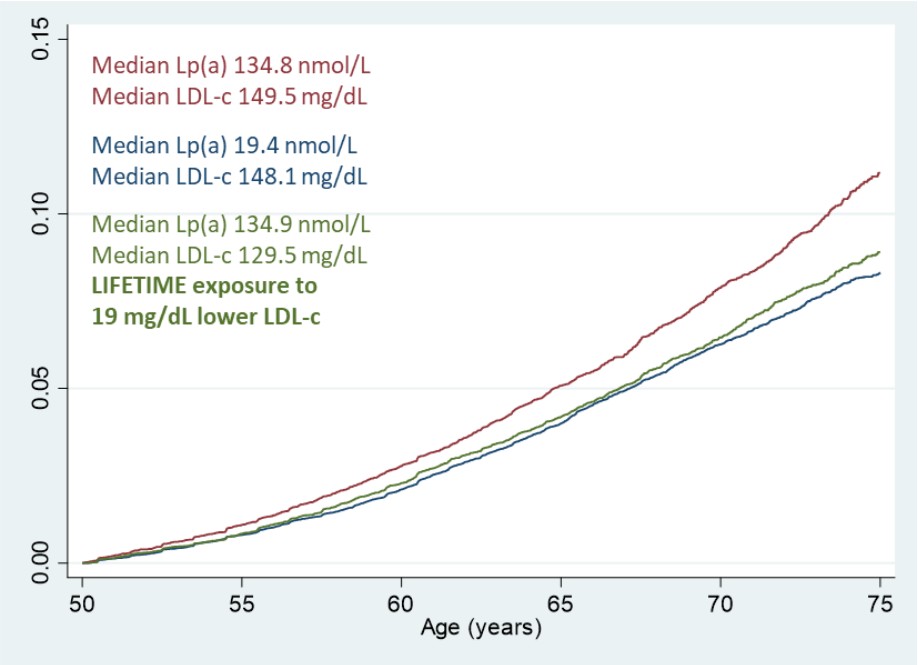

0.05
0.00
Median Lp(a) 146.9 nmol/L
Median LDL-c 155.8 mg/dL
Median Lp(a) 18.2 nmol/L
Median LDL-c 154.4 mg/dL
Median Lp(a) 148.0 nmol/L
Median LDL-c 134.9 mg/dL
LIFETIME exposure to
19 mg/dL lower LDL-c
60
65
70
75
Age (years)

**Appendix XII.** Event curves for lifetime risk of major coronary events by measured Lp(a) groups, LDL instrumental variable, at different enter time (40 years panel A, 50 years panel B, 60 years panel C).

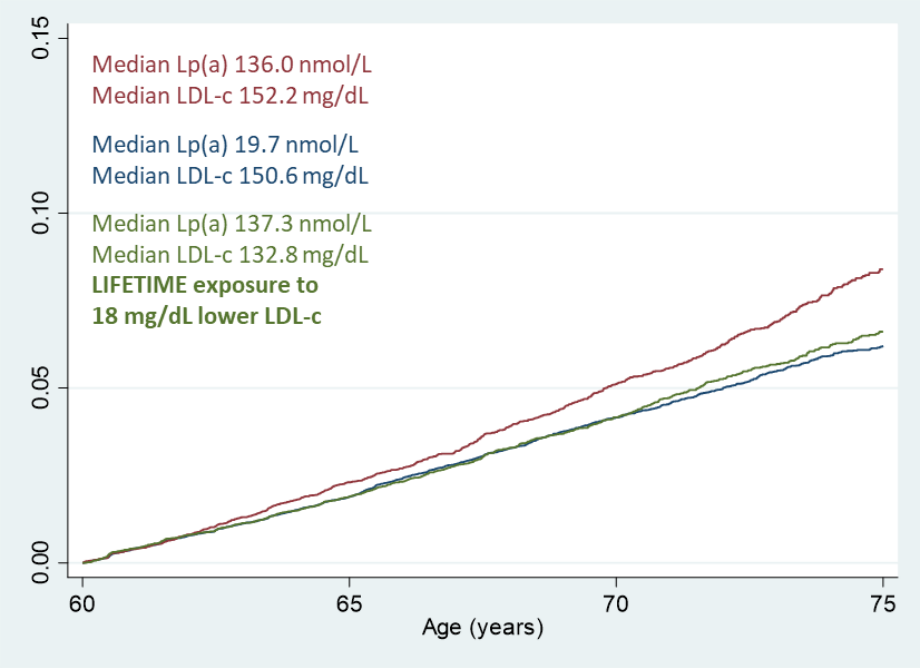
(A)




(B)



(C)


Median Lp(a) 136.0 nmol/L
Median LDL-c 152.2 mg/dL
Median Lp(a) 19.7 nmol/L
Median LDL-c 150.6 mg/dL
Median Lp(a) 137.3 nmol/L
Median LDL-c 132.8 mg/dL
LIFETIME exposure to
18 mg/dL lower LDL-c
0.00
0.05
0.10
0.15
60
65
70
75
Age (years)

## Appendix XIII.

Step 1:

In the population, participants with one copy of either rs10455872 or rs3798220 had a hazard ratio (HR) for MCE of 1.47 (95%CI: 1.42-1.51), compared with the reference group (no copies), attributable to a difference (Δ) in Lp(a) levels of 122.7 nmol/L. The HR standardized by an incremental of 100 nmol/L of Lp(a) is given by:

$$HR\ per\ 100\ nmol/L = HR^{100/\Delta} = 1.47^{100/122.7} = 1.37$$

Starting from this adjusted HR, all the estimate in the table (column "HR for MCE") were calculated based on specific difference (Δ) in Lp(a) levels in the population.

Step 2:

The relative risk reduction on log scale for each HR was calculated as:

$$\ln(RR) = \ln(HR) * (-1)$$

Step 3:

The duration of the treatment (X) was calculated as:

$$\mathrm{X} = 80 - Age\ start\ treatment\ (Rx)$$

where 80 years was set as maximum age in the survival analysis.

Step 4:

With these assumptions the following formula was applied to calculate the extra LDL cholesterol reduction needed (Y) to abolish the risk due to high Lp(a):

$$Y = \ln(RR) \,/\, [X * (-0.0126) - 0.1601\,]$$

where

- 0.0126 is the amount of each additional year LDL lowering reduce risk on log scale;
- 0.1601 is given by the sum of the natural log of HR 0.80, which is the risk reduction produces by 1 mmol/L decrease over 5 years (-0.2231), and the quantity described above multiply by 5 years of treatment (0.0126*5).

The reference for the formula applied in this study is:

*Ference BA et al., Low-density lipoproteins cause atherosclerotic cardiovascular disease. Evidence from genetic, epidemiologic, and clinical studies. A consensus statement from the European Atherosclerosis Society Consensus Panel. Eur Heart J. 2017 Aug 21;38(32):2459-2472. PMID: 28444290*.

Here an applied example of how to calculate the extra LDL-C reduction needed to abolish the risk due to high Lp(a), compared with the median value in the population, if the treatment is started at 40 years.

| **Δ nmol/L** | **HR for MCE** | **ln(RR)** | **Age start Rx** | **Rx duration (X)** | **Required LDL-C reduction (Y)** |
|---|---|---|---|---|---|
| 300 | 2.56 | -0.938 | 40 | 40 | 1.41 |
| 250 | 2.19 | -0.782 | 40 | 40 | 1.18 |
| 200 | 1.87 | -0.626 | 40 | 40 | 0.94 |
| 150 | 1.60 | -0.469 | 40 | 40 | 0.71 |
| 100 | 1.37 | -0.313 | 40 | 40 | 0.47 |
| 50 | 1.17 | -0.156 | 40 | 40 | 0.24 |
| ref | ref | ref | ref | ref | ref |

HR: hazard ratio. Estimates have been standardized for 100 nmol/L increment in Lp(a) level; MCE: major coronary events; ln(RR): natural log of the relative risk reduction. It represents how much the risk has to be reduced in a log scale in order to overcome the extra risk due to high Lp(a) values; Rx: treatment; the extra LDL-C reduction needed to abolish the risk due to high Lp(a) is expressed as mmol/L, to covert LDL cholesterol to milligrams per decilitre, multiply by 38.67.

# DOCTORIAL ACTIVITY REPORT

During my PhD programme, conducted at the Epidemiology and Preventive Pharmacology Service (SEFAP) of the Department of Pharmacological and Biomolecular Sciences, I have gain experiences and acquired/improved skills related to:

- integrate and use the pharmacological knowledge acquired in order to developing interdisciplinary pharmacoepidemiological studies in the field of health science;
- bibliographic search, review of scientific literature (through PubMed, EMBASE and Web of Science databases) and interpretation of epidemiological data;
- organization and management of large databases;
- design, conduction and analysis of pharmacoepidemiology and pharmacoutilization (drug use profiles) studies, mainly through healthcare utilization databases;
- evaluation of the risk/benefit profile of drugs in the context of real-world clinical practice, through the estimation of the association between their use and the reduction of the incidence of events (effectiveness) or the development of adverse events (safety), through both the use of databases (administrative, clinical or pharmacovigilance) and the application of meta-analytic methodologies based on the published results of clinical and/or experimental studies;
- understand both simple and more complex statistical methods for causal inference in Mendelian randomization studies, gaining sufficient knowledge to undertake my own Mendelian randomization analyses, to understand the assumptions on which causal inferences are based, and to critically appraise published studies using Mendelian randomization;
- mentoring activity for student theses.

I also acquired advanced knowledge of SPSS and STATA statistic software and improve even more my knowledge of R and SAS software.

My research activity during the PhD mainly focused on examining biomarkers and lifestyle factors for a causal effect on the risk of cardiovascular disease using the Mendelian randomization design. To understand the statistical methods for causal inference in Mendelian randomization studies, and the instrumental variable assumptions on which they are based, I attended a two-days course by Stephen Burgess at the MRC Biostatistics Unit/Cardiovascular Epidemiology Unit of the University of Cambridge (United Kingdom) in March 2018. I have been also involved in several projects at the Cardiovascular Epidemiology Unit of the Department of Public Health and Primary Care (University of Cambridge, UK), where I have been accepted to carry out a training period (six months, January-June 2020) with the supervision of Professor Brian A. Ference, director of Research (Professor) in Translational Therapeutics Executive and director of the Centre for Naturally Randomised Trials. In this context, I have also enhanced my knowledge on genomic and pharmacogenomic topic related to the cardiovascular system, and experienced the conduction of studies based on the principles of Mendelian randomization in an attempt to identify targets for early intervention, and to model "naturally randomized trials" that attempt to frame and answer clinical questions to fill evidence gaps when an actual clinical trial would be impractical or impossible to conduct. For this traineeship, I also won a scholarship supplied by The Italian Society of Pharmacology (SIF), which was made available for young researchers, to facilitate the stay in foreign laboratories for the advancement of research programs.

Besides this project, during my PhD course, I have also collaborated with:

- the research group of Professor Giovanni Corrao, at the Department of Statistics and Quantitative Methods (University of Milano-Bicocca). In this context, I gained experience in the design and in the conduction of epidemiological studies aimed at evaluating the prevalence of risk factors and their correlation with cardiovascular diseases in the Italian population, through the analysis of healthcare data from regional administrative databases of the outpatient drug prescriptions;
- the research group of Professor Enrica Menditto, director of the Center of Pharmacoeconomics and Drug Utilization Research (CIRFF) of the University of Naples Federico II. In this context, we designed and conducted a prospective, pragmatic, multicentre and open-label trial (EDU.RE.DRUG study) aiming to deeply investigate the prescribing practice among general practitioners and the appropriate drug use by their patients in two Italian regions;
- the research group of Francesco Barone Adesi, Professor of public health at the Università del Piemonte Orientale in Novara. In this context we collaborated to the following study: "Potentially inappropriate prescribing among elderly: evaluation of temporary trends 2012-2018 in Piedmont";
- the research group of Elisabetta Poluzzi, Professor at the department of Medical and Surgical Sciences of the University of Bologna. In this context we started a scientific collaboration to analyse the variation in drug consumption and/or use of health services in several cohorts of patients on chronic

treatments consequent to the severe acute respiratory syndrome coronavirus 2 and resultant COVID-19 pandemic. This project is still ongoing.

During my PhD programme, I attended a number of congresses (outlined below), at national and international level, in the belief that sharing experiences with other research groups working on the same topic of interest is a valuable key point to broaden knowledge and develop and optimize research practices.

| **Date** | **Title of contribution** | **Site** |
|---|---|---|
| 28th-30th November 2021 | Separate and combined effects of body mass index and polygenic predisposition to high BMI on the risk of developing type 2 diabetes **Oral presentation** | 35° Congresso Nazionale SISA. Virtual Edition |
| 7th-9th October 2021 | Effetto individuale e combinato dell'indice di massa corporea e della predisposizione poligenica ad elevati livelli di BMI sul rischio di sviluppo del diabete di tipo 2 **Oral presentation** | XV Congresso regionale SITeCS /giornata studio SISA Lombardia. Milan, Italy |
| 30th May-2nd June 2021 | Evaluating the distribution of a 12 LDL-C raising variants score in patients with familial hypercholesterolemia **Poster presentation** | The 89th European atherosclerosis society congress. Virtual Edition |

| | | |
|---|---|---|
| 21st-22nd May 2021 | Comparing the distribution of a 12 LDL-C raising single nucleotide polymorphisms score in patients with familial hypercholesterolemia enrolled in the LIPIGEN study **Oral presentation** | Spring Meeting Giovani Ricercatori SIIA-SIMI-SIPREC-SISA 2021. Virtual Edition |
| 9th -13th March 2021 | Effect of statins on Alzheimer's disease and dementia risk: a meta-analysis of observational studies **Oral presentation** | 40° Congresso nazionale della Società Italiana di Farmacologia. Virtual Edition |
| 22nd-24th November 2020 | Comparing the distribution of a 12 LDL-C raising single nucleotide polymorphisms score in patients with familial hypercholesterolemia enrolled in the LIPIGEN study **Oral presentation** | 34° Congresso Nazionale SISA. Virtual Edition |
| 15th-17th October 2020 | Individuazione di target terapeutici: randomizzazione mendeliana **Invited speaker** | XIV Congresso nazionale SITeCS /giornata studio SISA Lombardia. Milan, Italy |
| 4th-7th October 2020 | Cardiovascular outcomes with omega-3 polyunsaturated fatty acids | 88° European atherosclerosis society |

| | | |
|---|---|---|
| | supplementation: An updated meta-analysis of randomized controlled trials<br>**Poster presentation** | congress. Virtual Edition |
| 16th-17th September 2020 | Cumulative exposure to bisphosphonates and risk of cardiovascular and cerebrovascular events<br>**Oral presentation** | ICPE ALL ACCESS – 35 years of real-world science. Virtual Edition |
| 16th-17th September 2020 | Evaluation of the effect of omega-3 polyunsaturated fatty acids supplementation on cardiovascular outcomes: an updated meta-analysis of randomized controlled trials<br>**Oral presentation** | ICPE ALL ACCESS – 35 years of real-world science. Virtual Edition |
| 9th-10th December 2019 | Prescrizione di farmaci interagenti e di duplicati terapeutici negli adulti<br>**Oral presentation** | XXVIII Seminario Nazionale di Farmacologia, Rome, Italy |
| 24th-26th November 2019 | Association between a cumulative exposure to bisphosphonates and risk of incident cardio-cerebrovascular events: a retrospective cohort study<br>**Poster presentation** | 33° Congresso Nazionale SISA, Rome, Italy |
| 20th -23rd November 2019 | Sex differences in factors associated with adherence | 39° Congresso Nazionale della |

| | | |
|---|---|---|
| | to statin therapy: a population-based study<br>**Poster presentation** | Società Italiana di Farmacologia. Florence, Italy |
| 24th-26th October 2019 | Cumulative exposure to bisphosphonates and risk of cardio-cerebrovascular events: a population-based retrospective cohort study<br>**Oral presentation** | Convegno Regionale SISA Lombardia, XIII Congresso Nazionale SITeCS. Milan, Italy |
| 18th September 2019 | Cumulative exposure to bisphosphonates and risk of cardio-cerebrovascular events: a population-based retrospective cohort study<br>**Oral presentation** | Next Step X. Milan, Italy |
| 26th-29th May 2019 | Sex-differences in adherence to statin therapy in a real-world population<br>**Science at a glance** | 87th European atherosclerosis society congress. Maastricht, The Netherland |
| 28th February, 1st-2nd March 2019. | Sex differences in adherence to statin therapy in the clinical practice<br>**Poster presentation** | Spring Meeting dei gruppi Giovani Ricercatori SIIA, SIMI e SISA. Rimini, Italy |
| 25th-27th November 2018 | Valutazione della performance del Dutch Lipid Clinic Network score utilizzando un database italiano di pazienti con ipercolesterolemia familiare | 32° Congresso Nazionale SISA. Bologna, Italy |

| | | |
|---|---|---|
| | **Oral presentation** | |
| 4th-6th October 2018 | Valutazione della performance del Dutch Lipid Clinic Network score utilizzando un database italiano di pazienti con ipercolesterolemia familiare<br>**Oral presentation** | Convegno Regionale SISA Lombardia, XII Congresso Nazionale SITeCS. Milan, Italy |

Moreover, in the last three years, I followed a series of seminars, conferences and educational courses (listed below), in order to keep on training and updating my knowledge:

- IV Giornata della Ricerca del Centro E. Grossi Paoletti - La prevenzione cardiovascolare nell'era post-COVID, Milan, Italy, 26th November 2021.
- ICPE ALL ACCESS. Skills Courses, September 2020
- Spring Meeting Giovani Ricercatori SIIA-SIMI-SISA 2020 - Novità sulla prevenzione e cura della malattia cardiovascolare. 18th, 25th June, 2nd July 2020.
- XXVIII Seminario Nazionale di Farmaclogia, Istituto Superiore di Sanità - La valutazione dell'uso e della sicurezza dei farmaci: esperienze in Italia, 9th-10th December 201, Rome.
- Mendelian Randomization Course, Cambridge, United Kingdom, 20th-21st March 2019.
- "III Giornata del Centro E. Grossi Paoletti" Milan, Italy, 14th June 2019.
- "HOT NUT 2 – MICROBIOTA REVOLUTION: dove siamo oggi e quali risposte ci aspettiamo in futuro", Milan, Italy, 5th April 2019.

- "Essere cittadini tra scienza, sapere e decisione pubblica", Milan, Italy, 25th-26th March 2019.
- "HDL–Beyond Atherosclerosis", Milan, Italy, 1st February 2019.
- "Advance Course on Rare Dysplipidaemia and Atherosclerosis", Milan, Italy, 19th October 2018.

Finally, in the awareness that transferring scientific knowledge within the scientific community and, above all, to the public is one of the hardest challenges of the "research world" nowadays, I have carried out information and dissemination activities. In particular, I am a member of SEFAPnews editorial-board, that provides monthly newsletters published on SEFAP website (www.sefap.it), on the topic of pharmacovigilance, pharmacoepidemiology, pharmacoutilization and health policies. Furthermore, I was involving in dissemination activities through RicercaMix blog of the Department of Pharmacological and Biomolecular Sciences of the University of Milan (www.ricercamix.org). In addition, I published an article regarding Mendelian randomization in the Italian Journal of Atherosclerosis ("Cos'è uno studio di randomizzazione Mendeliana e quali sono le applicazioni in ambito di dislipidemie. GIA 2020; 11(2):26-40") and a short informative article about adherence to statin therapy ("Aderenza alla terapia con statine nella pratica clinica: differenze di genere. Medicina di Genere Newsletter, Ottobre 2020").

**List of publications**

I am the author of nineteen papers in international peer reviewer journals (four as "First Author", six as "Second Author"):

- Baragetti A, Casula M, Scarinzi P, Ristè F, Scicali R, Biolo M, Lugari S, Dall'Agata M, Gazzotti M, Olmastroni E, Alieva AS,

Nascimbeni F. Achilles tendon ultrasonography in familial hypercholesterolemia: A sub-study of the LIpid transPort disorders Italian GEnetic Network (LIPIGEN). J Intern Med. 2021 Dec 7. doi: 10.1111/joim.13421. Epub ahead of print. PMID: 34875114.

- Olmastroni E, Molari G, De Beni N, Colpani O, Galimberti F, Gazzotti M, Zambon A, Catapano AL, Casula M. Statin use and risk of dementia or Alzheimer's disease: a systematic review and meta-analysis of observational studies. Eur J Prev Cardiol. 2021 Dec 6:zwab208. doi: 10.1093/eurjpc/zwab208. Epub ahead of print. PMID: 34871380.

- EAS Familial Hypercholesterolaemia Studies Collaboration (FHSC). Global perspective of familial hypercholesterolaemia: a cross-sectional study from the EAS Familial Hypercholesterolaemia Studies Collaboration (FHSC). Lancet. 2021 Sep 7:S0140-6736(21)01122-3. doi: 10.1016/S0140-6736(21)01122-3.

- Cicolari S, Pavanello C, Olmastroni E, Del Puppo M, Bertolotti M, Mombelli G, Catapano AL., Calabresi L, Magni P. Interactions of oxysterols with atherosclerosis biomarkers in subjects with moderate hypercholesterolemia and effects of a nutraceutical combination (Bifidobacterium longum BB536, red yeast rice extract) (randomized, double-blind, placebo-controlled study). Nutrients. 2021 Jan 28;13(2):427. doi: 10.3390/nu13020427.

- Baragetti A, Severgnini M, Olmastroni E, Dioguardi CC, Mattavelli E, Angius A, Rotta L, Cibella J, Consolandi C, Grigore L, Pellegatta F, Giavarini F, Caruso D, Norata GD, Catapano AL, Peano C. Gut Microbiota Functional Dysbiosis Relates to Individual Diet in Subclinical Carotid Atherosclerosis. Nutrients. 2021 Jan 21;13(2):304. doi: 10.3390/nu13020304.

- Pirillo A, Casula M, Olmastroni E, Norata GD., Catapano AL.. Global epidemiology of dyslipidaemias. Nat Rev Cardiol. 2021 Apr 8. doi: 10.1038/s41569-021-00541-4

- Gazzotti M, Casula M, Olmastroni E, Averna M, Arca M, Catapano AL. How registers could enhance knowledge and characterization of genetic dyslipidaemias: the experience of the LIPIGEN in Italy and of other networks for familial hypercholesterolemia. Atheroscler Suppl. 2020 Dec; doi: 10.1016/j.atherosclerosissup.2021.01.007.

- Alieva AS, Olmastroni E, Reutova OV, Rotar OP, Konradi AO, Shlyakhto EV, Baragetti A, Grigore L, Pellegatta F, Casula M, Tragni E, Catapano AL. Prevalence and relationship between metabolic syndrome and risk of cardiovascular disease: evidence from two population-based studies. Atheroscler Suppl. 2020 Dec;42:e41-e48. doi: 10.1016/j.atherosclerosissup.2021.01.008.

- Casula M, Gazzotti M, Bonaiti F, Olmastroni E, Arca M, Averna M, Zambon A, Catapano AL, PROSISA Study Group. Reported muscle symptoms during statin treatment among Italian dyslipidemic patients in the real-life setting: The Prosisa Study. J Intern Med. 2020 Dec 1. doi: 10.1111/joim.13219. PMID: 33259671.

- Casula M, Olmastroni E, Gazzotti M, Galimberti F, Zambon A, Catapano AL. Omega-3 polyunsaturated fatty acids supplementation and cardiovascular outcomes: do formulation, dosage, and baseline cardiovascular risk matter? An updated meta-analysis of randomized controlled trials. Pharmacol Res. 2020 Jul 4;160:105060. doi: 10.1016/j.phrs.2020.105060. Epub ahead of print. PMID: 32634581.

- Casula M, Menditto E, Galimberti F, Russo V, Olmastroni E, Orlando V, Catapano A.L, Tragni E, on behalf of EDU.RE.DRUG

Group. A pragmatic controlled trial to improve the appropriate prescription of drugs in adult outpatients: design and rationale of the EDU.RE.DRUG study. Primary Health Care journal 2020 June.

- Russo V, Orlando V, Monetti VM, Galimberti F, Casula M, Olmastroni E, Tragni E, Menditto E; EDU.RE.DRUG Group. Geographical Variation in Medication Prescriptions: A Multiregional Drug-Utilization Study. Front Pharmacol. 2020 May 5;11:418. doi: 10.3389/fphar.2020.00418. PMID: 32536861; PMCID: PMC7269055.

- Casula M, Olmastroni E, Galimberti F, Tragni E, Corrao G, Scotti L, Catapano AL. Association between the cumulative exposure to bisphosphonates and hospitalization for atherosclerotic cardiovascular events: A population-based study. Atherosclerosis. 2020 May;301:1-7. doi:10.1016/j.atherosclerosis.2020.03.021. Epub 2020 Apr 7. PMID: 32289617.

- Olmastroni E, Boccalari MT, Tragni E, Rea F, Merlino L, Corrao G, Catapano AL, Casula M. Sex-differences in factors and outcomes associated with adherence to statin therapy in primary care: Need for customisation strategies. Pharmacol Res. 2020 May;155:104514. doi: 10.1016/j.phrs.2019.104514. Epub 2019 Oct 31. PMID: 31678211.

- Olmastroni E, Baragetti A, Casula M, Grigore L, Pellegatta F, Pirillo A, Tragni E, Catapano AL. Multilevel Models to Estimate Carotid Intima-Media Thickness Curves for Individual Cardiovascular Risk Evaluation. Stroke. 2019 Jul;50(7):1758-1765. doi: 10.1161/STROKEAHA.118.024692. Epub 2019 Jun 5. PMID:31164073.

- Casula M, Olmastroni E, Boccalari MT, Tragni E, Pirillo A, Catapano AL. Cardiovascular events with PCSK9 inhibitors: an updated meta-analysis of randomised controlled trials. Pharmacol Res. 2019 May;143:143-150. doi:10.1016/j.phrs.2019.03.021. Epub 2019 Mar 26. PMID: 30926528.

- Casula M, Olmastroni E, Pirillo A, Catapano AL; MEMBERS OF THE LIPIGEN STEERING COMMETTEE; PRINCIPAL INVESTIGATORS: Coordinator center; Participant Centers; Participant Laboratories; COLLABORATORS; STUDY CENTRAL LABORATORY AND ANALYSIS GROUP. Evaluation of the performance of Dutch Lipid Clinic Network score in an Italian FH population: The LIPIGEN study. Atherosclerosis. 2018 Oct;277:413-418. doi: 10.1016/j.atherosclerosis.2018.08.013. PMID: 30270079.

- Olmastroni E, Shlyakhto EV, Konradi AO, Rotar OP, Alieva AS, Boyarinova MA, Baragetti A, Grigore L, Pellegatta F, Tragni E, Catapano AL, Casula M. Epidemiology of cardiovascular risk factors in two population-based studies. Atheroscler Suppl. 2018 Sep;35:e14-e20. doi: 10.1016/j.atherosclerosissup.2018.08.003. Epub 2018 Aug 25. PMID: 30177370.

- Baragetti A, Grejtakova D, Casula M, Olmastroni E, Jotti GS, Norata GD, Catapano AL, Bellosta S. Proprotein Convertase Subtilisin-Kexin type-9 (PCSK9) and triglyceride-rich lipoprotein metabolism: Facts and gaps. Pharmacol Res. 2018 Apr;130:1-11. doi: 10.1016/j.phrs.2018.01.025. Epub 2018 Feb 8. PMID:29428206.